Under-Approximating Semantics in Clustered Assumption-Based Argumentation

Iosif Apostolakis, Johannes P. Wallner

Graz University of Technology
iosif.apostolakis@tugraz.at, johannes.p.wallner@tugraz.at

Abstract

Computational argumentation studies fundamental methods for reasoning within Artificial Intelligence (AI). Two prominent subfields in computational argumentation are abstract argumentation and structured argumentation. Abstract argumentation focuses on the interactions between arguments, ignoring their internal structure, while structured approaches utilize a given knowledge base to construct the arguments. Thus, the latter approach incorporates the internal structure of arguments into the reasoning process. In this work we introduce a form of abstraction on the well established structured approach of Assumption-Based Argumentation (ABA). Our goal is to provide methods to simplify complicated scenarios, by applying clustering over defeasible parts. Abstraction, particularly clustering, has been explored in recent research on abstract argumentation and in the adjacent field of logic programming. In fact, while clustering has also been applied to ABA, our approach takes a different, or rather dual, direction. In contrast to prior work on over-approximation on ABA, we propose the dual approach of under-approximation. We provide semantics for reasoning over clustered frameworks in a sound manner relative to original semantics, ensuring that any set deemed acceptable in the clustered scenario corresponds to an acceptable set. We show properties of the under-approximating semantics and illustrate our approach using a conceptual example based on medical recommendations.

Introduction

Argumentation is a key research area in Artificial Intelligence (AI), offering formal tools to model reasoning and debate in a structured, dialectical manner (Baroni et al. 2018; Gabbay et al. 2021). Over the years, this line of research has found a wide range of applications (Atkinson et al. 2017), including domains such as legal reasoning (Prakken and Sartor 2015), decision-making in healthcare (Čyras et al. 2021a), and in multi-agent systems (Amgoud, Dimopoulos, and Moraitis 2007; Dimopoulos, Mailly, and Moraitis 2019; Fan and Toni 2012).

Two major paradigms are standing out for formalizing argumentation processes, namely, structured argumentation (Čyras et al. 2018; Modgil and Prakken 2018; García and Simari 2018; Besnard and Hunter 2018; Gordon, Prakken, and Walton 2007; Kakas, Moraitis, and Spanoudakis 2019) and abstract argumentation (Dung 1995; Baroni et al. 2018). Structured approaches explicitly encode the way arguments are constructed from, potentially inconsistent, knowledge bases. In contrast, abstract argumentation abstracts away from internal structure and treats arguments as atomic entities linked by an attack relation, most commonly formalized as argumentation frameworks (AFs) (Dung 1995). In AFs arguments are represented as nodes in a graph, and directed edges indicate which arguments attack which others. To determine which sets of arguments are deemed acceptable, several semantical criteria have been proposed, among which admissibility and stable semantics are among the most prominent (Baroni, Caminada, and Giacomin 2011).

One of the central goals of argumentation is to support explanations, that is, reasoned justifications for or against certain claims. Accordingly, many methods have been developed over time to enhance the explainability of argumentation based reasoning (Čyras et al. 2021b; Vassiliades, Bassiliades, and Patkos 2021). A common approach in this setting is the simplification of argumentation frameworks to highlight relevant information (Boella, Kaci, and van der Torre 2009; Fan and Toni 2015; Baroni et al. 2014). Techniques such as abstraction are employed for this purpose, and a notable example is a recent method introducing clustering (Saribatur and Wallner 2021; Apostolakis, Saribatur, and Wallner 2024a,b).

In this paper, we focus on existential abstraction within structured argumentation. The idea is to cluster parts of a structured argumentation setting, enabling reasoning over a simplified version of the framework. This approach is developed within the well-established structured model of Assumption-Based Argumentation (Čyras et al. 2018), and is designed to preserve enough structure to allow meaningful conclusions while abstracting away detail that may be irrelevant or unwanted in specific reasoning contexts. ABA has been applied in areas such as medical decision-making (Čyras et al. 2021a) and coordination in multi-agent systems (Gao et al. 2016), offering a flexible setting for representing and resolving conflicts.

While recent work on abstraction in argumentation (Saribatur and Wallner 2021; Apostolakis, Saribatur, and Wallner 2024a,b) aims to ensure that all relevant behavior of the original framework is preserved in the abstracted ver-

sion, we take a different route. In this paper, we focus on under-approximation, in contrast to the aforementioned over-approximation, where the goal is not to capture all possible behaviors, but rather to provide sound reasoning outcomes that are guaranteed to be valid in the original framework. This perspective allows us to safely ignore certain details while ensuring that any conclusions drawn from the abstracted, clustered structure remain justified when mapped back to the full ABA setting. By under-approximating, we provide reliable reasoning in complex scenarios without introducing spurious argumentative conclusions.

A central concern in argumentation theory is determining which sets of arguments can be deemed acceptable. Acceptability is defined through argumentation semantics, which provide formal criteria such as admissibility, complete and stable semantics, for evaluating sets of arguments based on how they interact, particularly through attack and defense relations. In ABA these notions are instantiated with sets of assumptions, which are a core structural component of the framework. In ABA, conflicts arise when one set of assumptions leads to a conclusion that contradicts (i.e., attacks) an assumption in another set. Acceptability in ABA is defined over sets of assumptions rather than abstract arguments. A set of assumptions is said to be conflict-free if it does not derive the contrary of any of its own members. Such a set is admissible if it defends itself, that is, it must be able to attack any other assumption set that attacks it.

