A General Notion of Equivalence for Abstract Argumentation*

Ringo Baumann
University of Leipzig, Germany
baumann@informatik.uni-leipzig.de

Wolfgang Dvořák
TU Wien, Austria
dvorak@dbai.tuwien.ac.at

Thomas Linsbichler
TU Wien, Austria
linsbich@dbai.tuwien.ac.at

Stefan Woltran
TU Wien, Austria
woltran@dbai.tuwien.ac.at

Abstract

We introduce a parametrized equivalence notion for abstract argumentation that subsumes standard and strong equivalence as corner cases. Under this notion, two argumentation frameworks are equivalent if they deliver the same extensions under any addition of arguments and attacks that do not affect a given set of core arguments. As we will see, this notion of equivalence nicely captures the concept of local simplifications. We provide exact characterizations and complexity results for deciding our new notion of equivalence.

1 Introduction

Argumentation has become one of the major fields within AI over the last two decades [Rahwan and Simari, 2009; Bench-Capon and Dunne, 2007]. In particular, Dung's argumentation frameworks [Dung, 1995], AFs for short, are widely used and act as integral concepts in several advanced argumentation formalisms. They focus entirely on conflict resolution among arguments, treating the latter as abstract items without logical structure. Hence, the only information available in AFs is the so-called attack-relation that determines whether an argument is in a certain conflict with another one. As already outlined by Dung, AFs provide a formally simple basis to capture the essence of different nonmonotonic formalisms. Therefore, several so-called semantics are typically considered for AFs, see also [Baroni *et al.*, 2011]. A semantics delivers several sets of arguments (called extensions) that can be jointly accepted in order to satisfy certain properties. One such property is given by admissible sets which consist of arguments that do not attack each other and attack each argument attacking the set itself.

Bearing the nonmonotonic nature of AFs in mind, it is evident that the *standard* notion of *equivalence* (i.e., do two AFs possess the same sets of extensions?) is a rather weak concept. In particular, it is *not* the case that replacing an AF by an equivalent one is a faithful manipulation. As an example consider the AFs $F_{abc} = (\{a,b,c\}, \{(a,b),(b,c),(c,a)\})$ and $F_{ab} = (\{a,b\}, \{(a,a),(a,b)\})$, which are equivalent for

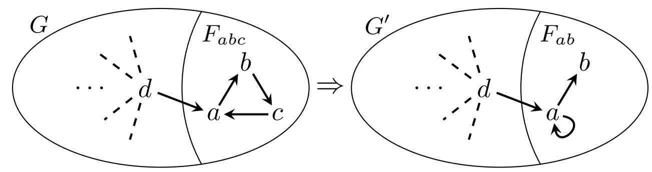


Figure 1: Replacing F_{abc} in G to obtain $G' = G[F_{abc}/F_{ab}]$.

most semantics, including admissible sets. However, replacing F_{abc} by F_{ab} in a larger AF G might not be an equivalence-preserving action. Suppose G expands F_{abc} via an attack from some argument d to b. Then, the mentioned replacement would change each admissible set $S \cup \{d,c\}$ into $S \cup \{d\}$. On the other hand, if F_{abc} is embedded in G only via an attack (d,a) – see Figure 1 – the replacement is faithful. More formally, we then have that the admissible sets of G and $G[F_{abc}/F_{ab}]$ are the same (a formal definition of replacements $G[\cdot/\cdot]$ is given in Section 6).

Observations of this kind gave rise to more restricted notions of equivalence [Oikarinen and Woltran, 2011; Baumann, 2012; Baumann and Woltran, 2016]. *Strong equivalence* (also called expansion equivalence) between two AFs F and F' holds (w.r.t. a semantics σ) iff for all AFs H the expanded AFs $F \cup H$ and $F' \cup H$ have the same σ-extensions. By definition, this notion of equivalence guarantees that F can be replaced by a strongly equivalent (w.r.t. σ) AF F' in any framework G without changing the σ-extensions of G. Interestingly, the characterization results for strong equivalence are surprisingly simple and can be given via so-called kernels, syntactic modifications of the involved AFs. From a theoretical perspective, it is thus open how this conceptual difference between standard and strong equivalence can be captured via a uniform formal characterization which has these two notions as corner cases.

From a computational point of view, strong equivalence (and related versions) seem to be an appealing notion, since checks for replacements, and thus also for simplifications in AFs, would become easy. However, strong equivalence is too restricted for practical purposes. Even obvious simplifications are not captured: an example are isolated self-loops, which can be safely removed from AFs for many standard semantics. However, AF $F = (\{a\}, \{(a,a)\})$ is not strongly equivalent to the empty AF $F' = (\emptyset, \emptyset)$ for admissible se-

*This research has been supported by DFG: BR 1817/7-2 and FWF: I2854.

mantics; just take $H = (\{a\}, \emptyset)$. Then, $\{a\}$ is admissible for $F' \cup H$ but not for $F \cup H$. This indicates that a suitable equivalence notion for replacement needs a particular treatment for those arguments which are directly involved in the change.

Hence, what we require is an equivalence notion that compares two AFs such that

1. the relations between *core* arguments are fixed, while
2. the remaining arguments are allowed to interact arbitrarily with possible expansions of the compared AFs.

Our proposal is to define, given a set of core arguments C and a semantics σ, *C-relativized equivalence* between two AFs F and F' w.r.t. σ (in symbols, $F \equiv^{\sigma}_{C} F'$) to hold, if $F \cup H$ and $F' \cup H$ have the same σ extensions, for each AF H not containing arguments from C. Observe that this notion indeed captures strong equivalence (set $C = \emptyset$) and standard equivalence (set C to be the universe of all arguments).

Coming back to our example with F_{abc} and F_{ab}, the idea is to set $C = \{a, b, c\}$ and compare the two AFs plus their interaction with the AF G where F_{abc} occurs in. In our case, we compare $F^G_{abc} = (\{a, b, c, d\}, \{(a, b), (b, c), (c, a), (d, a)\})$ and $F^G_{ab} = (\{a, b, d\}, \{(a, a), (a, b), (d, a)\})$. Then, $F^G_{abc} \equiv^{\sigma}_{C} F^G_{ab}$ implies that G and $G[F_{abc}/F_{ab}]$ are equivalent under σ, i.e., replacing F_{abc} by F_{ab} in G is safe for semantics σ.

