Fixpoints and Iterated Updates in Abstract Argumentation

Davide Grossi
Department of Computer Science
University of Liverpool, UK

Abstract

Fixpoints play a key role in the mathematical set up of abstract argumentation theory but, we argue, have been relatively underexamined in the literature. The paper studies the logical structure underlying the computation via approximation sequences of the sort of fixpoints relevant in argumentation. Concretely, it presents a number of novel results on the fixed point theory underpinning the main Dung's semantics and, inspired by recent literature on the logical analysis of equilibrium computation in games, it provides a characterization of those semantics in terms of iterated model updates.

Introduction

Abstract argumentation theory (Dung 1995) studies arguments as further unanalyzed elements in a graph of attacks. One of the key questions it addresses is when an argument can be considered 'justified' or 'tenable' given a graph of this type. To this aim, several structural properties of attack graphs have been defined and studied to account for different notions of justifiability. These properties are commonly called *extensions* and form the core of abstract argumentation theory (cf. (Baroni and Giacomin 2009) for a recent overview).

Dynamic epistemic logic (DEL, see (van Ditmarsch, Kooi, and van der Hoek 2007; van Benthem 2011) for recent overviews) is a broad family of logical systems extending epistemic logics *à la* Hintikka (Hintikka 1962) in order to account for processes in which agents engage while acquiring new information, revising beliefs, reasoning and learning. The present paper takes inspiration from recent influential applications of DEL, which have provided logical analyses of solution concepts of game theory, such as the iterated elimination of strictly dominated strategies (van Benthem 2007) or backwards induction (Baltag, Smets, and Zvesper 2009; van Benthem and Gheerbrant 2010).

The paper focuses on the fixpoint-theoretic underpinnings of extensions viewed as 'solution concepts' of attack graphs. It studies a number of theorems—some well-known, some, to the best of our knowledge, new—concerning the computation via approximation sequences of some of the main types of extensions (viz., grounded, complete, stable and preferred). Following the DEL methodology, these theorems are then analyzed in modal logic via iterated updates, making explicit their procedural nature and enabling precise epistemic interpretations of the approximation sequences.

Technically, the paper exploits the link between abstract argumentation and modal logic investigated in (Grossi 2009; 2010) and applies notions on iterated updates developed in (Baltag and Smets 2009). Some of the results we present on approximation sequences of Dung's extensions are inspired by old results on the logical analysis of self-reference (Yablo 1984).

All in all, the paper lays a first bridge between DEL and argumentation theory. It is our hope that the results presented here could spark future interaction between these two lively fields of research which, although by different mathematical means, pursue strictly related questions.

Structure of the paper: The first two sections introduce some preliminaries on abstract argumentation theory and the modal languages that can be used to study it. The third section presents an analysis in public announcement logic of the approximation sequence computing the grounded extensions. The fourth and fifth section pursue a similar analysis of the complete and stable extensions via an operation of update of the valuation function of a model. Those sections also present some original results on the fixpoint computation of complete and stable extensions.

Preliminaries

We introduce some key notions of abstract argumentation.

Attack graphs

We start by the key notion of (Dung 1995):

Definition 1 (Attack graph). *An attack graph—or Dung framework—is a tuple $\mathcal{A} = \langle A, \rightarrow \rangle$ where:*

- *A is a non-empty finite set—the set of arguments;*
- *$\rightarrow \subseteq A^2$ is a binary relation—the attack relation.*

The set of all Dung frameworks on a given set A is denoted $\mathfrak{A}(A)$. The set of all Dung frameworks is denoted $\mathfrak{A}$. With $x \rightarrow y$ we indicate that x attacks y, and with $X \rightarrow x$ we indicate that $\exists y \in X$ s.t. $y \rightarrow x$. Similarly, $x \rightarrow X$ indicates that $\exists y \in X$ s.t. $y \leftarrow x$.

These relational structures are the building blocks of abstract argumentation theory. Once A is taken to represent a set of arguments, and $\rightarrow$ an 'attack' relation between arguments (so that $a \rightarrow b$ means "a attacks b"), the study of these structures provides very general insights on how competing arguments interact and structural properties of subsets of A can be taken to formalize how collections of arguments form 'tenable' or 'justifiable' positions in an argumentation.

Characteristic functions of attack graphs

The formulation of all main argumentation theoretic properties makes use of two functions that can be naturally associated to each attack graph.

The first one is the function called in (Dung 1995) 'characteristic function', which we will call here defense function.

Definition 2 (Defense function). *Let $\mathcal{A} = \langle A, \rightarrow \rangle$ be a Dung framework. The defense function $\mathtt{d}_\mathcal{A} : \wp(A) \longrightarrow \wp(A)$ for $\mathcal{A}$ is so defined:*

$$\mathtt{d}_\mathcal{A}(X) \quad = \quad \{x \in A \mid \forall y \in A : \text{ IF } y \rightarrow x \text{ THEN } X \rightarrow y\}.$$

Given a set of arguments X, the n-fold iteration of $\mathtt{d}_\mathcal{A}$ is denoted $\mathtt{d}^n_\mathcal{A}$ for $0 \leq n < \omega$ and its infinite iteration is denoted $\mathtt{d}^\omega_\mathcal{A}$. For a given X, an infinite iteration generates an infinite sequence, or stream, $\mathtt{d}^0_\mathcal{A}(X), \mathtt{d}^1_\mathcal{A}(X), \mathtt{d}^2_\mathcal{A}(X), \ldots$. A stream is said to stabilize if and only if there exists $0 \leq n < \omega$ such that $\mathtt{d}^n_\mathcal{A}(X) = \mathtt{d}^{n+1}_\mathcal{A}(X)$. Such set $\mathtt{d}^n_\mathcal{A}(X)$ is then called the limit of the stream and is denoted $\mathtt{d}^\star_\mathcal{A}(X)$.

In other words, for a given $\mathcal{A}$, function $\mathtt{d}_\mathcal{A}$ encodes for each set of arguments X, which other arguments X is able to defend in $\mathcal{A}$.

The second function was first introduced in (Pollock 1987; 1991) and further studied in (Dung 1995). It is not known with a specific name in the literature. We call it here neutrality function.

Definition 3 (Neutrality function). *Let $\mathcal{A} = \langle A, \rightarrow \rangle$ be an attack graph. The neutrality function $\mathtt{n}_\mathcal{A} : \wp(A) \longrightarrow \wp(A)$ for $\mathcal{A}$ is so defined:*

$$\mathtt{n}_\mathcal{A}(X) \quad = \quad \{x \in A \mid \text{ NOT } X \rightarrow x\}$$

The definitions of n-fold iteration, stream, stabilization and limit are like in Definition 2.

Intuitively, given $\mathcal{A}$, function $\mathtt{n}_\mathcal{A}$ encodes for each set X of arguments in $\mathcal{A}$, the arguments about which X is neutral in the sense of not attacking any of those arguments.

Example 1 (Defensibility and neutrality in Figure 1). *The functions applied to the symmetric graph on the left of Figure 1 yield the following equations:*

$$\begin{array}{rclrcl}
\mathtt{d}(\emptyset) & = & \emptyset & \mathtt{n}(\emptyset) & = & \{a,b\} \\
\mathtt{d}(\{a\}) & = & \{a\} & \mathtt{n}(\{a\}) & = & \{a\} \\
\mathtt{d}(\{b\}) & = & \{b\} & \mathtt{n}(\{b\}) & = & \{b\} \\
\mathtt{d}(\{a,b\}) & = & \{a,b\} & \mathtt{n}(\{a,b\}) & = & \emptyset
\end{array}$$

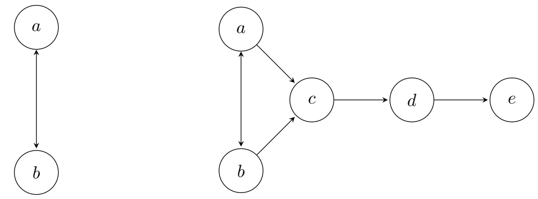


Figure 1: Two attack graphs.

Solving attack graphs

Table 1 recapitulates the basic notions of abstract argumentation which will be dealt with in the paper. They are all formulated either as fixpoints ($X = f(X)$) or post-fixpoints ($X \subseteq f(X)$) of the defense and neutrality functions, or as combinations of the two.[1]

Intuitively, conflict-freeness demands that the set of arguments at issue is not able to attack itself. Self-acceptability requires that the set of arguments is able to defend itself. An admissible set is then a set of arguments which is conflict-free and is able to defend all its attackers. So, as the name suggests, admissible sets can be thought of as 'admissible' positions within an attack graph. By considering those admissible sets which also contain all the arguments they are able to defend—viz., the admissible sets that are fixpoints of the defense function—we obtain the notion of complete extension. It formalizes the idea of a fully exploited admissible position, that is, a position which has no conflicts, and which consists exactly of all the arguments it can successfully defend. Stable extensions are the fixpoints of the neutrality function, that is, they are precisely the complement of the set of arguments they attack. The grounded extension represents what all complete extensions have in common. In a way, it formalizes what at least must be accepted as 'reasonable' within the graph. At the other end of the spectrum, preferred extensions are the maximal complete extensions which are also conflict-free. As such, they formalize the idea of a maximal coherent and admissible position in the graph.