The following example is adapted from a medical recommendation scenario presented by Zamborlini et al. (2017), with slight modifications to better align with the properties we aim to illustrate.

Example 1. *The example models a cancer treatment scenario involving chemotherapy and exercise therapy. Chemotherapy is the primary treatment, and two types of exercise therapy, namely standard and low-intensity, are recommended to mitigate side effects like fatigue and reduced physical fitness. However, these exercise therapies may be contraindicated if the patient has a fever or they may be unsuitable for patients with joint pain or joint inflammation.*

In a later section we construct a structured argumentation framework to represent this scenario. With such a framework we can argue whether a specific treatment could be acceptable for a given patient. That is, assuming that a patient has fever, we can analyze whether an exercise treatment would be a viable option for fighting fatigue. Conversely, if it is uncertain whether the patient has joint inflammation, we can leave this assumption open and let the reasoning process determine whether chemotherapy can be acceptable as a treatment in scenarios where the patient does or does not have joint inflammation.

Making decisions on whether a treatment is suitable for a patient is a task that requires sound reasoning. Also when making such decisions, one might be provided with more detailed information than what is required to draw their conclusions. For example, knowing that a patient suffers from arthritis is enough to rule out exercise therapy. Knowing in addition to arthritis whether the patient has joint pain or joint inflammation is information that in this scenario makes no difference to the decision making.

Abstraction via clustering allows us to group assumptions with similar argumentative behavior, enabling reasoning at a coarser level of granularity. Crucially, by using under-approximation, we ensure that any admissible (complete, stable) assumption set from such a clustered framework remains sound with respect to the original setting, even though some information has been omitted. This enables safe reasoning in settings where full information may be either unnecessary or unavailable.

Our main contributions are as follows.

- We introduce semantics for reasoning in clustered frameworks, abstracting classical semantics and concepts such as conflict-freeness, admissibility, complete, and stable semantics.
- We prove that for a set acceptable in an abstracted scenario, it holds that no matter what information was abstracted away, there always is a corresponding original acceptable set.
- We show that the under-approximating semantics have the same computational complexity of reasoning as classical semantics (Dimopoulos, Nebel, and Toni 2002).
- We illustrate the conceptual applicability of our approach via a use case inspired by medical treatment recommendations, demonstrating how abstraction may support sound reasoning in a medical setting.
- Our study complements earlier research on over-approximation in argumentation, by focusing on under-approximation.

Background

We recall terminology and foundational concepts for ABA.

Assumption-based Argumentation Our work is based on ABA frameworks, originally proposed in (Bondarenko et al. 1997). At the core of ABA lies a deductive system, denoted by a pair $(\mathcal{L}, \mathcal{R})$, where $\mathcal{L}$ is a formal language and $\mathcal{R}$ is a collection of inference rules over $\mathcal{L}$. Each rule $r \in \mathcal{R}$ has the form $a_0 \leftarrow a_1, \ldots, a_n$, where $a_i \in \mathcal{L}$. We write $head(r) = a_0$ for the rule's head and $body(r) = \{a_1, \ldots, a_n\}$ for its (possibly empty) body.

An ABAF enriches this deductive system by identifying a subset of $\mathcal{L}$ as assumptions and assigning to each a contrary.

Definition 1. *An ABA framework is a tuple $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, {}^-)$, with $(\mathcal{L}, \mathcal{R})$ a deductive system, $\mathcal{A} \subseteq \mathcal{L}$ a non-empty set of assumptions, and ${}^-$ a total function mapping assumptions $a \in \mathcal{A}$ to their contrary $\overline{a} \in \mathcal{L}$.*

This contrary mapping naturally extends to sets as follows: for $S \subseteq \mathcal{A}$, $\overline{S} = \{\overline{a} \mid a \in S\}$. We also focus on flat ABA frameworks, where no assumption can be derived via rules, which means that for every rule $r \in \mathcal{R}$, we have $head(r) \notin \mathcal{A}$. Additionally, we restrict attention to finite frameworks—$\mathcal{L}$ and $\mathcal{R}$ are assumed finite.

In case we omit to explicitly define D, we implicitly assume an ABAF $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, {}^-)$. An atom $s \in \mathcal{L}$ is derivable by a set of assumptions $A \subseteq \mathcal{A}$ if there is a sequence of rules $r_1, \ldots, r_n$, where $head(r_n) = s$ and the body of rule