Our main contributions are as follows:

- We first define restrictions for the main semantics of stable, admissible, preferred, complete and grounded extensions. These identify extensions of an AF F that are acceptable in some expansion $F \cup H$ and are integral for equivalence characterizations.
- We give exact characterizations of C-relativized equivalence for the five semantics mentioned above; in addition we also show results for conflict-free and naive sets.
- We provide a complexity analysis for deciding C-relativized equivalence; as corollaries we also obtain insight to the complexity of standard equivalence.
- Finally, we give a formal notion of replacement in AFs and illustrate how our equivalence notion can be employed for local simplifications within AFs.

Some proofs are only sketched or omitted due to space constraints. Full proofs are available in [Baumann *et al.*, 2017].

2 Preliminaries

In this section, we introduce argumentation frameworks [Dung, 1995] and recall the semantics we study (for an overview, see [Baroni *et al.*, 2011]). We fix U as countably infinite domain of arguments.

Definition 2.1. *An* argumentation framework (AF) *is a pair $F = (A, R)$ where $A \subseteq U$ is a finite set of arguments and $R \subseteq A \times A$ is the attack relation. The pair $(a, b) \in R$ means that a attacks b. We use $A(F)$ to refer to A and $R(F)$ to refer to R. We say that an AF is given over a set B if $A(F) \subseteq B$.*

Given an AF F and $S \subseteq U$, we define $S^+_F = \{x \mid \exists y \in S : (y, x) \in R(F)\}$, $S^-_F = \{x \mid \exists y \in S : (x, y) \in R(F)\}$, and the range *of S in F as $S^{\oplus}_F = (S \cap A(F)) \cup S^+_F$.*

Given $F = (A, R)$, argument $a \in A$ is defended *(in F) by a set $S \subseteq U$ if $\{a\}^-_F \subseteq S^+_F$. The* characteristic function *$\mathcal{F}_F : 2^A \to 2^A$ of F is defined as $\mathcal{F}_F(S) = \{a \in A \mid a$ is defended by S in $F\}$.*

Given AFs $F = (A, R)$, $F' = (A', R')$, and $S \subseteq U$, we denote the union of AFs as $F \cup F' = (A \cup A', R \cup R')$, and define $F \setminus S = (A \setminus S, R \cap ((A \setminus S) \times (A \setminus S)))$ and $F \cap S = (A \cap S, R \cap ((A \cap S) \times (A \cap S)))$.

Semantics for argumentation frameworks are defined as functions σ which assign to each AF F a set $\sigma(F) \subseteq 2^{A(F)}$ of extensions. We consider for σ the functions $naive$, grd, stb, adm, com, and prf, which stand for naive, grounded, stable, admissible, complete, and preferred extensions, respectively.

Definition 2.2. *Let $F = (A, R)$ be an AF. A set $S \subseteq A$ is* conflict-free (in F), *if there are no $a, b \in S$, such that $(a, b) \in R$. $cf(F)$ denotes the collection of conflict-free sets of F. For a conflict-free set $S \in cf(F)$, it holds that*

- *$S \in naive(F)$, if there is no $T \in cf(F)$ with $T \supset S$;*
- *$S \in stb(F)$, if $S^{\oplus}_F = A$;*
- *$S \in adm(F)$, if $S \subseteq \mathcal{F}_F(S)$;*
- *$S \in com(F)$, if $S = \mathcal{F}_F(S)$;*
- *$S \in grd(F)$, if $S \in com(F)$ and $\nexists T \subset S$ s.t. $T \in com(F)$;*
- *$S \in prf(F)$, if $S \in adm(F)$ and $\nexists T \supset S$ s.t. $T \in adm(F)$.*

We recall that for each AF F, the grounded semantics yields a unique extension, which is the least fixed-point $\mathcal{F}^{\infty}_F(\emptyset)$ of the characteristic function $\mathcal{F}_F$.

3 Notions of Equivalence

We first review two equivalence notions for AFs from the literature, namely standard and strong equivalence.

Definition 3.1. *Given a semantics σ. Two AFs F and G are* (standard) equivalent *w.r.t. σ ($F \equiv^{\sigma} G$) iff $\sigma(F) = \sigma(G)$.*

Definition 3.2. *Given a semantics σ. Two AFs F and G over U are* strongly equivalent *w.r.t. σ ($F \equiv^{\sigma}_{S} G$) iff $F \cup H \equiv^{\sigma} G \cup H$ holds for each AF H over U.*

In this work we introduce the new notion of C-relativized equivalence, which is parametrized by the set C of *core* arguments which will not be directly touched by the possible expansions (i.e., AFs H added to the compared AFs are not arbitrary anymore).

Definition 3.3. *Given a semantics σ and $C \subseteq U$. Two AFs F and G are C-*relativized equivalent *w.r.t. σ ($F \equiv^{\sigma}_{C} G$) iff $F \cup H \equiv^{\sigma} G \cup H$ holds for each AF H over $U \setminus C$.*

Notice that (i) for $C = \emptyset$ the C-relativized equivalence coincides with strong equivalence and (ii) when $C = U$ then C-relativized equivalence is just standard equivalence (the only AF over $U \setminus C = \emptyset$ is $(\emptyset, \emptyset)$ and $F \cup (\emptyset, \emptyset) = F$ for all AFs F).

The following observation expresses the fact that C-relativized equivalence survives if we extend the core C with further untouchable arguments. Since in general standard equivalence ($C = U$) does not imply strong equivalence ($C = \emptyset$) the assertion does not hold for shrinking the core.

Observation 3.4. *For any two AFs F, G, any two sets $C, D \subseteq U$ and any semantics σ, if $C \subseteq D$ and $F \equiv^\sigma_C G$, then $F \equiv^\sigma_D G$.*

An immediate consequence of the observation above is that strong (standard) equivalence is more (less) demanding than relativized equivalence, no matter which core C is considered. This is simply due to the fact that for any core C, $\emptyset \subseteq C \subseteq U$. The next proposition gives more refined conditions for the coincidence between C-relativized equivalence and strong or standard equivalence, respectively. Since we consider finite AFs only we restrict our considerations to finite cores too.

