Journal of Artificial Intelligence Research 66 (2019) 1099-1145 Submitted 10/2019; published 12/2019

If Nothing Is Accepted – Repairing Argumentation Frameworks

Ringo Baumann BAUMANN@INFORMATIK.UNI-LEIPZIG.DE
Markus Ulbricht MULBRICHT@INFORMATIK.UNI-LEIPZIG.DE
Department of Computer Science,
Leipzig University, Germany

Abstract

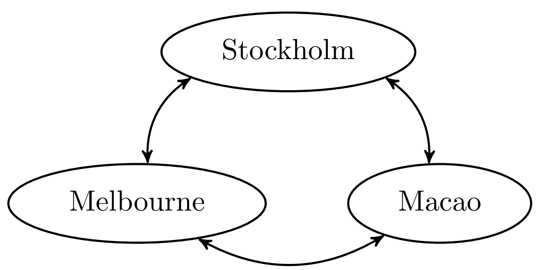
Conflicting information in an agent's knowledge base may lead to a semantical defect, that is, a situation where it is impossible to draw any plausible conclusion. Finding out the reasons for the observed inconsistency (so-called *diagnoses*) and/or restoring consistency in a certain minimal way (so-called *repairs*) are frequently occurring issues in knowledge representation and reasoning. In this article we provide a series of first results for these problems in the context of abstract argumentation theory regarding the two most important reasoning modes, namely credulous as well as sceptical acceptance. Our analysis includes the following problems regarding minimal repairs/diagnoses: existence, verification, computation of one and enumeration of all solutions. The latter problem is tackled with a version of the so-called *hitting set duality* first introduced by Raymond Reiter in 1987. It turns out that grounded semantics plays an outstanding role not only in terms of complexity, but also as a useful tool to reduce the search space for diagnoses regarding other semantics.

1. Introduction

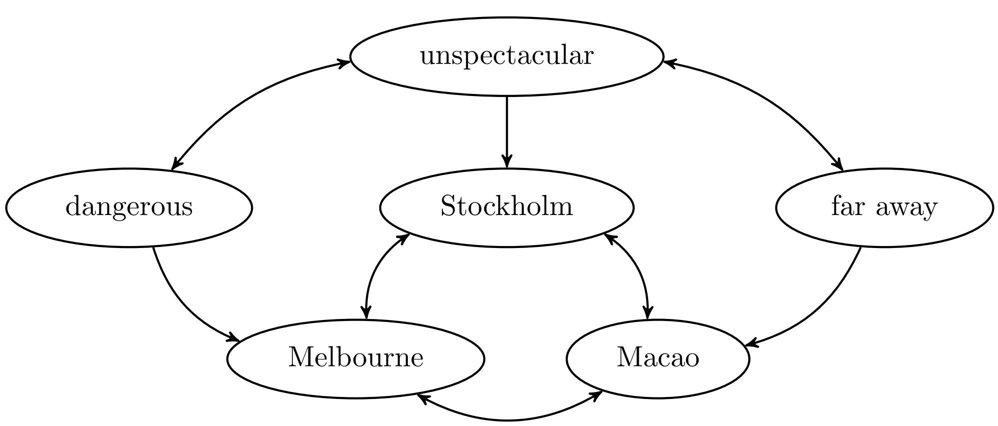
A well-known problem in knowledge representation and reasoning is the semantical collapse of an agent's knowledge base $\mathcal{K}$, i.e. $\mathcal{K}$ is inconsistent and thus does not allow any plausible conclusion. Hansson coined the term *consolidation* and defined it as an operation that withdraws parts of $\mathcal{K}$ in such a way that, first, the resulting knowledge base $\mathcal{K}'$ is consistent and secondly, the change is as small as possible (Hansson, 1994). Even earlier, Reiter introduced the by now presumably best-known formal treatment of this problem in his seminal paper (Reiter, 1987). The so-called *diagnostic problem* for a given system arises whenever we observe that the system does not behave as it should. Reiter used first-order logic as representation formalism and his definition of a diagnosis contains the concepts of consistency as well as minimality. We mention that even before Reiter first approaches to handling inconsistencies in formal systems appeared (da Costa, 1974). However, over time, the topic of restoring consistency under the requirement of minimal change received a lot of attention in many different knowledge representation formalisms like situation calculus (Mcilraith, 1999), logic programs (Sakama & Inoue, 2003), description logic including non-monotonic versions (Lembo, Lenzerini, Rosati, Ruzzi, & Savo, 2011; Bienvenu, 2012; Eiter, Fink, & Stepanova, 2013) as well as probabilistic conditional logic (Potyka & Thimm, 2014) to mention a few.

In this paper we focus on the non-monotonic theory of abstract argumentation (Dung, 1995). More precisely, we consider an abstract argumentation framework (AF) as an agent's knowledge base and the associated extensions correspond to her beliefs (Coste-Marquis, Konieczny, Mailly, & Marquis, 2014; Nouioua & Würbel, 2014; Diller, Haret, Linsbichler, Rümmele, & Woltran, 2018). In brief, Dung-style AFs consist of arguments and attacks which are treated as primitives, i.e. the internal structure of arguments is not considered. The major focus is on resolving conflicts. To this end a variety of semantics have been defined, each of them specifying acceptable sets of arguments, so-called *extensions*, in a particular way. The starting point of our study is a semantical defect of an agent's AF which prevents her from drawing any plausible conclusion in the sense that no argument is accepted. That may mean, for example, that no argument is contained in each extension, so-called *sceptical acceptance*. Our aim is to obtain an agent which is able to act. Therefore we want to know what are minimal diagnoses of the given knowledge base, i.e. which parts are causing the semantical defect. For instance, a certain minimal diagnosis may consist of arguments which are somehow out of date or not as significant in comparison to the others. Consequently, one may tend to discard these arguments. To illustrate a situation like this, consider the following example from Ulbricht (2019b).

Example 1.1. Assume an agent is planning her vacation. The agent's preferred travel destinations are Macao, Stockholm and Melbourne. She only wants to visit one of them:



The agent is aware of the many poisonous animals in Australia and hence believes Melbourne is quite dangerous. Macao is very far away. On the other hand, she visits Europe quite often and thus finds Stockholm less spectacular than the other two options. She deems being at a dangerous place or pretty far away as spectacular since it is unusual. There is no relation between the distance and potential risks since dangerous places can be found anywhere in the world.



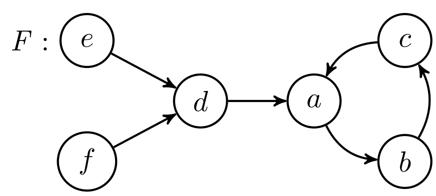
Let us consider the most prominent semantics, namely the stable one. Informally, a set of arguments is a stable extension if there are no conflicts between them and moreover, all other arguments are attacked by at least one argument of the set. Although this AF possesses some stable extensions, for example $E_1 = \{\text{Stockholm}, \text{ dangerous}, \text{ far away}\}$ and $E_2 = \{\text{Macao}, \text{ unspectacular}\}$, our agent is not satisfied since when it comes to reasoning about travel destinations, she is quite sceptical. However, there is no sceptically accepted argument since obviously $E_1 \cap E_2 = \emptyset$. Our agent is thus unable to formally decide which city to visit. As a possible solution, she tries to consider other semantics like the grounded one which does not really help since the unique grounded extension is the empty set. Moreover, other semantics induce similar problems. She would thus be interested in techniques to modify this AF in a reasonable way until it possesses accepted arguments.

The main aim of this article is to study such semantical defects with regard to the following naturally arising questions:

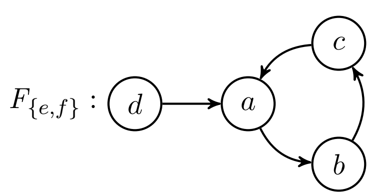
- *Diagnosis* – Which sets of arguments are causing the collapse?
- *Properties* – Do diagnoses always exist? Are there certain preferred diagnoses? How computationally costly is it to verify a candidate diagnosis?
- *Computation* – How to compute one or even all diagnoses?
- *Repair* – How to use this information to obtain an agent which is indeed able to act?

In order to get a first impression of some mentioned points let us consider a more abstract example.

Example 1.2 (Diagnosis and Repair)**.** The AF F does not possess any stable extension. More formally, $stb(F) = \emptyset$.



One may argue that the arguments e and f together can be seen as a diagnosis for the semantical defect of F since ignoring these arguments and their corresponding attacks result in a meaningful AF denoted by $F_{\{e,f\}}$. In particular, $stb\big(F_{\{e,f\}}\big) = \{\{d, b\}\}$.



Note that neither of both arguments can be omitted since the resulting frameworks would collapse too. In this sense, the presented diagnosis $\{e, f\}$ and the corresponding repair $F_{\{e,f\}}$ are minimal.

The structure of the article can be summarized as follows: After discussing necessary preliminaries for abstract argumentation in Section 2 we consider in Section 3 the question of existence of minimal diagnoses and we provide relations between diagnoses wrt. different semantics. Grounded semantics plays a central role here since its minimal diagnoses represent bounds for diagnosis of other semantics. We tackle the second question in Section 4, namely: How to systematically find all minimal diagnoses of a given AF under a given semantics. In particular, we will see that the considered AF F indeed possesses further minimal repairs. These formal results are mainly due to the well-known *hitting set duality* first introduced by Reiter (1987) and a recently generalized version of it (Brewka, Thimm, & Ulbricht, 2017). We also briefly discuss subclasses of AFs, namely *symmetric*, *compact* and *acyclic* frameworks in Section 5. This section also demonstrates how to infer stronger results regarding existence of repairs when given a *splitting* of an AF and discusses special issues regarding *infinite* frameworks. Afterwards, we study the computational complexity for the associated existence and verification problem. In the subsequent Section 7 we briefly discuss repair strategies for stable and preferred semantics. Finally, we conclude as well as discuss some related work in Section 8.

A preliminary version (Baumann & Ulbricht, 2018) of this paper appeared in the Proceedings of the 16th International Conference on Principles of Knowledge Representation and Reasoning.

2. Background in Abstract Argumentation

In the original formulation (Dung, 1995), an *abstract argumentation framework* is a directed graph $F = (A, R)$ where nodes in A represent arguments and the relation R models *attacks*, i.e. for $a, b \in A$, if $(a, b) \in R$ we say that *a attacks b* or *a* is *an attacker of b*. We say that F is *self-controversial* if any argument attacks itself. If not stated otherwise, we restrict ourselves to non-empty finite AFs. Formally, we introduce an infinite reference set $\mathcal{U}$, so-called *universe of arguments* and require for any possible AF, $A \subseteq \mathcal{U}$. The collection of all possible AFs is abbreviated by $\mathcal{F}$. Moreover, for a set E we use E^+ for $\{b \mid (a, b) \in R, a \in E\}$ and define $E^\oplus = E \cup E^+$. In case we need to be specific about the AF under consideration, we use the more informative notation $E^\oplus_F$. A further essential notion in argumentation is *defense*. More precisely, an argument b is *defended by* a set A if each attacker of b is counter-attacked by some $a \in A$. Then, the *characteristic function* of the AF F is given via $\Gamma_F : 2^A \to 2^A$ with $E \mapsto \{a \in A \mid a \text{ is defended by } E\}$.

An *extension-based semantics* $\sigma : \mathcal{F} \to 2^{2^{\mathcal{U}}}$ is a function which assigns to any AF $F = (A, R)$ a set of sets of arguments $\sigma(F) \subseteq 2^A$. Each one of them, so-called *σ-extension*, is considered to be acceptable with respect to F. Besides conflict-free and admissible sets (abbr. *cf* and *ad*) we consider stable, semi-stable, complete, preferred, grounded, ideal and eager semantics (abbr. *stb*, *ss*, *co*, *pr*, *gr*, *il* and *eg*, respectively). Recent overviews are given by Baroni, Caminada, and Giacomin (2011, 2018).

Definition 2.1. Let $F = (A, R)$ be an AF and $E \subseteq A$.

1. $E \in cf(F)$ iff there are no $a, b \in E$ satisfying $(a, b) \in R$,

2. $E \in ad(F)$ iff $E \in cf(F)$ and E defends all its elements,

3. $E \in stb(F)$ iff $E \in cf(F)$ and $E^{\oplus} = A$,

4. $E \in ss(F)$ iff $E \in ad(F)$ and there is no $\mathcal{I} \in ad(F)$ satisfying $E^{\oplus} \subset \mathcal{I}^{\oplus}$,

5. $E \in co(F)$ iff $E \in ad(F)$ and for any $a \in A$ defended by E, $a \in E$ (equivalently, $E \in cf(F)$ and $\Gamma(E) = E$),

6. $E \in pr(F)$ iff $E \in co(F)$ and there is no $\mathcal{I} \in co(F)$ satisfying $E \subset \mathcal{I}$ (equivalently, $E \in cf(F)$ and $\Gamma(E) = E$ and $\subseteq$-maximal wrt. the conjunction of both properties),

7. $E \in gr(F)$ iff $E \in co(F)$ and there is no $\mathcal{I} \in co(F)$ satisfying $\mathcal{I} \subset E$ (equivalently, $E \in cf(F)$ and $\Gamma(E) = E$ and $\subseteq$-minimal wrt. the conjunction of both properties),

8. $E \in il(F)$ iff $E \in co(F)$ and $E \subseteq \bigcap pr(F)$ and $\subseteq$-maximal wrt. the conjunction of both properties,

9. $E \in eg(F)$ iff $E \in co(F)$ and $E \subseteq \bigcap ss(F)$ and $\subseteq$-maximal wrt. the conjunction of both properties.

We say that a semantics σ is *universally defined* if $\sigma(F) \neq \emptyset$ for any $F \in \mathcal{F}$. If even $|\sigma(F)| = 1$ we say that σ is *uniquely defined*. All semantics apart from stable are universally defined. In addition, grounded, ideal and eager semantics are examples of uniquely defined semantics. Stable semantics may *collapse*, i.e. there are AFs F, s.t. $stb(F) = \emptyset$ (cf. running example F depicted in Example 1.2). For two semantics σ and τ we write $\sigma \subseteq \tau$ if for any AF F, $\sigma(F) \subseteq \tau(F)$. For instance, it is well-known that $stb \subseteq ss \subseteq pr \subseteq co \subseteq ad \subseteq cf$.

In the present article we are interested in situations where a given AF $F = (A, R)$ does not possess accepted arguments. To make the notion of acceptance precise, we utilize the usual two alternative reasoning modes, namely *credulous* as well as *sceptical acceptance*. We require $\sigma(F)$ to be non-empty for sceptical reasoning in order to avoid the (for our purpose) unintended situation that every argument is sceptically accepted due to technical reasons. Set-theoretically the intersection over the empty family of sets would yield any argument.[1] However, in our setting it makes sense to define it as the empty set since the (sceptical or credulous) acceptance of an argument a should imply the existence of at least one extension containing a.

Definition 2.2. Given a semantics σ, an AF $F = (A, R)$ and an argument $a \in A$. We say that a is

1. *credulously accepted wrt.* σ if $a \in \bigcup \sigma(F)$,

2. *sceptically accepted wrt.* σ if $a \in \bigcap \sigma(F)$ and $\sigma(F) \neq \emptyset$.

As already mentioned, our motivation for a concept of *inconsistency* is a semantical defect of an agent's AF which prevents her from drawing any plausible conclusion in the sense that nothing is accepted. This is clearly not only relevant for credulous reasoning

1. Applying the standard set-theoretical definition leads to $\bigcap \emptyset = \{x \in \mathcal{U} \mid \forall E \in \emptyset : x \in E\} = \mathcal{U}$ (Baumann & Spanring, 2015, Section 2).

(at least one extension is non-empty), but also sceptical reasoning (there are undisputed arguments). Since we require $\sigma(F) \neq \emptyset$ for sceptical acceptance, we naturally obtain the following notion of inconsistent argumentation frameworks.

Definition 2.3. Given a semantics σ, an AF $F = (A, R)$. We say that F is

1. *inconsistent wrt. credulous reasoning and* σ if $\bigcup \sigma(F) = \emptyset$,
2. *inconsistent wrt. sceptical reasoning and* σ if $\bigcap \sigma(F) = \emptyset$ where we let $\bigcap \emptyset = \emptyset$.

We omit the specifications "wrt. credulous reasoning" and/or "wrt. σ" whenever the reasoning mode, the semantics or both are implicitly clear or do not matter.

Having established this background including our notion of *inconsistent* AFs we are now ready to tackle the problem of *repairing* AFs via moving to an appropriate subframework. In the following section, we will discuss different notions of repairs, whether they exist and connections between them.

3. On the Existence of Repairs

Clearly, before computing potential repairs one may wonder what types of repairs exist and whether there are minimal diagnoses at all. In this section we provide the formal notions and results wrt. this problem and in particular, we give an affirmative answer for nearly all considerable cases. We also investigate the relationship between different repairs. Unfortunately, the existence of a least repair is not guaranteed which leads to follow-up question of *how to repair?* which will be considered in the subsequent section.

3.1 Notions for Repairs

Our repair approach involves moving to subgraphs of a given AF. So let us start by introducing the required notions and formalizing the concepts of diagnoses and repairs. Consider an AF $F = (A, R)$. For a given set $\mathcal{S} \subseteq A$ of arguments we use $F_{\mathcal{S}}$ as a shorthand for the restriction of F to the set $A \setminus \mathcal{S}$, i. e.

$$F|_{A\setminus\mathcal{S}} := \big(A|_{A\setminus\mathcal{S}}, R|_{A\setminus\mathcal{S}\times A\setminus\mathcal{S}}\big) = (A \setminus \mathcal{S}, \{(a, b) \in R \mid a, b \in A \setminus \mathcal{S}\}).$$

In other words, for $\mathcal{S} \subseteq A$, $F_{\mathcal{S}}$ is the subframework of F induced by the removal of arguments in $\mathcal{S}$. Analogously, for a given set $\mathcal{S} \subseteq R$ of attacks we use $F_{\mathcal{S}}$ or sometimes $F \setminus \mathcal{S}$ as a shorthand for $(A, R \setminus \mathcal{S})$. We will also sometimes abuse notation and write $F \cup \{a\}$ and $F \cup \{(a, b)\}$ instead of $F \cup (\{a\}, \emptyset)$ and $F \cup (\{a, b\}, \{(a, b)\})$, respectively. As usual, the latter unions are understood pointwise. Consider the following frameworks.

Example 3.1. Recall the AF F from Example 1.2, i. e. $F = (A, R)$ with

$$A = \{a, b, c, d, e, f\} \qquad R = \{(a, b), (b, c), (c, a), (d, a), (e, d), (f, d)\}.$$

Let $\mathcal{S} = \{d\} \subseteq A$. Then,

$$\begin{aligned} F_{\mathcal{S}} &= (A \setminus \{d\},\ \{(a, b) \in R \mid a, b \in A \setminus \{d\}\}) \\ &= \big(\{a, b, c, e, f\},\ \{(a, b), (b, c), (c, a)\}\big). \end{aligned}$$

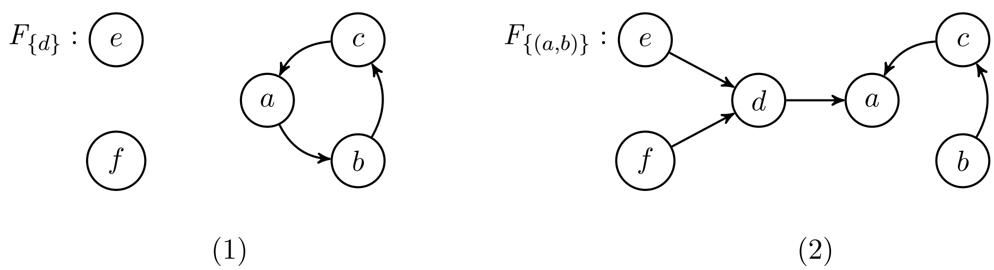


Figure 1: AFs $F_{\{d\}}$ and $F_{\{(a,b)\}}$ discussed in Example 3.1

Let $\mathcal{S}' = \{(a, b)\} \subseteq R$. Then,

$$\begin{aligned} F_{\mathcal{S}'} &= (A, R \setminus \{(a, b)\}) \\ &= \big(\{a, b, c, d, e, f\},\ \{(b, c), (c, a), (d, a), (e, d), (f, d)\}\big). \end{aligned}$$

The AFs $F_{\{d\}}$ and $F_{\{(a,b)\}}$ are depicted in Figure 1 (1) and (2), respectively.