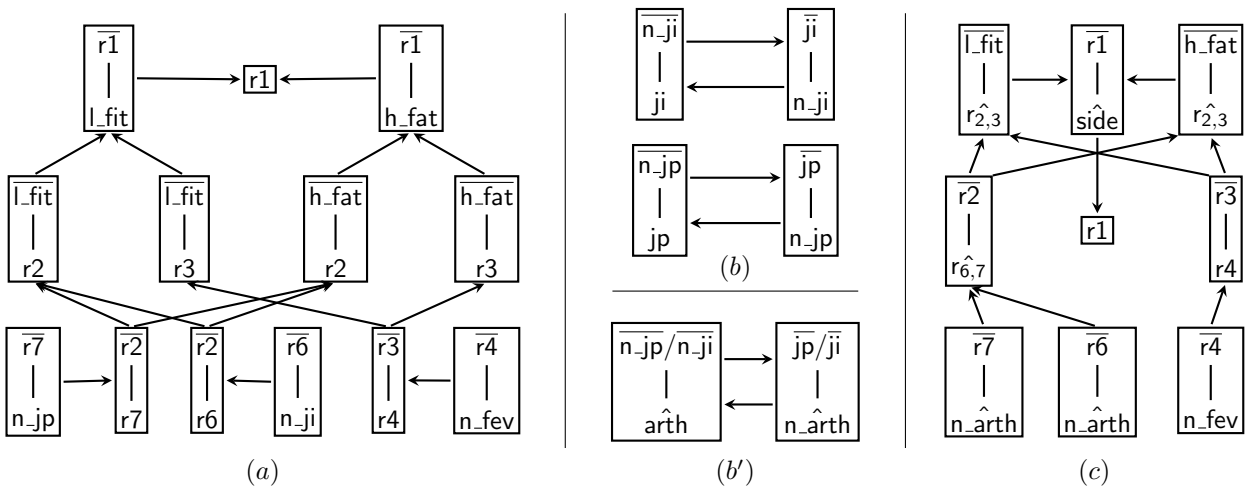

Figure 1: This figure shows part of an ABA framework (a), which also includes rule structures for each assumption representing a condition such as those shown in (b). In (c), we present a clustered version of (a), where some assumptions are replaced by their corresponding clusters. Part (b') shows the clustered counterpart of (b). Due to space constraints, a single argument with two conclusions is depicted, in place of two arguments each with a single conclusion.

r_i is a subset of $A \cup \bigcup_{j<i}\{head(r_j)\}$. For a set of assumptions A we define the deductive closure of A, $Th_D(A)$, to be the set of atoms that are derivable from A in D. We omit the subscript D when the context is clear. Note that we define $A \subseteq Th_D(A)$ to hold (assumptions can derive themselves).

A set of assumptions A attacks an assumption a if the contrary of a lies in $Th_D(A)$.

Definition 2. *Let $A, B \subseteq \mathcal{A}$. Then A attacks B in D if there exists some $b \in B$ such that $\overline{b} \in Th_D(A)$.*

This attack relation is used to define key semantic concepts such as conflict-freeness and defense.

Definition 3. *A set $A \subseteq \mathcal{A}$ is conflict-free in D, written $A \in cf(D)$, if A does not attack itself. Moreover, A defends a set $B \subseteq \mathcal{A}$ if, for every $C \subseteq \mathcal{A}$ that attacks B, the set A also attacks C.*

We can utilize those notions to define standard ABA semantics such as admissible, complete, and stable semantics.

Definition 4. *Let $A \in cf(D)$. Define that A is admissible in D ($A \in adm(D)$) if A defends itself, A is complete in D ($A \in com(D)$) if $A \in adm(D)$ and A defends no $a \in \mathcal{A} \setminus A$, A is stable in D ($A \in stb(D)$) if A attacks each $a \in \mathcal{A} \setminus A$.*

Under a given semantics σ (e.g., admissible, complete, or stable), an atom $s \in \mathcal{L}$ is said to be credulously accepted in D if there exists an $A \in \sigma(D)$ s.t. $s \in Th_D(A)$, and skeptically accepted if $s \in Th_D(A)$ for all $A \in \sigma(D)$.

Example 2. *We now elaborate on the technical details of our running example shown in Figure 1(a). In this framework, assumptions consist of the recommendations $r_1, \ldots, r_7$, the condition-related assumptions l_fit, h_fat, n_fev, n_jp, n_ji, along with additional assumptions opposing the presence of each condition-related assumption, i.e. h_fit,l_fat, fev, jp, and ji respectively. Indicatively, h_fat corresponds to high fatigue while l_fat to low fatigue, modelling opposing conditions. Assumptions $r_1, \ldots, r_7$ correspond to medical recommendations. Specifically, r_1 represents chemotherapy, while r_2 and r_3 denote two alternative types of exercise therapy. Recommendation r_4 advises against performing the exercise in r_3 when the patient has a high fever. In addition, r_6 and r_7 state that any form of exercise should be avoided in the presence of joint inflammation or joint pain, respectively. We consider the language to be the set of assumptions $\mathcal{A}$ along with a contrary for each assumption, i.e. $\mathcal{L} = \mathcal{A} \cup \overline{\mathcal{A}}$. The rule set includes all rules depicted in boxes in Figure 1(a), along with those rules used to model the mutual attacks between assumptions representing opposing conditions, such as the ones depicted in Figure 1(b). Despite not showing rules representing mutual attacks between other opposing conditions in Figure 1, we do consider them to be part of the framework. Indicatively, our framework also contains rules $\overline{n\_fev} \leftarrow fev$ and $\overline{fev} \leftarrow n\_fev$, although they are visually omitted. The contrary relations follow directly from the naming of each atom.*