Proposition 3.5. *Let F, G be AFs, $C \subseteq U$ a finite core, $\sigma \in \{stb, adm, com, grd, prf\}$, and $B = C \cap (A(F) \cup A(G))$.*

1. If $B = \emptyset$, then $F \equiv^\sigma_C G$ iff $F \equiv^\sigma_S G$.

2. If $B = A(F) \cup A(G)$, then $F \equiv^\sigma_C G$ iff $F \equiv^\sigma G$.

4 Characterization Results

In what follows, we aim for giving characterizations for deciding $F \equiv^\sigma_C G$ with finite $C \subset U$, such that an explicit consideration of all possible expansions is avoided. In other words, we need semantical concepts that are solely defined on the AFs F and G, but take the core C into account. To this end, we start with the concept of C-restricted semantics. Our main result for exactly characterizing $F \equiv^\sigma_C G$ then requires that the C-restricted extensions coincide for the compared AFs. As we will see in Section 4.2, some further semantics-dependent conditions must be met.

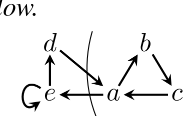
4.1 C-restricted Semantics

For C-restricted semantics, we restrict the relevant properties of the original semantics to the core arguments.

Definition 4.1. *Let F be an AF, $C \subseteq U$ and $E \subseteq A(F)$.*

- *$E \in stb_C(F)$ if $E \in cf(F)$ and $A(F) \cap C \subseteq E^\oplus_F$.*
- *$E \in adm_C(F)$ if $E \in cf(F)$ and $E^-_F \cap C \subseteq E^+_F$.*
- *$E \in prf_C(F)$ if $E \in adm_C(F)$ and for all $D \in adm_C(F)$ with $E \setminus C = D \setminus C$, $E^+_F \setminus C \subseteq D^+_F \setminus C$, and $E^-_F \setminus E^+_F \supseteq D^-_F \setminus D^+_F$ we have $E \cap C \not\subset D \cap C$.*

Example 4.2. *For AF F^G_{abc} from the introduction and $C = \{a, b, c\}$, we have $stb_C(F^G_{abc}) = \{\{d, b\}\}$, $adm_C(F^G_{abc}) = \{\emptyset, \{d\}, \{d, b\}\}$. In this particular case, standard extensions and restricted ones coincide. Let us thus extend F^G_{abc} to the AF F as depicted below.*



We observe that $stb_C(F) = stb_C(F^G_{abc})$ although $stb(F) = \emptyset$; likewise $adm_C(F) = adm_C(F^G_{abc})$ but $adm(F) = \{\emptyset\}$. Finally, we have $prf_C(F^G_{abc}) = prf_C(F) = \{\emptyset, \{d, b\}\}$. $\Diamond$

In order to define the C-restricted complete and grounded semantics we need the concept of the *C-restricted characteristic function* $\mathcal{F}_{F,C,E}(S)$ for AF F and $E, S \subseteq A(F)$.

$$\mathcal{F}_{F,C,E}(S) = \{a \in E \mid \forall c \in C : (c, a) \in R(F) \to c \in S^+_F\}$$
$$\cup \{c \in C \cap A(F) \mid \forall (b, c) \in R(F) : b \in S^+_F \cup (S^-_F \setminus C)\}$$

Definition 4.3. *Let F be an AF, $C \subseteq U$ and $E \subseteq A(F)$.*

- *$E \in com_C(F)$ if $E \in cf(F)$ and $E = \mathcal{F}_{F,C,E}(E)$.*
- *$E \in grd_C(F)$ if $E \in cf(F)$ and $E = \mathcal{F}^\infty_{F,C,E}(\emptyset)$.*

Notice that in case $A(F) \subseteq C$, C-restricted semantics as given in Defs. 4.1 and 4.3 reduce to the original semantics, i.e., $\sigma_C(F) = \sigma(F)$ for any C with $A(F) \subseteq C$.

Another crucial feature of C-restricted semantics is that $\sigma_C(F)$ returns all the argument sets that are projections of σ-extensions in some $F \cup H$ with H defined over $U \setminus C$.

Proposition 4.4. *Let F be an AF, $\sigma \in \{stb, adm, com, grd, prf\}$, and $C \subseteq U$. $E \in \sigma_C(F)$ iff there exists an AF H over $U \setminus C$ and $T \in \sigma(F \cup H)$ such that $T \cap A(F) = E$.*

Proof Sketch (stable/admissible). $\Rightarrow$: Let $B = (A(F) \setminus (E \cup C))$ and consider the AF $H = (\{t\} \cup B, \{(t, b) \mid b \in B\})$ with $t \in U \setminus C$ a fresh argument (not occurring in F). Then H is given over $U \setminus C$ and $E \cup \{t\} \in \sigma(F \cup H)$, for $\sigma \in \{stb, adm\}$. $\Leftarrow$: Let $\sigma \in \{stb, adm\}$ and suppose $T \in \sigma(F \cup H)$ for some H over $U \setminus C$. For $E = T \cap A(F)$, we clearly have $E \in cf(F)$. As all attacks on $c \in C$ are already present in F, for $\sigma = stb$ we have that $A(F) \cap C \subseteq E^+_F$ and thus $E \in stb_C(F)$. Likewise for $\sigma = adm$, since each $c \in E^-_F \cap C$ is already attacked by E in F, $E \in adm_C(F)$. $\square$

The proposition above establishes a close relationship with the *enforcing problem* [Baumann and Brewka, 2010]. More precisely, the C-restricted σ-extensions E are exactly the sets enforceable without touching the core arguments.

Example 4.5. *Recall F from Example 4.2. For $C = \{a, b, c\}$ we had $\{b, d\} \in stb_C(F)$ and $\{b, d\} \in adm_C(F)$. The construction in the proof of Prop. 4.4 just adds an argument t attacking e (note that t and e are not from C). For the resulting AF it is easily checked that $\{t, b, d\}$ is among its admissible sets, resp. is its only stable extension.* $\Diamond$

Next we consider properties that will appear in the C-relativized equivalence characterizations of all semantics $\sigma \in \{stb, adm, com, grd, prf\}$.

Lemma 4.6. *If $F \equiv^\sigma_C G$ then $A(F) \setminus C = A(G) \setminus C$ or $\sigma_C(F) = \sigma_C(G) = \emptyset$.*

Proof Sketch (stable). If $stb_C(F) = \emptyset$, $stb(F \cup H) = \emptyset$ for all H over $U \setminus C$ (by Prop. 4.4), and thus also $stb_C(G) = \emptyset$.