Following the usual notions of "repairs" and "diagnoses" of knowledge bases we define:

Definition 3.2. Given a semantics σ and an AF F. We call $\mathcal{S} \subseteq A$ ($\mathcal{S} \subseteq R$) an *argument-based* (*attack-based*) *σ-cred-diagnosis* of F iff $F_{\mathcal{S}}$ is consistent wrt. credulous reasoning. Moreover, we call the AF $F_{\mathcal{S}}$ an *argument-based* (*attack-based*) *σ-cred-repair* of F. [2] We use the terms *minimal* and *least* for $\subseteq$-minimal or $\subseteq$-least σ-diagnosis as well as the associated σ-repairs. We define (minimal, least) *σ-scep-diagnoses* and *σ-scep-repairs* analogously.

If clear from context or irrelevant we drop the considered semantics or reasoning mode.

Example 3.3. Again consider F and $\mathcal{S}$ from Example 3.1, i. e. $F = (A, R)$ with the set of attacks $A = \{a, b, c, d, e, f\}$ and $R = \{(a, b), (b, c), (c, a), (d, a), (e, d), (f, d)\}$ as well as $\mathcal{S}, \mathcal{S}'$ with $\mathcal{S} = \{d\}$ and $\mathcal{S}' = \{(a, b)\}$. Let $\sigma = stb$. We see that $F_{\mathcal{S}}$ possesses no stable extension due to the (still existing) odd loop. Thus, $\mathcal{S}$ is no *stb*-diagnosis of F. However, $\mathcal{S}'$ is a *stb*-diagnosis, since $F_{\mathcal{S}'}$ possesses the unique stable extension $\{a, b, e, f\}$. So, $F_{\mathcal{S}'}$ is a *stb*-repair. Since only one attack is removed, it is quite easy to see that $\mathcal{S}'$ is even a minimal diagnosis.

3.2 Relations between Credulous and Sceptical Reasoning Mode

We start with some general relations between credulous and sceptical diagnoses. The following theorem applies to any semantics. It states that minimal credulous diagnoses can be found as subsets of sceptical diagnoses.

Theorem 3.4. *Given an AF F and a semantics σ. If $\mathcal{S}$ is a scep-σ-diagnosis of F, then there is a minimal cred-σ-diagnosis $\mathcal{S}'$ of F, s.t. $\mathcal{S}' \subseteq \mathcal{S}$.*

Proof. Let $\mathcal{S}$ be a scep-σ-diagnosis of F. This means, $\bigcap \sigma(F_{\mathcal{S}}) \neq \emptyset$. Consequently, $\sigma(F_{\mathcal{S}}) \neq \emptyset$ and therefore $\bigcup \sigma(F_{\mathcal{S}}) \neq \emptyset$. Thus, $\mathcal{S}$ is a cred-σ-diagnosis of F. Moreover, by finiteness of $\mathcal{S}$ we deduce the existence of a minimal cred-σ-diagnosis $\mathcal{S}'$ of F with $\mathcal{S}' \subseteq \mathcal{S}$ concluding the proof. □

2. We do not fix the order of the specifications, so we also speak of *cred-σ-diagnoses*, *cred-σ-repairs* etc.

Observe that the above theorem holds for both argument-based as well as attack-based diagnoses.

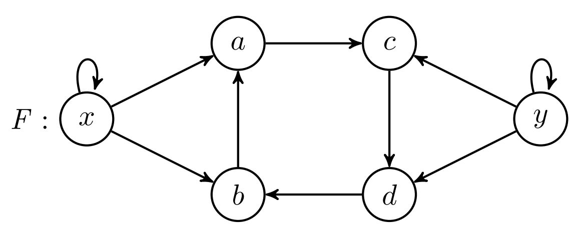
Vice versa, sceptical diagnoses can be found as supersets of credulous ones. We want to mention two issues. First, in contrast to the assertion before, the proof of Theorem 3.5 requires semantics specific properties and thus, does not hold for any argumentation semantics. Secondly, it is not quite clear whether even minimality can be shown. More precisely, the situation differs depending on the semantics and whether argument-based or attack-based diagnoses are considered. Let us start with the existence of an arbitrary repair:

Theorem 3.5. *Given an AF F and a semantics $\sigma \in \{stb, ss, co, pr, gr, il, eg\}$. If $\mathcal{S}$ is a cred-σ-diagnosis of F, then there is a scep-σ-diagnosis $\mathcal{S}'$ of F, s.t. $\mathcal{S} \subseteq \mathcal{S}'$.*

Proof. We give two different proofs depending on whether $\mathcal{S}$ is argument- or attack-based.
Argument-based: We argue as in the proof of Theorem 3.5: Assume $\mathcal{S}$ is a cred-σ-diagnosis of F. Then, there is an argument $a \in \bigcup \sigma(F_{\mathcal{S}}) \neq \emptyset$. Since a σ-extension is conflict-free, $(a, a) \notin F_{\mathcal{S}}$. So define $\mathcal{S}' = (A \setminus \{a\}, \emptyset)$ yielding $F_{\mathcal{S}'} = (\{a\}, \emptyset)$. Thus, we obtain $\sigma(F_{\mathcal{S}'}) = \{\{a\}\}$ for any semantics σ. Hence, $\bigcap \sigma(F_{\mathcal{S}'}) = \{a\}$ so $\mathcal{S}'$ is a scep-σ-diagnosis of F.
Attack-based: This is trivial because $\mathcal{S}' = R$ yields an AF with *no* attacks, so $\sigma(F_{\mathcal{S}'}) = \{A\}$ for any considered semantics σ. Moreover, $\mathcal{S} \subseteq \mathcal{S}'$ for any attack-based diagnosis $\mathcal{S}$. □

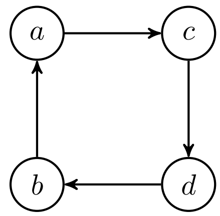
Now we turn to minimality, i. e. the following problem: Given an AF F, a semantics $\sigma \in \{stb, ss, co, pr, gr, il, eg\}$ and a *minimal* cred-σ-diagnosis $\mathcal{S}$ of F, is there a *minimal* scep-σ-diagnosis $\mathcal{S}'$ of F, s.t. $\mathcal{S} \subseteq \mathcal{S}'$? Let us first mention the trivial cases $\sigma \in \{gr, eg, il\}$ where the reasoning modes (due to uniqueness) coincide. Clearly, minimality is given as noted in Corollary 3.13. For the other semantics the answer differs depending on the type of diagnoses under consideration. So let us start with argument-based ones. Here we have the following counterexample for $\sigma \in \{ss, co, pr\}$.

Example 3.6. Consider the following AF $F = (A, R)$:



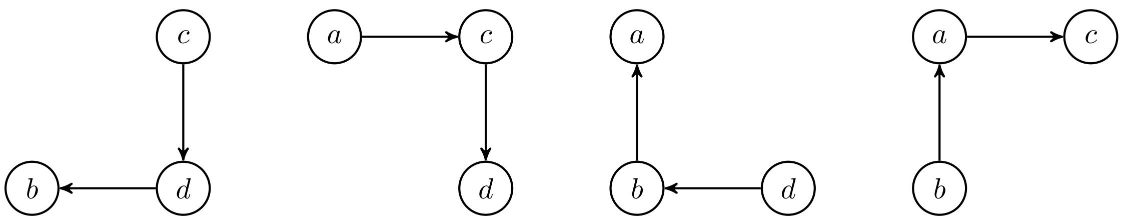
Please observe the structure of this AF: We have an even loop consisting of arguments a, b, c and d which is disturbed by two dummy arguments x and y. Let $\sigma \in \{ss, pr, co\}$. Clearly, there is no way to defend anything from x and y and thus $\emptyset$ is the only σ-extension. Hence, F is inconsistent wrt. σ. Now consider the AF $F_{\{x\}}$, i. e. the argument x is removed. We still have $\emptyset$ as the only extension: The two candidates $\{a, d\}$ and $\{b, c\}$ still lack defense against y. However, removal of x and y yields the following AF $F_{\{x,y\}}$ possessing the two non-empty extensions $\{a, d\}$ and $\{b, c\}$:

$F_{\{x,y\}}$:



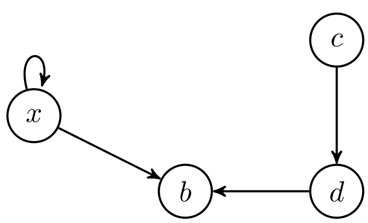
We thus found the minimal σ-*cred* diagnosis $\{x, y\}$. In order to extend this diagnosis to a σ-*scep* diagnosis, we need to remove a, b, c or d as well, so we have σ-*scep* diagnoses $\{x, y, a\}$, $\{x, y, b\}$, $\{x, y, c\}$ and $\{x, y, d\}$ and their corresponding repairs as depicted below:

$F_{\{x,y,a\}}$: $F_{\{x,y,b\}}$: $F_{\{x,y,c\}}$: $F_{\{x,y,d\}}$:



However, there is no *minimal* σ-*scep* diagnosis among them. For example, in $F_{\{x,y,a\}}$ the argument a is removed in order to render c sceptically accepted. This, in turn, does not depend on x, so there is no harm in moving to $F_{\{y,a\}}$ instead; c is still sceptically accepted:

$F_{\{y,a\}}$:



Due to symmetry, we see in addition the minimal σ-*scep*-diagnoses $\{y, c\}$, $\{x, b\}$ and $\{x, d\}$. There is thus no minimal σ-*scep*-diagnosis $\mathcal{S}'$ with $\{x, y\} \subseteq \mathcal{S}'$.

For stable semantics, this is an open problem. We conjecture that minimality can be guaranteed, but did not find a proof so far.

Conjecture 3.7. *Let F be an AF. If $\mathcal{S}$ is a minimal argument-based stb-cred-diagnosis of F, then there is a minimal stb-scep-diagnosis $\mathcal{S}'$ of F, s.t. $\mathcal{S} \subseteq \mathcal{S}'$.*

Let us now turn to attack-based diagnoses. They are more fine-grained since removing a single attack is just removing an attack, where removing an argument yields an arbitrary amount of removed attacks. We can thus answer the question affirmatively for preferred, stable and semi-stable semantics. We have the following, even stronger result. The proof of the theorem below illustrates how precise attack-based diagnoses operate: Given $(a,b) \in \mathcal{S}$ for a minimal σ-diagnosis ($\sigma \in \{ss, pr, stb\}$), then b is guaranteed to be sceptically accepted.

Theorem 3.8. *Let $F = (A,R)$ be an AF and let $\sigma \in \{ss, pr, stb\}$. Assume F is inconsistent wrt. σ and credulous reasoning. If $\mathcal{S} \subseteq R$ is a minimal σ-cred-diagnosis of F, then $\mathcal{S}$ is a minimal σ-scep-diagnosis as well.*

Proof. Let $(a,b) \in \mathcal{S}$. By assumption, $F_{\mathcal{S}} \cup \{(a,b)\} = F_{\mathcal{S}\setminus\{(a,b)\}}$ is inconsistent wrt. σ and credulous reasoning and $F_{\mathcal{S}}$ is consistent, i. e. there is a non-empty extension E, so we have $\emptyset \neq E \in \sigma(F_{\mathcal{S}})$. We claim that b is sceptically accepted in $F_{\mathcal{S}}$.
Stable: Let E be a stable extension of $F_{\mathcal{S}}$. By definition, $E^{\oplus}_{F_{\mathcal{S}}} = A$ and E is conflict-free in $F_{\mathcal{S}}$. Assume for the sake of contradiction $b \notin E$. We claim that in this case, E must be a stable extension of $F_{\mathcal{S}} \cup \{(a,b)\}$ as well:

- Since E was conflict-free in $F_{\mathcal{S}}$ and $b \notin E$, E is also conflict-free in $F_{\mathcal{S}} \cup \{(a,b)\}$,
- due to $E^{\oplus}_{F_{\mathcal{S}}} = A$ we immediately infer $A = E^{\oplus}_{F_{\mathcal{S}}} \subseteq E^{\oplus}_{F_{\mathcal{S}}\cup\{(a,b)\}}$, and hence we infer $A = E^{\oplus}_{F_{\mathcal{S}}\cup\{(a,b)\}}$ since the other inclusion is clear.

The two items above are the properties a stable extension requires. Now, since E is a stable extension of $F_{\mathcal{S}} \cup \{(a,b)\}$, we see that $\mathcal{S}$ is not a *minimal stb-cred*-diagnosis of F contradicting our assumption. We thus conclude $b \in E$. Since E was an arbitrary σ-extension of $F_{\mathcal{S}}$, b is sceptically accepted. We thus infer that $\mathcal{S}$ is even a *stb-scep*-diagnosis. Minimality will be discussed below.
Preferred: Now let E be a non-empty preferred extension of $F_{\mathcal{S}}$. Hence $E = \Gamma_{F_{\mathcal{S}}}(E)$ with $E \neq \emptyset$. Again assume $b \notin E$. In this case, b is not defended by E, otherwise we had $b \in \Gamma_{F_{\mathcal{S}}}(E)$. This means, there is an argument $c \in A$ with $(c,b) \in R \setminus \mathcal{S}$ (hence, $c \neq a$) with either $c \in E$ or c is not attacked by E. Now consider $F_{\mathcal{S}} \cup \{(a,b)\}$. Since $b \notin E$, E is still conflict-free. Moreover, there is the argument c as above, so we have $b \notin \Gamma_{F_{\mathcal{S}}\cup\{(a,b)\}}(E)$ as well. This means the additional attack (a,b) is irrelevant for the characteristic function Γ applied to E, i. e.

$$E = \Gamma_{F_{\mathcal{S}}}(E) = \Gamma_{F_{\mathcal{S}}\cup\{(a,b)\}}(E)$$

and hence, $E \neq \emptyset$ is a complete extension of $F_{\mathcal{S}} \cup \{(a,b)\}$. Thus, $F_{\mathcal{S}} \cup \{(a,b)\}$ is consistent wrt. complete semantics and credulous reasoning, implying it is consistent wrt. preferred semantics and credulous reasoning. As above, this contradicts minimality of $\mathcal{S}$. We thus conclude $b \in E$ for any *non-empty* preferred extension E. Since there is at least one non-empty preferred extension, $\emptyset \notin pr(F_{\mathcal{S}})$ and we thus see again that b is sceptically accepted.
Semi-Stable: As before, assume we are given an non-empty semi-stable extension E of $F_{\mathcal{S}}$ with $b \notin E$. We may argue as above since E is also a preferred extension of $F_{\mathcal{S}}$. We thus infer that E is a complete extension of $F_{\mathcal{S}} \cup \{(a,b)\}$. This is the same contradiction as before, implying $b \in E$. Again, E was an arbitrary non-empty extension and $\emptyset \notin ss(F_{\mathcal{S}})$, so b is sceptically accepted.
Minimality: In all three cases we observed that $\mathcal{S}$ is a σ-*cred*-diagnosis as well. As a final remark we note that $\mathcal{S}$ must be *minimal* since $\mathcal{S}$ was assumed to be minimal for credulous reasoning already: a proper subset $\mathcal{S}' \subsetneq \mathcal{S}$ cannot be a σ-*scep*-diagnosis. □

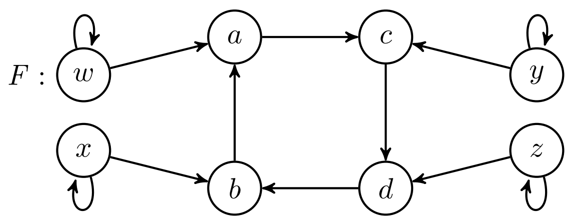
Observe that inconsistency of F wrt. credulous reasoning was a premise of the above theorem. In case F possesses a non-empty σ-extension, it is clear that a minimal σ-*scep*-diagnosis can be found as a superset of the single minimal σ-*cred*-diagnosis $\emptyset$. So we summarize:

Corollary 3.9. *Let F be an AF and let $\sigma \in \{ss, pr, stb\}$. If $\mathcal{S}$ is a minimal attack-based σ-cred-diagnosis of F, then there is a minimal σ-scep-diagnosis $\mathcal{S}'$ of F, s.t. $\mathcal{S} \subseteq \mathcal{S}'$.*

Proof. If F is consistent wrt. σ and credulous reasoning, then $\mathcal{S}$ must be empty, so the claim is trivial. Otherwise, if $\mathcal{S}$ is a minimal attack-based σ-*cred*-diagnosis of F, then due to Theorem 3.8 we may set $\mathcal{S}' = \mathcal{S}$, i. e. we can even guarantee equality. □

When considering complete semantics, we do *not* have minimality in Theorem 3.5 in general. At a first glance, this might be surprising considering the affirmative answer for preferred semantics. However, in the proof of Theorem 3.8 we could exclude $\emptyset$ as a possible preferred extension once we found an arbitrary non-empty fixed point of Γ. This does not work for complete semantics. Hence, we find the following counterexample:

Example 3.10. Consider the following AF $F = (A, R)$:



Clearly, $\emptyset$ is the only complete extension. The reader may verify that $\mathcal{S} = \{(w, a), (z, d)\}$ is a minimal *co*-*cred*-diagnosis of F, yielding the complete extension $\{a, d\}$. Since $\emptyset$ is a complete extension of $F_{\mathcal{S}}$ as well, this is no *co*-*scep*-diagnosis of F. One may check that $\mathcal{S}$ cannot be extended to a *minimal* sceptical diagnosis: A minimal *co*-*scep*-diagnosis must ensure that at least one argument is unattacked. Thus, they are given as $\{(w, w)\}, \ldots, \{(z, z)\}$ and $\{(w, a), (b, a)\}, \ldots, \{(z, d), (c, d)\}$. So there is no minimal *co*-*scep*-diagnoses $\mathcal{S}'$ with $\mathcal{S} \subseteq \mathcal{S}'$.

Finally, we show two helpful, but not unexpected relations between different semantics and their reasoning modes.

Theorem 3.11. *Given two semantics σ and τ, s.t. $\sigma \subseteq \tau$ and σ is universally defined. Let F be an AF.*

1. *If $\mathcal{S} \subseteq F$ is a cred-σ-diagnosis of F, then there is a minimal cred-τ-diagnosis $\mathcal{S}'$ of F, s.t. $\mathcal{S}' \subseteq \mathcal{S}$.*

2. *If $\mathcal{S} \subseteq F$ is a scep-τ-diagnosis of F, then there is a minimal scep-σ-diagnosis $\mathcal{S}'$ of F, s.t. $\mathcal{S}' \subseteq \mathcal{S}$.*

Proof. We prove the second item only. Let $\mathcal{S}$ be a *scep*-τ-diagnosis of F. This means, $\bigcap \tau(F_{\mathcal{S}}) \neq \emptyset$. Since $\sigma \subseteq \tau$ is assumed we deduce $\emptyset \neq \bigcap \tau(F_{\mathcal{S}}) \subseteq \bigcap \sigma(F_{\mathcal{S}})$. Since σ is universally defined we have $\sigma(F_{\mathcal{S}}) \neq \emptyset$ which implies that $\mathcal{S}$ is a *scep*-σ-diagnosis of F. Moreover, by finiteness of $\mathcal{S}$ we deduce the existence of a minimal *scep*-σ-diagnosis $\mathcal{S}'$ of F with $\mathcal{S}' \subseteq \mathcal{S}$ concluding the proof. □

3.3 Uniquely Defined Semantics

We now focus on uniquely defined semantics, i.e. we have $|\sigma(F)| = 1$ for any $F \in \mathcal{F}$. Considering the semantics we investigate in this article this means $\sigma \in \{gr, eg, il\}$.

Please note that in case of uniquely defined semantics we have that any (minimal) sceptical diagnosis is a (minimal) credulous one and vice versa.