In this running example, we focus on reasoning about acceptable sets that contain r_1, corresponding to chemotherapy. This reflects the case in which a patient is known to have cancer and wishes to determine whether there exists a scenario in which some form of exercise can be performed alongside chemotherapy. According to admissibility, to have an accepted set that contains r_1, we must defend it from any received attacks. Thus, we need to also attack assumptions l_fit and h_fat, which in turn means that we need to have either r_2 or r_3 (or both) in our set. Finally to have those, we need to be defending them. To defend r_2 we ought to have both n_jp and n_ji, while for r_3 we just need n_fev. To conclude, some admissible sets that contain r_1 are, $A_1 = \{r_1, r_2, n\_jp, n\_ji\}$, $A_2 = \{r_1, r_3, n\_fev\}$, and $A_3 = \{r_1, r_2, r_3, n\_jp, n\_ji, n\_fev\}$. The set A_1 repre-

sents a scenario in which the side effects of chemotherapy are mitigated through the exercise therapy recommended by r_2. This is possible because the patient does not exhibit joint pain or joint inflammation. Set A_2 supports recommendation r_3 instead, as in this scenario there is no information about joint pain or inflammation, but we know that the patient has no fever. The latter set A_3 corresponds to a case where the patient is free of fever and shows no signs of arthritis, thereby allowing either (or both) exercise therapies to be performed.

Clustering Assumptions To enable reasoning over simplified versions of structured argumentation frameworks, we adopt existential abstraction through clustering of assumptions in ABA frameworks. This subsection recalls main definitions and key properties from earlier work (Apostolakis, Saribatur, and Wallner 2024a).

Clustering is formalized as a surjective mapping $m : \mathcal{A} \to \hat{\mathcal{A}}$, where $\mathcal{A}$ is the original set of assumptions in an ABAF, and $\hat{\mathcal{A}}$ is a set of clusters. Each cluster $\hat{a} \in \hat{\mathcal{A}}$ is a non-empty subset of $\mathcal{A}$, and together they partition $\mathcal{A}$. For instance, if $\mathcal{A} = \{a, b, c\}$, we might map a and b into the same cluster $\hat{a}$, while keeping c "unclustered", resulting in $m(a) = m(b) = \hat{a}$, $m(c) = c$. Unclustered assumptions are called singletons, mapped to themselves, and the set $Single(\hat{A})$ denotes the singletons inside some set of clusters $\hat{A}$. The mapping m extends pointwise to sets of assumptions: for any $S \subseteq \mathcal{A}$, we write $m(S) = \{m(a) \mid a \in S\}$. We also define the inverse mapping $m^{-1} : \hat{\mathcal{A}} \to 2^{\mathcal{A}}$, which maps clusters on the sets of their assumptions, and extend it in order to also map sets of clusters to the union of their contents: for any $\hat{S} \subseteq \hat{\mathcal{A}}$, we write $m^{-1}(\hat{S}) = \bigcup_{\hat{a} \in \hat{S}} \hat{a}$. For example, if $\hat{S} = \{\{a, b\}, \{c\}\}$, then $m^{-1}(\hat{S}) = \{a, b, c\}$. A set $S \subseteq \mathcal{A}$ is said to be a preimage of $\hat{S}$ iff $m(S) = \hat{S}$.

This transformation affects rules and contraries of the ABAF, giving rise to a clustered ABAF (cABAF).

Definition 5. *Let $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-)$ be an ABAF, and let $m : \mathcal{A} \to \hat{\mathcal{A}}$ be a clustering. Then the clustered ABAF $\hat{D} = m(D) = (\hat{\mathcal{L}}, \hat{\mathcal{R}}, \hat{\mathcal{A}}, \hat{^-})$ is defined as:*

- *$\hat{\mathcal{A}} = m(\mathcal{A})$, and $\hat{\mathcal{L}} = \mathcal{L} \setminus \mathcal{A} \cup \hat{\mathcal{A}}$,*
- *each rule $r \in \mathcal{R}$ is mapped to $\hat{r} = head(r) \leftarrow m(body(r))$, and*
- *the contrary function $\hat{^-} : \hat{\mathcal{A}} \to 2^{\hat{\mathcal{L}}}$ is defined as*

$$\hat{\overline{\hat{a}}} = \{\hat{b} \in \hat{\mathcal{A}} \mid \exists a \in \hat{a}, \exists b \in \hat{b}, \text{ such that } \overline{a} = b\} \cup \{\hat{x} \in \hat{\mathcal{L}} \setminus \hat{\mathcal{A}} \mid \exists a \in \hat{a}, \overline{a} = x\}.$$

Note that a cABAF differs from a classical ABAF in that the contrary function may now map to a set of atoms or clusters. When context allows, we omit the double-hat notation for readability (i.e. $\overline{\hat{a}}$ instead of $\hat{\overline{\hat{a}}}$). In case $body(r)$ contains an $x \notin A$ then we assume $m(x) = x$.

Assuming a clustered framework $\hat{D} = (\mathcal{L}, \hat{\mathcal{A}}, \hat{\mathcal{R}}, ^-)$ implicitly requires the existence of an original assumption set $\mathcal{A}$ that maps to $\hat{\mathcal{A}}$ through some partition mapping m. When the context is clear, we might omit referring to $\mathcal{A}$ or m. Also, if not defined otherwise, D should always refer to some preimage of $\hat{D}$, i.e., a framework mapping to $\hat{D}$ through m.