Example 2 (Extensions in Figure 1). *Consider the graph on the right of Figure 1. The grounded extension is $\emptyset$. There are two complete extensions which are also preferred and stable: $\{a, d\}$ and $\{b, d\}$. An example of a conflict-free set which is not admissible is $\{c, e\}$.*

Argumentation in modal logic

This section presents a modal logic for abstract argumentation theory first introduced in (Grossi 2009; 2010). The logic is based on the intuition that attack graphs (Definition 1) can be viewed as Kripke frames and can therefore be used to interpret modal formulae.

A simple modal language for argumentation

Once an attack graph is viewed as a Kripke frame, the addition of a function assigning names to sets of arguments—a

[1]We are not aware of any work in argumentation which, like we do, formulates conflict-freeness as a post-fixpoint and stable extensions as fixpoints of the neutrality function.

X is conflict-free in $\mathcal{A}$	iff	$X \subseteq \mathtt{n}_{\mathcal{A}}(X)$
X is self-acceptability in $\mathcal{A}$	iff	$X \subseteq \mathtt{d}_{\mathcal{A}}(X)$
X is admissible in $\mathcal{A}$	iff	$X \subseteq \mathtt{n}_{\mathcal{A}}(X)$ and $X \subseteq \mathtt{d}_{\mathcal{A}}(X)$
X is a complete extension of $\mathcal{A}$	iff	$X \subseteq \mathtt{n}_{\mathcal{A}}(X)$ and $X = \mathtt{d}_{\mathcal{A}}(X)$
X is a stable extension of $\mathcal{A}$	iff	$X = \mathtt{n}_{\mathcal{A}}(X)$
X is the grounded extension of $\mathcal{A}$ ($\mathtt{Grn}_{\mathcal{A}}$)	iff	$X = \mathsf{lfp}.\mathtt{d}_{\mathcal{A}}$
X is a preferred extension of $\mathcal{A}$	iff	X is a largest complete extension of $\mathcal{A}$

Table 1: Some of the key notions of abstract argumentation theory from (Dung 1995). Expression $\mathsf{lfp}.\mathtt{d}_{\mathcal{A}}$ denotes the least fixpoint of the defense function.

labeling or valuation function—yields a Kripke model.

Definition 4 (Attack models). *Let $\mathbf{P}$ be a set of atoms. An attack model is a tuple $\mathcal{M} = \langle \mathcal{A}, \mathcal{V} \rangle$ where $\mathcal{A} = \langle A, \rightarrow \rangle$ is an attack graph and $\mathcal{V} : \mathbf{P} \longrightarrow \wp(A)$ is a valuation function. A pointed attack model is a pair $\langle \mathcal{M}, a \rangle$ with $a \in A$. The set of all attack models is denoted $\mathfrak{M}$.*

Intuitively, an attack model is nothing but an attack graph where arguments are labeled by propositional atoms or, equivalently, where sets of arguments are named. So, the fact that an argument a belongs to the set $\mathcal{V}(p)$ in a given model $\mathcal{M}$ reads in logical notation as $\langle \mathcal{A}, \mathcal{V} \rangle, a \models p$. By using the language of propositional logic we can then form 'complex' labels φ for sets of arguments stating, for instance, that "a belongs to both the sets called p and q": $\langle \mathcal{A}, \mathcal{V} \rangle, a \models p \wedge q$. Consider now statements such as: "*there exists* an argument in a set named φ attacking argument a" or "*for all* attackers of argument a there exist some attackers in a set named φ". These are statements involving a bounded quantification and they can be naturally formalized by a modal operator $\Diamond$ whose reading is: "there exists an attacking argument such that ...". To this we turn in the next section.

Syntax & semantics Language $\mathcal{L}$ is a standard modal language with two modalities: $\Diamond$ and $\langle \mathsf{U} \rangle$ (the universal modality). It is built on the set of atoms $\mathbf{P}$ by the following BNF:

$$\mathcal{L}(\mathbf{P}) : \varphi ::= p \mid \bot \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Diamond\varphi \mid \langle \mathsf{U} \rangle\varphi$$

where p ranges over $\mathbf{P}$. Standard definitions for the remaining Boolean operators and the duals $\Box$ and $[\mathsf{U}]$ are assumed.

Definition 5 (Satisfaction for $\mathcal{L}$). *Let $\varphi \in \mathcal{L}$. The satisfaction of φ by a pointed attack model $(\mathcal{M}, a)$ is inductively defined as follows:*

$$\mathcal{M}, a \models \Diamond\varphi \iff \exists b \in A : a \leftarrow b \text{ AND } \mathcal{M}, b \models \varphi$$
$$\mathcal{M}, a \models \langle \mathsf{U} \rangle\varphi \iff \exists b \in A : \mathcal{M}, b \models \varphi$$

Boolean clauses are omitted. As usual, φ is valid in an attack model $\mathcal{M}$ iff it is satisfied in all pointed models of $\mathcal{M}$, i.e., $\mathcal{M} \models \varphi$. The truth-set of φ, i.e., $\{a \in A \mid \mathcal{M}, a \models \varphi\}$, is denoted $[\![\varphi]\!]_{\mathcal{M}}$. The set of $\mathcal{L}$-formulae which are true in the class $\mathfrak{A}$ of all attack models is called $\mathsf{K_U}$.

These are fully standard clauses for modal logic semantics, but let us see what their intuitive reading is in argumentation-theoretic terms. The first clause states that argument a belongs to the set called $\Diamond\varphi$ iff some argument b is reachable via the inverse of the attack relation and b belongs to φ or, more simply, iff a is attacked by some argument in φ. The second clause states that argument a belongs to the set called $\langle \mathsf{U} \rangle\varphi$ iff there exists some argument b in φ, in other words, iff the set called φ is non-empty. Similarly, for the duals: $\Box\varphi$ expresses that all attackers have property φ, and $[\mathsf{U}]\varphi$ expresses that all arguments have property φ.

Example 3. *Consider the graphs in Figure 1 and call the graph on the left $\mathcal{A}_L$ and the one on the right $\mathcal{A}_R$. Let then $\mathcal{V} : \{p\} \longrightarrow \wp(A)$ be a valuation function interpreting atom p on a set of arguments such that $\mathcal{V}(p) = \{b\}$. Here are a few illustrative modal statements:*

$$\langle \mathcal{A}_L, \mathcal{V} \rangle, b \models \neg\Box\bot \qquad \langle \mathcal{A}_R, \mathcal{V} \rangle, b \models \langle \mathsf{U} \rangle(\Diamond\Diamond p \wedge \Diamond\Diamond\Diamond p)$$
$$\langle \mathcal{A}_L, \mathcal{V} \rangle, b \models \Box\Diamond p \qquad \langle \mathcal{A}_R, \mathcal{V} \rangle, b \models \langle \mathsf{U} \rangle\Box(p \vee \Diamond p)$$

The two on the left state that b in $\mathcal{A}_L$ is not 'unattacked' and, respectively, that all its attackers are attacked by some argument labeled p (in this case b itself). The one at the top right corner states that in $\mathcal{A}_R$ there exists an argument (namely d) which has both a chain of two and three attackers ending in some argument labeled p. The last one states that there exists an argument (c) such that all its attackers (a and b) are either in p or are attacked by some argument in p.

Properties of the logic Logic $\mathsf{K_U}$ is a well-studied and well-behaved system: it has a simple strongly complete axiomatics, a polynomial model checking problem and an EXPTIME-complete satisfiability problem (cf. (Blackburn, de Rijke, and Venema 2001, Ch. 7)).

Defense and neutrality in modal logic

We show that functions $\mathtt{d}_{\mathcal{A}}$ and $\mathtt{n}_{\mathcal{A}}$ correspond to the functions denoted in $\mathcal{L}$ by the modal expressions $\Box\Diamond$ and, respectively, $\neg\Diamond$ on a given graph $\mathcal{A}$.[2]

Lemma 1 (Defence and neutrality in modal logic). *Let $\mathcal{A}$ be an attack graph and $\mathcal{V}$ a valuation function.*

$$\langle \mathcal{A}, \mathcal{V} \rangle, a \models \Box\Diamond\varphi \iff a \in \mathtt{d}_{\mathcal{A}}([\![\varphi]\!]_{\langle \mathcal{A}, \mathcal{V} \rangle})$$
$$\langle \mathcal{A}, \mathcal{V} \rangle, a \models \neg\Diamond\varphi \iff a \in \mathtt{n}_{\mathcal{A}}([\![\varphi]\!]_{\langle \mathcal{A}, \mathcal{V} \rangle})$$

[2]Technically, the claim is a direct consequence of the existence of a homomorphism from the term algebra $\mathsf{Term} = \langle \mathcal{L}, \wedge, \neg, \bot, \Diamond \rangle$ of language $\mathcal{L}$ (without universal modality) to the complex algebra $\mathsf{Set}_{\mathcal{A}} = \langle 2^A, \cap, -, \emptyset, f \rangle$ where $f : \wp(A) \longrightarrow \wp(A)$ such that $f(A) = \{a \in A \mid \exists b \in A : a \leftarrow b\}$ (Blackburn, de Rijke, and Venema 2001, Ch. 5). Nonetheless, we find it worthwhile to state it explicitly for $\Box\Diamond$- and $\neg\Diamond$-formulae.