Lemma 3.12. *If F is an AF and $\sigma \in \{gr, eg, il\}$, then $\mathcal{S}$ is a cred-σ-diagnosis of F iff it is a scep-σ-diagnosis of F.*

Proof. If $|\sigma(F)| = 1$, then $\bigcap \sigma(F) = \bigcup \sigma(F)$. □

This implies in particular that minimality in Theorem 3.5 can be shown.

Corollary 3.13. *Given an AF F and a semantics $\sigma \in \{gr, eg, il\}$. If $\mathcal{S}$ is a minimal cred-σ-diagnosis of F, then there is a minimal scep-σ-diagnosis $\mathcal{S}'$ of F, s.t. $\mathcal{S} \subseteq \mathcal{S}'$.*

Proof. Set $\mathcal{S}' = \mathcal{S}$ and apply Lemma 3.12. □

We proceed with grounded semantics since these results will play a central role for all other semantics considered in this article. Dung originally defined the grounded extension of an AF $F = (A, R)$ as the $\subseteq$-least fixpoint of the so-called *characteristic function* $\Gamma_F : 2^A \to 2^A$ with $E \mapsto \{a \in A \mid a \text{ is defended by } E\}$. Moreover, he showed that this definition coincides with the $\subseteq$-least complete extension (Dung, 1995, Theorem 25) as introduced in Definition 2.1. Since Γ_F is shown to be $\subseteq$-monotonic we may compute the unique grounded extension G stepwise, i.e. applying Γ_F iteratively starting from the empty set. More precisely, $G = \bigcup_{i=1}^{|A|} \Gamma_F^i(\emptyset)$ (Baumann & Spanring, 2017, Section 3.2). For instance, the unique grounded extensions of $F_{\{c\}}$ and $F_{\{a\}}$ are $\{e, f, a\} = \Gamma^2_{F_{\{c\}}}(\emptyset)$ and $\{e, f, b\} = \Gamma^1_{F_{\{a\}}}(\emptyset)$, respectively. Consequently, an AF possesses a non-empty grounded extension if and only if there exists at least one unattacked argument. This renders argument-based diagnoses weaker in some cases since there is no way to remove a single argument. It is thus clear that a self-controversial AF does not possess an argument-based diagnosis. More precisely, we find the following:

Fact 3.14. *Let $\sigma = gr$. Let F be an AF.*

1. *There exists a minimal argument-based gr-repair for F iff F is not self-controversial.*
2. *There exists a minimal attack-based gr-repair for F.*

Proof. For the first item, assume $F = (A, R)$ and $a \in A$ does not attack itself. Then, $\mathcal{S} = A \setminus \{a\}$ is a *gr*-diagnosis of F. Due to finiteness, we find a minimal one $\mathcal{S}'$ with $\mathcal{S}' \subseteq \mathcal{S}$. For the second item observe that $\mathcal{S} = R$ is a diagnosis. Again we can move to a minimal one. □

In addition to Fact 3.14, we observe that diagnoses for ideal and eager semantics can be found as subsets of a grounded diagnosis. The intuitive reason is the fact that ideal semantics accepts more arguments than grounded semantics and eager semantics is even more credulous than ideal semantics.

Lemma 3.15. *Let $\sigma \in \{il, eg\}$. Let F be an AF. If $\mathcal{S}'$ is a gr-diagnosis of F, then there is a minimal σ-diagnosis $\mathcal{S}$ of F, s.t. $\mathcal{S} \subseteq \mathcal{S}'$.*

Proof. As already mentioned, $\mathcal{S}'$ is a gr-diagnosis of F if and only if $F_{\mathcal{S}'}$ contains an unattacked argument, say a. In this case we see that a also occurs in the unique ideal as well as eager extension. Hence, $\mathcal{S}'$ is a σ-diagnosis for $\sigma \in \{il, eg\}$. Due to finiteness, there is a minimal σ-diagnosis $\mathcal{S}$ with $\mathcal{S} \subseteq \mathcal{S}'$. □

The subsequent main theorem claims the existence of minimal σ-diagnoses for the considered uniquely defined semantics (recall that we do not need to distinguish between credulous and sceptical reasoning for those semantics). Due to the above fact, we find the claim for grounded semantics. Equipped with a grounded diagnosis, we find the other two by applying Lemma 3.15. Moreover, the restriction to finite AFs even gives us the existence of minimal ones.

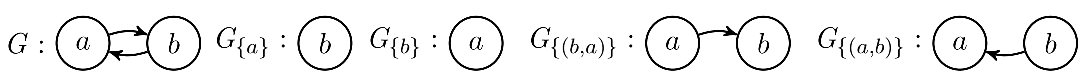
Theorem 3.16. *Let $\sigma \in \{gr, eg, il\}$. Let F be an AF.*

1. *There exists a minimal argument-based σ-repair for F iff F is not self-controversial.*

2. *There exists a minimal attack-based σ-repair for F.*

Proof. For $\sigma = gr$, this is Fact 3.14. For $\sigma \in \{eg, il\}$, this can be inferred from Lemma 3.15 after applying Fact 3.14. □

Example 3.17. The following simple framework G demonstrates that least σ-repairs do not necessarily exist. For $\sigma \in \{gr, eg, il\}$ we have $\sigma(G) = \{\emptyset\}$, i.e. nothing is credulously/sceptically accepted.



Observe that all four depicted given diagnoses are minimal, i. e. $\{a\}$, $\{b\}$, $\{(a, b)\}$ and $\{(b, a)\}$. This example thus illustrates that a least repair does not necessarily exist. This is true for both argument-based as well as attack-based diagnoses.

This finishes our discussion on uniquely defined semantics. In the subsequent section, we turn to universally defined semantics.

3.4 Universally Defined Semantics

Let us consider now semantics which provide us with at least one acceptable position. The following lemma shows that for these semantics minimal credulous as well as sceptical diagnoses are guaranteed, whenever there is a grounded diagnosis.

Lemma 3.18. *Let $\sigma \in \{ss, pr, co\}$. For any AF F there exists a minimal σ-diagnosis $\mathcal{S}$, whenever there exists a gr-diagnosis $\mathcal{S}'$ of F. Moreover, even $\mathcal{S} \subseteq \mathcal{S}'$ can be guaranteed.*

Proof. Let $\sigma \in \{ss, pr, co\}$ and $\mathcal{S}'$ a gr-diagnosis of F. Hence, $gr(F_{\mathcal{S}'}) = \{G\}$ with $G \neq \emptyset$. Since G is the $\subseteq$-least fixpoint of $\Gamma_{F_{\mathcal{S}'}}$ we deduce $G \subseteq C$ for any $C \in co(F_{\mathcal{S}'})$. Due to $ss \subseteq pr \subseteq co$ and the universal definedness of σ we have $\emptyset \neq G \subseteq \bigcap \sigma(F_{\mathcal{S}'})$ as well as $\emptyset \neq G \subseteq \bigcup \sigma(F_{\mathcal{S}'})$. Hence, $\mathcal{S}'$ is a sceptical as well as credulous σ-diagnosis of F. Due to finiteness of F, there exists a minimal σ-diagnosis $\mathcal{S} \subseteq \mathcal{S}'$ concluding the proof. □

Combining Theorem 3.16 and Lemma 3.18 yields the subsequent main theorem for the considered universally defined semantics. As usual, there is a slight difference between argument-based and attack-based diagnoses.

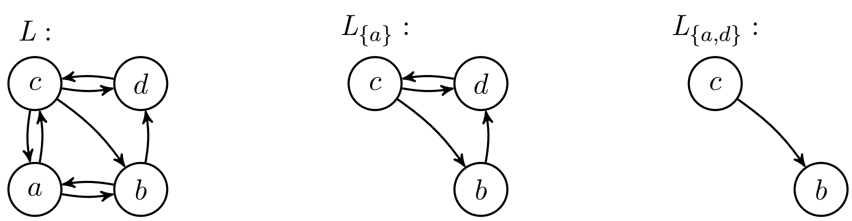
Theorem 3.19. *Let $\sigma \in \{ss, pr, co\}$. Let F be an AF.*

1. *There exists a minimal argument-based σ-repair for F iff F is not self-controversial.*
2. *There exists a minimal attack-based σ-repair for F.*

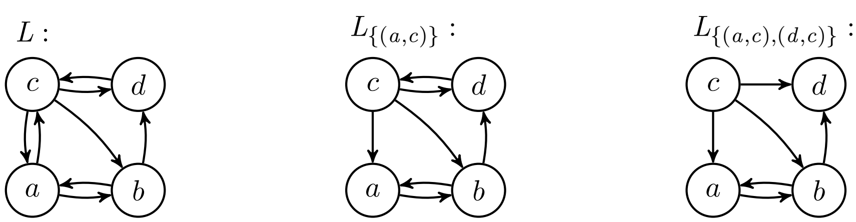
Proof. Apply Theorem 3.16 and Lemma 3.18. □

The following example shows, as promised in Lemma 3.18, that already computed grounded diagnoses can be used to find minimal preferred diagnoses. Moreover, in contrast to uniquely defined semantics we observe that minimal sceptical and minimal credulous diagnoses do not necessarily coincide. This is the case for both argument-based and attack-based diagnoses.

Example 3.20. Consider the following AF L. Since we have no unattacked arguments we deduce $gr(L) = \{\emptyset\}$, i.e. nothing is accepted.
Argument-based repairs: Please observe that $L_{\{a\}}$ and $L_{\{d\}}$ do not possess a grounded extension, either. Consequently, $L_{\{a,d\}}$ is a minimal argument-based gr-repair since we have $gr(L_{\{a,d\}}) = \{\{c\}\}$. Note that $\{a, d\}$ is even a sceptical as well credulous preferred diagnosis of L. These diagnoses are not minimal for preferred semantics since $pr(L) = \{\{a, d\}, \{c\}\}$ implies $\bigcup pr(L) \neq \emptyset$ as well as $pr(L_{\{a\}}) = \{\{c\}\}$ entails $\bigcap pr(L_{\{a\}}) \neq \emptyset$. Altogether, we have strict subset relation ($\emptyset \subsetneq \{a\} \subsetneq \{a, d\}$) between minimal credulous preferred, minimal sceptical preferred and minimal grounded diagnoses.



Attack-based repairs: Regarding attack-based diagnoses, we make similar observations with the chain $\emptyset \subsetneq \{(a, c)\} \subsetneq \{(a, c), (d, c)\}$.



3.5 Collapsing Semantics

Stable semantics is the only prominent semantics which may collapse even for finite AFs. Consider e. g. the AF F from Example 1.2, which does not possess a stable extension as we already explained. However, in terms of existence of repairs we do not observe any differences to all other considered semantics.

Fact 3.21. *Let F be an AF.*

1. *There exists a minimal argument-based stb-repair for F iff F is not self-controversial.*
2. *There exists a minimal attack-based stb-repair for F.*

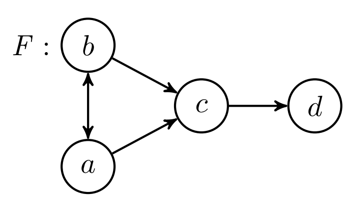
In contrast to all other considered semantics we have that stable diagnoses can not be necessarily found as subsets of an already computed grounded one (Lemmata 3.15, 3.18). For instance, the AF F from Example 1.2 possesses the unique grounded extension $\{e, f\}$. Consequently, we have the trivial (least) *gr*-diagnosis, namely the empty set. As F does not possess a stable extension, all minimal *stb*-diagnoses are non-empty. Nevertheless, credulous as well as sceptical diagnoses for stable semantics can be found as supersets of grounded ones.

Lemma 3.22. *Let $F = (A, R)$ be an AF. If $\mathcal{S}'$ is a gr-diagnosis of F, then there is a stb-diagnosis $\mathcal{S}$ of F, s.t. $\mathcal{S}' \subseteq \mathcal{S}$.*

Proof. Argument-based: Given $\mathcal{S}'$ as *gr*-diagnosis of $F = (A, R)$, i.e. $gr(F_{\mathcal{S}'}) = \{E\}$ with $E \neq \emptyset$. Consider now $E^{\oplus}$ wrt. the attack-relation of $F_{\mathcal{S}'}$. Obviously, $\mathcal{S}' \subseteq A \setminus E^{\oplus}$ and moreover, $gr\left(F_{A \setminus E^{\oplus}}\right) = \{E\}$. Obviously, by construction $E \in stb\left(F_{A \setminus E^{\oplus}}\right)$. Furthermore, since E is non-empty we deduce that there is at least one unattacked argument $a \in E$. Hence, for any $E' \in stb\left(F_{A \setminus E^{\oplus}}\right)$ we have $a \in E'$. Consequently, $A \setminus E^{\oplus}$ serves as a credulous as well as sceptical diagnosis for stable semantics.
Attack-based: Trivial since we may set $\mathcal{S} = R$. □

Please note that Lemma 3.22 does not claim minimality of the *stb*-diagnosis $\mathcal{S}$. Indeed, the following example illustrates that existence of a minimal *stb*-diagnosis with $S' \subseteq S$ as above is not obtained in general.

Example 3.23. Consider the following AF F



Since every argument is attacked we infer that $\emptyset$ cannot be a *gr*-diagnosis. A possible *gr*-diagnosis is $\{a\}$. Indeed, this is also a *stb*-diagnosis (wrt. credulous as well as sceptical reasoning), but not minimal since F itself possesses the sceptically accepted argument d. Moreover we make the same observations for the *gr*-diagnosis $\{(a, b)\}$. Hence, this is a counterexample for both types of diagnoses.

This finishes our discussion regarding rather general results pertaining to the existence and relationships of repairs. In the next section, we show how to characterize *all* repairs of a given AF. This is achieved via a generalized version of Reiter's well-known hitting set duality.

4. Characterizing all Diagnoses: A Hitting Set Duality for AFs

In his seminal paper, Reiter (1987) establishes – for his setting of a given system description – a duality result between the set of all minimal repairs and the minimal conflicts. Recently, it was shown by Brewka et al. (2019) that Reiter's duality can be generalized to arbitrary logics given that the knowledge bases in question can be modeled as a finite set of formulas. In order to capture even non-monotonic logics a refinement of the notion of inconsistency was necessary. In this section we will demonstrate how to utilize this generalized duality for the non-monotonic theory of abstract argumentation.

4.1 Hitting Set Duality for General Logics

The result by Brewka et al. (2019) uses a generic definition of a logic adapted from Brewka and Eiter (2007). In a nutshell, a logic L consists of syntax and semantics of formulas. To model the syntax properly, we stipulate a set $\mathcal{WF}$ of so-called well-formed formulas. Any knowledge base $\mathcal{K}$ consists of a subset of $\mathcal{WF}$. To model the semantics, we let $\mathcal{BS}$ be a set of so-called belief sets. Intuitively, given a knowledge base $\mathcal{K}$, the set of all that can be inferred from $\mathcal{K}$ is $B \subseteq \mathcal{BS}$. To formalize this, a mapping $\mathcal{ACC}$ assigns the set B of corresponding belief sets to each knowledge base $\mathcal{K}$. Finally, some belief sets are considered inconsistent. We call the set of all inconsistent belief sets $\mathcal{INC}$. Hence, our definition of a logic is as follows.

Definition 4.1. A logic L is a tuple $L = (\mathcal{WF}, \mathcal{BS}, \mathcal{INC}, \mathcal{ACC})$ where $\mathcal{WF}$ is a set of well-formed formulas, $\mathcal{BS}$ is the set of belief sets, $\mathcal{INC} \subseteq \mathcal{BS}$ is an upward closed[3] set of inconsistent belief sets, and $\mathcal{ACC} : 2^{\mathcal{WF}} \to 2^{\mathcal{BS}}$ is a mapping. A *knowledge base* $\mathcal{K}$ of L is a finite subset of $\mathcal{WF}$. A knowledge base $\mathcal{K}$ is called *inconsistent* iff $\mathcal{ACC}(\mathcal{K}) \subseteq \mathcal{INC}$.

Let us briefly discuss the notion of *strong inconsistency* and how it induces a hitting set duality for our setting.

Definition 4.2. Let $\mathcal{K}$ be any knowledge base. For $\mathcal{H} \subseteq \mathcal{K}$, $\mathcal{H}$ is called *strongly $\mathcal{K}$-inconsistent* if $\mathcal{H} \subseteq \mathcal{H}' \subseteq \mathcal{K}$ implies $\mathcal{H}'$ is inconsistent. $\mathcal{H}$ is *minimal strongly $\mathcal{K}$-inconsistent* if $\mathcal{H}$ is strongly $\mathcal{K}$-inconsistent and $\mathcal{H}' \subsetneq \mathcal{H}$ implies that $\mathcal{H}'$ is not strongly $\mathcal{K}$-inconsistent. Let $SI_{min}(\mathcal{K})$ denote the set of all minimal strongly $\mathcal{K}$-inconsistent subsets of $\mathcal{K}$.

We proceed with the well-known concepts of (minimal) hitting sets and (maximal) consistent subsets.

Definition 4.3. Let $\mathcal{M}$ be a set of sets. We call $\mathcal{S}$ a *hitting set* of $\mathcal{M}$ if $\mathcal{S} \cap M \neq \emptyset$ for each $M \in \mathcal{M}$. A hitting set $\mathcal{S}$ of $\mathcal{M}$ is a *minimal hitting set* of $\mathcal{M}$ if $\mathcal{S}' \subsetneq \mathcal{S}$ implies $\mathcal{S}'$ is not a hitting set of $\mathcal{M}$.

3. S is upward closed if $B \in S$, $B \subseteq B'$ implies $B' \in S$.

Definition 4.4. We say $\mathcal{H} \subseteq \mathcal{K}$ is a *maximal consistent* subset of $\mathcal{K}$ if $\mathcal{H}$ is consistent and $\mathcal{H} \subsetneq \mathcal{H}' \subseteq \mathcal{K}$ implies $\mathcal{H}'$ is inconsistent. We denote the set of all maximal consistent subsets of $\mathcal{K}$ by $C_{max}(\mathcal{K})$.

Now we are ready to phrase the duality result from Brewka et al. (2017).

Theorem 4.5. *Let $\mathcal{K}$ be a knowledge base. Then, $\mathcal{S}$ is a minimal hitting set of $SI_{min}(\mathcal{K})$ if and only if $\mathcal{K} \setminus \mathcal{S} \in C_{max}(\mathcal{K})$.*

4.2 Strong Inconsistency in Abstract Argumentation

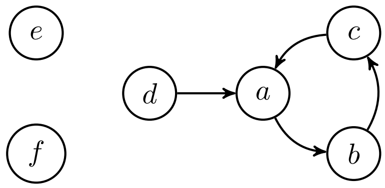
Our goal is to apply Theorem 4.5 to the theory of abstract argumentation frameworks. So let us rephrase the notions we require appropriately. Clearly, the concept of *maximal* consistency does not require a translation to AFs since these are the sub-AFs corresponding to minimal repairs. We thus consider strong inconsistency. As usual, this can be done for argument-based and attack-based diagnoses.

Definition 4.6. Let $F = (A, R)$ be an AF and let σ be any semantics.
Argument-based: We call $\mathcal{H} \subseteq A$ a *strongly inconsistent* set of arguments of F wrt. σ and credulous (sceptical) reasoning if for each $\mathcal{H}'$ with $\mathcal{H} \subseteq \mathcal{H}' \subseteq A$ the AF $F_{A\setminus\mathcal{H}'} = (\mathcal{H}', R_{|\mathcal{H}'})$ is inconsistent wrt. σ and credulous (sceptical) reasoning. Let $SI^{A}_{min}(F, \sigma, cred)$ ($SI^{A}_{min}(F, \sigma, scep)$) be the set of all minimal strongly inconsistent sets of arguments of F wrt. σ and credulous (sceptical) reasoning.
Attack-based: We call $\mathcal{H} \subseteq R$ a *strongly inconsistent* set of attacks of F wrt. σ and credulous (sceptical) reasoning if for each $\mathcal{H}'$ with $\mathcal{H} \subseteq \mathcal{H}' \subseteq R$ the AF $F_{R\setminus\mathcal{H}'} = (A, \mathcal{H}')$ is inconsistent wrt. σ and credulous (sceptical) reasoning. Let $SI^{R}_{min}(F, \sigma, cred)$ ($SI^{R}_{min}(F, \sigma, scep)$) be the set of all minimal strongly inconsistent sets of attacks of F wrt. σ and credulous (sceptical) reasoning.