Example 3. *As an example, Figure 1(c) shows a clustered version of Figure 1(a). Here, the assumptions r_2 and r_3 are grouped into the cluster $\hat{r_{2,3}}$,representing some form of exercise therapy, while r_6 and r_7 form the cluster $\hat{r_{6,7}}$, representing recommendations against exercise. The assumptions n_ji and n_jp are clustered as $n\_\hat{arth}$, and ji and jp as $\hat{arth}$, representing the absence and presence of arthritis, respectively. Finally, the assumptions l_fit and h_fat as well as assumptions h_fit and l_fat are grouped into the clusters $\hat{side}$ and $n\_\hat{side}$ respectively, representing presence and absence of side effects associated with chemotherapy. Remaining assumptions are mapped to themselves.*

Clustering preserves certain derivability relationships. For example, the set of atoms that are derivable by a set of assumptions is a subset to the atoms derived by their clustered images i.e., $\forall x \in Th_D(S) \setminus \mathcal{A}$ we find that $x \in Th_{\hat{D}}(\hat{S})$. Moreover, if an atom is derivable from some set of clusters, then it is certain that there exist a preimage set that derives this atom, i.e., $x \in Th_{\hat{D}}(\hat{S})$, then $x \in Th_D(S')$ for some $S' \subseteq \mathcal{A}$. Flatness is also preserved.

To analyze argumentation semantics, we distinguish two kinds of attacks that can occur between clusters in a cABAF.

Definition 6. *Let $\hat{A}, \hat{B} \subseteq \hat{\mathcal{A}}$ in a cABAF $\hat{D}$.*

- *$\hat{A}$ (normally) attacks $\hat{B}$ if $\exists \hat{b} \in \hat{B}$ such that $\hat{\overline{b}} \cap Th_{\hat{D}}(\hat{A}) \neq \emptyset$,*
- *$\hat{A}$ fully attacks $\hat{B}$ if $\exists \hat{b} \in \hat{B}$ such that $\hat{\overline{b}} \subseteq Th_{\hat{D}}(\hat{A})$,*

Intuitively, normal attacks generalize classical attacks and full attacks require that all vulnerabilities of the attacked cluster are derived.

Example 4. *Assumption r_4 derives the contrary of r_3, which is part of the contraries of the cluster $\hat{r_{2,3}}$. Hence r_4 normally attacks the cluster $\hat{r_{2,3}}$. In addition to that, $\hat{r_{6,7}}$ also normally attacks $\hat{r_{2,3}}$, and if we combine these attacks, we see that $\{r_4, \hat{r_{6,7}}\}$ fully attack $\hat{r_{2,3}}$, since $Th_{\hat{D}}(\{r_4, \hat{r_{6,7}}\}) = \{\overline{r_2}, \overline{r_3}\} = \hat{\overline{r_{2,3}}}$. Also the cluster $\hat{side}$ fully attacks r_1, while the singleton n_fev also fully attacks r_4.*

Under-Approximation

In this section, we formalize the notion of under-approximation in the context of clustered Assumption-Based Argumentation frameworks (cABAFs).

A semantics $\hat{\sigma}$ on a cABAF $\hat{D}$ is a set $\hat{\sigma}(\hat{D}) \subseteq 2^{\hat{\mathcal{A}}}$, representing the sets of clustered assumptions that satisfy the given criteria. To distinguish between semantics on ABAFs and cABAFs, we refer to the former as classical semantics and the latter as clustered (or abstract) semantics, unless the context is clear. We next define the desired relationship between an abstract semantics $\hat{\sigma}$ and a classical semantics σ.

Definition 7. *An abstract semantics $\hat{\sigma}$ under-approximates a semantics σ iff for all frameworks D s.t. $m(D) = \hat{D}$, it holds that $\hat{\sigma}(\hat{D}) \subseteq m(\sigma(D))$.*

The intuition behind under-approximation is to ensure that any set of clustered assumptions accepted under abstract semantics corresponds to at least one acceptable set in every original framework. In other words, we restrict the abstract semantics to include only those sets that are guaranteed to preserve acceptability across all the possible preimages. This notion stands in contrast to over-approximating semantics, which accept a set of clusters if there exists at least one preimage framework in which a corresponding set of assumptions is acceptable under the original semantics. However, over-approximation allows for "spurious" sets to be accepted, i.e. sets mistakenly deemed accepted due to the abstracted information. Under-approximation, by contrast, is more cautious, as it avoids introducing spurious sets. Therefore, under-approximation trades off completeness for soundness as it may exclude some valid sets, but never includes spurious ones. Formally, for some over-approximating semantics σ_o it holds that $\sigma_o(\hat{D}) \supseteq m(\sigma(D))$. A clustering with no spurious set under some semantics, is called faithful w.r.t. this semantics. Intuitively, the connection between the over- and under-approximation is that given a specific clustering, a set of clusters that is under-approximated under some abstract semantics σ_u must also be accepted by any over-approximating semantics σ_o. That means that $\sigma_u(\hat{D}) \subseteq \sigma_o(\hat{D})$. If $\sigma_u(\hat{D}) = \sigma_o(\hat{D})$ holds, we can prove that $\hat{D}$ is a faithful clustering w.r.t. any of its preimages. From a reasoning perspective, under-approximation ensures that every credulously accepted cluster has a credulously accepted assumption in all preimages, while over-approximation preserves soundness for skeptical acceptance.