Sketch of proof. For $\Box\Diamond$ we have these equivalences:

$$\begin{aligned}[\![\Box\Diamond\varphi]\!]_{\langle \mathcal{A},\mathcal{V}\rangle} &= \{a \mid \forall b: \text{ IF } a \leftarrow b \text{ THEN } b \leftarrow [\![\varphi]\!]_{\langle \mathcal{A},\mathcal{V}\rangle}\} \\ &= \mathtt{d}_{\mathcal{A}}([\![\varphi]\!]_{\langle \mathcal{A},\mathcal{V}\rangle}).\end{aligned}$$

The equations hold by the semantics of $\Box\Diamond$ and Definition 2. The reasoning for $\neg\Diamond\varphi$ is analogous. $\Box$

A direct consequence of this lemma is that $\mathcal{L}$ can express some of the properties of Table 1 in the following way: atom p denotes a set of arguments which is admissible (Formula 1 below), complete (Formula 2), and stable (Formula 3):[3]

$$[\mathsf{U}](p \to \Box\Diamond p) \wedge [\mathsf{U}](p \to \neg\Diamond p) \quad (1)$$

$$[\mathsf{U}](p \leftrightarrow \Box\Diamond p) \wedge [\mathsf{U}](p \to \neg\Diamond p) \quad (2)$$

$$[\mathsf{U}](p \leftrightarrow \neg\Diamond p) \quad (3)$$

Modal principles Emphasizing the modal nature of $\mathtt{d}_{\mathcal{A}}$ and $\mathtt{n}_{\mathcal{A}}$ has the advantage of allowing us to use available modal principles in our proofs without having to resort to longer direct arguments. All the theorems of logic K^{U} concerning $\Box\Diamond$- and $\neg\Diamond$-formulae can legitimately be seen as theorems of abstract argumentation. Here we list a sample of these theorems which will be used in the paper. They all express known fundamental properties of defense and neutrality functions, albeit in a more concise way. Proofs are standard and are omitted.

Fact 1. *The following are theorems of* K^{U}*:*

$$\Box\Diamond\varphi \leftrightarrow \neg\Diamond\neg\Diamond\varphi \quad (4)$$

$$\Box\Diamond\Box\Diamond\bot \leftrightarrow \Box\Diamond\bot \vee \Box\Diamond\Box\Diamond\bot \quad (5)$$

$$[\mathsf{U}](\varphi \to \psi) \to [\mathsf{U}](\Box\Diamond\varphi \to \Box\Diamond\psi) \quad (6)$$

$$[\mathsf{U}](\varphi \to \psi) \to [\mathsf{U}](\neg\Diamond\psi \to \neg\Diamond\varphi) \quad (7)$$

Formula (4) is the modal counterpart of the equivalence of the defense function and the 2-fold iteration of the neutrality function, i.e., for any X and graph $\mathcal{A}$: $\mathtt{d}_{\mathcal{A}}(X) = \mathtt{n}_{\mathcal{A}}(\mathtt{n}_{\mathcal{A}}(X))$. Formula (5) states that, for any $\mathcal{A}$, the finite union of subsequent iterations of $\mathtt{d}_{\mathcal{A}}$ over $\emptyset$ is equivalent to the longest iteration; Formula (6) expresses, for any $\mathcal{A}$, the monotonicity of $\mathtt{d}_{\mathcal{A}}$ while Formula (7) expresses the antitonicity of $\mathtt{n}_{\mathcal{A}}$.

In the remaining of the paper, in order to concisely express the n^{th} iteration of $\Box\Diamond$ (resp., $\neg\Diamond$) we will write $(\Box\Diamond)^n$ (resp., $(\neg\Diamond)^n$).

Dynamics of the grounded extension

The present section is devoted to a logical analysis of the process of computation of the grounded extension.

Computing the grounded extension

As we have seen in Table 1, the grounded extension is the least fixpoint of the defense function. By the monotonicity of the defense function (Fact 1), and the Knaster-Tarski theorem[4], we know that the grounded extension exists for any attack graph and is equal to the intersection of all pre-fixpoints of the defense function, i.e., $\bigcap\{X \subseteq A \mid \mathtt{d}_{\mathcal{A}}(X) \subseteq X\}$.

Our starting point here is a known result, due to (Dung 1995), according to which, under some specific conditions such as the finiteness of A, the grounded extension of an attack graph can be 'approximated from below', that is, it can be obtained as the limit of a sequence of iterations of the defense function starting at the empty set.[5]

The result relies on the following lemma expressing a general stabilization property of the streams generated by the defense function.

Lemma 2 (Stabilization). *Let $\mathcal{A} = \langle A, \rightarrow\rangle$ be an attack graph and $X \subseteq A$. If $X \subseteq \mathtt{d}_{\mathcal{A}}(X)$ then the stream generated by $\mathtt{d}^{\omega}_{\mathcal{A}}$ stabilizes.*

Sketch of proof. Function $\mathtt{d}_{\mathcal{A}}$ is monotonic (Fact 1) hence the stream generated by $\mathtt{d}^{\omega}_{\mathcal{A}}$ is increasing. By the finiteness of A (Definition 1), the stream cannot be strictly increasing, hence there exists a limit. $\Box$

Theorem 1 (Approximating the grounded extension). *For any attack graph $\mathcal{A}$:*

$$\mathsf{lfp}.\mathtt{d}_{\mathcal{A}} = \bigcup_{0 \le n < \omega} \mathtt{d}^n_{\mathcal{A}}(\emptyset)$$

Proof. By Lemma 2 the stream stabilizes at limit $\mathtt{d}^{\star}_{\mathcal{A}}(\emptyset)$. By the monotonicity of $\mathtt{d}_{\mathcal{A}}$ (Fact 1), and the finiteness assumption on A (Definition 1) we have that $\mathtt{d}^{\star}_{\mathcal{A}}(\emptyset) = \bigcup_{0\le n<\omega} \mathtt{d}^n_{\mathcal{A}}(\emptyset)$ since $\mathtt{d}_{\mathcal{A}}\left(\bigcup_{0\le n<\omega} \mathtt{d}^n_{\mathcal{A}}(\emptyset)\right) = \bigcup_{0\le n<\omega} \mathtt{d}^n_{\mathcal{A}}(\emptyset)$. Then, by monotonicity again, we conclude that $\mathtt{d}^{\star}_{\mathcal{A}}(\emptyset)$ is the smallest fixpoint.[6] $\Box$

Intuitively, the grounded extension is the set of arguments built in the limit of the process that starts by taking the set of arguments defended by the empty set, then adding the set of arguments that are defended by the set defended by the empty set, and so on.

Approximation and hard updates

Basic intuition The idea behind this section is that the approximation sequence computing the grounded extension can be viewed as a dual process of iterated removal of 'unfeasible' arguments. Intuitively, an agent confronted with an attack graph will at first consider all arguments as equally cogent. She will then engage in a step-wise process of scrutiny of the graph, which reduces her uncertainty about which arguments are to be considered justified by removing arguments that are clearly unjustified. The upshot is that, at the end, she will remain with only justified arguments.

[3]To express also the grounded and the preferred extensions richer modal languages are needed. See (Grossi 2010; 2011).

[4]See (Davey and Priestley 1990, Ch. 10).

[5]The condition in (Dung 1995) is weaker than finiteness, namely that no argument is attacked by infinitely many attackers. Here we limit ourselves to the version of the result for finite graphs, since the logical structure of the approximation sequence is the same in the finite or infinite case. This allows us to avoid some complications that would make proofs lengthier.

[6]In the case of infinite graphs the assumption that no argument is attacked by infinitely many attackers still guarantees that the limit of the stream $\mathtt{d}^{\omega}_{\mathcal{A}}(\emptyset)$ is a fixpoint, and hence the smallest.

The "being defended" formula In the case of the grounded extension this process of elimination concerns the iterated elimination of arguments satisfying the modal formula $\Diamond\Box\bot$, that is, arguments that are attacked by an unattacked argument.