Now suppose $stb_C(F) \neq \emptyset$ and $A(F) \setminus C \neq A(G) \setminus C$. W.l.o.g. there is an $a \in A(F) \setminus C$ and $a \notin A(G)$. Suppose there is $E \in stb_C(F)$ with $a \in E$. By Prop. 4.4, we can give an H such that there is a $T \in stb(F \cup H)$ with $T \cap A(F) = E$. Notice that H does not contain arguments from E. Thus $a \notin A(G \cup H)$ and hence $T \notin stb(G \cup H)$, a contradiction to $F \equiv^{stb}_C G$. Thus there is no $E \in stb_C(F)$ with $a \in E$.

Let $E \in stb_C(G)$. By Prop. 4.4, there is an AF H such that there is a $T \in stb(G \cup H)$ with $T \cap A(G) = E$ and we can build this H with $a \notin A(H)$. Let now $H' = H \cup (\{a\}, \emptyset)$ and observe that H' is still given over $U \setminus C$. Then, $T \cup \{a\} \in stb(G \cup H')$ but it cannot be that $T \cup \{a\} \in stb(F \cup H')$ as this, by Prop. 4.4, would give rise to an $E \in stb_C(F)$ with $a \in E$. But this is in contradiction to $F \equiv^{stb}_C G$. $\square$

Lemma 4.7. *If $F \equiv^\sigma_C G$ then $\sigma_C(F) = \sigma_C(G)$.*

Proof. $\sigma_C(F) \subseteq \sigma_C(G)$: By Prop. 4.4, for each $E \in \sigma_C(F)$ there is an H over $U \setminus C$ and $T \in \sigma(F \cup H)$, such that $T \cap A(F) = E$. By assumption, $T \in \sigma(G \cup H)$ and, by Prop. 4.4, $E' = T \cap A(G) \in \sigma_C(G)$. As $A(H) \cap C = \emptyset$, we have that $T \cap C = E \cap C = E' \cap C$, and, by Lem. 4.6, $A(F) \setminus C = A(G) \setminus C$ and thus $E \setminus C = E' \setminus C$. Hence, $E = E'$ and $E \in \sigma_C(G)$. $\sigma_C(F) \supseteq \sigma_C(G)$ is by symmetry. □

4.2 Characterizations

In the following we give the characterizations for all semantics under consideration. We already have that two AFs can only be C-relativized equivalent w.r.t. σ if $A(F) \setminus C = A(G) \setminus C$ (or $\sigma_C(F) = \emptyset$) and $\sigma_C(F) = \sigma_C(G)$. Now depending on the concrete semantics we have to appoint additional conditions for the sets $E \in \sigma_C(F)$ to ensure that they appear in the same expansions of F and G.

Stable Semantics. For stb semantics we require for each $E \in stb_C(F)$ that the range of E coincides on $F \setminus C$ and $G \setminus C$. That is, the arguments that have to be attacked by H to make E stable in $F \cup H$ coincide with the arguments that have to be attacked by H to make E stable in $G \cup H$.

Theorem 4.8. *Let F, G be AFs and $C \subseteq U$. Then, $F \equiv_C^{stb} G$ iff the following conditions jointly hold:*

1. *if $stb_C(F) \neq \emptyset$, $A(F) \setminus C = A(G) \setminus C$;*
2. *$stb_C(F) = stb_C(G)$; and*
3. *for all $E \in stb_C(F)$, $E_F^+ \setminus C = E_G^+ \setminus C$.*

Proof Sketch. $\Rightarrow$: The conditions (1) and (2) are immediate by Lem. 4.6 and Lem. 4.7. Now let $stb_C(F) = stb_C(G) \neq \emptyset$ and $A(F) \setminus C = A(G) \setminus C$, and assume there is an $E \in stb_C(F)$ s.t. $E_F^+ \setminus C \neq E_G^+ \setminus C$. W.l.o.g. let $a \in A(F) \setminus E_F^\oplus$ and $a \notin A(G) \setminus E_G^\oplus$. We observe that $a \notin E$ and $a \notin C$ and thus $a \in E_G^+$. Let $H = (\{t\} \cup A(G) \setminus E_G^\oplus, \{(t, b) \mid b \in A(G) \setminus E_G^\oplus\})$ where t is fresh argument from $U \setminus C$. Observe that H does not contain arguments from C since $E \in stb_C(G)$ and thus each $a \in C$ occurring in G is attacked by E. We have $E \cup \{t\} \in stb(G \cup H)$, while $E \cup \{t\} \notin stb(F \cup H)$, a contradiction to $F \equiv_C^{stb} G$.

$\Leftarrow$: Suppose $F \not\equiv_C^{stb} G$. W.l.o.g. there is an AF H over $U \setminus C$ and a set S such that $S \in stb(F \cup H)$ but $S \notin stb(G \cup H)$. By Prop. 4.4, $E = S \cap A(F) \in stb_C(F)$. If now $E \notin stb_C(G)$ or $A(F) \setminus C \neq A(G) \setminus C$, we are done, i.e., condition (1) or (2) is already violated. So suppose $E \in stb_C(G)$, and $A(F) \setminus C = A(G) \setminus C$. We have to show $E_F^+ \setminus C \neq E_G^+ \setminus C$. Recall that $S \notin stb(G \cup H)$. Since $E \in stb_C(G)$ there exists an $a \in A(G \cup H) \setminus C$ not attacked by S in $G \cup H$, thus in particular $a \notin E_G^+$. Since S does not attack a via H and $S \in stb(F \cup H)$ we conclude that either $a \in E_F^+$ or $a \notin A(F)$. However, since $a \notin C$ and $A(F) \setminus C = A(G) \setminus C$, it follows that $a \in E_F^+$, thus violating (3). □

Example 4.9. *Recall F from Example 4.2 and let $F' = F_{ab}^G \cup (\{a, d, e\}, \{(a, e), (e, e), (e, d)\})$, i.e. instead of the cycle through a, b, c, we have just two arguments a, b where a attacks itself and b. For $C = \{a, b, c\}$, it is easily checked that F and F' satisfy all three conditions, i.e., we have $F \equiv_C^{stb} F'$. In fact, even for the AF $F'' = (\{a, b, d, e\}, \{(a, a), (a, e), (e, e), (e, d)\})$, i.e., F' without the attack from a to b, $F \equiv_C^{stb} F''$ holds.*