As usual the semantics as well as the reasoning mode will sometimes be clear from the context or irrelevant. In this case we will them implicit and simply write $SI^{A}_{min}(F)$ resp. $SI^{R}_{min}(F)$.

Example 4.7 (Strong inconsistency for AFs)**.** Consider the running example F.
Argument-based: For our running example AF $F = (A, R)$ we had $A = \{a, b, c, d, e, f\}$. Let us focus on credulous reasoning. We already observed that F has no stable extension, i.e. A itself is a strongly inconsistent set of arguments wrt. stable semantics. The subset $\mathcal{H}_1 \subseteq A$ with $\mathcal{H}_1 = \{a, b, c\}$ induces the AF $F_1 = (\{a, b, c\}, \{(a, b), (b, c), (c, a)\})$ corresponding to the odd circle contained in F. However, $\mathcal{H}_1$ is not a strongly inconsistent set of arguments since the framework induced by $\mathcal{H}_2$ with $\mathcal{H}_1 \subseteq \mathcal{H}_2 \subseteq A$ given as $\mathcal{H}_2 = \{a, b, c, d\}$ has the stable extension $\{b, d\}$ (cf. AF $F_{\{e,f\}}$ depicted in Example 1.2). One may easily verify that $SI^{A}_{min}(\mathcal{K}) = \{\{a, b, c, e\},\ \{a, b, c, f\}\}$.
Attack-based: We have $R = \{(a, b),\ (b, c),\ (c, a),\ (d, a),\ (e, d),\ (f, d)\}$. Again consider credulous reasoning and stable semantics. Now the subset $\mathcal{H}_1 \subseteq R$ with $\mathcal{H}_1 = \{(a, b),\ (b, c),\ (c, a)\}$ induces to the odd circle contained in F. Again, $\mathcal{H}_1$ is not strongly inconsistent since the superset $\mathcal{H}_2$ with $\mathcal{H}_1 \subseteq \mathcal{H}_2 \subseteq R$ given as $\mathcal{H}_2 = \{(a, b),\ (b, c),\ (c, a),\ (d, a)\}$ induces the AF

$F_{R\setminus\mathcal{H}_2} = F_{\{(e,d),(f,d)\}}$ with the stable extension $\{b,d,e,f\}$. Note however that the represented framework is now different, namely F with two *attacks* removed:



We obtain

$$SI^{R}_{min}(F) = \{\{(a,b),\ (b,c),\ (c,a),\ (e,d)\},\ \{(a,b),\ (b,c),\ (c,a),\ (f,d)\}\}.$$

Now we are ready to apply Theorem 4.5 to diagnoses of AFs. The result holds for all semantics and both reasoning modes. It can be stated for both argument-based as well as attack-based repairs. To prove this, we simply construct a logic where any formula α of a knowledge base $\mathcal{K}$ corresponds to an argument (for argument-based diagnoses) resp. attack (for attack-based diagnoses). Since an AF is a tuple consisting of both arguments and attacks, one of them will be fixed.

Proposition 4.8. *[Duality: argument-based] Let $F = (A,R)$ be an AF. Let σ be any semantics and consider any reasoning mode. Then, $\mathcal{S}$ is a minimal hitting set of $SI^{A}_{min}(F)$ if and only if $\mathcal{S}$ is a minimal σ-diagnosis of F.*

Proof. Assume the set R of attacks is fixed. Define a logic $L = (\mathcal{WF}, \mathcal{BS}, \mathcal{INC}, \mathcal{ACC})$ with $\mathcal{WF} = A$ (that is, a knowledge base $\mathcal{K}$ is a finite set of arguments in A), $\mathcal{BS} = A$, $\mathcal{INC} = \emptyset$ and $\mathcal{ACC}(\mathcal{K}) = \bigcup \sigma(F)$ resp. $\mathcal{ACC}(\mathcal{K}) = \bigcap \sigma(F)$ where F is the AF $F = (\mathcal{K}, R_{|\mathcal{K}})$. Now given an AF $F = (A,R)$ a minimal argument-based repair corresponds to a maximal consistent subset $\mathcal{H} \subseteq \mathcal{K} = A$ and a minimal strongly inconsistent set of arguments of F corresponds to a minimal strongly inconsistent subset of $\mathcal{K}$. Thus, the claim can be seen by applying Theorem 4.5 to the logic L we defined here. □

Proposition 4.9. *[Duality: attack-based] Let $F = (A,R)$ be an AF. Let σ be any semantics and consider any reasoning mode. Then, $\mathcal{S}$ is a minimal hitting set of $SI^{R}_{min}(F)$ if and only if $\mathcal{S}$ is a minimal σ-diagnosis of F.*

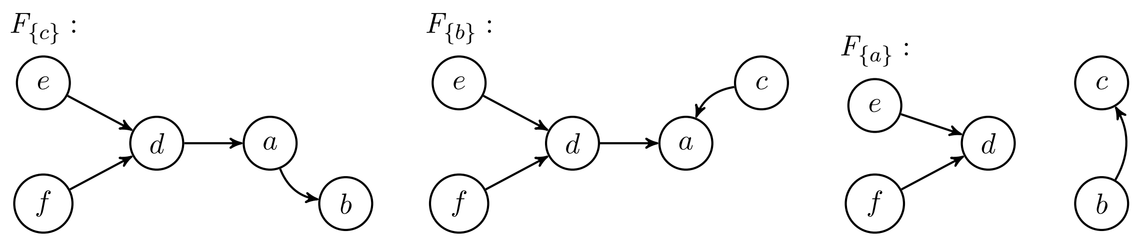
Proof. Assume the set A of arguments is fixed. Define a logic $L = (\mathcal{WF}, \mathcal{BS}, \mathcal{INC}, \mathcal{ACC})$ with $\mathcal{WF} = A \times A$ (that is, a knowledge base $\mathcal{K}$ is a finite set of attacks over A), $\mathcal{BS} = A$, $\mathcal{INC} = \emptyset$ and $\mathcal{ACC}(\mathcal{K}) = \bigcup \sigma(F)$ resp. $\mathcal{ACC}(\mathcal{K}) = \bigcap \sigma(F)$ where F is the AF $F = (A, \mathcal{K})$. Now given an AF $F = (A,R)$ a minimal attack-based repair corresponds to a maximal consistent subset $\mathcal{H} \subseteq \mathcal{K} = R$ and a minimal strongly inconsistent set of attacks of F corresponds to a minimal strongly inconsistent subset of $\mathcal{K}$. Thus, the claim can be seen by applying Theorem 4.5 to the logic L we defined here. □

Example 4.10 (Maximal Consistent Subsets via Hitting Set Duality)**.** Consider again the running example F with stable semantics and credulous reasoning.

Argument-based: For the argument-based diagnoses we checked that

$$SI^{A}_{min}(\mathcal{K}) = \{\{a, b, c, e\}, \{a, b, c, f\}\}.$$

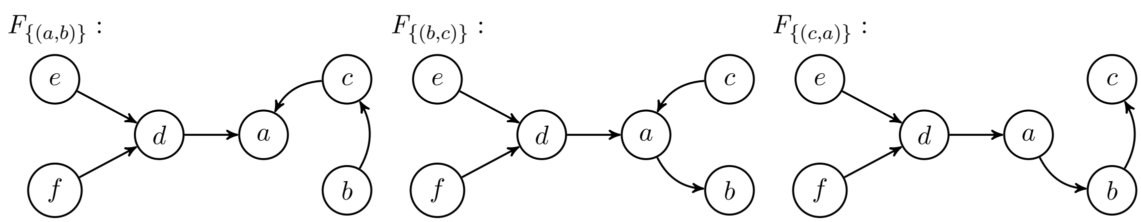
According to Definition 4.3 we obtain four minimal hitting sets, namely $\{a\}$, $\{b\}$, $\{c\}$ and $\{e, f\}$. Observe that $\{e, f\}$ is the already found *stb*-diagnosis presented in Example 1.2. The minimal hitting sets for F can be interpreted as follows: Either one argument from the odd circle needs to be removed or both e and f to facilitate d. These sets correspond to the *stb*-repairs $F_{\{e,f\}}$ (considered in Example 1.2) as well as $F_{\{c\}}$, $F_{\{b\}}$ and $F_{\{a\}}$ depicted below. We obtain $stb\left(F_{\{c\}}\right) = \{\{e, f, a\}\}$, $stb\left(F_{\{b\}}\right) = \{\{e, f, c\}\}$ and $stb\left(F_{\{a\}}\right) = \{\{e, f, b\}\}$.



Attack-based: Recall

$$SI^{R}_{min}(F) = \{\{(a, b),\ (b, c),\ (c, a),\ (e, d)\},\ \{(a, b),\ (b, c),\ (c, a),\ (f, d)\}\}$$

for the attack-based diagnoses. We find the hitting sets $\{(a, b)\}$, $\{(b, c)\}$, $\{(c, a)\}$ as well as $\{(e, d),\ (f, d)\}$. The latter one yields $F_{\{(e,d),(f,d)\}}$ from Example 4.7. The former ones correspond to the following AFs:



This finishes our discussion regarding a characterization of *all* diagnoses of a given AF. The results of this section indicate that finding repairs can be achieved in a straightforward way or by consideration of the strongly inconsistent arguments resp. attacks. This might not be the most efficient approach if one is just interested in *one* diagnosis, but it helps representing *all* repairs in a concise way.

In his seminal paper, Reiter (1987) establishes a duality result between hitting sets of minimal inconsistent subsets ("conflict sets") and maximal consistent subsets. Reiter's paper is also concerned about computing hitting sets. In fact, many algorithms and systems for enumerating minimal inconsistent sets build on the duality between minimal inconsistent sets, maximal consistent sets, and their respective hitting sets (Bailey & Stuckey, 2005; Liffiton & Sakallah, 2008; Liffiton, Previti, Malik, & Marques-Silva, 2016). For example, Liffiton et al. (2016) take turns in computing minimal unsatisfiable sets and maximal consistent sets and uses the duality between the two to compute remaining sets of either type.

Hitting sets are also utilized in computation of causes and responsibilities of inconsistency in databases (Bertossi & Salimi, 2017). It would thus be interesting to see whether approaches like these can be adapted to repair inconsistent AFs.

The motivation of Reiter (1987) for establishing the duality result is finding diagnoses. Many further approaches consider this problem as well and apply and extend his work. To name just a few, Stern et al. (2012) exploit the duality between minimal inconsistent sets and maximal consistent sets and, similarly as discussed above for the task of enumerating these sets, interleaves construction of these two sets with each other, in order to solve the diagnosis problem. The work by Metodi et al. (2014) casts the problem into SAT and reports on to-date significant performance improvements. Similarly, Marques-Silve et al. (2015) solve the diagnosis problem by casting it into a MaxSAT problem and leveraging SAT solvers.

In the next section, we focus on particular aspects and properties of AFs in order to infer tailored, more advanced properties.

5. Special Cases For AFs

We may obtain more insightful results when focusing on certain aspects of the AF under consideration. We will start our investigation with symmetric AFs (Coste-Marquis, Devred, & Marquis, 2005). In a nutshell, an AF is symmetric if the attack relation is. Symmetry is a rather strong property, yielding a variety of additional connections between repairs and diagnoses. Moreover, complexity results for symmetric AFs are quite encouraging (see Section 6 below). The same is true for compact (Baumann, Dvorák, Linsbichler, Strass, & Woltran, 2014) and acyclic AFs. We then continue with a splitting method (Baumann, 2011). Splitting methods are an important concept in non-monotonic reasoning which allows for a certain modularization of the knowledge base under consideration which is usually not given due to non-monotonic interactions between formulas (Baroni, Giacomin, & Liao, 2018). The structural properties that come along with splitting can be utilized to infer results for diagnoses. The last part of this section will be devoted to *infinite* AFs (Baumann & Spanring, 2017). Allowing the underlying set of arguments to be infinite possesses additional challenges since the behavior of an infinite AF is less intuitive and *existence* and *uniqueness* of certain extensions as well as diagnoses is no longer guaranteed.

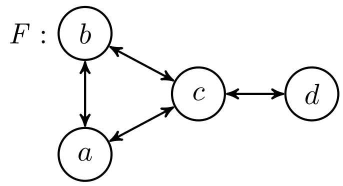
5.1 Symmetric, Compact and Acyclic Frameworks

According to Coste-Marquis et al. (2005) an AF $F = (A, R)$ is *symmetric* if R is symmetric, nonempty and irreflexive.

Definition 5.1. If $F = (A, R)$ is an AF, then we call F *symmetric* if

- $R \neq \emptyset$,
- $(a, b) \in R$ implies $(b, a) \in R$ for any $a, b \in A$,
- $(a, a) \notin R$ for any $a \in A$.

Example 5.2. The following AF F is symmetric:



Now assume we are given a symmetric AF F. Clearly, when moving from F to $F_{\mathcal{S}}$ for $\mathcal{S} \subseteq A$ resp. $\mathcal{S} \subseteq R$, we do not change the fact that the attack relation is irreflexive. We might end up with an AF with no attack (violating non-emptiness), but this does not concern us since an AF of the form $F = (A, \emptyset)$ is consistent for both reasoning modes and any considered semantics. However, we need to make sure that the symmetry of the attack relation is preserved, otherwise we lose the properties we want to utilize. This is clearly no issue for argument-based diagnoses:

Lemma 5.3. *If $F = (A, R)$ is symmetric and $\mathcal{S} \subseteq A$, then the attack-relation of $F_{\mathcal{S}}$ is either empty or symmetric and irreflexive.*

Of course, we want to consider attack-based diagnoses as well. Fortunately, the restriction we need to make for the removal of attacks is quite natural. We simply need to ensure that the diagnosis operates symmetric in the sense that removal of (a, b) implies removal of (b, a) as well. Formally:

Definition 5.4. Let $F = (A, R)$ be a symmetric AF. An attack-based diagnosis $\mathcal{S} \subseteq R$ of F is *symmetric* iff $(a, b) \in \mathcal{S} \Leftrightarrow (b, a) \in \mathcal{S}$.

Clearly, we now have:

Lemma 5.5. *If $F = (A, R)$ is symmetric and $\mathcal{S} \subseteq R$ a symmetric diagnosis, then the attack relation of $F_{\mathcal{S}}$ is either empty or symmetric.*

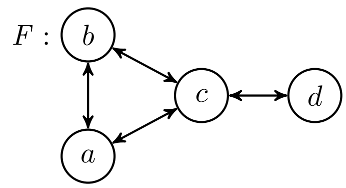
In case of stable, semi-stable and preferred semantics we obtain a very useful property, namely any argument $a \in A$ belongs to at least one extension (Coste-Marquis et al., 2005, Proposition 6). Consequently, we may show the following properties.

Proposition 5.6. *Given a semantics $\sigma \in \{stb, ss, pr\}$ and a symmetric AF $F = (A, R)$.*

1. *$\emptyset$ is the least cred-σ-diagnosis and*
2. *$\mathcal{S} \subseteq F$ is a (minimal) symmetric scep-σ-diagnosis iff $\mathcal{S}$ is a (minimal) symmetric gr-diagnosis.*

Let us assume that our current knowledge base underlies further external revision processes (Coste-Marquis et al., 2014; Baumann & Brewka, 2015; Diller et al., 2018). Both items can be gainfully used if we know that certain types of revision do not affect the symmetry of an AF. More precisely, the items 1 and 2 ensure that we have either nothing to do (if interested in credulous reasoning) or we may act according to grounded semantics instead of σ (if sceptical reasoning is chosen).

Example 5.7. Consider again the symmetric AF F introduced in Example 5.2.



We have $stb(F) = \{\{a, d\}, \{b, d\}, \{c\}\}$. This means, no argument is sceptically accepted. In order to repair regarding grounded semantics we have to ensure the existence of at least one unattacked argument. Consequently, the least *scep-gr*-repair is given as $F_{\{c\}}$. As promised by Item 2 in Proposition 5.6 this indeed coincides with the least *scep-stb*-repair.

Let us briefly consider two further classes of frameworks, namely so-called *compact* and *acyclic* ones. The first one is semantically defined and characterized by the feature that each argument of the AF occurs in at least one extension of the AF (Baumann et al., 2014; Baumann, 2018). For instance, the AF F depicted in Example 5.2 is compact wrt. stable semantics. Compact frameworks obviously fulfill Item 1 of Proposition 5.6 and thus build an interesting subclass of AFs if interested in credulous reasoning. The second class is syntactically defined and as expected an AF is acyclic if it does not contain any cycles. Such frameworks are known to be *well-founded* (Dung, 1995) which means, they possess exactly one complete extension which is grounded, preferred and stable (Coste-Marquis et al., 2005, Propositions 1 and 2). This means, the agent is able to act (in both reasoning modes) whenever we are faced with an acyclic AF.

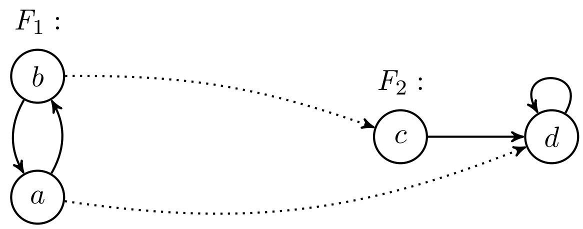
5.2 Splitting

Let us now investigate situations where we are given a *splitting* of the AF under consideration. Splitting is an important concept in non-monotonic reasoning as it abuses structural properties of a knowledge base in order to identify a certain monotonic behavior. More precisely, splitting methods try to divide a theory in subtheories such that the formal semantics of the entire theory can be obtained by constructing the semantics of the subtheories. For AFs, splitting was considered in several works (Baumann, 2011, 2014; Baroni et al., 2018). We briefly recall the required notions here and then demonstrate how to infer properties of repairs and diagnoses.

Definition 5.8. Let $F_1 = (A_1, R_1)$ and $F_2 = (A_2, R_2)$ be two AFs with $A_1 \cap A_2 = \emptyset$. Let $R_3 \subseteq A_1 \times A_2$. We call (F_1, F_2, R_3) a *splitting* of the AF $F = (A_1 \cup A_2, R_1 \cup R_2 \cup R_3)$.

In a nutshell, if (F_1, F_2, R_3) is a splitting of F, then extensions of F_1 can be computed as a first step to compute an extensions of F. The AF F_2 does not influence F_1 and can thus be considered later.

Example 5.9. Let $F_1 = (A_1, R_1)$ with $A_1 = \{a, b\}$ and $R_1 = \{(a, b), (b, a)\}$, $F_2 = (A_2, R_2)$ with $A_2 = \{c, d\}$ and $R_2 = \{(c, d), (d, d)\}$ and let $R_3 = \{(a, d), (b, c)\}$. Then, (F_1, F_2, R_3) is a splitting of the following AF:



The idea of splitting is as follows: Once we are given an extension E_1 of the AF F_1, based on E_1 we want to construct a reduced version of F_2. Then we compute an extension E_2 of this reduced AF to obtain an extension $E_1 \cup E_2$ of F. In the following, we define how to reduce F_2 when considering stable semantics. The other semantics will be discussed afterwards.

Definition 5.10 (Reduct)**.** Let $F_2 = (A_2, R_2)$ be an AF and A_1 such that $A_1 \cap A_2 = \emptyset$. Let $S \subseteq A_1$ and $L \subseteq A_1 \times A_2$. The (S, L)-reduct of F_2, denoted by $F_2^{S,L}$, is the AF

$$F_2^{S,L} = \left(A_2^{S,L}, R_2^{S,L}\right) \qquad \text{with} \qquad A_2^{S,L} = \{a_2 \in A_2 \mid \nexists a_1 \in S : (a_1, a_2) \in L\}$$

$$R_2^{S,L} = \left\{(a, b) \in R_2 \mid a, b \in A_2^{S,L}\right\}.$$

Example 5.11. Consider again our previous example. The AF $F_1 = (A_1, R_1)$ as above has two stable extensions $E_1 = \{a\}$ and $E_1' = \{b\}$. So we are interested in the (E_1, R_3)- and (E_1', R_3)-reduct of F_2 which are $F_2^{E_1,R_3} = (\{c\}, \emptyset)$ and $F_2^{E_1',R_3} = (\{d\}, \{(d, d)\})$:

The former has the stable extension $E_2 = \{c\}$, the latter none. Indeed, the unique stable extension of the whole AF F is $\{a, c\} = E_1 \cup E_2$.