From this point on, unless explicitly stated otherwise, any reference to an abstract semantics shall be understood as referring to an under-approximating semantics. We begin by introducing a clustered version of conflict-freeness.

Definition 8. *A set $\hat{A} \subseteq \hat{\mathcal{A}}$ is abstract conflict-free, i.e. $\hat{A} \in \hat{cf}(\hat{D})$, if $\hat{A}$ does not fully attack itself.*

Example 5. *As stated in Example 4, the set $\{r_4, r_{\hat{6},7}, r_{\hat{2},3}\}$ is not conflict-free since $\{r_4, r_{\hat{6},7}\}$ fully attacks $r_{\hat{2},3}$. Not considering this set as abstract conflict-free is expected as Figure 1(a) is actually a preimage framework, where no preimage of $\{r_4, r_{\hat{6},7}, r_{\hat{2},3}\}$ is conflict-free. On the contrary, the set $\{r_{\hat{6},7}, r_{\hat{2},3}\}$ is abstract conflict-free since there is no full attack within it. It also holds that there is always a preimage of the set that is conflict-free. This is because, the contraries $\overline{r_3}, \overline{r_6}, \overline{r_7} \notin Th_{\hat{D}}(\{r_{\hat{6},7}, r_{\hat{2},3}\})$, and thus the sets $\{r_3, r_7\}$ and $\{r_3, r_6\}$ are always conflict-free in every preimage.*

In abstract contexts, we refer to abstract conflict-freeness simply as conflict-freeness, distinguishing it from strong conflict-freeness, which excludes all attacks among clusters. Having a strongly conflict-free set of clusters $\hat{S}$ essentially means that the entire preimage $m^{-1}(\hat{S})$ is conflict-free in any corresponding preimage framework. However, just avoiding full attacks among clusters is enough to ensure that, within each cluster, there exists at least one assumption whose contrary is not derived by the set of clusters. Moreover, if the set of clusters does not derive the contrary, then neither does any of its preimages. Thus this assumption remains unattacked in every preimage.

This leads to our next result, that this notion of conflict-freeness achieves to under-approximate the original semantics. That is, if $\hat{D}$ is a clustered ABA, then for any D preimage of $\hat{D}$, it holds that $\hat{cf}(\hat{D}) \subseteq m(cf(D))$.

Theorem 1. *It holds that $\hat{cf}$ under-approximates cf.*

An under-approximating semantics is considered optimal if, whenever a clustered set is not accepted, there exists a preimage of the framework in which all preimages of the set are also not accepted.

Definition 9. *Let $\hat{\sigma}$ be an abstract semantics that under-approximates σ. Then $\hat{\sigma}$ is optimal if for all clustered sets $\hat{A}$ s.t. in any preimage framework D there is an original set of assumptions $A \in \sigma(D)$ mapping to $\hat{A}$, we have $\hat{A} \in \hat{\sigma}(\hat{D})$.*

Under this notion of optimality, the following result shows that abstract conflict-freeness is optimal. Intuitively, no other definition of abstract conflict-freeness is both under-approximating and accepts more sets as $\hat{cf}$.

Theorem 2. *The abstract semantics $\hat{cf}$ is optimal.*

Next we define abstract admissibility. To this end, we first introduce the notion of defense in the abstract setting.

Definition 10. *Let $\hat{S}$ be a set of clusters. Then we say that:*

- *$\hat{S}$ defeats an atom x iff for every set $\hat{B}$ deriving x there is a cluster $\hat{b} \in \hat{B}$, such that $\overline{\hat{b}} \subseteq Th_{\hat{D}}(\hat{S})$,*
- *$\hat{S}$ weakly defeats x iff for every set $\hat{B}$ deriving x there is a cluster $\hat{b} \in \hat{B}$, such that $\overline{\hat{b}} \cap Th_{\hat{D}}(\hat{S}) \neq \emptyset$.*
- *$\hat{S}$ fully defends a cluster $\hat{a}$ iff $\hat{S}$ defeats the set $\overline{\hat{a}}$.*
- *$\hat{S}$ partially defends a cluster $\hat{a}$ iff $\hat{S}$ defeats part of the set of contraries of $\hat{a}$.*

Example 6. *Consider the atom $\overline{r_2}$. This atom is derived once, by the cluster $r_{\hat{2},3}$. Thus, the cluster $n\_\hat{arth}$ defeats $\overline{r_2}$, since it fully attacks $r_{\hat{2},3}$. Additionally, the above entails that $n\_\hat{arth}$ partially defends the cluster $r_{\hat{2},3}$, since it defeats part of its contrary set. In the same manner we can see that n_fev defeats $\overline{r_3}$ and thus the set $n\_\hat{arth}, n\_fev$, fully defends cluster $r_{\hat{2},3}$. Finally the set $\hat{S} = \{\hat{side}, r_4\}$ weakly defeats $\overline{l\_fit}$. This is because $\overline{l\_fit}$ is derived in two ways. One is by $r_{\hat{2},3}$ and the other is by $n\_\hat{side}$ (which is not shown in Figure 1 but is part of the visually omitted rules mentioned earlier). The later cluster is fully attacked by $\hat{S}$, however the former cluster is only normally attacked by $\hat{S}$.*