This is, in essence, the same process used in game theory to compute equilibria via iterated elimination of strictly dominated strategies, which has been object of logical investigation in, amongst others, (van Benthem 2007). In general, processes of this type have been extensively studied in the last decade in that branch of epistemic logic that has come to be known as DEL (Dynamic Epistemic Logic, (van Ditmarsch, Kooi, and van der Hoek 2007)). The iterated elimination of $\Diamond\Box\bot$-arguments corresponds to what in DEL is called a *public announcement* or *hard update* of the negation of $\Diamond\Box\bot$, i.e., of $\Box\Diamond\top$ which we will abbreviate by $\mathtt{DF}$. The formula expresses the property that "all attackers (of the current evaluation point in the graph) are attacked" or, roughly, that "the current argument is defended by at least one argument". We will also abbreviate $\Box\bot$ by $\mathtt{UA}$ ('unattacked'). Arguments satisfying $\mathtt{UA}$ will be sometimes called dead ends, according to the standard modal logic terminology.

Remark 1 (Frame language). *Properties* $\mathtt{DF}$ *and* $\mathtt{UA}$ *are both expressible in a limited fragment of the language* $\mathcal{L}$ *introduced above. The fragment, which we call* $\mathcal{L}^{frame}$*, is defined by the following BNF:*

$$\varphi ::= \bot \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Diamond\varphi$$

Notice that this is a so-called frame language *(Blackburn, de Rijke, and Venema 2001, Ch. 3.1), which does not use propositional atoms. In fact this language does not need models to be interpreted, but simply attack graphs (Definition 1). In the current section we will work only with this fragment, coming back to* $\mathcal{L}$ *in the last two sections dedicated to complete and stable extensions. This might look like a radical restriction, but we will see that this is really all we need for a dynamic analysis of the grounded extension.*

Hard updates The intuition behind the hard update of a formula φ over an attack graph is that, after φ has been publicly announced (or learnt), all arguments satisfying the negation of φ will be removed from the graph as deemed unplausible. Here is the formal definition:

Definition 6 (Hard updates and their iteration). *Let* $\varphi \in \mathcal{L}^{frame}$*. The hard update by* φ *is a function* $!\varphi : \{\mathcal{A} \in \mathfrak{A} \mid [\![\varphi]\!]_{\mathcal{A}} \neq \emptyset\} \longrightarrow \mathfrak{A}$*. The value of* $!\varphi$ *given* $\mathcal{A}$ *is* $\mathcal{A}_{!\varphi} = \langle A_{!\varphi}, \rightarrow_{!\varphi}\rangle$ *where:*

- $A_{!\varphi} := [\![\varphi]\!]_{\mathcal{A}}$*, i.e., the new set of arguments are the ones that satisfy the announced formula;*
- $\rightarrow_{!\varphi} := \rightarrow \cap A^2_{!\varphi}$*, i.e., the restriction of the attack relation to* $A_{!\varphi}$*.*

When defined, we denote the n*-fold finite iteration of a hard update* $!\varphi$ *by* $!\varphi^n$ *and its infinite iteration by* $!\varphi^\omega$*. An infinite iteration* $!\varphi^\omega$ *generates a stream of graphs:* $\mathcal{A}_{!\varphi^0}, \mathcal{A}_{!\varphi^1}, \mathcal{A}_{!\varphi^2}, \ldots$*. A stream is said to* stabilize *if and only if there exists* $1 \leq n < \omega$ *such that* $\mathcal{A}_{!\varphi^n} = \mathcal{A}_{!\varphi^{n+1}}$*.* $\mathcal{A}_{!\varphi^n}$ *is then called the* limit *of the stream and denoted* $\mathcal{A}_{!\varphi^\star}$*.*

So, the hard update of a formula φ transforms a given attack graph $\mathcal{A}$ in which φ has a non-empty denotation, into a new graph $\mathcal{A}'$ where all arguments that did not satisfy φ in $\mathcal{A}$ have been eliminated. Such updates can be iterated giving rise to streams of attack graphs which stabilize when no further hard update by φ changes the graph any more.

Hard update of 'being defended' We study here some relevant properties of $\mathtt{DF}$ and its hard update $!\mathtt{DF}$.

Fact 2. *For any attack graph* $\mathcal{A}$*: i) if* $\mathcal{A}, a \models \mathtt{UA}$ *then* $\mathcal{A}, a \models \mathtt{DF}$*; ii)* $[\![\mathtt{DF}]\!]_{\mathcal{A}} \neq \emptyset$*; iii)* $!\mathtt{DF}$ *is a function* $!\mathtt{DF} : \mathfrak{A} \longrightarrow \mathfrak{A}$*; iv)* $\mathtt{d}_{\mathcal{A}_{!\mathtt{DF}^n}}(X) = \mathtt{d}_{\mathcal{A}}(X)$ *where* $X \subseteq A_{\mathcal{A}_{!\mathtt{DF}^n}}$ *for any* $0 \leq n < \omega$*. v)* $[\![\mathtt{DF}]\!]_{\mathcal{A}_{!\mathtt{DF}}} = [\![\Box\Diamond\mathtt{DF}]\!]_{\mathcal{A}}$*; vi)* $[\![\mathtt{UA}]\!]_{\mathcal{A}_{!\mathtt{DF}}} = [\![\Box\Diamond\mathtt{UA}]\!]_{\mathcal{A}}$*.*

Sketch of proof. i) If a is a dead-end, then it trivially satisfies $\Box\Diamond\top$. ii) Suppose towards a contradiction that for all a, $\mathcal{A}, a \models \neg\Box\Diamond\top$ and, by modal principles, for all a, $\mathcal{A}, a \models \Diamond\Box\bot$. Hence $\exists b$ such that $\mathcal{A}, b \models \Box\bot$. But by i) it follows that $[\![\mathtt{UA}]\!] = \emptyset$. Contradiction. iii) follows directly from ii) as $\{\mathcal{A} \in \mathfrak{A} \mid [\![\mathtt{DF}]\!]_{\mathcal{A}} \neq \emptyset\} = \mathfrak{A}$. iv) is proven by a simple induction on the length n of the iteration of the hard update; v) is proven by the following series of equations:

$$\begin{aligned} [\![\Box\Diamond\top]\!]_{\mathcal{A}_{!\Box\Diamond\top}} &= \mathtt{d}_{\mathcal{A}_{!\Box\Diamond\top}}([\![\top]\!]_{\mathcal{A}_{!\Box\Diamond\top}}) \\ &= \mathtt{d}_{\mathcal{A}}([\![\Box\Diamond\top]\!]_{\mathcal{A}}) \\ &= [\![\Box\Diamond\Box\Diamond\top]\!]_{\mathcal{A}} \end{aligned}$$

The first equation holds by Lemma 1, the second by Definition 6 and item iv, and the third again by Lemma 1. vi) is proven in a similar fashion. □

These are the intuitive readings of the above properties: i) if an argument has no attackers, than it is (trivially) defended; ii) in every attack graph there exist arguments that are defended by some argument; iii) the hard update of 'being defended' is a total function; iv) the defense function after n hard updates coincides with the defense function of the original graph restricted to the subsets of the domain at the n^{th} update; v) and vi) the arguments that are defended (resp., unattacked) after update $!\mathtt{DF}$, are the same arguments that are acceptable with respect to the defended (resp. unattacked) arguments in the original graph.

Remark 2. *In the DEL jargon,* $\mathtt{DF}$ *is what is sometimes called a successful formula, that is, a formula* φ *for which it holds that: for any pointed frame* $(\mathcal{A}, a)$*, if* $\mathcal{A}, a \models \varphi \wedge \Diamond\varphi$ *then* $\mathcal{A}_{!\varphi}, a \models \varphi$ *(Holliday and Icard 2010). However,* $\mathtt{DF}$ *can change its truth value from true to false after an update* $!\mathtt{DF}$*. This is the case of argument* b *in Figure 2.*

Grounded extension via iterated update

We can now present a characterization of the grounded extension via the iteration of hard update $!\mathtt{DF}$. First, the stream generated by this update always stabilizes:

Lemma 3. *Function* $!\mathtt{DF}$ *stabilizes for any attack graph.*

Sketch of proof. First observe that the stream of sets $A_{!\mathtt{DF}^0}, A_{!\mathtt{DF}^1}, \ldots$ generated by $\mathcal{A}_{!\mathtt{DF}^\omega}$ is decreasing since $A_{!\Box\Diamond\top} = A - [\![\neg\Box\Diamond\top]\!]_{\mathcal{A}}$ (Definition 6). However, by the finiteness assumption on A, the stream cannot be strictly decreasing, hence there exists a limit. □

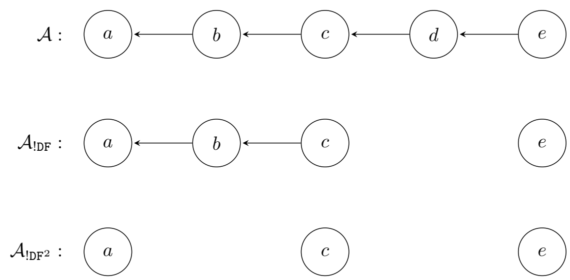


Figure 2: Example of iterated hard update of !DF. After the first update by !DF argument d is eliminated since $\mathcal{A}, d \models \neg\mathtt{DF}$. Similarly, after the second update b is eliminated since $\mathcal{A}_{!\mathtt{DF}}, b \models \neg\mathtt{DF}$.