Now the following theorem states that we can indeed find extensions E of F by considering an extension E_1 of F_1 and then reduce F_2 and continue computing. More precisely, we find *all* extensions of F this way:

Theorem 5.12. *[(Baumann, 2011)] Let (F_1, F_2, R_3) be a splitting of the AF F, i.e. we have $F = (A, R) = (A_1 \cup A_2, R_1 \cup R_2 \cup R_3)$ where $F_1 = (A_1, R_1)$ and $F_2 = (A_2, R_2)$.*

- *If E_1 is a stable extension of F_1 and E_2 a stable extension of the (E_1, R_3)-reduct of F_2, then $E_1 \cup E_2$ is a stable extension of F.*

- *Vice versa, if E is a stable extension of F, then $E_1 = E \cap A_1$ is a stable extension of F_1 and $E_2 = E \cap A_2$ a stable extension of the (E_1, R_3)-reduct of F_2.*

We can utilize this in order to find properties of repairs. Any stable extension contains a stable extension of F_1. So F_1 has to be consistent wrt. credulous reasoning. This means when trying to find repairs for F, one may start with repairs of F_1. In the following, we show that one can extend repairs of F_1 even to *minimal* repairs of F.

Proposition 5.13. *Let (F_1, F_2, R_3) be a splitting of the AF $F = (A_1 \cup A_2, R_1 \cup R_2 \cup R_3)$ where $F_1 = (A_1, R_1)$ and $F_2 = (A_2, R_2)$. If $\mathcal{S}_1$ is a minimal stb-cred-diagnosis of F_1, then there is a minimal stb-cred-diagnosis $\mathcal{S}$ of F with $\mathcal{S}_1 \subseteq \mathcal{S}$.*

Proof. Let $\mathcal{S}_1$ be a minimal *stb-cred*-diagnosis of F_1. Let E_1 be a stable extension of $(F_1)_{\mathcal{S}_1}$. Now consider $F_2^{E_1,R_3}$, i. e. the (E_1, R_3)-reduct of F_2. If $F_2^{E_1,R_3}$ possesses a stable extension, then we are done. If this is not the case, we need to be careful since two different extensions of $(F_1)_{\mathcal{S}_1}$ might induce two reducts where the minimal repairs are in a subset relation; so we cannot just take the minimal repair of $F_2^{E_1,R_3}$. So assume for the moment there is an extension of $(F_1)_{\mathcal{S}_1}$ such that the reduct is not self-controversial and let

$$\mathcal{S}_2 \in \min_{E_1 \in stb\left((F_1)_{\mathcal{S}_1}\right)} \left\{ \mathcal{S} \mid \mathcal{S} \text{ is a minimal repair of } F_2^{E_1,R_3} \right\}$$

where we quickly observe that the minimum exists since we are dealing with finite AFs. Now let $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$. We claim that $\mathcal{S}$ is a minimal diagnosis of F. For this, we observe that we cannot remove any element from $\mathcal{S}_1$ since this was assumed to be a minimal diagnosis of F_1 and from Theorem 5.12 we know that it needs to possess a stable extension. Moreover, if we are given $\mathcal{S}_2$ and the extension E_1 in which the minimum is attained, we see that $\left(F_2^{E_1,R_3}\right)_{\mathcal{S}_2}$ possesses a stable extension ensuring that $\mathcal{S}$ is a diagnosis. Now, minimality is due to construction of $\mathcal{S}_2$. Finally, if the reduct is self-controversial for any extension of $(F_1)_{\mathcal{S}_1}$, we can set $\mathcal{S}_2 = A_2$ for argument-based and $\mathcal{S}_2 = R_2$ for attack-based diagnoses. □

Even though being rather simple, the most important observation in the previous proposition was that F_1 needs to be consistent in order for F to be consistent. We can phrase this observation in terms of $SI^A_{min}(F)$ and $SI^R_{min}(F)$, i. e. the strongly inconsistent sets of arguments and attacks.

Proposition 5.14. *If (F_1, F_2, R_3) is a splitting of $F = (A, R) = (A_1 \cup A_2, R_1 \cup R_2 \cup R_3)$, then*

- $SI^A_{min}(F_1, stb, cred) \subseteq SI^A_{min}(F, stb, cred)$,
- $SI^R_{min}(F_1, stb, cred) \subseteq SI^R_{min}(F, stb, cred)$.

Proof. We prove the first item only. Assume $\mathcal{H}_1 \in SI_{min}(F_1, stb, cred) = SI_{min}(F_1)$. Then, for any set $\mathcal{H}'_1$ with $\mathcal{H}_1 \subseteq \mathcal{H}'_1 \subseteq A_1$ the AF $\left(\mathcal{H}'_1, (R_1)_{|\mathcal{H}'_1}\right)$ has no stable extension. Due to Theorem 5.12, this implies $\left(\mathcal{H}'_1, (R_1)_{|\mathcal{H}'_1}\right) \cup F_2 \cup R_3$ has no stable extension, either. Since we can also apply the splitting theorem after moving to a sub-AF of F_2, we see that the AF

$$\left(\mathcal{H}'_1, (R_1)_{|\mathcal{H}'_1}\right) \cup \left(\mathcal{H}'_2, (R_2)_{|\mathcal{H}'_2}\right) \cup R_3$$

is inconsistent for any $\mathcal{H}'_1$ with $\mathcal{H}_1 \subseteq \mathcal{H}'_1 \subseteq A_1$ and any $\mathcal{H}'_2$ with $\mathcal{H}'_2 \subseteq A_2$. Thus, $\mathcal{H}$ is a strongly inconsistent sets of arguments of F. Minimality can be inferred from the splitting theorem in a similar way. Hence, $\mathcal{H}_1 \in SI^A_{min}(F)$. □

Let us now assume we are given $SI^A_{min}(F_1)$. Due to the hitting set duality, i. e. Propositions 4.8 and 4.9 we find a minimal diagnosis of F_1 by removing a minimal hitting set $\mathcal{S}_1$ of $SI^A_{min}(F_1)$. In general it is not quite clear whether we can now extend $\mathcal{S}_1$ to a *minimal* hitting set $\mathcal{S}$ of $SI^A_{min}(F)$. Due to $SI^A_{min}(F_1) \subseteq SI^A_{min}(F)$ we can surely extend $\mathcal{S}_1$ to a hitting set of $SI^A_{min}(F)$, but minimality is not clear. We can however prove it via Proposition 5.13:

Proposition 5.15. *Let (F_1, F_2, R_3) be a splitting of $F = (A, R) = (A_1 \cup A_2, R_1 \cup R_2 \cup R_3)$. If $\mathcal{S}_1$ is a minimal hitting set of $SI^A_{min}(F_1, stb, cred)$ ($SI^R_{min}(F_1, stb, cred)$), then there is a minimal hitting set $\mathcal{S}$ of $SI^A_{min}(F, stb, cred)$ ($SI^R_{min}(F, stb, cred)$) with $\mathcal{S}_1 \subseteq \mathcal{S}$.*

Proof. Due to Propositions 4.8 and 4.9, $\mathcal{S}_1$ is a minimal hitting set of $SI^A_{min}(F_1, stb, cred)$ ($SI^R_{min}(F_1, stb, cred)$) iff $\mathcal{S}_1$ is a minimal diagnosis of F_1. Due to Proposition 5.13 there is a minimal diagnosis $\mathcal{S}$ of F with $\mathcal{S}_1 \subseteq \mathcal{S}$. Again due to Propositions 4.8 and 4.9, $\mathcal{S}$ is a minimal hitting set of $SI^A_{min}(F, stb, cred)$ ($SI^R_{min}(F, stb, cred)$). □

We want to mention that the situation differs when considering sceptical reasoning. If F_1 is such that $stb(F_1) \neq \emptyset$ but no argument is sceptically accepted, then it might still be the case that F possesses a sceptically accepted argument due to F_2. So consistency of F_1 is not a necessary condition anymore. It is, however, *almost* sufficient:

Proposition 5.16. *Let (F_1, F_2, R_3) be a splitting of $F = (A, R) = (A_1 \cup A_2, R_1 \cup R_2 \cup R_3)$ where $F_1 = (A_1, R_1)$ and $F_2 = (A_2, R_2)$. If F_1 is consistent wrt. sceptical reasoning and stable semantics, then so is F iff there is at least one extension $E_1 \in stb(F_1)$ such that $stb\left(F_2^{E_1,R_3}\right) \neq \emptyset$.*

Proof. Immediate from Theorem 5.12. □

So far, our investigation was restricted to stable semantics. The reason is quite simple: Theorem 5.12 is based on the (E_1, R_3)-reduct which requires further adjustments to obtain the desired result for other semantics. The intuitive reason is that there might be arguments $a \in A$ which are neither in a σ-extension E nor attacked by E if we consider $\sigma \neq stb$ (Baumann, 2011). Formally, we consider the set of *undefined* arguments wrt. E (also known as *undecided* arguments) as follows.

Definition 5.17. If $F = (A, R)$ is an AF, σ any semantics and $E \in \sigma(F)$, then the set of *undefined* arguments wrt. E is

$$U_E = \{b \in A \mid b \notin E, \nexists a \in E : (a, b) \in R\}.$$

To obtain the splitting result for other semantics, we use the (E_1, R_3)-reduct as usual, but in addition we introduce dummy attacks to those arguments in F_2 which are attacked by an undefined argument from F_1 wrt. E_1. More precisely, the (S, L)-modification is defined as follows:

Definition 5.18 (Modification)**.** Let $F_2 = (A_2, R_2)$ be an AF and A_1 such that $A_1 \cap A_2 = \emptyset$. Let $S \subseteq A_1$ and $L \subseteq A_1 \times A_2$. The (S, L)-modification of F_2, denoted by F_2, is the AF

$$mod_{S,L}(F_2) = (A_2, R_2 \cup \{(b, b)\} \mid \exists a \in S : (a, b) \in L).$$

Please observe that we are not interested in $mod_{U_{E_1},R_3}(F_2)$, but in $mod_{U_{E_1},R_3}\left(F_2^{E_1,R_3}\right)$, so we consider the (U_{E_1}, R_3)-modifcation of the (E_1, R_3)-reduct of F_2.

Theorem 5.19. *[(Baumann, 2011)] Let* (F_1, F_2, R_3) *be a splitting of the AF* F*, i.e. we have* $F = (A, R) = (A_1 \cup A_2, R_1 \cup R_2 \cup R_3)$ *where* $F_1 = (A_1, R_1)$ *and* $F_2 = (A_2, R_2)$*. Let* $\sigma \in \{stb, ad, pr, co, gr\}$*.*

- *If* E_1 *is a* σ*-extension of* F_1 *and* E_2 *a* σ*-extension of the* (U_{E_1}, R_3)*-modification of* $F_2^{E_1,R_3}$*, then* $E_1 \cup E_2$ *is a* σ*-extension of* F*.*
- *Vice versa, if* E *is a* σ*-extension of* F*, then* $E_1 = E \cap A_1$ *is a* σ*-extension of* F_1 *and* $E_2 = E \cap A_2$ *a* σ*-extension of the* (U_{E_1}, R_3)*-modification of* $F_2^{E_1,R_3}$*.*

For stable semantics and credulous reasoning, the splitting theorem can be used to infer that consistency of F_1 is *necessary* for consistency of F. For the other semantics, it is not a necessary, but a *sufficient* condition.

Proposition 5.20. *Let* (F_1, F_2, R_3) *be a splitting of* $F = (A, R) = (A_1 \cup A_2, R_1 \cup R_2 \cup R_3)$*. Let* $\sigma \in \{ad, pr, co, gr\}$*. If* F_1 *is consistent wrt.* σ*, then* F *is consistent wrt.* σ *as well. This holds for both credulous and sceptical reasoning.*

Proof. Immediate from Theorem 5.19. □

The treatment of an AF using splitting is convenient since the structural properties induce strong results. It is thus not surprising, yet encouraging to see that this principle is capable of improving the investigation of diagnoses and repairs of AFs. As it turns out, splitting can be used to reduce the search space for repairs (see e.g. Proposition 5.20) or compute minimal diagnoses stepwise (as in Proposition 5.13). Moreover, splitting is also meaningful when looking for strongly inconsistent arguments resp. attacks. We believe this is a promising research direction for further investigation, including concrete algorithms to compute repairs.

5.3 Infinite AFs

Until now, our investigation was restricted to finite AFs, i.e. $F = (A, R)$ where A is a finite set of arguments (and thus R a finite set of attacks). Within this section we want to drop this restriction and investigate which results still hold. In recent times, infinite AFs receive more and more attention. One main reason was the observation that studying finite AFs only is a limitation from a theoretical, conceptual, as well as practical perspective (Baroni, Cerutti, Dunne, & Giacomin, 2013). In order to overcome this problem the authors introduced deterministic finite automaton which represent infinite structures and provided a rigorous study of their properties. To give an example, infinite set of arguments frequently occur in case of structured argumentation which usually rely on the evaluation of induced AFs (Besnard & Hunter, 2008). For instance, for rule-based argumentation formalisms like ASPIC (Prakken, 2010) we may obtain infinite AFs even if we have a finite number of rules only (Caminada & Oren, 2014; Strass, 2018). Consequently, results regarding diagnoses and repairs are worth studying in the non-finite case too.

As usual, when moving from the finite to the infinite case, we are concerned about *existence* and *uniqueness* of certain sets as this might not be clear anymore (Baumann & Spanring, 2015, 2017).

In order to keep this section concise, we focus on a few semantics σ only, i.e. we consider $\sigma \in \{co, pr, ss, gr\}$. We will see that *gr*-diagnoses play an important role similar to the finite case described in Section 3.

So let us start with the grounded one. In the finite case, we observed that an F is consistent wrt. grounded semantics iff there is at least one unattacked argument. Let us formally state that this is also the case for infinite AFs.

Proposition 5.21. *If $F = (A, R)$ is an infinite AF, then F possesses a non-empty grounded extension iff there is at least one unattacked argument.*

Proof. Observe that $G = \bigcup_{i=1}^{|A|} \Gamma_F^i(\emptyset)$ is the grounded extension of F, not only for finite AFs. Due to monotony of the characteristic function Γ_F, we have $G \neq \emptyset$ iff $\Gamma_F(\emptyset) \neq \emptyset$, i.e. iff there is an unattacked argument. □

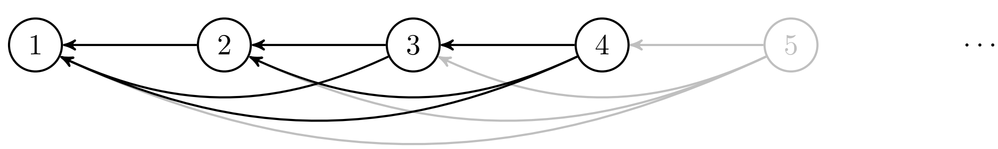
This also means we find *gr*-diagnoses as before, namely by removing arguments resp. attacks until at least one argument is unattacked. Hence:

Fact 5.22. *If $F = (A, R)$ is an infinite AF, then F possesses a*

- *argument-based grounded diagnosis iff it is not self-controversial,*
- *attack-based grounded diagnosis.*

Whether a *minimal* *gr*-diagnosis exists for a given AF $F = (A, R)$ is no longer trivial since we cannot just move from a diagnosis $\mathcal{S}$ to a minimal one $\mathcal{S}'$. Indeed, when considering argument-based diagnoses, there is no minimal *gr*-diagnosis in general.

Example 5.23. Consider the AF $F = (A, R) = (\mathbb{N}, \{(i, j) \mid i > j\})$:



It is easy to check that this AF possesses *gr*-diagnoses: Let $j \in \mathbb{N}$. Then, there is an attack $(i, j) \in R$ for each $i \in \mathbb{N}$ with $i > j$. So in order to obtain a *gr*-diagnosis of F we may remove each argument i with $i > j$, i.e. we let $\mathcal{S}_j = \{i > j \mid i \in \mathbb{N}\}$. This is a diagnosis since no argument attacks j within $F_{\mathcal{S}_j}$, so the grounded extension is non-empty. It is however not minimal since j was arbitrary. More precisely, for any $j' > j$, the set $\mathcal{S}_{j'}$ is a *gr*-diagnosis of F as well satisfying $\mathcal{S}_{j'} \subseteq \mathcal{S}_j$. Since we can always move to a smaller diagnosis, we see that there is no minimal one.

In case of attack-based diagnoses, the situation is different. We can guarantee the existence of minimal repairs and their structure is rather simple as we see in the proof of the following proposition.

Proposition 5.24. *If $F = (A, R)$ is an infinite AF, then there is a minimal attack-based gr-diagnosis.*

Proof. Assume the *gr* extension of F is empty. Let $a \in A$ and $B = \{(b, a) \in R \mid b \in A\}$. Observe $B \neq \emptyset$ due to Proposition 5.21. We claim that B is a minimal *gr*-diagnosis of F. Since a is unattacked in F_B, it is a diagnosis due to Proposition 5.21. Now consider $\mathcal{S} \subsetneq B$. Then, $F_{\mathcal{S}}$ contains an attack of the form (c, a) with $(c, a) \in B \setminus \mathcal{S}$. Hence, a is attacked in $F_{\mathcal{S}}$. Since the grounded extension of F is empty and $\mathcal{S} \subseteq B$, any other argument in $F_{\mathcal{S}}$ is attacked as well. We thus see that $F_{\mathcal{S}}$ is no *gr*-diagnosis of F. Since $\mathcal{S}$ was arbitrary, B is minimal. □

Now let $\sigma \in \{pr, co\}$. It has been noted (Dung, 1995) that those semantics are also universally defined when considering infinite AFs. Thus given a *gr*-repair, we also have a σ-repair as in the finite case.

Proposition 5.25. *Let $F = (A, R)$ be an infinite AF and $\sigma \in \{pr, co\}$. If $\mathcal{S}$ is a gr-diagnosis of F, then $\mathcal{S}$ is a σ-diagnosis of F as well. This holds for both reasoning modes.*

Proof. Consider complete semantics and sceptical reasoning. Since the *gr*-extension is the least *co*-extension and nonempty for $F_{\mathcal{S}}$ by assumption, we have that at least one argument, say $a \in A$, is sceptically accepted wrt. complete semantics. This also implies the claim for credulous reasoning. Now consider $\sigma = pr$. Recall that if E is a preferred extension, then it is a complete extension as well. Hence,

$$a \in \bigcap co(F_{\mathcal{S}}) \subseteq \bigcap pr(F_{\mathcal{S}})$$

implying a is also sceptically accepted for preferred semantics. This finishes our proof. □

Please observe that it is not trivial in general whether we can turn the σ-diagnosis we found in Proposition 5.25 into a minimal one. We note that Example 5.23 shows that minimal argument-based σ-diagnoses do not necessarily exist for $\sigma \in \{pr, co\}$ since this example works analogously for $\sigma \in \{pr, co\}$.

Let us conclude this section we a short remark regarding so-called *finitary* AFs. In a finitary AF, each argument is only allowed to have finitely many attackers. This solves nearly all issues we had during this section with *minimal* diagnoses at once: Now, any *gr*-diagnosis is necessarily finite and given a finite diagnosis, we can easily move to a minimal one. Hence, we obtain existence of minimal complete and preferred diagnoses. Moreover, any finitary AF possesses as semi-stable extension (Weydert, 2011; Baumann & Spanring, 2015), so the same can be guaranteed here.

Definition 5.26. The AF $F = (A, R)$ is called *finitary* if $\{a \in A \mid (a, b) \in R\}$ is finite for each $b \in A$.