Towards the definition of admissibility, we first introduce the set $\hat{A}^*$ which denotes the maximal subset of $\hat{A} \in \hat{cf}(\hat{D})$ that fully defends itself. We observe that a set of clusters that fully defends itself has two key properties. First, if it is abstract conflict-free, then it is necessarily strongly conflict-free. Second, its entire preimage defends itself in every corresponding preimage framework. Therefore $\hat{A}^*$ plays an essential role in admissibility as any cluster that is (partially) defended by $\hat{A}^*$ is guaranteed to contain an assumption that

is defended by $m^{-1}(\hat{A}^*)$ in any preimage framework. In any conflict-free $\hat{A} \in \hat{cf}(\hat{D})$, the set $\hat{A}^*$ is unique.

Definition 11. *A set $\hat{A} \subseteq \hat{\mathcal{A}}$ is abstract admissible, i.e. $\hat{A} \in \hat{adm}(\hat{D})$, iff $\hat{A} \in \hat{cf}(\hat{D})$, and if $\hat{A}^*$ partially defends $\hat{A} \setminus \hat{A}^*$.*

Example 7. *The set $n\_\hat{arth}$ receives a full attack from the cluster $\hat{arth}$. However, $\{n\_\hat{arth}\}$ fully attacks $\hat{arth}$, thus it fully defends itself. This means that in this case $\{n\_\hat{arth}\}^* = \{n\_\hat{arth}\}$, hence $\{n\_\hat{arth}\}$ is abstract admissible, as $\{n\_\hat{arth}\}^*$ partially defends $\{n\_\hat{arth}\} \setminus \hat{A}^* = \emptyset$.*

Recall from Example 2 that the sets $A_1 = \{r_1, r_2, n\_jp, n\_ji\}$, $A_2 = \{r_1, r_3, n\_fev\}$, and $A_3 = \{r_1, r_2, r_3, n\_jp, n\_ji, n\_fev\}$ are admissible in the original framework. Let us consider the set $\hat{A}_1 = \{r_1, r_{\hat{2},3}, n\_\hat{arth}\}$. Here $\hat{A}_1^* = \{n\_\hat{arth}\}$, as it is the maximal subset that fully defends itself. Then according to Example 4 the cluster $r_{\hat{2},3}$ is partially defended, but r_1 is not defended by $\hat{A}_1^*$. Hence $\hat{A}_1$ is not abstract admissible. Not accepting this set is justified by the fact that if we removed from the original framework the rule $\overline{l\_fit} \leftarrow r_2$, then we obtain a new preimage D', where no preimage of $\hat{A}_1$ is admissible, because then r_1 cannot be defended. On the contrary $\hat{A}_3$ fully defends itself, and thus $\hat{A}_3^* = \hat{A}_3$. This means that its entire preimage is admissible under any preimage framework.

The intuition of admissibility lies in $\hat{A}^*$. This set collects the clusters in $\hat{A}$ that are strongly conflict-free and whose contraries are fully defeated by itself. As such, $\hat{A}^*$ acts as a "safe" subset since its entire preimage is guaranteed to be admissible, and any atom defeated by $\hat{A}^*$ ensures that all assumptions having this atom as a contrary are defended.

Theorem 3. *It holds that $\hat{adm}$ under-approximates adm.*

Definition 12. *A set $\hat{A} \subseteq \hat{\mathcal{A}}$ is abstract stable, i.e. $\hat{A} \in \hat{stb}(\hat{D})$, iff $\hat{A} \in \hat{cf}(\hat{D})$, and if $\hat{A}^*$ fully attacks $\hat{\mathcal{A}} \setminus \hat{A}$.*

Example 8. *In our running example, a stable set of clusters would be $\hat{A} = \{n\_\hat{arth}, n\_fev, r_{\hat{2},3}, r_1, n\_\hat{side}\}$. Here $\hat{A}^* = \hat{A}$ and all clusters in the complement of $\hat{A}$, such as r_4 or $r_{\hat{6},7}$ are fully attacked. On the contrary, $\hat{A} = \{n\_\hat{arth}, n\_fev, r_{\hat{2},3}, r_1\}$ is not stable, as $n\_\hat{side}$ is neither contained, nor fully attacked by $\hat{A}^*$.*

The intuition here closely parallels that of admissibility. A cluster being fully attacked by the set $\hat{A}^*$ guarantees that all assumptions within the cluster are attacked in any original framework. In addition, abstract conflict-freeness ensures the existence of assumptions within the set that are not attacked and are thus reliably defended.

Theorem 4. *It holds that $\hat{stb}$ under-approximates stb.*

As in classical semantics, a stable set is also admissible.

Proposition 5. *It holds that any abstract stable set is also an admissible set, i.e. $\hat{stb}(\hat{D}) \subseteq \hat{adm}(\hat{D})$.*

In contrast to over-approximation, where identifying a non-trivial semantics that abstracts complete semantics remains elusive, under-approximation allows for a straightforward definition of such a semantics.