It is instructive to notice that the stream may stabilize trivially and without any argument removal, e.g., in Figure 1 where the depicted graphs are already limits of the stream. This is in general the case whenever the graph does not contain any dead end. We now get to the theorem we are after.

Theorem 2 (The grounded extension via iterated updates). *For any graph $\mathcal{A} = \langle A, \rightarrow \rangle$:*

$$\mathsf{lfp.d}_{\mathcal{A}} \quad = \quad [\![\mathtt{UA}]\!]_{\mathcal{A}_{!\mathtt{DF}^\star}}$$

Proof. We first prove the following lemma:

$$[\![\bigvee_{0<i\leq n} (\Box\Diamond)^i \bot]\!]_{\mathcal{A}} \quad = \quad [\![\mathtt{UA}]\!]_{\mathcal{A}_{!\mathtt{DF}^{n-1}}}$$

By induction on the length of the $\Box\Diamond$-iteration:
B: For $n = 1$: $[\![\Box\bot]\!]_{\mathcal{A}} = [\![\Box\bot]\!]_{\mathcal{A}}$.
S: Assume (IH) that $[\![\bigvee_{0<i\leq n}(\Box\Diamond)^i\bot]\!]_{\mathcal{A}} = [\![\Box\bot]\!]_{\mathcal{A}_{!\mathtt{DF}^{n-1}}}$. We show that $[\![\bigvee_{0<i\leq n+1}(\Box\Diamond)^i\bot]\!]_{\mathcal{A}} = [\![\Box\bot]\!]_{\mathcal{A}_{!\mathtt{DF}^n}}$. The claim is proven by the following series of equations:

$$\begin{aligned}
[\![\Box\bot]\!]_{\mathcal{A}_{!\mathtt{DF}^n}} &= [\![\Box\Diamond\Box\bot]\!]_{\mathcal{A}_{!\mathtt{DF}^{n-1}}} \\
&= \mathtt{d}_{\mathcal{A}_{!\mathtt{DF}^{n-1}}}([\![\Box\bot]\!]_{\mathcal{A}_{!\mathtt{DF}^{n-1}}}) \\
&= \mathtt{d}_{\mathcal{A}_{!\mathtt{DF}^{n-1}}}([\![\bigvee_{0<i\leq n} (\Box\Diamond)^i\bot]\!]_{\mathcal{A}}) \\
&= \mathtt{d}_{\mathcal{A}_{!\mathtt{DF}^{n-1}}}([\![(\Box\Diamond)^n\bot]\!]_{\mathcal{A}}) \\
&= [\![(\Box\Diamond)^{n+1}\bot]\!]_{\mathcal{A}} \\
&= [\![\bigvee_{0<i\leq n+1} (\Box\Diamond)^{n+1}\bot]\!]_{\mathcal{A}}
\end{aligned}$$

The first equation holds by Fact 2, the second by Lemma 1, the third by IH, the fourth and last by modal principles (Fact 1), and the fifth by Lemma 1 and Fact 2. By Lemma 1 and Theorem 1 we then get the following series of equations:

$$[\![\mathtt{UA}]\!]_{\mathcal{A}_{!\mathtt{DF}^\star}} = \bigcup_{0\leq n<\omega} [\![(\Box\Diamond)^n\bot]\!]_{\mathcal{A}} = \bigcup_{0\leq n<\omega} \mathtt{d}^n_{\mathcal{A}}(\emptyset) = \mathsf{lfp.d}_{\mathcal{A}}$$

thereby completing the proof. □

Intuitively, the theorem states that the grounded extension coincides with the unattacked arguments in the graph obtained at the limit of the iteration of the !DF update. In other words: $a \in \mathsf{lfp.d}_{\mathcal{A}}$ if and only if $\mathcal{A}_{!\mathtt{DF}^\star}, a \models \mathtt{UA}$. Figure 2 illustrates this process on a simple graph.

Remark 3 (Largest fixpoint of the defense function). *The largest fixpoint of the defense function can be obtained in the same fashion, and it coincides with the whole domain of the limit graph: $A_{!\mathtt{DF}^\star} = \mathsf{gfp.d}_{\mathcal{A}}$.*

On the epistemic interpretation of !DF

The intuitions that drove our analysis in this section have been mainly epistemic: an agent confronted with an attack graph engages in a process of reduction of her uncertainty concerning which arguments are to be considered justified. Although it is not our aim here to provide a fully fledged rendering of our results in epistemic logic it is worth spending a few words on the issue.

Logic K^{U} can be viewed as an extension of an epistemic S5 system with the attack modality $\Diamond$, where the universal modality $\langle\mathsf{U}\rangle$ is taken to be the epistemic operator. The interaction axiom between the two modalities is the inclusion principle: $\Diamond\varphi \rightarrow \langle\mathsf{U}\rangle\varphi$. So the update !DF reduces the range of the $\langle\mathsf{U}\rangle$ modality, thereby modelling the acquisition of knowledge. In this view, Theorem 2 has also the following interesting consequence: $a \in \mathsf{lfp.d}_{\mathcal{A}}$ if and only if $\mathcal{A}_{!\mathtt{DF}^\star}, a \models [\mathsf{U}]\mathtt{UA}$. That is, an argument belongs to the grounded extension if and only if, in the announcement limit, *it is known* that there are no attackers.

We conclude with one DEL-related remark. Although we have used the notion of hard update, we have not extended the language of K^{U} to talk about such operations, viz. modalities $[!\varphi]$ for φ in the extended language. An extension of this type is obviously possible and an axiomatization of these operators could be obtained, as usual in DEL, via reduction axioms. All we have to do is to extend one of the existing complete axiomatizations of Public Announcement Logic (see (Wang 2011) for a recent overview) with axiom:

$$[!\varphi]\Box\psi \quad \leftrightarrow \quad (\varphi \rightarrow \Box(\varphi \rightarrow [!\varphi]\psi)$$

Roughly, the axiom states that $\Box\psi$ holds after the hard update $!\varphi$ if and only if, if the current argument satisfies φ then all attackers satisfying φ will also satisfy ψ after the update. The soundness of this axiom with respect to models on attack graphs is easily proven.

Dynamics of complete extensions

We turn now to a similar analysis, but with different logical tools, of the process of computation of complete extensions.

Fixpoint computation of complete extensions

The section moves from a slight generalization of Theorem 1, which accounts for the computation via approximation sequences of any complete extension—hence including the grounded. Although applying the very same ideas behind the approximation sequence for the grounded extension, this result has, to the best of our knowledge, never been reported in the literature.

We start with a lemma that relates admissibility of a set of arguments to the generation of a conflict-free fixpoint of the defense function as the limit of a stream starting at that set:

Lemma 4. *Let $\mathcal{A} = \langle A, \rightarrow \rangle$ be an attack graph and $X \subseteq A$. If X is admissible then:*

i) $\mathtt{d}^\star_{\mathcal{A}}(X)$ *is the smallest fixpoint of* $\mathtt{d}_{\mathcal{A}}$ *containing* X*;*

ii) $\mathtt{d}^\star_{\mathcal{A}}(X)$ *is conflict-free.*

Sketch of proof. i) By Lemma 2 we know that the limit exists. To prove that $\mathtt{d}^\star_{\mathcal{A}}(X)$ is the smallest fixpoint of $\mathtt{d}_{\mathcal{A}}$ containing X, assume towards a contradiction that there exists Y s.t. $X \subseteq Y = \mathtt{d}_{\mathcal{A}}(Y) \subset \mathtt{d}^\star_{\mathcal{A}}(X)$. By the monotonicity of $\mathtt{d}_{\mathcal{A}}$, $\mathtt{d}^n_{\mathcal{A}}(X) \subseteq Y$ for all $0 \leq n < \omega$. Hence $\mathtt{d}^\star_{\mathcal{A}}(X) \subseteq Y$, against the assumption. ii) Assume towards a contradiction that $\mathtt{d}^\star_{\mathcal{A}}$ is not conflict-free. By the finiteness assumption over A (Definition 1), and the assumption that X is admissible, it follows that there exists a $0 \leq n < \omega$ such that $\mathtt{d}^{n-1}_{\mathcal{A}}(X)$ is conflict-free and $\mathtt{d}^n_{\mathcal{A}}(X)$ is not. Let $a, b \in \mathtt{d}^n_{\mathcal{A}}(X) = \mathtt{d}_{\mathcal{A}}(\mathtt{d}^{n-1}_{\mathcal{A}}(X))$ such that $a \leftarrow b$. Hence $\exists c, d \in \mathtt{d}^{n-1}_{\mathcal{A}}(X)$ such that $b \leftarrow c$ and $c \leftarrow d$, contradicting the conflict-freeness of $\mathtt{d}^{n-1}_{\mathcal{A}}(X)$.[7] □

Theorem 3 (Approximating complete extensions). *Let* $\mathcal{A} = \langle A, \rightarrow \rangle$ *be an attack graph and* $X \subseteq A$ *be admissible:*

$$\mathtt{cmp}_{\mathcal{A}}(X) \quad = \quad \bigcup_{0 \leq n < \omega} \mathtt{d}^n_{\mathcal{A}}(X)$$

where $\mathtt{cmp}_{\mathcal{A}}(X)$ *denotes the smallest complete extension of* $\mathcal{A}$ *containing* X.