If we had $C = \{a, b\}$, condition (1) would be violated; indeed $F \not\equiv_C^{stb} F'$ is then witnessed by adding $H = (\{c, e, t\}, \{(t, e)\})$, as $stb(F \cup H) = \{\{t, d, b\}\}$ and $stb(F' \cup H) = \{\{t, d, b, c\}\}$. On the other hand, for $C = \{a, b, c\}$, the role of b and c is indeed different: if we use in F' argument c instead of b, we have $stb_C(F') = \{\{d, c\}\}$; thus condition (2) would be violated. Finally, consider F''' given by F' plus an additional attack (b, e). Note that we still have $stb_C(F''') = \{\{d, b\}\}$, but now $E_F^+ \setminus C \neq E_{F'''}^+ \setminus C$, hence condition (3) is violated here. Even without expanding the AFs, we obtain different stable extensions, i.e., $stb(F) = \emptyset$ while $stb(F''') = \{\{d, b\}\}$. ◊

When considering $C = \emptyset$ the above characterization boils down to (1) $A(F) = A(G)$, (2) $cf(F) = cf(G)$ and (3) for all $E \in cf(F)$, $E_F^+ = E_G^+$. That is, the two AFs F and G have to coincide except for attacks from self-attacking arguments, i.e., we end up with the concept of stable kernels from [Oikarinen and Woltran, 2011], which characterize strong equivalence for stb. For $C = A(F \cup G)$, only condition (2) remains which is equivalent to $stb(F) = stb(G)$. Similar observations can be made for the forthcoming results.

Admissible Semantics. For adm semantics we have the additional condition that for each $E \in adm_C(F)$ the attackers of E that are not already attacked by E coincide in F and G.

Theorem 4.10. *Let F, G be AFs and $C \subseteq U$. Then, $F \equiv_C^{adm} G$ iff the following conditions jointly hold: (1) $A(F) \setminus C = A(G) \setminus C$; (2) $adm_C(F) = adm_C(G)$; and (3) for all $E \in adm_C(F)$, (3a) $E_F^+ \setminus C = E_G^+ \setminus C$ and (3b) $E_F^- \setminus E_F^+ = E_G^- \setminus E_G^+$.*

Example 4.11. *Let us first consider F, F', and F'' from Example 4.9, again with $C = \{a, b, c\}$. It can be shown that all three conditions then hold, i.e., $F \equiv_C^{adm} F'$. However, F'' is a too drastic simplification for admissible semantics, since $\{b\} \in adm_C(F'')$ but $\{b\} \notin adm_C(F)$.*

To show the role of condition (3b), consider the AFs $F_1 = F \cup (\{g\}, \{(g, g)\})$ and $F_2 = F \cup (\{g, b\}, \{(g, g), (g, b)\})$; conditions (1), (2), and (3a) are fulfilled. However, for $E = \{d, b\} \in adm_C(F_1)$, we have $E_{F_1}^- \setminus E_{F_1}^+ = \{e\}$, while $E_{F_2}^- \setminus E_{F_2}^+ = \{e, g\}$. Hence condition (3b) is violated, witnessed by the expansion $H = (\{t, e\}, \{(t, e)\})$, which yields $\{t, d, b\} \in adm(F_1 \cup H)$, but $\{t, d, b\} \notin adm(F_2 \cup H)$. ◊

Preferred Semantics. The characterization for prf is very much like for adm, the only difference being that one considers $prf_C(\cdot)$ instead of $adm_C(\cdot)$. This similarity reflects the fact that $F \equiv_C^{prf} G$ whenever $F \equiv_C^{adm} G$.

Theorem 4.12. *Let F, G be AFs and $C \subseteq U$. Then, $F \equiv_C^{prf} G$ iff the following conditions jointly hold: (1) $A(F) \setminus C = A(G) \setminus C$; (2) $prf_C(F) = prf_C(G)$; and (3) for all $E \in prf_C(F)$, (3a) $E_F^+ \setminus C = E_G^+ \setminus C$ and (3b) $E_F^- \setminus E_F^+ = E_G^- \setminus E_G^+$.*

Table 1: Complexity of C-relativized Equivalence.

σ	*naive*	*grd*	*adm*	*com*	*stb*	*prf*
$F \equiv^{\sigma}_{C} G$	in P	coNP-c.	coNP-c.	coNP-c.	coNP-c.	Π_2^{P}-c.

Complete Semantics. For *com* semantics we have all the conditions we had for *adm*, but also the additional condition (3c) that ensures that the same arguments are defended in $F \cup H$ and $G \cup H$, for all AFs H over $U \setminus C$.

Theorem 4.13. *Let F, G be AFs and $C \subseteq U$. Then, $F \equiv^{com}_{C} G$ iff the following conditions jointly hold: (1) $A(F) \setminus C = A(G) \setminus C$; (2) $com_C(F) = com_C(G)$; and (3) for all $E \in com_C(F)$, (3a) $E^+_F \setminus C = E^+_G \setminus C$, (3b) $E^-_F \setminus E^+_F = E^-_G \setminus E^+_G$, and (3c) for all S with $E^-_F \setminus E^+_F \subseteq S \subseteq A(F) \setminus (C \cup E)$, if $\mathcal{F}_{F\setminus S}(E) \cap C = E \cap C$ or $\mathcal{F}_{G\setminus S}(E) \cap C = E \cap C$ then $\mathcal{F}_{F\setminus S}(E) = \mathcal{F}_{G\setminus S}(E)$.*

Grounded Semantics. For the characterization of *grd* we make use of the following variant of the characteristic function, $\mathcal{F}_{F,E}(S) = \{a \in E \mid S \text{ defends } a \text{ in } F\}$ for $E \subseteq A(F)$.