Theorem 5.27. *Let F be finitary. Any minimal gr-diagnosis of F is finite. If $\mathcal{S}$ is a gr-diagnosis of F, then there is a minimal gr-diagnosis $\mathcal{S}'$ of F with $\mathcal{S}' \subseteq \mathcal{S}$. If $\mathcal{S}'$ is a minimal gr-diagnosis of F, then there is a minimal σ-diagnosis $\mathcal{S}''$ of F with $\mathcal{S}'' \subseteq \mathcal{S}'$ for any $\sigma \in \{co, pr, ss\}$.*

Proof. The claims about *gr*-diagnoses are clear. Given a finite *gr*-diagnosis, apply Proposition 5.25 to obtain a finite σ-diagnosis for $\sigma \in \{co, pr\}$. Due to finiteness, one can turn this diagnosis into a minimal one. Moreover, due to Weydert (2011), any finitary AF F possesses a semi-stable extension. Now if $F_{\mathcal{S}'}$ is a minimal *gr*-repair, then there is at least one unattacked argument. It is easy to see that this occurs in each semi-stable extension, so $\mathcal{S}'$ is a *ss*-diagnosis as well. Again we can move to a minimal one due to finiteness. □

When investigating infinite instead of finite AFs, one needs to accept the possibility that certain results are not conveyed. In our case, the investigation showed that argument-based diagnoses are not well-behaved and minimality can almost never be guaranteed. It is worth mentioning that a reasonably simple example suffices to demonstrate this for all considered semantics and both reasoning modes. When considering attack-based diagnoses minimal *gr*-diagnoses always exist (and thus, minimal *scep*-*co*-diagnoses as well) and their structure is quite simple.

6. Computational Complexity

In this section we discuss the computational complexity of two decision problems that naturally arise, namely the existence problem as well as the verification problem regarding minimal repairs. To keep this section varied and within a reasonable space, we mostly restrict our investigation to $\sigma \in \{gr, co, pr, stb\}$. We also consider argument-based diagnoses only. We believe that this suffices to demonstrate the reader how to derive the expected results.

We assume the reader to be familiar with the polynomial hierarchy. Furthermore, we consider the differences classes $\mathsf{D}^p_m = \{L_1 \cap L_2 \mid L_1 \in \Sigma^p_m,\ L_2 \in \Pi^p_m\}$ as defined by Papadimitriou (1994). In the following let σ be a semantics and $\diamond$ one of the considered reasoning modes, i. e. $\diamond \in \{cred, scep\}$.

Ex-Min-Repair$_{\sigma,\diamond}$
Input: An AF F
Output: TRUE iff there is a minimal σ-$\diamond$-diagnosis for F

Ver-Min-Repair$_{\sigma,\diamond}$
Input: $(F, \mathcal{S})$ where $F = (A, R)$ is an AF and $\mathcal{S} \subseteq A$
Output: TRUE iff $\mathcal{S}$ minimal σ-$\diamond$-diagnosis for F

We start with the problem of deciding whether a minimal repair exists. As we know from Theorems 3.16, 3.19 and Fact 3.21 it suffices to perform a simple syntactical check, which can be done in linear time. In particular, it does not matter which reasoning mode is considered. We thus find:

Proposition 6.1. *For $\sigma \in \{ad, gr, eg, il, ss, pr, co, stb\}$ and $\diamond \in \{cred, scep\}$ the problem* Ex-Min-Repair$_{\sigma,\diamond}$ *can be solved in linear time.*

We turn to the problem Ver-Min-Repair$_{\sigma,\diamond}$, which will turn out to be more demanding in most cases. Hardness results of the subsequent subsections are oftentimes adjustments of existing constructions (Dvorák & Dunne, 2018). Membership results are a corollary of the following observation.

Proposition 6.2. *Let $\diamond \in \{cred, scep\}$. Let σ be any semantics. If deciding whether an AF F is consistent wrt. σ and $\diamond$ is in Σ_m^p for any integer $m \geq 1$, then* Ver-Min-Repair$_{\sigma,\diamond}$ *is in D_m^p. If deciding whether an AF F is consistent wrt. σ and $\diamond$ is in Π_m^p for any integer $m \geq 1$, then* Ver-Min-Repair$_{\sigma,\diamond}$ *is in Π_{m+1}^p.*

Proof. If checking consistency is in Σ_m^p, then we check in Σ_m^p whether we are given a repair $F_{\mathcal{S}}$ and for minimality we non-deterministically guess a subset $\mathcal{S}'$ and verify that $F_{\mathcal{S}'}$ is not a repair in Π_m^p, which needs to be the case for every $\mathcal{S}' \subseteq \mathcal{S}$. In summary, this procedure is in D_m^p. The other case is similar. □

The decision problem Ver-Min-Repair$_{\sigma,scep}$ involves checking all subsets of $\mathcal{S}$. In the first case of Proposition 6.2 this results in moving to the corresponding difference class. This happens e.g. in Theorem 6.5 below. However, the reason is that the underlying decision problem, i. e. "is the framework consistent wrt. credulous reasoning?" is in Σ_m^p for an m (in case of Theorem 6.5 we have $m = 1$). Thus, verifying inconsistency is in Π_m^p. Here, verifying inconsistency for all subsets $\mathcal{S}'$ of $\mathcal{S}$ does not induce a quantifier alternation. That is why the upper bounds are (under standard assumptions) below Π_{m+1}^p. In case of sceptical reasoning, however, we naturally face coNP resp. Π_2^p lower bounds. Here, oftentimes the second case of Proposition 6.2 applies.

More importantly, most of the corresponding hardness results can indeed be shown as we will see via the subsequent constructions.

6.1 Grounded Semantics

Given an AF (and a potential diagnosis), we know that the grounded extension is non-empty if and only if there is an argument which is not attacked. Thus, verifying that a given set is a *gr*-diagnosis is quite easy. It turns out that minimality is tractable as well.

Proposition 6.3. *For $\diamond \in \{cred, scep\}$, the problem* Ver-Min-Repair$_{gr,\diamond}$ *is in P.*

Proof. If $F_{\mathcal{S}}$ contains no unattacked argument, we reject.
Argument-based: If $F_{\mathcal{S}}$ contains an unattacked argument, we check whether this is the case for each $F_{\mathcal{S}\setminus\{\alpha\}}$ with $\alpha \in \mathcal{S}$. If this is not the case, then $\mathcal{S}$ must be minimal.
Attack-based: $\mathcal{S}$ is minimal iff F itself is inconsistent and $\mathcal{S}$ is of the form $\{(b, a) \in R \mid b \in A\}$ for an $a \in A$. □

We want to mention that we can even compute *all* *gr*-diagnoses in P. We believe this observation is relevant since the grounded repairs play an essential role as the results from Section 3 suggest. Assume we are given the AF $F = (A, R)$ with $gr(F) = \{\emptyset\}$. Since a grounded diagnosis needs to ensure that at least one argument $a \in A$ is not attacked anymore, we can successively look at any $a \in A$ and consider $\mathcal{S} = \{b \in A \mid (b, a) \in R\}$ for argument-based diagnoses and $\mathcal{S} = \{(b, a) \in R \mid b \in A\}$ for attack-based diagnoses. If $\mathcal{S}$ is a minimal *gr*-diagnosis (verification is in P due to Proposition 6.3), we add $\mathcal{S}$ to our list, otherwise we delete it. Since there are at most $|A|$ *gr*-diagnoses, this procedure is in P. So:

Proposition 6.4. *Computing all gr-diagnoses of a given AF F can be done in P.*

Now, even though finding a σ-diagnosis may become rather hard depending on σ, we can efficiently compute all gr-diagnoses and then utilize Lemmata 3.15, 3.18 and 3.22 in order to reduce the search space. This approach explains the central role of grounded semantics. In a nutshell, the gr-repairs can be seen as a (polynomial time computable) starting point in order to find minimal repairs for other semantics. A thorough investigation of this approach is part of future work. Moreover, the investigation of further subclasses of AFs seems rather promising considering computational complexity. For example, the ones we considered trivialize credulous diagnoses in almost any case. Other restrictions might ensure tractability of certain problems we considered here while being less restrictive.

6.2 Universally Defined Semantics

Considering the computational complexity of different reasoning problems in AFs, it is quite unsurprising that Ver-Min-Repair$_{\sigma,cred}$ is intractable for most semantics σ as it requires checking whether a *non-empty* extension exists. Due to the additional minimality check we require, our problem turns out to be in the corresponding difference class.

Theorem 6.5. Ver-Min-Repair$_{\sigma,cred}$ *is* D_1^p*-complete for* $\sigma \in \{pr, co\}$.

Proof. Membership is due to Proposition 6.2. For hardness, we reduce the problem MC which is defined as follows: Given a formula Φ in 3-CNF, i. e. $\Phi = C_1 \wedge, \ldots, \wedge C_r$. We identify a formula with a set of clauses, i. e. $\Phi = \{C_1, \ldots, C_r\}$. Let ϑ be a subformula, i. e. ϑ is w. l. o. g. of the form $\vartheta = \{C_1, \ldots, C_s\}$ with $\{C_1, \ldots, C_s\} \subseteq \{C_1, \ldots, C_r\}$. The problem MC is given via

MC
Input: (Φ, ϑ) where Φ is a formula in 3-CNF and $\vartheta \subseteq \Phi$
Output: TRUE iff ϑ is satisfiable, but any formula ϑ' with $\vartheta \subsetneq \vartheta' \subseteq \Phi$ is not.

The problem MC is D_1^p-complete (Papadimitriou, 1994).

So let Φ be as above. Let $x_1, ..., x_n$ be the literals occurring in Φ, set $\neg\neg x_i = x_i$. We can prove hardness utilizing a minor adjustment of the standard construction (Dvorák & Dunne, 2018) depicted in Figure 2 (a). Let F be the AF $F = (A, R)$ with $A = \{x_1, \neg x_1, \ldots, x_n, \neg x_n, C_1, \ldots, C_r, \Phi, \overline{\Phi}\}$ and

$$\begin{aligned} R = &\{(x_i, \neg x_i) \mid i \in \{1, \ldots, n\}\} \cup \{(\neg x_i, x_i) \mid i \in \{1, \ldots, n\}\} \\ &\cup \{(x_i, C_j) \mid x_i \text{ occurs in } C_j\} \cup \{(C_j, \Phi) \mid j \in \{1, \ldots, r\}\} \\ &\cup \{(\Phi, \overline{\Phi})\} \cup \{(\overline{\Phi}, x_i) \mid x_i \text{ occurs in } \Phi\} \cup \{(\overline{\Phi}, C_j) \mid j \in \{1, \ldots, r\}\}. \end{aligned}$$

Consider $\sigma = ad$. It is well-known that Φ is satisfiable iff there is a non-empty admissible extension of the framework depicted in Figure 2 (a). The reader may verify that a non-empty admissible extension E needs to contain some of the X arguments. In order to defend them, $\Phi \in E$ is required. In order to find an admissible set of arguments defending Φ, the formula needs to be satisfiable. Now let ϑ be a subformula of Φ, i. e. ϑ is w. l. o. g. of the form $\vartheta = \{C_1, \ldots, C_s\}$. Then, (Φ, ϑ) is "yes" instance of MC iff $(F, \mathcal{S})$ with $\mathcal{S} = \{C_{s+1}, \ldots, C_r\}$ is a "yes" instance of Ver-Min-Repair$_{ad,cred}$.

Clearly, any framework possesses a non-empty admissible extension iff this is the case for $\sigma = co$ and $\sigma = pr$ □

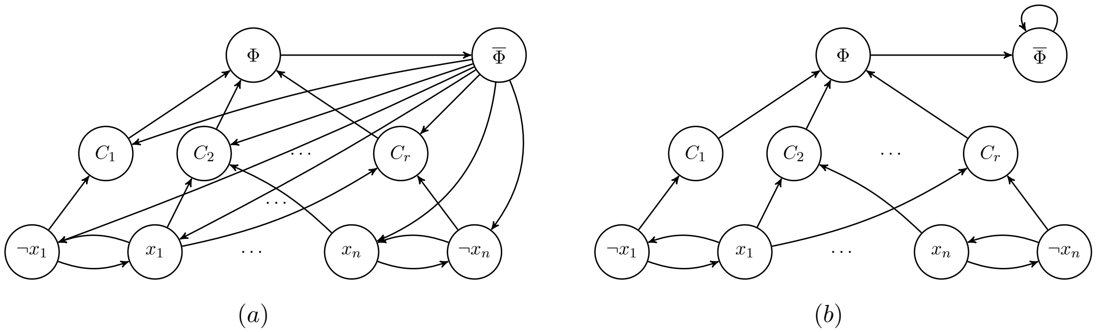


Figure 2: Standard Constructions for 3-SAT reductions

We turn to sceptical reasoning. Recall that the unique grounded extension of an AF F is complete as well. Moreover we have $gr(F) \subseteq \bigcap co(F)$. Hence, any framework F possesses a sceptically accepted argument wrt. grounded semantics if and only if this is the case for complete semantics. Hence, applying Proposition 6.3 yields:

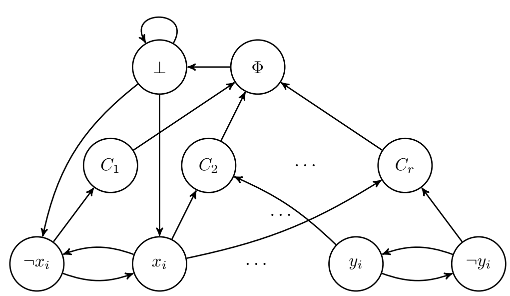
Corollary 6.6. Ver-Min-Repair$_{co,scep}$ *is in* P.

Now we consider $\sigma = pr$. Recall that deciding whether an argument is sceptically accepted is Π_2^p-complete (Dvorák & Dunne, 2018). Thus, given a framework F and a set $\mathcal{S} \subseteq A$ of arguments, the decision problem Ver-Min-Repair$_{pr,scep}$ involves checking whether for *all* $\mathcal{S}'$ with $\mathcal{S}' \subseteq \mathcal{S}$ the framework $F_{\mathcal{S}'}$ does *not* possess a sceptically accepted argument. Since the latter check is in Σ_2^p for each $\mathcal{S}'$, we immediately see a Π_3^p upper bound for Ver-Min-Repair$_{pr,scep}$ due to Proposition 6.2. So the main work for the following theorem is the lower bound:

Theorem 6.7. Ver-Min-Repair$_{pr,scep}$ *is* Π_3^p*-complete.*

Proof. Membership is due to Proposition 6.2. Recall the construction from Dvorák and Dunne (2018) with the property that the AF sceptically accepts an argument wrt. preferred semantics if and only if a formula $\Phi = \forall Y \, \exists X : \phi(X, Y)$ in CNF evaluates to true.

$\mathcal{H}_{\Pi_2^p}$:



In order to prove hardness in Π_3^p for our problem we the task is to simulate an additional quantifier. This, however, comes natural since the decision problem Ver-Min-Repair$_{pr,scep}$ involves consideration of *all* subsets $\mathcal{S}'$ of a given set $\mathcal{S} \subseteq A$ of arguments.

So let us assume we are given a formula $\Psi = \exists Z \, \forall Y \, \exists X : \psi(X, Y, Z)$ in CNF. we augment the construction from Dvorák and Dunne (2018) with the intention that Ψ evaluates to true if and only if $(F, \mathcal{S})$ is a "no" instance of Ver-Min-Repair$_{pr,scep}$ (that is, there exists a subset $\mathcal{S}'$ with $\mathcal{S}' \subsetneq \mathcal{S}$ such that $F_{\mathcal{S}'}$ possesses a sceptically accepted argument wrt. preferred semantics).

We augment the construction $\mathcal{H}_{\Pi_2^p}$. We will construct an AF $F = (A, R)$ with a set $\mathcal{S} \subseteq A$ of arguments with the following properties:

- $F_{\mathcal{S}}$ itself is consistent, i. e. there is a sceptically accepted argument,
- subsets of $\mathcal{S}$, i. e. the sets $\mathcal{S}' \subseteq \mathcal{S}$ may correspond to assignments to the Z-variables,
- there is one $\mathcal{S}'$ with $\mathcal{S}' \subsetneq \mathcal{S}$ such that $F_{\mathcal{S}'}$ is consistent if and only if the formula $\Psi = \exists Z\, \forall Y\, \exists X : \psi(X, Y, Z)$ evaluates to true.

The first and the last item together ensure that $(F, \mathcal{S})$ is a "yes" instance of the problem Ver-Min-Repair$_{pr,scep}$ if and only if the formula Ψ is false.

Before depicting and explaining our construction, we name all arguments occurring in the AF. We hope this helps readability of the proof. Which attacks we include will be explained later. Our framework is $F = (A, R)$ with

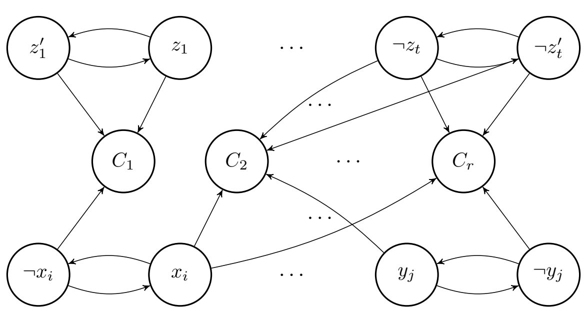
$$A = \{x_1, \neg x_1, \ldots, x_n, \neg x_n, y_1, \neg y_1, \ldots, y_m, \neg y_m, z_1, \neg z_1, z_1', \neg z_1', \ldots, z_t, \neg z_t, z_t', \neg z_t', C_1, \ldots, C_r, D_{1,1}, \ldots, D_{1,4}, \ldots, D_{t,1}, \ldots, D_{t,4}, \Phi, \top, \bot\}$$

Moreover, we set

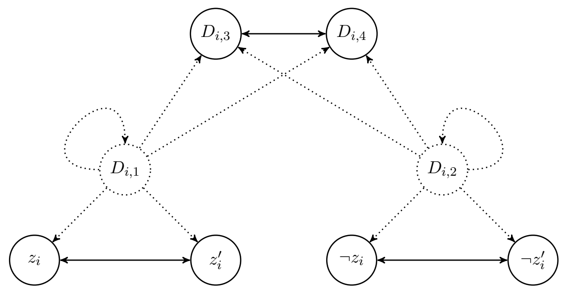
$$\mathcal{S} = \{D_{1,1}, D_{1,2}, \ldots, D_{t,1}, D_{t,2}\}$$

We will see that sets $\mathcal{S}'$ with $\mathcal{S}' \subseteq \mathcal{S}$ induce AFs $F_{\mathcal{S}'}$ with the intuition that we assign values to Z variables.

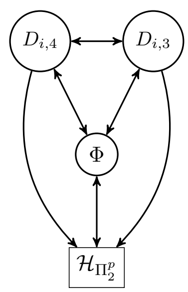
So first we consider Z variables $z_1, \neg z_1, \ldots, z_t, \neg z_t$ which attack the $C_1, \ldots, C_r$ in the natural way: We have $(z_j, C_i) \in R$ iff z_j occurs in the clause C_i and $(\neg z_j, C_i) \in R$ iff $\neg z_j$ occurs in the clause C_i. We also consider copies $z_1', \neg z_1', \ldots, z_t', \neg z_t'$.



The reason for the copies $z_1', \ldots, \neg z_t'$ is to ensure that the Z arguments themselves are not sceptically accepted. Now consider the following gadget, which will be included for any Z variable.



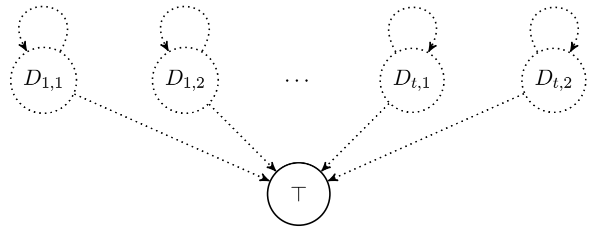
The dummy arguments $D_{i,1}$ and $D_{i,2}$ as well as their attacks are depicted with dotted lines to illustrate that they do *not* occur in $F_{\mathcal{S}}$, as those arguments belong to $\mathcal{S}$. Augmenting $F_{\mathcal{S}}$ with $D_{i,1}$, for example, ensures that z_i and z_i' are never defended and thus occur in no preferred extension. Hence, this choice corresponds to letting z_i be *false*. The role of the $D_{i,3}$ and $D_{i,4}$ becomes apparent considering the following arguments:



Now, for any $i \in \{1, \ldots, t\}$ we observe: Since $D_{i,3}$ and $D_{i,4}$ attack all arguments in $\mathcal{H}_{\Pi_2^p}$ we have that $\{D_{1,3}, \ldots, D_{t,3}\}$ and $\{D_{1,4}, \ldots, D_{t,4}\}$ are two preferred extensions for any $F_{\mathcal{S}'}$ with $\mathcal{S}' \subseteq \mathcal{S}$ which contains *both* $D_{i,1}$ *and* $D_{i,2}$, i. e. in $F_{\mathcal{S}'}$ occur *neither* $D_{i,1}$ *nor* $D_{i,2}$. Hence, the intersection $\bigcap pr(F_{\mathcal{S}'})$ is empty. We thus see: A framework $F_{\mathcal{S}'}$ with $\mathcal{S}' \subseteq \mathcal{S}$ can only possess a sceptically accepted argument if for each $i \in \{1, \ldots, t\}$, $D_{i,1}$ or $D_{i,2}$ occur in $\mathcal{S}'$.