Proof. Like in for Theorem 1, the finiteness of A guarantees that $\mathtt{d}^\star_{\mathcal{A}}(X) = \bigcup_{0 \leq n < \omega} \mathtt{d}^n_{\mathcal{A}}(X)$. The fact that $\mathtt{d}^\star_{\mathcal{A}}(X) = \mathtt{cmp}_{\mathcal{A}}(X)$ follows then directly from Lemma 4. □

Intuitively, starting with an admissible set, we can obtain the smallest complete extension containing it by a finite iteration of the defense function.

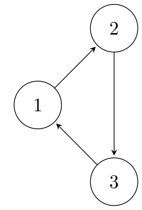
Remark 4 (Preferred extensions). *Notice that, under some specific assumptions on* X*,* $\mathtt{d}^\star_{\mathcal{A}}(X)$ *generates the preferred extension containing* X*. Such an assumption is that* X *is 'big enough' in the precise sense that it contains enough arguments to be able, from some argument in* X*, to reach any argument in the graph via the attack relation, i.e., if* $A = \{a \mid \exists b \in X \text{ S.T. } a \leftarrow^+ b\}$*, where* $\leftarrow^+$ *denotes the transitive closure of the* $\leftarrow$ *relation.*

Approximation and valuation updates

Basic intuition Notice, first of all, that Theorem 2 is the special case of Theorem 3 where $X = \emptyset$. The process of approximation at issue in both these theorems starts with a given admissible set and iteratively extends it to incorporate the arguments it defends. However, to analyze Theorem 2 we have resorted to a dual process of elimination of unjustified arguments which we have modeled via hard updates.

In this section, we analyze the more general Theorem 3 not through argument elimination, but rather directly through an iterated reinterpretation of a designated atom P of our language. We call it 'position atom' as we take it to denote a position in an argumentation, i.e., a designated set of arguments to which an agent commits during a step-wise process of solution of the attack graph.

[7] It might be instructive to mention that this is also, in essence, the argument behind the *fundamental lemma* in (Dung 1995).



$$\begin{array}{lcl} \mathcal{V}_{:=\mathtt{DF}^0_{\mathsf{P}}}(\mathsf{P}) & = & \{1\} \\ \mathcal{V}_{:=\mathtt{DF}^1_{\mathsf{P}}}(\mathsf{P}) & = & \{3\} \\ \mathcal{V}_{:=\mathtt{DF}^2_{\mathsf{P}}}(\mathsf{P}) & = & \{2\} \\ \mathcal{V}_{:=\mathtt{DF}^3_{\mathsf{P}}}(\mathsf{P}) & = & \{1\} \\ \ldots & & \ldots \end{array}$$

Figure 3: A 3-cycle attack graph, and the first four elements of the stream $:= \mathtt{DF}^\omega_{\mathsf{P}}$ generated at the model whose valuation is $\mathcal{V}(\mathsf{P}) = \{1\}$. The stream cycles every three steps.

Valuation update The tool we borrow from DEL for our analysis consists in a model-change operation which targets the valuation function, called valuation update or propositional change (e.g., (van Benthem, van Eijck, and Kooi 2006)). Here we focus on a version of this operation that only modifies the valuation of a designated atom P.

Definition 7 (Valuation update). *Let* $\mathcal{M} = \langle A, \rightarrow, \mathcal{V} \rangle$ *be an attack model and* $\varphi \in \mathcal{L}^{\mathsf{U}}$*. The valuation update* $:= \varphi$ *is a function* $:= \varphi : \mathfrak{M} \longrightarrow \mathfrak{M}$*. The value of* $:= \varphi$ *given* $\mathcal{M}$ *is* $\mathcal{M}_{:=\varphi} = \langle \mathcal{A}_{:=\varphi}, \mathcal{V}_{:=\varphi} \rangle$ *where:*

- $A_{:=\varphi} = A$ *and* $\rightarrow_{:=\varphi} = \rightarrow$*, i.e., the set of arguments and the attack relation remain the same;*
- $\mathcal{V}_{:=\varphi}$ *is such that* $\mathcal{V}_{:=\varphi}(\mathsf{P}) = [\![\varphi]\!]_{\mathcal{M}}$*, i.e., the designated atom* P *is given as value the truth-set of* φ *in the model before the update.*

Iteration, streams, stabilization and limits are defined just like for the case of hard updates (Definition 6).

Iterated valuation update of "being defended by P**"** The natural candidate for the property driving the update iteration is expressed by formula $\Box\Diamond\mathsf{P}$, which we will abbreviate by $\mathtt{DF}_{\mathsf{P}}$, and which denotes the set of arguments defended by the arguments in position P. The update we will focus on is, therefore, $:= \mathtt{DF}_{\mathsf{P}}$.

From a DEL point of view, the first thing to notice by looking at the stream generated by $:= \mathtt{DF}^\omega_{\mathsf{P}}$ is that the stream does not stabilize in general. Figure 3 provides a simple example, well-known to the argumentation theorists as typical instance of an odd loop of attacks.[8] However, our modal language has the necessary resources to recast Lemma 2 obtaining a stabilization result and, consequently, a characterization of complete extensions via iterated valuation updates.

Theorem 4 (Complete extensions via iterated updates). *For any model* $\mathcal{M}$*, if* $\mathcal{M} \models (\mathsf{P} \rightarrow \Box\Diamond\mathsf{P}) \wedge (\mathsf{P} \rightarrow \neg\Diamond\mathsf{P})$ *then:*

$$\mathtt{cmp}_{\mathcal{A}}([\![\mathsf{P}]\!]_{\mathcal{M}}) \quad = \quad [\![\mathsf{P}]\!]_{\mathcal{M}_{:=\mathtt{DF}^\star_{\mathsf{P}}}}$$

Sketch of proof. Observe first of all that $\mathcal{M}_{:=\mathtt{DF}^\omega_{\mathsf{P}}}$ generates the stream $[\![\mathsf{P}]\!]_{\mathcal{M}_{:=\mathtt{DF}^0_{\mathsf{P}}}}, [\![\mathsf{P}]\!]_{\mathcal{M}_{:=\mathtt{DF}^1_{\mathsf{P}}}}, \ldots$. This stream, by Lemma 1 and Definition 7, is identical to the stream

[8] Recent work in DEL has dedicated quite some attention to updates that do not necessarily stabilize (e.g. (Baltag and Smets 2009)). To the best of our knowledge, however, no work in that literature has studied oscillatory behavior of valuation updates.

$\mathtt{d}^0_{\mathcal{A}}([\![\mathsf{P}]\!]_{\mathcal{M}}), \mathtt{d}^1_{\mathcal{A}}([\![\mathsf{P}]\!]_{\mathcal{M}}), \ldots$. Hence the stream reaches a limit by Lemma 2. The claim follows then from a direct application of Theorem 3 and Lemma 1. □

From this it directly follows that $a \in \mathtt{cmp}_{\mathcal{A}}([\![\mathsf{P}]\!]_{\mathcal{M}})$ if and only if $\mathcal{M}_{:=\mathtt{DF}^*_{\mathsf{P}}}, a \models \mathsf{P} \wedge [\mathsf{U}](\mathsf{P} \rightarrow \neg\Diamond\mathsf{P}) \wedge [\mathsf{U}](\mathsf{P} \leftrightarrow \Box\Diamond\mathsf{P})$. Intuitively, argument a belongs to the smallest complete extension containing the truth-set of P if and only if the truth-set of P in limit of the valuation update by $\Box\Diamond\mathsf{P}$ is a complete extension and a belongs to it.