Theorem 4.14. *Let F, G be AFs and $C \subseteq U$. Then, $F \equiv^{grd}_{C} G$ iff the following holds: (1) $A(F) \setminus C = A(G) \setminus C$; (2) $grd_C(F) = grd_C(G)$; and (3) for all $E \in grd_C(F)$ and all $S \subseteq A(F) \setminus (C \cup E)$ (3a) $\mathcal{F}^{\infty}_{F\setminus S, E\cup C}(\emptyset) = \mathcal{F}^{\infty}_{G\setminus S, E\cup C}(\emptyset)$, (3b) if $\mathcal{F}^{\infty}_{F\setminus S, E\cup C}(\emptyset) = E$ then $E^+_F \setminus (C \cup S) = E^+_G \setminus (C \cup S)$, and (3c) if $\mathcal{F}^{\infty}_{F\setminus S, E\cup C}(\emptyset) = E$ then $\mathcal{F}_{F\setminus S}(E) = \mathcal{F}_{G\setminus S}(E)$.*

Conflict-free and Naive Semantics. Notice that two AFs possess the same conflict-free sets iff they possess the same naive extensions and thus $\equiv^{cf}_{C}$ and $\equiv^{naive}_{C}$ coincide. Moreover, the C-restricted semantics of *cf* is just *cf* itself.

Theorem 4.15. *Let F, G be AFs and $C \subseteq U$. Then, $F \equiv^{cf}_{C} G$ ($F \equiv^{naive}_{C} G$) iff the following conditions jointly hold: (1) $cf(F) = cf(G)$ and (2) $A(F) \setminus C = A(G) \setminus C$.*

5 Computational Properties

While strong equivalence can be efficiently decided (cf. [Oikarinen and Woltran, 2011]), testing standard equivalence is coNP-hard for $\sigma \in \{stb, adm, prf, com\}$ as it generalizes the problem of deciding whether an AF has a (non-empty) extension [Dunne and Wooldridge, 2009]. These hardness results extend to C-relativized equivalence. Upper bounds are given by the characterizations presented in Section 4.2. Our complexity results are summarized in Table 1 ($\mathcal{C}$-c. stands for $\mathcal{C}$-complete). Grounded semantics has a special behavior: while both standard and strong equivalence are tractable, C-relativized equivalence is coNP-complete as we show next.

Theorem 5.1. *Deciding $F \equiv^{grd}_{C} G$ is* coNP*-complete.*

Proof. The membership in coNP is due the characterization in Thm. 4.14. coNP-hardness is by a reduction from deciding whether two CNF formulas are equivalent. Let φ and ψ be two CNF formulas over atoms X and let C_φ, C_ψ be the sets of clauses. Moreover, we can assume that φ and ψ do not have partial models. For a CNF formula ϕ with clauses C_ϕ we define the AF $F_\phi = (A, R)$ with $A = X \cup \bar{X} \cup C_\phi \cup \{t\}$ and $R = \{(c,t),(c,c) \mid c \in C_\phi\} \cup \{(x,\bar{x}),(\bar{x},x) \mid x \in X\} \cup \{(x,c) \mid x \in c \in C_\phi\} \cup \{(\bar{x},c) \mid \neg x \in c \in C_\phi\}$. We now have that $\varphi \equiv \psi$ iff $F_\varphi \equiv^{grd}_{C} F_\psi$ for $C = C_\varphi \cup C_\psi \cup \{t\}$. $\square$

It remains to show Π_2^{P}-hardness of $F \equiv^{prf}_{C} G$. We prove the result for $F \equiv^{prf} G$ by reduction from the Π_2^{P}-complete problem of deciding whether an AF F is coherent [Dunne and Bench-Capon, 2002], i.e., whether $stb(F) = prf(F)$.

Theorem 5.2. *Deciding $F \equiv^{prf}_{C} G$ is Π_2^{P}-complete.*

Proof. Membership in Π_2^{P} follows from Thm. 4.12. We show hardness for testing $F \equiv^{prf} G$ (a corner case of $F \equiv^{prf}_{C} G$) by a reduction from the problem of testing whether an AF F is coherent, where we can assume that $\emptyset \notin prf(F)$. We transform F to an AF $F' = (A', R')$ with $A' = A(F) \cup \{t\}$ and $R' = R(F) \cup \{(t,a),(a,t) \mid a \in A(F)\}$, yielding $stb(F') = stb(F) \cup \{\{t\}\}$ and $prf(F') = prf(F) \cup \{\{t\}\}$. That is, F' is coherent iff F is coherent but we have $stb(F') \neq \emptyset$. Now we can apply Translation 4 from [Dvořák and Woltran, 2011] which maps F' to an AF G such that $stb(F') = prf(G)$ (given that $stb(F') \neq \emptyset$) and can be efficiently computed. Hence $stb(F) = prf(F)$ iff $prf(F') = prf(G)$. $\square$

Recall that for $C = \emptyset$, testing $\equiv^{\sigma}_{C}$ equivalence is computationally easy, while it is hard in the general case. Thus, one promising approach towards practical feasible algorithms is to consider characterizations whose performance depends on the set C. In other words, given AFs F and G to be compared under $\equiv^{\sigma}_{C}$, we aim to restrict the comparison of the C-restricted extensions (which is indeed the most expensive test in all characterizations). We give a first result into that direction for stable semantics. Let F be an AF and $B, E \subseteq U$. We define the *stable reduct* of F w.r.t. E and B as the AF $F^*_{B,E} = (A(F) \setminus E^{\oplus}_F, R^*)$ with $R^* = \{(a,b) \in R(F) \mid a, b \in A(F) \setminus E^{\oplus}_F\} \cup \{(a,a) \mid a \in (A(F) \cap B) \setminus E^{\oplus}_F\}$.

Theorem 5.3. *Let F, G be AFs, $C \subseteq A(F \cup G)$, and $B = C^{\oplus}_{F\cup G} \cup C^{-}_{F\cup G}$. Then, $F \equiv^{stb}_{C} G$ iff the following conditions jointly hold: (1) if $stb_C(F \cap B) \neq \emptyset$, $A(F) \setminus C = A(G) \setminus C$; (2) $stb_C(F \cap B) = stb_C(G \cap B)$; and (3) for all $E \in stb_C(F \cap B)$, $F^*_{B,E} \equiv^{stb}_{S} G^*_{B,E}$.*

In the above characterization the number of C-restricted sets we have to consider in (1) and (2) does not depend on the number of total arguments but only on the number of arguments that are either in $C \cap A(F \cup G)$ or neighbors of such arguments. Moreover, the strong equivalence in (3) can be tested in polynomial time.

6 Simplifications

We come back to the issue of simplification raised in the introduction. We begin by defining the notion of replacement.