Now assume this is given, i. e. we have a framework $F_{\mathcal{S}'}$ with $\mathcal{S}' \subseteq \mathcal{S}$ as described. Recall that the choice of the $D_{i,1}$ and $D_{i,2}$ naturally corresponds to a (partial) assignment $\omega : Z \to \{0, 1\}$. As in the original construction $\mathcal{H}_{\Pi_2^p}$ we see that Φ is sceptically accepted iff $\forall Y\, \exists X : \psi(X, Y, Z)$ evaluates to true under the assignment $\omega : Z \to \{0, 1\}$. In this case, $F_{\mathcal{S}'}$ is consistent. Since this applies to any $\mathcal{S}'$ of the form described above we see: Every $F_{\mathcal{S}'}$ with $\mathcal{S}' \subseteq \mathcal{S}$ is inconsistent iff for any assignment $\omega : Z \to \{0, 1\}$ the formula $\forall Y\, \exists X : \psi(X, Y, Z)$ evaluates to false iff the formula $\exists Z\, \forall Y\, \exists X : \psi(X, Y, Z)$ evaluates to false.

To summarize, we have established: $\exists Z\, \forall Y\, \exists X : \psi(X, Y, Z)$ is false iff $F_{\mathcal{S}'}$ is inconsistent for all $\mathcal{S}' \subsetneq \mathcal{S}$. The latter *nearly* means that $(F, \mathcal{S})$ for $\mathcal{S} = A \setminus H$ is a positive instance of Ver-Min-Repair$_{pr,scep}$. What we have left to do is to make sure that $F_{\mathcal{S}}$ itself is consistent, i. e. there is at least one sceptically accepted argument. The following final gadget does the job:



There is no other argument attacking $\top$. Hence, as long as no $D_{i,1}$ resp. $D_{i,2}$ argument is chosen, $\top$ is sceptically accepted. As soon as a proper superset of H is under consideration, $\top$ can never be defended and is thus rendered pointless. □

6.3 Stable Semantics

Let us now turn to our only example for collapsing semantics, namely $\sigma = stb$. If we are interested in credulous reasoning, we face a similar situation as in Theorem 6.5.

Theorem 6.8. Ver-Min-Repair$_{stb,cred}$ *is* D_1^p*-complete.*

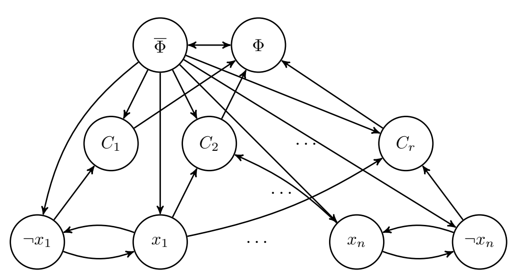
Proof. Membership is due to Proposition 6.2. For hardness, utilize the construction depicted in Figure 2 (b) as in the proof of Theorem 6.5. □

We turn to sceptical reasoning. Since finding a stable extension is NP-complete it is not hard to see that there is a coNP lower bound for sceptical reasoning. However, as the framework in question might collapse, we also need to verify that there is *at least one* stable extension of a given framework. The result is a D_1^p lower bound (Rahwan & Simari, 2009). Interestingly, however, this observation does not change anything in our case. The coNP lower bound is already responsible for Ver-Min-Repair$_{stb,scep}$ to have a Π_2^p lower bound. Given $H \subseteq \mathcal{K}$ the decision problem Ver-Min-Repair$_{stb,scep}$ involves checking whether *all* sets H' with $H \subseteq H' \subseteq \mathcal{K}$ do *not* possess any sceptically accepted argument. Since the latter test has a NP lower bound, we have a Π_2^p lower bound for Ver-Min-Repair$_{stb,scep}$. More precisely:

Theorem 6.9. Ver-Min-Repair$_{stb,scep}$ *is* Π_2^p*-complete.*

Proof. Membership is due to Proposition 6.2. For hardness, recall how to prove that sceptical reasoning is coNP-complete for stable semantics. Given a formula $\Phi = \exists X : \phi(X)$ where ϕ is a formula over variables in $X = \{x_1, \dots, x_n\}$ in 3-CNF with $\phi(x) = C_1 \wedge \dots \wedge C_r$ recall the following construction $\mathcal{H}_\Phi$ from Rahwan and Simari (2009). It has the property that the AF accepts $\overline{\Phi}$ sceptically wrt. stable semantics if and only if $\Phi = \exists X : \phi(X)$ evaluates to false.

$\mathcal{H}_\Phi$:



We augment this construction in order to show that Ver-Min-Repair$_{stb,scep}$ is Π_2^p-hard. We proceed as in the proof of Theorem 6.7. We construct an AF $F = (A, R)$ and consider $\mathcal{S} \subseteq A$ as well as the induces sub-AF $F_{\mathcal{S}}$. We need to check whether in $F_{\mathcal{S}}$ any argument is sceptically accepted, while this is not the case for any $F_{\mathcal{S}}$ with $\mathcal{S}' \subsetneq \mathcal{S}$. Hence, we may also take possible subsets of $\mathcal{S}$ into account. Assume we are given $\Psi = \forall Y \, \exists X : \psi(X, Y)$ with X as above and $Y = \{y_1, \ldots, y_m\}$. Assume $\psi(X, Y) = \psi$ is in CNF with $\psi = C_1 \wedge \ldots \wedge C_r$. We will construct a sub-AF $F_{\mathcal{S}}$ with the following properties:

- $F_{\mathcal{S}}$ itself is consistent, i. e. there is a sceptically accepted argument,
- subsets of $\mathcal{S}$, i. e. the sets $\mathcal{S}' \subseteq \mathcal{S}$ may correspond to assignments to the Y-variables,
- there is no AF $F_{\mathcal{S}'}$ with $\mathcal{S}' \subsetneq \mathcal{S}$ which is consistent if and only if the given formula $\Psi = \forall Y \, \exists X : \psi(X, Y)$ evaluates to true.

The first and the last item together ensure that $(F, \mathcal{S})$ with $\mathcal{S}$ is a "yes" instance of Ver-Min-Repair$_{stb,scep}$ if and only if the formula Ψ is true.

Before depicting and explaining our construction we give the arguments A of the AF $F = (A, R)$ as we did in the proof of Theorem 6.7. We have

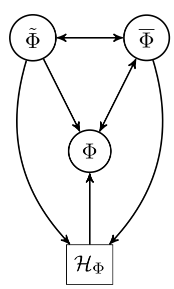
$$A = \{x_1, \neg x_1, \ldots, x_n, \neg x_n, y_1, \neg y_1, \ldots, y_m, \neg y_m, C_1, \ldots, C_r, \Phi, \overline{\Phi}, \tilde{\Phi}, y_1?, \ldots, y_m?, all?, \top\}$$

Moreover, our subset $\mathcal{S} \subseteq A$ is

$$\mathcal{S} = \{y_1, \neg y_1, \ldots, y_m, \neg y_m\}.$$

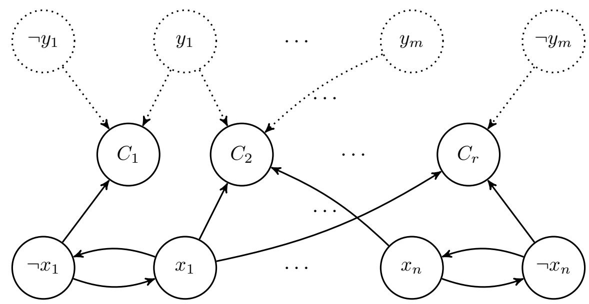
For the moment it suffices to observe that sets $\mathcal{S}'$ with $\mathcal{S}' \subseteq \mathcal{S}$ correspond to choosing y arguments.

Please observe that $\mathcal{H}_\Phi$ from above possesses a stable extension containing only $\overline{\Phi}$. Our first step is consideration of a similar argument which will be called $\tilde{\Phi}$. Similar to $\overline{\Phi}$, the argument $\tilde{\Phi}$ attacks all arguments. Moreover, $\tilde{\Phi}$ and $\overline{\Phi}$ attack each other.

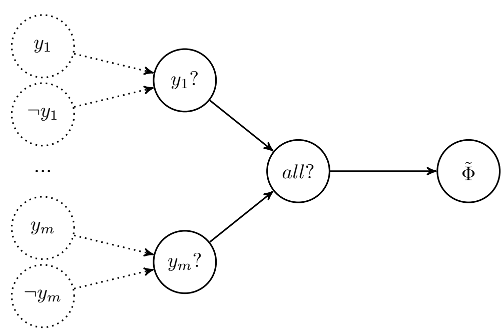


This framework has at least two stable extensions, namely $\overline{\Phi}$ and $\tilde{\Phi}$. Note the intended asymmetry between the two arguments: $\overline{\Phi}$ is attacked by Φ while $\tilde{\Phi}$ is not. The reason is as follows: The purpose of $\overline{\Phi}$ is as in the original construction to control whether there is a satisfying assignment to the given formula or not. This is why it needs to be attacked by Φ. However, $\tilde{\Phi}$ is utilized to render some sub-AFs $F_{\mathcal{S}'}$ with $\mathcal{S}' \subseteq \mathcal{S}$ inconsistent as we will see later. This is why there is no attack from Φ to $\tilde{\Phi}$.

As our next step, we consider arguments $y_1, \neg y_1, \ldots, y_m, \neg y_m$ which shall correspond to the Y-variables in the given formula $\Psi = \forall Y \, \exists X : \psi(X, Y)$. As already pointed out, they do not occur in the AF $F_{\mathcal{S}}$ as they belong to $\mathcal{S}$. So, a subset $\mathcal{S}'$ with $\mathcal{S}' \subseteq \mathcal{S}$ somewhat corresponds to a partial assignment $\omega : Y \to \{0,1\}$. Similar to the X arguments, they attack the arguments $C_1, \ldots, C_r$ in the natural way: We have $(y_j, C_i) \in R$ iff y_j occurs in the clause C_i and $(\neg y_j, C_i) \in R$ iff $\neg y_j$ occurs in the clause C_i. Note that the Y arguments do *not* attack each other.

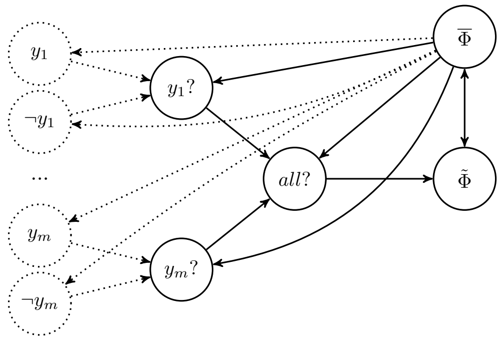


The Y arguments as well as their attacks are depicted with dotted lines to illustrate that they do *not* occur in $F_{\mathcal{S}}$. Now let us make sure that only intended subsets of $\{y_1, \neg y_1, \ldots, y_m, \neg y_m\}$ are relevant. Recall that they shall correspond to assignments $\omega : Y \to \{0,1\}$. Interestingly, we only need to prune away *partial* assignments, i. e. cases where there is an index j such that neither y_j nor $\neg y_j$ occurs in $F_{\mathcal{S}'}$. The case that both y_j and $\neg y_j$ occur in $F_{\mathcal{S}'}$ –actually not corresponding to a well-defined assignment– does no harm as we will explain later. Consider the following additional arguments and attacks:



Observe that the auxiliary argument "*all*?" attacks $\tilde{\Phi}$ only and not $\overline{\Phi}$. The meaning of this construction is as follows: In case for any j there is y_j or $\neg y_j$ occurring in $F_{\mathcal{S}'}$, each "y_j?" is attacked and hence, "*all*?" is defended from $\{y_1?, \ldots, y_m?\}$. It may thus occur in a stable extension. In this case, $\tilde{\Phi}$ does *not* occur in any stable extension. Otherwise, it does.

However, $\overline{\Phi}$ keeps attacking all arguments and is thus still a given possibility to find a stable extension (otherwise some of the Y or auxiliary arguments could be sceptically accepted):



We thus see the following: For any $F_{\mathcal{S}'}$ with $\mathcal{S}' \subseteq \mathcal{S}$ there is always a stable extension containing $\overline{\Phi}$. There is another stable extension containing $\tilde{\Phi}$ iff for at least one index j neither y_j nor $\neg y_j$ occur in $F_{\mathcal{S}'}$. Then, no argument is sceptically excepted. The AF $F_{\mathcal{S}'}$ is thus inconsistent.

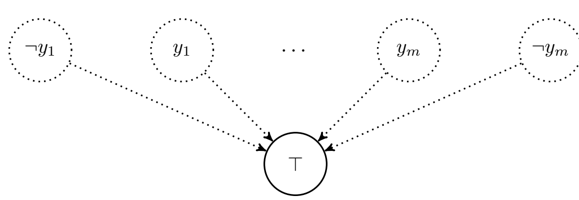
Now let us assume that for any index j there is *either* y_j *or* $\neg y_j$ occurring in $F_{\mathcal{S}'}$. Then we see that there are two cases:

1. The formula $\exists X : \psi(X, Y)$ evaluates to true under the corresponding assignment $\omega : Y \to \{0, 1\}$. Then, as in the construction from Rahwan and Simari (2009) there is a stable extension containing the corresponding X arguments as well as Φ, but not $\overline{\Phi}$. Hence, no argument is sceptically accepted.

2. The formula $\exists X : \psi(X, Y)$ evaluates to false under the corresponding assignment $\omega : Y \to \{0, 1\}$. Then, $\overline{\Phi}$ is sceptically accepted.

Hence, the formula $\forall Y \, \exists X : \psi(X, Y)$ is true iff the former case always occurs. Assume this is the case. Now consider a choice of the Y-variables which does not correspond to a well-defined assignment, i. e. there is j such that both y_j and $\neg y_j$ occur in $F_{\mathcal{S}'}$. It is clear that for this AF $F_{\mathcal{S}'}$ we also have the former case, i. e. no argument is sceptically accepted since this was already the case with only y_j or $\neg y_j$ occurring in the AF.

To summarize, we have established: $\forall Y \, \exists X : \psi(X, Y)$ is true iff $F_{\mathcal{S}'}$ is inconsistent for all $\mathcal{S}' \subsetneq \mathcal{S}$. The latter almost means that $(F, \mathcal{S})$ is a "yes" instance of Ver-Min-Repair$_{stb,scep}$. What we have left to do is to make sure that $F_{\mathcal{S}}$ itself is consistent, i. e. there is at least one sceptically accepted argument. The following final gadget does the job:



There is no other argument attacking $\top$. Hence, as long as no Y argument is chosen, $\top$ is sceptically accepted. As soon as a proper superset of $\mathcal{S}$ is under consideration, $\top$ is rendered pointless. $\square$

6.4 Subclasses

Let us briefly discuss the subclasses of AFs we mentioned in Section 3. If the AF in question is *symmetric*, Ex-Min-Repair$_{\sigma,cred}$ trivializes for any σ we considered in this article since irreflexivity ensures that the AF is not self-controversial. The problem Ver-Min-Repair$_{\sigma,cred}$ is trivial for $\sigma \in \{ad, stb, pr, co\}$ since any symmetric AF possesses a credulously accepted argument. Hence, $(F, \mathcal{S})$ is a positive instance of Ver-Min-Repair$_{\sigma,cred}$ (with $\sigma \in \{ad, stb, pr, co\}$) iff $\mathcal{S} = \emptyset$. It is easy to see that Ver-Min-Repair$_{gr,cred}$ can be solved in linear time. If the framework is *compact* or *acyclic*, there is inherently nothing to solve wrt. credulous reasoning. Sceptical reasoning is similarly easy.

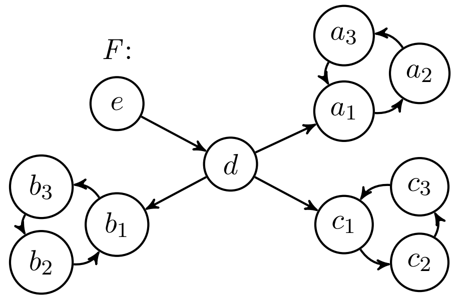
The results of this section show that the problem of verifying a given minimal diagnosis is rather demanding from a computational point of view. More precisely, in every case the upper bound given in Proposition 6.1 -which is rather generic and thus appears to be weak- is either a lower bound as well or there is a quite trivial reason why this is not the case. An example for the latter is $\sigma = ad$ and $\diamond = scep$. This observation emphasizes the importance of grounded diagnoses as they reduce the search space for *every* other case except $\sigma = stb$. In view of that, tractability of computing *all* grounded diagnoses is quite encouraging.

7. How to Repair? - A Short Case Study

As mentioned before, due to Lemmata 3.15, 3.18 and 3.22 we may reduce the search space for diagnoses as long as we are equipped with an already computed grounded one. If one is interested in *all* diagnoses, the notion of strong inconsistency in order to use the hitting set duality is proven to be useful. The aim of this section is to briefly demonstrate how to repair a given AF. We discuss both credulous and sceptical reasoning.

First let us consider an example with stable semantics. Let us start with credulous reasoning. It is well-known that in case of finite AFs the non-existence of acceptable positions implies the existence of odd-cycles. This means, by contraposition, one possible strategy for repairing AFs in case of stable semantics is to break odd-cycles. This approach corresponds to the minimal *stb*-repairs $F_{\{a\}}$, $F_{\{b\}}$ and $F_{\{c\}}$ from Example 1.2. Since possessing odd-cycles is not sufficient for the collapse of stable semantics further considerations are required. Indeed, in case of our running example, we have seen that eliminating the arguments e and f results in a minimal *stb*-repair, namely $F_{\{e,f\}}$, too. Regarding the principle of minimal change one may argue that breaking the odd-cycle in F has to be preferred over the latter strategy since less arguments are involved. The following slightly modified version of this example shows that this observation is not true in general. A further intensive study of this issue will be part of future work.

Example 7.1. Consider the following AF F. One may easily confirm that there are 9 minimal *cred-stb*-diagnoses, namely $\{a_i, b_j, c_k\}$ with $i, j, k \in \{1, 2, 3\}$. They comply with the idea to break all odd loops of the given AF. However, $\{e\}$ is a minimal *cred-stb*-diagnosis as well, and arguably the most immediate one.

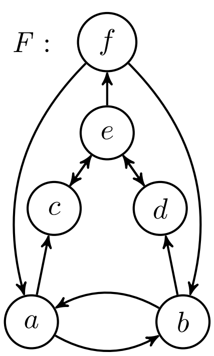


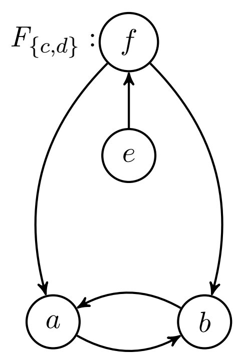
As the reader may have already observed, the same applies to attack-based diagnoses. For example, $\{(a_1, a_2), (b_1, b_2), (c_1, c_2)\}$ is an attack-based diagnosis breaking all odd loops. However, consideration of $\{(e, d)\}$ suffices.

The subsequent example considers a semantical defect wrt. preferred semantics which is tackled via grounded repairs.