On the epistemic interpretation of $:= \mathtt{DF}_{\mathsf{P}}$

The analysis proposed makes use of a designated atom P. We have chosen this formal set up for conciseness. However, there are strict relationships with the literature in epistemic logic. Roughly, once we cast the valuation of P into a designated subset of the support of the frame, thus working with frames $\langle A, \mathsf{P}, \rightarrow\rangle$,[9] formulae of the form $[\mathsf{U}](\mathsf{P} \rightarrow \varphi)$ simulate a modal operator $[\mathsf{P}]\varphi$ with the following semantics:

$$\mathcal{M}, a \models [\mathsf{P}]\varphi \iff \forall b \in \mathsf{P} : \mathcal{M}, b \models \varphi$$

This operator has been proposed for instance in (Meyer and van der Hoek 1995), in order to extend S5 with an operator modeling a "working belief", i.e., a set of assumptions an agent can entertain before evaluating a formula. The same operator has also been used in (Grossi, Meyer, and Dignum 2006) to give a simple modal rendering of a notion of context. Finally, the set P can be viewed as an accessibility relation R_{P} such that: $aR_{\mathsf{P}}b$ iff $b \in \mathsf{P}$. It is easy to see that this is a transitive and euclidean relation (with the extra constraint that the set of accessible states does not vary across the model). Hence, these are models of logic K45, the subsystem of the basic doxastic logic KD45 allowing for possibly inconsistent beliefs.[10]

In this perspective, P can legitimately be viewed as modeling one's belief within an argumentation in terms of the arguments she is ready to accept. Consequently, the update $:= \mathtt{DF}_{\mathsf{P}}$ can be viewed as the expansion of one's argumentative position as a process of inference of what follows from a position, where 'to follow' does not denote a logical consequence relation, but the argumentation theoretic relation of defense. Implicit in the definition of complete extension is, therefore, that one must accept what one can defend.

We conclude with a few words on the dynamic operator $[:= \varphi]$ that can be associated to the $:= \varphi$ update. The extension of K^{U} with these operators can be axiomatized via the following simple axioms:[11]

$$\langle := \varphi\rangle\Diamond\psi \quad \leftrightarrow \quad \Diamond\langle := \varphi\rangle\psi$$
$$\langle := \varphi\rangle\langle \mathsf{U}\rangle\psi \quad \leftrightarrow \quad \langle \mathsf{U}\rangle\langle := \varphi\rangle\psi$$

[9] Slightly abusing notation we take P here to be a subset of A rather than an atom.

[10] A detailed study of the relations between [P] and the modal operator of KD45 is carried out in (Grossi, Meyer, and Dignum 2008), where it is shown that the logic of [P] is, in fact, K45.

[11] An equivalent system extending the context logic of (Grossi, Meyer, and Dignum 2006) with dynamic operators of this sort has been studied in (Aucher et al. 2009).

The dynamics of stable extensions

In this final section we will focus on the issues arising when attempting a computation of stable extensions via approximation sequences.

Computing stable extensions

In studying complete extensions we saw that, although not stabilizing in general (recall Figure 3), $\mathtt{d}^{\omega}_{\mathcal{A}}(X)$ would generate a stabilizing stream on any graph $\mathcal{A}$ provided X be self-acceptable, i.e, be a post-fixpoint of $\mathtt{d}_{\mathcal{A}}$: $X \subseteq \mathtt{d}_{\mathcal{A}}(X)$ (Lemma 2). This good behavior relied on the monotonicity of the defense function.

Monotonicity is, however, not satisfied by $\mathtt{n}_{\mathcal{A}}$, which is instead antitone (recall Fact 1). Besides, even initiating an approximation sequence at a post-fixpoint of $\mathtt{n}_{\mathcal{A}}$ (i.e., a conflict-free set) will not, in general, yield a stabilizing stream (see example below). The question arises then of which conditions can still generate some form of well-behaved stream.

Example 4. *Consider again the graph in Figure 3. The stream generated by* $\mathtt{n}^{\omega}_{\mathcal{A}}$ *starting at the conflict-free set* $\{1\}$ *is* $\{1\}, \{1,2\}, \{2\}, \{2,3\}, \{3\}, \{1,3\}, \{1\}, \ldots$ *and loops with an orbit containing all the sets of arguments of cardinality one and two.*

Decomposition, stabilization and limits The interesting result, inspired by work on logical semantics for self-referential statements (Yablo 1984), is that if X is taken to be a post-fixpoint of both $\mathtt{d}_{\mathcal{A}}$ and $\mathtt{n}_{\mathcal{A}}$ then the stream generated by $\mathtt{n}^{\omega}_{\mathcal{A}}(X)$, although not always stabilizing, does exhibit some good behavior:

Lemma 5. *For any attack graph* $\mathcal{A} = \langle A, \rightarrow\rangle$ *and* $X \subseteq A$, *if* X *is admissible then:*

i) the stream $\mathtt{n}^0_{\mathcal{A}}(X), \mathtt{n}^2_{\mathcal{A}}(X), \ldots, \mathtt{n}^{2n}_{\mathcal{A}}(X), \ldots$ *for* $0 \leq n < \omega$ *stabilizes;*

ii) the stream $\mathtt{n}^1_{\mathcal{A}}(X), \mathtt{n}^3_{\mathcal{A}}(X), \ldots, \mathtt{n}^{2n+1}_{\mathcal{A}}(X), \ldots$ *for* $0 \leq n < \omega$ *stabilizes.*

Proof. Recall that $\mathtt{d}_{\mathcal{A}}(X) = \mathtt{n}^2_{\mathcal{A}}(X)$ for any $\mathcal{A}$ and $X \subseteq A$ (Fact 1). i) The stream of even iterations is stream $\mathtt{d}^{\omega}_{\mathcal{A}}(X)$ and hence, by Lemma 2, it stabilizes. ii) The stream of odd iterations is stream $\mathtt{d}^{\omega}_{\mathcal{A}}(\mathtt{n}_{\mathcal{A}}(X))$. By a simple induction we can prove that the stream is decreasing, since $\mathtt{n}_{\mathcal{A}}$ is antitone (Fact 1). The base case $\mathtt{n}^3_{\mathcal{A}} \subseteq \mathtt{n}_{\mathcal{A}}(X)$ follows from X being such that $X \subseteq \mathtt{n}^2_{\mathcal{A}}$. The claim follows by A's finiteness. □

The proof makes explicit several interesting features of the stream generated by $\mathtt{n}^{\omega}_{\mathcal{A}}$ on an admissible set X: first, the stream can be split in two parts, the part consisting of even and, respectively, odd iterations of $\mathtt{n}_{\mathcal{A}}$; second, the two parts grow towards each other as the stream of even iterations is increasing, while the one of odd iterations is decreasing; third, the two streams can actually be viewed as streams of the defense function $\mathtt{d}_{\mathcal{A}}$ applied to X and to $\mathtt{n}_{\mathcal{A}}(X)$; finally—and that is the consequence of these observations, both parts stabilize. These features are depicted in Figure 4.

Based on this lemma, we can identify the limits of the stream as special fixpoints of the defense function.

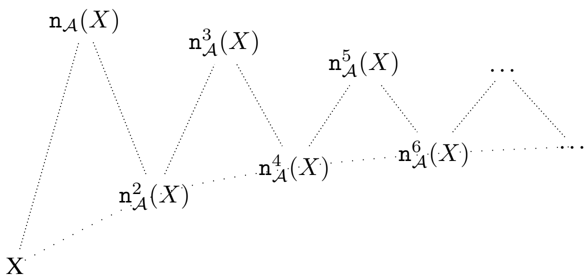


Figure 4: Decomposability of the stream generated by $n^{\omega}_{\mathcal{A}}(X)$. The horizontal axis indicates the number of iterations, while the vertical axis indicates set theoretic inclusion.

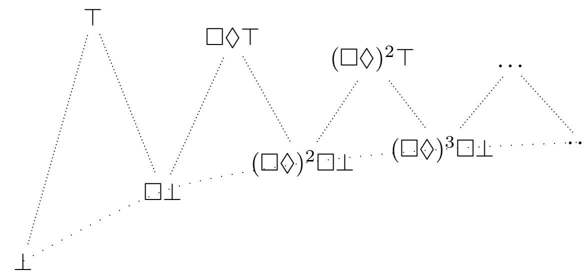


Figure 5: Decomposability of the stream $:= \neg\Diamond\bot^{\omega}$.

Theorem 5. *For any attack graph $\mathcal{A} = \langle A, \rightarrow\rangle$ and $X \subseteq A$, if X is admissible then:*

i) the limit $(n^2_{\mathcal{A}})^{\star}(X)$ of the stream of even iterations of $n_{\mathcal{A}}$ is the smallest fixpoint of $d_{\mathcal{A}}$ containing X;

ii) the limit $(n^2_{\mathcal{A}})^{\star}(n_{\mathcal{A}}(X))$ of the stream of odd iterations $n_{\mathcal{A}}$ is the largest fixpoint of $d_{\mathcal{A}}$ containing X.