Definition 6.1. *Given AFs F, F', G such that $A(F') \subseteq A(F) \cup (U \setminus A(G))$ and F is a sub-AF of G (i.e., $A(F) \subseteq A(G)$ and $R(F) = R(G) \cap (A(F) \times A(F))$), let $A = (A(G) \setminus A(F)) \cup A(F')$. The* replacement of F by F' in G *is defined as $G[F/F'] = (A, ((R(G) \setminus R(F)) \cap (A \times A)) \cup R(F'))$.*

As it turns out, faithfulness of the replacement of a sub-AF by another within a larger AF G follows from C-relativized equivalence of the the sub-AFs conjoined with their immediate neighborhood in G.

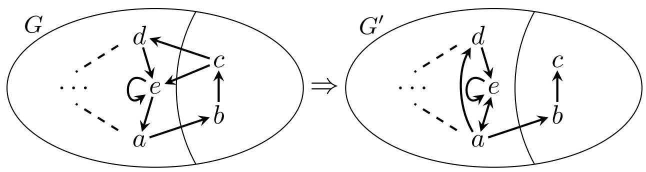
Proposition 6.2. *For AFs F, F', G and $C \subseteq U$ such that $A(F) \cup A(F') \subseteq C$, $(A(G) \setminus A(F)) \cap C = \emptyset$, and F a sub-AF of G, let $B = (A(F))^{\oplus}_{G} \cup (A(F))^{-}_{G}$ and $F^G = (B, R(G) \cap (B \times B))$. Then, $F^G \equiv^{\sigma}_{C} F^G[F/F']$ implies $G \equiv^{\sigma} G[F/F']$.*

A key feature of Def. 6.1 is that the attacks connecting the AFs F and F' to G are not changed, unless the involved argument in F is removed in F' (then the attack is also removed). Therefore the condition for C-relativized equivalence boils down to $stb_C(F^G) = stb_C(F^G[F/F'])$, since the other conditions from Thm. 4.8 are trivially satisfied (similar observations can be given for the other semantics).

Example 6.3. *Recalling the introductory example, faithfulness of replacing F_{abc} by F_{ab} in an arbitrary larger AF G being connected to F_{abc} by an attack (d, a) (cf. Figure 1), is then verified by $stb_C(F^G_{abc}) = \{\{d, b\}\} = stb_C(F^G_{ab})$. In other words we have that cycles of length 3 can be simplified under the stable semantics to two arguments, whenever the cycle has exactly one incoming attack. This kind of simplification can be generalized to arbitrary odd-length cycles in C, allowing for potential deletion of several arguments.* $\Diamond$

The replacement of sub-AFs with fixed connections to the rest-AF is a particular application of the results of Section 4.2. The notion of C-relativized equivalence is, however, more general and gives rise to simplifications of the following kind.

Example 6.4. *Consider the AFs G and G' depicted below.*



Note the single strongly connected component in G is split into three (smaller) components in G'. Let F, F' be the sub-AFs of G, G' with arguments $\{a, b, c, d, e\}$. To prove $G \equiv^{stb} G'$ we show $F \equiv^{stb}_{C} F'$ for $C = \{b, c\}$: (1) $A(F) \setminus C = \{a, d, e\} = A(F') \setminus C$, (2) $stb_C(F) = \{\{a, c\}, \{b\}, \{b, d\}\} = stb_C(F')$, and (3) $\{a, c\}^{+}_{F} \setminus C = \{d, e\} = \{a, c\}^{+}_{F'} \setminus C$, $\{b\}^{+}_{F} \setminus C = \emptyset = \{b\}^{+}_{F'} \setminus C$, $\{b, d\}^{+}_{F} \setminus C = \{e\} = \{b, d\}^{+}_{F'} \setminus C$. Again, this result can be generalized to arbitrary even-length paths among arguments in C. $\Diamond$

7 Related Work

Strong equivalence as well as further related notions have been thoroughly studied in the literature (cf. [Baumann, 2014; Baumann and Brewka, 2015; Baumann and Woltran, 2016]). Almost all of these notions are somehow disappointing regarding potential space for simplification. In fact, for most of these notions no arguments are redundant and deletions of attacks rely on the presence of self-loops. In particular, in case of self-loop-free AFs nothing can be simplified.

The concept of restricted admissible and stable semantics has been considered in dynamic programming algorithms based on tree-decompositions [Dvořák *et al.*, 2012; Charwat, 2012]. An investigation on the amount of neighborhood (in a graph-theoretical sense) needed to verify acceptability for different semantics is [Baumann *et al.*, 2016].

The issue of local evaluation of AFs was also tackled in the work on input/output AFs [Baroni *et al.*, 2014a; Giacomin *et al.*, 2016]. There the behavior of AFs (with dedicated input- and output-arguments) is described by the possible valuations of the output-arguments for each possible input. For the most prominent semantics it is shown whether having the same I/O behavior is sufficient for replacing one AF by another without affecting the evaluation of the rest of the AF. Our notion of equivalence is stronger as it is w.r.t. the whole AF and not only the part not affected by the replacement. Moreover, our work differs as we do not explicitly model I/O arguments and are more focused on finding exact conditions for the faithfulness of replacements.

The works on splitting [Baumann, 2011; Baroni *et al.*, 2014b] and SCC-recursiveness [Baroni *et al.*, 2005] allow for local evaluations but require that strongly connected components are considered in a specific order. Baumann *et al.* [2012] relaxed these conditions for stable semantics.

The concept of relativized equivalence was also studied for other nonmonotonic formalisms, in particular for Answer-Set Programming, see e.g. [Eiter *et al.*, 2007]. As well, simplification strategies have been suggested on basis of equivalence notions. Such replacements are typically defined as an exchange of rules in a logic program. This already indicates the main difference to our work, since replacing sub-graphs in AFs provides some subtle issues to be taken into consideration (cf. Section 6). This also might explain why in abstract argumentation the relation between equivalence notions and simplifications has been underexplored so far.

8 Discussion

In this paper, we introduced a general notion of equivalence for AFs and studied their characterizations and complexity.

There are several ways to pursue the presented research. First, an inclusion of other extension-based and labelling-based semantics is an immediate objective. Another direction to consider are weaker versions of C-relativized equivalence, for instance in analogy to normal expansion equivalence [Baumann, 2012], altering Def. 3.3 such that attacks between the original arguments of F and G cannot be changed. This situation is typical in the instantiation-based context (where AFs are constructed from an underlying knowledge base) since usually one can decide whether there is a conflict between arguments by solely considering these arguments.