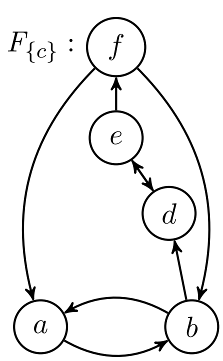
Example 7.2. The following AF F exemplifies a situation where preferred semantics do not possess any sceptically accepted argument. More precisely, $\bigcap pr(F) = \emptyset$ due to

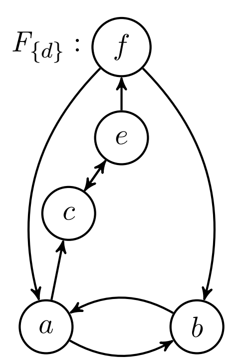
$$pr(F) = \{\{a, e\}, \{b, e\}, \{c, d, f\}\}.$$





Our goal is to find a minimal *scep*-*pr*-diagnosis $\mathcal{S}$, i.e. a set $\mathcal{S}$ such that $\bigcap pr(F_{\mathcal{S}}) \neq \emptyset$ and $pr(F_{\mathcal{S}}) \neq \emptyset$. Lemma 3.18 suggests that looking for *gr*-repairs is a reasonable starting point. In order to guarantee at least one unattacked argument one finds $\{c, d\}$ as minimal *gr*-diagnosis. Let $F_{\{c,d\}}$ denote the associated minimal repair. We have $gr\left(F_{\{c,d\}}\right) = \{\{e\}\}$. Hence, $\bigcap pr(F_{\{c,d\}}) \neq \emptyset$ is implied. This means, $\{c, d\}$ is a *scep*-*pr*-diagnosis. Moreover, $\{c, d\}$ is even minimal proven by the following two AFs $F_{\{c\}}$ and $F_{\{d\}}$.





Indeed, we have $\bigcap pr\big(F_{\{c\}}\big) = \bigcap pr\big(F_{\{d\}}\big) = \emptyset$ since $\{a, e\}, \{d, f\} \in pr\big(F_{\{c\}}\big)$ are disjoint which is also the case for $\{a, e\}, \{c, f\} \in pr\big(F_{\{d\}}\big)$.

8. Related Work and Future Directions

In this paper, we investigated approaches aiming at repairing argumentation frameworks which are *inconsistent* in the sense that they do not possess any accepted argument. We considered a reasonable range of semantics, the standard reasoning modes, namely *credulous* and *sceptical* reasoning, and two different tools to repair, namely removal of certain (minimal) sets of arguments or attacks. We identified repairs wrt. grounded semantics as the arguably most important case: They can be utilized as a starting point in order to calculate repairs wrt. other semantics, coincide for both reasoning modes and are tractable from an algorithmic point of view. We illustrated how to derive stronger results for specific situations like restricting the AFs to certain subclasses or if AFs allows for a splitting. We also investigated infinite AFs. We studied the computational complexity of two natural arising reasoning problems, i. e. existence and verification of repairs. Our hardness results confirm the central role of the tractable *gr*-diagnoses. In a short case study, we illustrated how to repair a concrete AF under consideration.

The topic of *diagnoses* and *repairs* as introduced by Reiter (1987) is less developed in the area of abstract argumentation. The closest one to our work is by Nouioua and Würbel (2014). The authors define an operator and provide an algorithm, s.t. the resulting framework does not collapse. The mentioned work considers a semantical defect as the absence of any extension. Consequently, only stable semantics can be considered in contrast to our setup which additionally includes a treatment of semantics which may provide the empty set as unique extension. Moreover, restoring consistency is achieved via dropping a minimal set of arguments or attacks. In the latter case, all arguments survive the revision process.

The concept of maximal consistency in non-monotonic logics, is not novel (Sakama & Inoue, 2003, Definition 5.3). Extensions of consistency removal to non-monotonic logics can also be found in the literature. An example in autoepistemic logic has been analyzed by Inoue and Sakama (1995). The closest to our approach in Section 4 is probably the one by Eiter et al. (2014) where the authors have studied ways of restoring consistency in multi-context systems (Brewka & Eiter, 2007). They focus on the case where the source of inconsistency can be attributed to the bridge rules of a multi-context system.

The very first and basic works which are dealing with dynamics in abstract argumentation can be traced back to the beginnings of 2010. For instance, the so-called *enforcing problem* and the related *minimal change problem* (Baumann & Brewka, 2010; Baumann, 2012; Baumann & Brewka, 2013) are still studied in further variations (de Saint-Cyr, Bisquert, Cayrol, & Lagasquie-Schiex, 2016; Wallner, Niskanen, & Järvisalo, 2017). More precisely, these problems are dealing with the question whether it is possible (and if yes, as little effort as possible) to add new information in such a way that a desired set of arguments becomes an extension or at least a subset of one. Kim et al. (2013) studied a similar problem under the name *σ-repair* and provided parameterized complexity results. Although adding information as well as desired sets are not the focus of our study there is at least one interesting similarity to our work, namely: given an AF where nothing is credulously accepted, then enforcing a certain non-empty set can be seen as a special kind of repairing. Other works are case studies of what happens with the set of extensions if one deletes or adds one argument (Cayrol, de Saint-Cyr, & Lagasquie-Schiex, 2010; Bisquert, Cayrol, de Saint-Cyr, & Lagasquie-Schiex, 2011). The so-called *destructive change* is somehow the inverse of our notion of credulous repair since the initial framework possesses at least one credulously accepted argument whereas the result does not. Quite recently, the so-called *extension removal problem* was studied (Baumann & Brewka, 2019). That is, is it possible to modify a given AF in such a way that certain undesired extensions are no longer generated?

We want to mention that the introduced notions of repairs do have a correspondence in the general multistage reasoning process for knowledge bases (KBs) (Caminada & Amgoud, 2007). An argument-based repair can be seen as the result of forgetting some pieces of knowledge in the original KB. In contrast, attack-based repairs correspond to changing the underlying notion of attack in the overall instantiation process (Baumann, 2014, Section 1.1.1). Several future directions are already mentioned in the text. The most apparent one stems from the observation that our approach involves *removal* of attacks resp. arguments only. One could extend this approach to *adding* information similar in spirit to Ulbricht (2019a) or one could simply consider a hybrid notion of repairs, involving the simultaneous removal of arguments and attack relations. It will be interesting to see to which extent shown results may change if a more liberal way of removal is allowed. For instance, for specific AFs it might be the case that instead of removing three arguments, just removing one argument and one attack could restore consistency too. Another direction would be to generalize the considered notions of repairs in the sense that in addition to sceptical or credulous acceptance we require the existence of at least/exactly/at most n extensions. Such an requirement is directly linked to numerical aspects of argumentation semantics like maximal and possible numbers of extensions (Baumann & Strass, 2013; Baumann, Dvorák, Linsbichler, Spanring, Strass, & Woltran, 2016).

Moreover, a further intensive study of subclasses of AFs seems to be very promising since certain useful semantical properties are already ensured by syntactic properties. Furthermore, it is already known that AFs can be seen as a restricted class of logic programs (LPs). More precisely, there is a standard translation T from AFs to LPs, s.t. for any AF F, $\sigma(F)$ coincides with $\tau(T(F))$ for certain pairs of semantics σ and τ (Strass, 2013, Theorem 4.13). This means, one interesting research question is to which extent our results can be conveyed to repairing in logic programming.

The present article contributes to a thorough understanding of inconsistency in abstract argumentation, which might also benefit from the research area of measuring inconsistency (Hunter & Konieczny, 2004). Measuring inconsistency in non-monotonic logics has recently been studied (Ulbricht, Thimm, & Brewka, 2018a, 2018b) and could be extended in a similar fashion to abstract argumentation.

Acknowledgments

This work was partially funded by Deutsche Forschungsgemeinschaft DFG (Research Training Group 1763; project BR 1817/7-2 as well as project 406289255) and a postdoc fellowship of the Deutsche Akademische Austauschdienst DAAD (57407370).

References

Bailey, J., & Stuckey, P. J. (2005). Discovery of minimal unsatisfiable subsets of constraints using hitting set dualization. In *Practical Aspects of Declarative Languages, 7th International Symposium, PADL 2005, Long Beach, CA, USA, January 10-11, 2005, Proceedings*, pp. 174–186.

Baroni, P., Giacomin, M., & Liao, B. (2018). Locality and modularity in abstract argumentation. In Baroni, P., Gabbay, D., Giacomin, M., & van der Torre, L. (Eds.), *Handbook of Formal Argumentation*, chap. 19. College Publications.

Baroni, P., Caminada, M., & Giacomin, M. (2011). An introduction to argumentation semantics. *The Knowledge Engineering Review*, *26*, 365–410.

Baroni, P., Caminada, M., & Giacomin, M. (2018). Abstract argumentation frameworks and their semantics. In Baroni, P., Gabbay, D., Giacomin, M., & van der Torre, L. (Eds.), *Handbook of Formal Argumentation*, chap. 4. College Publications.

Baroni, P., Cerutti, F., Dunne, P. E., & Giacomin, M. (2013). Automata for infinite argumentation structures. *Artificial Intelligence*, *203*, 104–150.

Baumann, R. (2018). On the nature of argumentation semantics: Existence and uniqueness, expressibility, and replaceability. In Baroni, P., Gabbay, D., Giacomin, M., & van der Torre, L. (Eds.), *Handbook of Formal Argumentation*, chap. 14. College Publications.

Baumann, R. (2011). Splitting an argumentation framework. In *Logic Programming and Nonmonotonic Reasoning - 11th International Conference, LPNMR 2011, Vancouver, Canada, May 16-19, 2011. Proceedings*, pp. 40–53.

Baumann, R. (2012). What does it take to enforce an argument? minimal change in abstract argumentation. In *ECAI 2012 - 20th European Conference on Artificial Intelligence. Including Prestigious Applications of Artificial Intelligence (PAIS-2012) System Demonstrations Track, Montpellier, France, August 27-31 , 2012*, pp. 127–132.

Baumann, R. (2014). *Metalogical Contributions to the Nonmonotonic Theory of Abstract Argumentation*. College Publications - Studies in Logic.

Baumann, R., & Brewka, G. (2010). Expanding argumentation frameworks: Enforcing and monotonicity results. In *Computational Models of Argument: Proceedings of COMMA 2010, Desenzano del Garda, Italy, September 8-10, 2010*, pp. 75–86.

Baumann, R., & Brewka, G. (2013). Spectra in abstract argumentation: An analysis of minimal change. In *LPNMR, Proceedings of 12th International Conference in Logic Programming and Nonmonotonic Reasoning,*, pp. 174–186.

Baumann, R., & Brewka, G. (2015). AGM meets abstract argumentation: Expansion and revision for Dung frameworks. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, pp. 2734–2740. AAAI Press.

Baumann, R., & Brewka, G. (2019). Extension removal in abstract argumentation - an axiomatic approach. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Ar-*

tificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019., pp. 2670–2677.

Baumann, R., Dvorák, W., Linsbichler, T., Spanring, C., Strass, H., & Woltran, S. (2016). On rejected arguments and implicit conflicts: The hidden power of argumentation semantics. *Artificial Intelligence*, *241*, 244–284.

Baumann, R., Dvorák, W., Linsbichler, T., Strass, H., & Woltran, S. (2014). Compact argumentation frameworks. In *ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic - Including Prestigious Applications of Intelligent Systems (PAIS 2014)*, pp. 69–74.

Baumann, R., & Spanring, C. (2015). Infinite argumentation frameworks - On the existence and uniqueness of extensions. In *Advances in Knowledge Representation, Logic Programming, and Abstract Argumentation - Essays Dedicated to Gerhard Brewka on the Occasion of His 60th Birthday*, pp. 281–295.

Baumann, R., & Spanring, C. (2017). A study of unrestricted abstract argumentation frameworks. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pp. 807–813.

Baumann, R., & Strass, H. (2013). On the maximal and average numbers of stable extensions. In *Theory and Applications of Formal Argumentation - Second International Workshop, TAFA 2013, Beijing, China, August 3-5, 2013, Revised Selected papers*, pp. 111–126.

Baumann, R., & Ulbricht, M. (2018). If nothing is accepted - repairing argumentation frameworks. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, 30 October - 2 November 2018.*, pp. 108–117.

Bertossi, L. E., & Salimi, B. (2017). From causes for database queries to repairs and model-based diagnosis and back. *Theory of Computing Systems*, *61*(1), 191–232.

Besnard, P., & Hunter, A. (2008). *Elements of Argumentation*. MIT Press.

Bienvenu, M. (2012). On the complexity of consistent query answering in the presence of simple ontologies. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, July 22-26, 2012, Toronto, Ontario, Canada.*

Bisquert, P., Cayrol, C., de Saint-Cyr, F. D., & Lagasquie-Schiex, M. (2011). Change in argumentation systems: Exploring the interest of removing an argument. In *Scalable Uncertainty Management - 5th International Conference, SUM 2011, Dayton, OH, USA, October 10-13, 2011. Proceedings*, pp. 275–288.

Brewka, G., & Eiter, T. (2007). Equilibria in heterogeneous nonmonotonic multi-context systems. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, July 22-26, 2007, Vancouver, British Columbia, Canada*, pp. 385–390.

Brewka, G., Thimm, M., & Ulbricht, M. (2017). Strong inconsistency in nonmonotonic reasoning. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pp. 901–907.

Brewka, G., Thimm, M., & Ulbricht, M. (2019). Strong inconsistency. *Artificial Intelligence*, *267*, 78–117.

Caminada, M., & Amgoud, L. (2007). On the evaluation of argumentation formalisms. *Artificial Intelligence*, *171*, 286–310.

Caminada, M. W. A., & Oren, N. (2014). Grounded semantics and infinitary argumentation frameworks. In *Proceedings of the 26rd Benelux Conference on Artificial Intelligence (BNAIC 2014)*.

Cayrol, C., de Saint-Cyr, F. D., & Lagasquie-Schiex, M. (2010). Change in abstract argumentation frameworks: Adding an argument. *Journal of Artificial Intelligence Research*, *38*, 49–84.

Coste-Marquis, S., Devred, C., & Marquis, P. (2005). Symmetric argumentation frameworks. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 8th European Conference, ECSQARU 2005, Barcelona, Spain, July 6-8, 2005, Proceedings*, pp. 317–328.

Coste-Marquis, S., Konieczny, S., Mailly, J., & Marquis, P. (2014). On the revision of argumentation systems: Minimal change of arguments statuses. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference, KR 2014, Vienna, Austria, July 20-24, 2014*.

da Costa, N. C. A. (1974). On the theory of inconsistent formal systems.. *Notre Dame Journal of Formal Logic*, *15*(4), 497–510.

de Saint-Cyr, F. D., Bisquert, P., Cayrol, C., & Lagasquie-Schiex, M. (2016). Argumentation update in YALLA (yet another logic language for argumentation). *International Journal of Approximate Reasoning*, *75*, 57–92.

Diller, M., Haret, A., Linsbichler, T., Rümmele, S., & Woltran, S. (2018). An extension-based approach to belief revision in abstract argumentation. *International Journal of Approximate Reasoning*, *93*, 395–423.

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.

Dvorák, W., & Dunne, P. E. (2018). Computational problems in formal argumentation and their complexity. In Baroni, P., Gabbay, D., Giacomin, M., & van der Torre, L. (Eds.), *Handbook of Formal Argumentation*, chap. 13. College Publications.

Eiter, T., Fink, M., Schüller, P., & Weinzierl, A. (2014). Finding explanations of inconsistency in multi-context systems. *Artificial Intelligence*, *216*, 233–274.

Eiter, T., Fink, M., & Stepanova, D. (2013). Data repair of inconsistent dl-programs. In *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*, pp. 869–876.

Hansson, S. O. (1994). *Taking Belief Bases Seriously*, pp. 13–28. Springer Netherlands, Dordrecht.

Hunter, A., & Konieczny, S. (2004). Approaches to Measuring Inconsistent Information. In *Inconsistency Tolerance*, Vol. 3300 of *Lecture Notes in Computer Science*, pp. 189–234. Springer International Publishing.

Inoue, K., & Sakama, C. (1995). Abductive framework for nonmonotonic theory change. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, IJCAI 95, Montréal Québec, Canada, August 20-25 1995, 2 Volumes*, pp. 204–210.

Kim, E. J., Ordyniak, S., & Szeider, S. (2013). The complexity of repairing, adjusting, and aggregating of extensions in abstract argumentation. In *Theory and Applications of Formal Argumentation - Second International Workshop, TAFA*, pp. 158–175.

Lembo, D., Lenzerini, M., Rosati, R., Ruzzi, M., & Savo, D. F. (2011). Inconsistency-tolerant semantics for description logic ontologies (extended abstract). In *Sistemi Evoluti per Basi di Dati - SEBD 2011, Proceedings of the Nineteenth Italian Symposium on Advanced Database Systems*, pp. 287–294.

Liffiton, M. H., Previti, A., Malik, A., & Marques-Silva, J. (2016). Fast, flexible MUS enumeration. *Constraints*, *21*(2), 223–250.

Liffiton, M. H., & Sakallah, K. A. (2008). Algorithms for computing minimal unsatisfiable subsets of constraints. *Journal of Automated Reasoning*, *40*(1), 1–33.

Marques-Silva, J., Janota, M., Ignatiev, A., & Morgado, A. (2015). Efficient model based diagnosis with maximum satisfiability. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pp. 1966–1972.

Mcilraith, S. (1999). Towards a theory of diagnosis, testing and repair. In *Proceedings of the 5th International Workshop on Principles of Diagnosis (DX)*.

Metodi, A., Stern, R., Kalech, M., & Codish, M. (2014). A novel sat-based approach to model based diagnosis. *Journal of Artificial Intelligence Research*, *51*, 377–411.

Nouioua, F., & Würbel, E. (2014). Removed set-based revision of abstract argumentation frameworks. In *26th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2014*, pp. 784–791.

Papadimitriou, C. (1994). *Computational Complexity*. Addison-Wesley.

Potyka, N., & Thimm, M. (2014). Consolidation of probabilistic knowledge bases by inconsistency minimization. In *ECAI 2014 - 21st European Conference on Artificial Intelligence*, pp. 729–734.

Prakken, H. (2010). An abstract framework for argumentation with structured arguments. *Argument and Computation*, *1*(2), 93–124.

Rahwan, I., & Simari, G. R. (2009). *Argumentation in artificial intelligence*, Vol. 47. Springer.

Reiter, R. (1987). A theory of diagnosis from first principles. *Artificial Intelligence*, *32*(1), 57–95.

Sakama, C., & Inoue, K. (2003). An abductive framework for computing knowledge base updates. *Theory and Practice of Logic Programming*, *3*(6), 671–713.

Stern, R. T., Kalech, M., Feldman, A., & Provan, G. M. (2012). Exploring the duality in conflict-directed model-based diagnosis. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, July 22-26, 2012, Toronto, Ontario, Canada.*

Strass, H. (2013). Approximating operators and semantics for abstract dialectical frameworks. *Artificial Intelligence*, *205*, 39–70.

Strass, H. (2018). Instantiating rule-based defeasible theories in abstract dialectical frameworks and beyond. *Journal of Logic and Computation*, *28*(3), 605–627.

Ulbricht, M. (2019a). Repairing non-monotonic knowledge bases. In *European Conference on Logics in Artificial Intelligence*, pp. 151–167. Springer.

Ulbricht, M. (2019b). *Understanding Inconsistency – A Contribution to the Field of Nonmonotonic Reasoning*. Ph.D. thesis, Leipzig University.

Ulbricht, M., Thimm, M., & Brewka, G. (2018a). Inconsistency measures for disjunctive logic programs under answer set semantics. In Grant, J., & Martinez, M. V. (Eds.), *Measuring Inconsistency in Information*, Vol. 73 of *Studies in Logic*. College Publications.

Ulbricht, M., Thimm, M., & Brewka, G. (2018b). Measuring strong inconsistency. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pp. 1989–1996.

Wallner, J. P., Niskanen, A., & Järvisalo, M. (2017). Complexity results and algorithms for extension enforcement in abstract argumentation. *Journal of Artificial Intelligence Research*, *60*, 1–40.

Weydert, E. (2011). Semi-stable extensions for infinite frameworks. In *Proceedings of the 23rd Benelux Conference on Artificial Intelligence (BNAIC 2011)*, pp. 336–343.