Proof. i) The claim follows directly from Lemma 4 and the fact that $n_{\mathcal{A}}(n_{\mathcal{A}}(X)) = d_{\mathcal{A}}(X)$ (Fact 1). ii) By Fact 1 and the finiteness of A we have that: $(n^2_{\mathcal{A}})^{\star}(n_{\mathcal{A}}(X)) = \bigcap_{0\leq n<\omega} d^n_{\mathcal{A}}(n_{\mathcal{A}}(X))$. □

Stable extensions We finally arrive at a characterization of stable extensions as limits of streams generated by the neutrality function:

Theorem 6 (Approximating stable extensions). *For any attack graph $\mathcal{A} = \langle A, \rightarrow\rangle$ and $X \subseteq A$, if X is admissible then:*

$$\mathtt{stb}_{\mathcal{A}}(X) = n^{\star}_{\mathcal{A}}(X) \iff d^{\star}_{\mathcal{A}}(X) = d^{\star}_{\mathcal{A}}(n_{\mathcal{A}}(X))$$

where $\mathtt{stb}_{\mathcal{A}}(X)$ denotes the unique stable extension containing X.

Proof. [RIGHT TO LEFT] By the finiteness of A, from $d^{\star}_{\mathcal{A}}(X) = d^{\star}_{\mathcal{A}}(n_{\mathcal{A}}(X))$ follows that $\bigcap_{0\leq n<\omega} d^n_{\mathcal{A}}(n_{\mathcal{A}}(X)) = \bigcup_{0\leq n<\omega} d^n_{\mathcal{A}}(X)$ hence there exists one unique limit of the stream generated by $n^{\omega}_{\mathcal{A}}$ which equals the only fixpoint of $n_{\mathcal{A}}$ containing X. [LEFT TO RIGHT] Straightforward. □

The theorem provides a novel characterisation of stable extensions in terms of the behavior in the limit of the iteration of the neutrality function. It states that, by starting with an admissible set, the infinite iteration of the neutrality function will reach a fixpoint, viz., the stable extension containing the original set, if and only if the streams of even and odd iterations of the function converge to the same limit. Two remarks on this theorem are in order:

Remark 5 (Preferred extensions (continued)). *An interesting consequence of the theorem is that stable extensions can be approximated from below starting from admissible sets if and only if they actually coincide with a preferred extension. This is so because the limits of the even and odd streams in $n^{\omega}_{\mathcal{A}}$ coincide only if there exists only one fixpoint of $d_{\mathcal{A}}$ containing the admissible set, and hence that fixpoint must be a preferred extension. Notice that this convergence condition is implied by the condition introduced in Remark 4.*

Remark 6 (A special case). *The case of Theorem 6 for $X = \emptyset$ is reported in (Lifschitz 1996) and, in the context of abstract argumentation, in (Dung 1995) where formal relationships between the grounded extension and a semantics for argumentation proposed in (Pollock 1987; 1991) are discussed. That semantics is based on the stream generated from the empty set by the neutrality function, i.e., the limit of the even iterations stream in $n^{\omega}_{\mathcal{A}}(\emptyset)$. Figure 5 depicts this special case in its modal formulation.*

Iterated update of "not being attacked by P"

The logical set up of valuation updates introduced for the analysis of complete extensions can be fully reused here. Valuation updates will now be driven by the formula $\neg\Diamond\mathsf{P}$, which we abbreviate by $\mathtt{N}_{\mathsf{P}}$ and which denotes the set of arguments which are not attacked by position P.

Theorem 6 could then be recast in DEL exactly like Theorem 3 has been recast in Theorem 4 in the previous section, and the stream generated by $:= \mathtt{N}^{\omega}_{\mathsf{P}}$ will be decomposable in one stream of even and one stream of odd iterated updates with possibly distinct limits.

The epistemic considerations of the previous section also carry over: P can be taken to denote one agent's belief over the to-be-accepted arguments, and $:= \mathtt{N}_{\mathsf{P}}$ as the basis of a process of update of that belief. The stream generated by $:= \mathtt{N}^{\omega}_{\mathsf{P}}$ can then be interpreted as an attempt at expanding one's argumentative position P by incorporating what is not attacked by that position. If the initial position is admissible, in the general case, this incorporation process produces two limits. If they coincide, then the process has led the truth-set of P to be a stable extension. If they do not, the difference between the truth-set of P in the limit of odd iterations and the truth-set of P in the limit of even iterations represents the set of arguments about which the iterated update is unable to decide, letting the agent's belief oscillate between the two positions.

Conclusions

The paper has provided a novel understanding of the process of computation of Dung's semantics via fixpoint computation. It has presented a DEL analysis of some of the main semantics for argumentation: grounded, complete, stable and preferred (although the latter has only collaterally been dealt with in Remarks 4 and 5). This has allowed us to provide an epistemic interpretation of the process of approximation which, we hope, could bring abstract argumentation closer to some of the concepts and techniques of modern epistemic logic. On the argumentation theory side, the paper has also proven novel results on the fixpoint computation of complete (Theorem 3) and stable (Theorem 6) extensions.

Acknowledgments

The author would like to thank the anonymous reviewers of KR 2012 for the very useful comments that helped improving the paper to its current version.

References

Aucher, G.; Grossi, D.; Herzig, A.; and Lorini, E. 2009. Dynamic context logic. In He, X.; Horty, J.; and Pacuit, E., eds., *Proceedings of LORI 2009*, volume 5834 of *LNAI*. Springer.

Baltag, A., and Smets, S. 2009. Group belief dynamics under iterated revision: Fixed-points and cycles of joint upgrades. In *Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge (TARK'09)*, 41–50. ACM.

Baltag, A.; Smets, S.; and Zvesper, J. 2009. Keep 'hoping' for rationality: a solution to the backward induction paradox. *Synthese* 169:301–333.

Baroni, P., and Giacomin, M. 2009. Semantics of abstract argument systems. In Rahwan, I., and Simari, G. R., eds., *Argumentation in Artifical Intelligence*. Springer.

Blackburn, P.; de Rijke, M.; and Venema, Y. 2001. *Modal Logic*. Cambridge: Cambridge University Press.

Davey, B. A., and Priestley, H. A. 1990. *Introduction to Lattices and Order*. Cambridge University Press.

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

Grossi, D.; Meyer, J.; and Dignum, F. 2006. Classificatory aspects of counts-as: An analysis in modal logic. *Journal of Logic and Computation* 16(5):613–643. Oxford University Press.

Grossi, D.; Meyer, J.; and Dignum, F. 2008. The many faces of counts-as: A formal analysis of constitutive-rules. *Journal of Applied Logic* 6(2):192–217.

Grossi, D. 2009. Doing argumentation theory in modal logic. ILLC Prepublication Series PP-2009-24, Institute for Logic, Language and Computation.

Grossi, D. 2010. On the logic of argumentation theory. In van der Hoek, W.; Kaminka, G.; Lespérance, Y.; and Sen, S., eds., *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, 409–416. IFAAMAS.

Grossi, D. 2011. An application of model checking games to abstract argumentation. In van Ditmarsch, H.; Lang, J.; and Ju, S., eds., *Logic, Rationality and Interaction: Third Internalional Workshop (LORI 2011)*, volume 6953 of *LNAI*, 74–86.

Hintikka, J. 1962. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca: Cornell University Press.

Holliday, W., and Icard, T. 2010. Moorean phenomena in epistemic logic. In Beklemishev, L.; Goranko, V.; and Shehtman, V., eds., *Advances in Modal Logic*. College Publications. 178–199.

Lifschitz, V. 1996. Foundations of logic programming. In *Principles of Knowledge Representation*. CSLI Publications. 69–127.

Meyer, J., and van der Hoek, W. 1995. *Epistemic Logic for AI and Computer Science*, volume 41 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press.

Pollock, J. L. 1987. Defeasible reasoning. *Cognitive Science* 11:481–518.

Pollock, J. L. 1991. A theory of defeasible reasoning. *International Journal of Intelligent Systems* 6(1):33–54.

van Benthem, J., and Gheerbrant, A. 2010. Game solution, epistemic dynamics and fixed-point logics. *Fundamenta Informaticae* 1(4):19–41.

van Benthem, J.; van Eijck, J.; and Kooi, B. 2006. Logics of communication and change. *Information and Computation* 204(11):1620–1662.

van Benthem, J. 2007. Rational dynamics and epistemic logic in games. *International Game Theory Review* 9(1):13–45.

van Benthem, J. 2011. *Logical Dynamics of Information and Interaction*. Cambridge University Press.

van Ditmarsch, H.; Kooi, B.; and van der Hoek, W. 2007. *Dynamic Epistemic Logic*, volume 337 of *Synthese Library Series*. Springer.

Wang, Y. 2011. On axiomatizations of pal. In van Ditmarsch, H.; Lang, J.; and Ju, S., eds., *Proceedings of the Third international conference on Logic, rationality, and interaction (LORI'11)*, volume 6953 of *LNCS*, 314–327. Springer.

Yablo, S. 1984. Truth and reflection. *Journal of Philosophical Logic* 14:297–349.