On the practical side, we plan to employ our notion of equivalence for a systematic investigation of possible simplifications and to implement these findings in a preprocessing tool for abstract argumentation systems.

Finally, we plan to study restricted equivalence in the general setting of graph problems (as it was already done for strong equivalence by Lonc and Truszczyński [2011]) which might yield results that go beyond the field of argumentation.

References

[Baroni *et al.*, 2005] Pietro Baroni, Massimiliano Giacomin, and Giovanni Guida. SCC-recursiveness: A general schema for argumentation semantics. *Artif. Intell.*, 168(1-2):162–210, 2005.

[Baroni *et al.*, 2011] Pietro Baroni, Martin Caminada, and Massimiliano Giacomin. An introduction to argumentation semantics. *Knowledge Eng. Review*, 26(4):365–410, 2011.

[Baroni *et al.*, 2014a] Pietro Baroni, Guido Boella, Federico Cerutti, Massimiliano Giacomin, Leendert W. N. van der Torre, and Serena Villata. On the input/output behavior of argumentation frameworks. *Artif. Intell.*, 217:144–197, 2014.

[Baroni *et al.*, 2014b] Pietro Baroni, Massimiliano Giacomin, and Beishui Liao. On topology-related properties of abstract argumentation semantics. A correction and extension to dynamics of argumentation systems: A division-based method. *Artif. Intell.*, 212:104–115, 2014.

[Baumann and Brewka, 2010] Ringo Baumann and Gerhard Brewka. Expanding argumentation frameworks: Enforcing and monotonicity results. In Pietro Baroni, Federico Cerutti, Massimiliano Giacomin, and Guillermo R. Simari, editors, *Proc. COMMA*, volume 216 of *Frontiers in Artificial Intelligence and Applications*, pages 75–86. IOS Press, 2010.

[Baumann and Brewka, 2015] Ringo Baumann and Gerhard Brewka. The equivalence zoo for Dung-style semantics. *J. Log. Comput.*, 2015. doi.org/10.1093/logcom/exv001.

[Baumann and Woltran, 2016] Ringo Baumann and Stefan Woltran. The role of self-attacking arguments in characterizations of equivalence notions. *J. Log. Comput.*, 26(4):1293–1313, 2016.

[Baumann *et al.*, 2012] Ringo Baumann, Gerhard Brewka, Wolfgang Dvořák, and Stefan Woltran. Parameterized splitting: A simple modification-based approach. In Esra Erdem, Joohyung Lee, Yuliya Lierler, and David Pearce, editors, *Correct Reasoning - Essays on Logic-Based AI in Honour of Vladimir Lifschitz*, volume 7265 of *Lecture Notes in Computer Science*, pages 57–71. Springer, 2012.

[Baumann *et al.*, 2016] Ringo Baumann, Thomas Linsbichler, and Stefan Woltran. Verifiability of argumentation semantics. In Pietro Baroni, Thomas F. Gordon, Tatjana Scheffler, and Manfred Stede, editors, *Proc. COMMA*, volume 287 of *Frontiers in Artificial Intelligence and Applications*, pages 83–94. IOS Press, 2016.

[Baumann *et al.*, 2017] Ringo Baumann, Wolfgang Dvořák, Thomas Linsbichler, and Stefan Woltran. A general notion of equivalence for abstract argumentation. Technical Report DBAI-TR-2017-105, TU Wien, 2017.

[Baumann, 2011] Ringo Baumann. Splitting an argumentation framework. In James P. Delgrande and Wolfgang Faber, editors, *Proc. LPNMR*, volume 6645 of *Lecture Notes in Computer Science*, pages 40–53. Springer, 2011.

[Baumann, 2012] Ringo Baumann. Normal and strong expansion equivalence for argumentation frameworks. *Artif. Intell.*, 193:18–44, 2012.

[Baumann, 2014] Ringo Baumann. Context-free and context-sensitive kernels: Update and deletion equivalence in abstract argumentation. In Torsten Schaub, Gerhard Friedrich, and Barry O'Sullivan, editors, *Proc. ECAI*, volume 263 of *Frontiers in Artificial Intelligence and Applications*, pages 63–68. IOS Press, 2014.

[Bench-Capon and Dunne, 2007] Trevor J. M. Bench-Capon and Paul E. Dunne. Argumentation in artificial intelligence. *Artif. Intell.*, 171(10-15):619–641, 2007.

[Charwat, 2012] Günther Charwat. Tree-decomposition based algorithms for abstract argumentation frameworks. Master's thesis, TU Wien, 2012.

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

[Dunne and Bench-Capon, 2002] Paul E. Dunne and Trevor J. M. Bench-Capon. Coherence in finite argument systems. *Artif. Intell.*, 141(1/2):187–203, 2002.

[Dunne and Wooldridge, 2009] Paul E. Dunne and Michael Wooldridge. Complexity of abstract argumentation. In Iyad Rahwan and Guillermo R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 85–104. Springer, 2009.

[Dvořák and Woltran, 2011] Wolfgang Dvořák and Stefan Woltran. On the intertranslatability of argumentation semantics. *J. Artif. Intell. Res. (JAIR)*, 41:445–475, 2011.

[Dvořák *et al.*, 2012] Wolfgang Dvořák, Reinhard Pichler, and Stefan Woltran. Towards fixed-parameter tractable algorithms for abstract argumentation. *Artif. Intell.*, 186:1 – 37, 2012.

[Eiter *et al.*, 2007] Thomas Eiter, Michael Fink, and Stefan Woltran. Semantical characterizations and complexity of equivalences in answer set programming. *ACM Trans. Comput. Log.*, 8(3), 2007.

[Giacomin *et al.*, 2016] Massimiliano Giacomin, Thomas Linsbichler, and Stefan Woltran. On the functional completeness of argumentation semantics. In Chitta Baral, James P. Delgrande, and Frank Wolter, editors, *Proc. KR*, pages 43–52. AAAI Press, 2016.

[Lonc and Truszczyński, 2011] Zbigniew Lonc and Mirosław Truszczyński. On graph equivalences preserved under extensions. *Discrete Mathematics*, 311(12):966–977, 2011.

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

[Rahwan and Simari, 2009] Iyad Rahwan and Guillermo R. Simari. *Argumentation in Artificial Intelligence*. Springer, 2009.