Definition 13. *A set $\hat{A} \subseteq \hat{\mathcal{A}}$ is abstract complete, i.e. $\hat{A} \in \hat{com}(\hat{D})$, iff $\hat{A} \in \hat{adm}(\hat{D})$, and if $\hat{A}$ does not weakly defend any cluster $\hat{c} \notin \hat{A}$.*

Example 9. *Let us revisit the abstract admissible set $\{n\_fev, n\_\hat{arth}, r_{\hat{2},3}, r_1\}$ from previous examples. This set fully attacks the cluster $\hat{side}$, and thus fully defends the cluster $n\_\hat{side}$. This set is not complete as it weakly defends a cluster it does not contain (in fact it normally defends this cluster, but weak defense suffices in this case). The set $\{n\_fev, n\_\hat{arth}, r_{\hat{2},3}, r_1, n\_\hat{side}\}$ is abstract complete.*

Theorem 6. *It holds that $\hat{com}$ under-approximates com.*

If a clustered ABA contains only singleton clusters then the abstract semantics coincide with the classical semantics. A key observation for this is that when all clusters are singletons, any abstract attack corresponds to a concrete attack.

Proposition 7. *In the special scenario where $Singles(\hat{\mathcal{A}}) = \mathcal{A}$ or in other words the mapping function is the identity, abstract semantics $\hat{cf}, \hat{adm}, \hat{stb}, \hat{com}$ coincide with their classical counterparts.*

To compare abstract semantics across different levels of abstraction, we introduce the notion of refinement.

Definition 14. *Let m_1 and m_2 be two clusterings over a set of assumptions $\mathcal{A}$. Let D be an ABAF over $\mathcal{A}$ and D' and D'' be the clustered frameworks through m_1 and m_2. Then D' is called a refinement of D'', if for all $a \in \mathcal{A}$ it holds that $m_1(a) \subseteq m_2(a)$.*

The following result captures a key property of abstract semantics: more refined clusterings offer better precision.

Proposition 8. *Let $\hat{D}'$ be a refinement of $\hat{D}''$ associated with mappings m_1 and m_2 respectively. Then, for any $\hat{A}_2 \in \hat{adm}(\hat{D}'') \Rightarrow \exists \hat{A}_1 \in \hat{adm}(\hat{D}')$ s.t. $m_2(m_1^{-1}(\hat{A}_1)) = \hat{A}_2$.*

This result is somewhat expected, as a more refined framework corresponds to less possible preimages (the abstract semantics more constrained). In the special case when m_1 is the identity, Prop. 7 recovers the classical semantics.

We show computational complexity of reasoning tasks.

Proposition 9. *Let $\hat{A}$ be a set of clustered assumptions, $x \in \hat{\mathcal{L}} \setminus \hat{\mathcal{A}}$ an atom, and $\hat{b}$ a clustered assumption. Decision problems such as whether x is derived by $\hat{A}$ in $\hat{D}$, whether $\hat{A}$ normally or fully attacks $\hat{b}$, or whether $\hat{A}$ (weakly) defeats or (partially) defends $\hat{b}$, are all decidable in polytime.*

We show the complexity of verifying whether a given set of clustered assumptions is part of an abstract semantics.

Proposition 10. *Let $\hat{A}$ be a set of clustered assumptions. For $\hat{\sigma} \in \{\hat{cf}, \hat{adm}, \hat{com}, \hat{stb}\}$ one can check whether $\hat{A} \in \hat{\sigma}(\hat{D})$, for a given $\hat{D}$, holds in polynomial time.*

Similarly, complexity of credulous and skeptical acceptance coincides with the classical setting.

Proposition 11. *Let $\hat{\sigma} \in \{\hat{adm}, \hat{com}, \hat{stb}\}$. It is NP-complete to decide whether a given atom is credulously accepted under $\hat{\sigma}$ and it is coNP-complete to decide skeptical acceptance under $\hat{stb}$.*

Abstraction on Medical Recommendations

In this section we formalize the translation of the structured modeling of clinical recommendations proposed by Zamborlini et al. (2017) into ABA. This translation forms the basis of our running example and illustrates how these models can be represented within ABA.

Treatment recommendation reasoning (TMR) is a conceptual model that is used to represent and reason about clinical knowledge, particularly to detect interactions among recommendations from different clinical guidelines. Based on this model, we extract the ABA framework in Figure 1(a). Our method for constructing an ABA from TMR differs from the existing approach in the literature (Čyras et al. 2021a).

In the TMR model, recommendations are depicted as structured objects that link a recommended action to the effects it is expected to produce, see Table 1. Each recommendation is labeled and associated with effects such as "chemotherapy decreases the presence of a breast malignant tumor". Each effect is evaluated in terms of its contribution to the overall treatment goal, which can be positive, negative, or neutral. For instance, decreasing tumor presence is considered a positive contribution and increasing fatigue due to chemotherapy is viewed as a negative one. In the TMR model recommendations can interact. They may conflict, offer alternative therapies, or address and repair the side effects caused by other recommendations.

We construct an ABA framework from the given TMR model as follows.

- Each recommendation is represented as an assumption. An acceptable set containing a recommendation represents a scenario in which performing the recommended action is permissible.
- Each condition, such as fever or high fatigue, is represented by an assumption, paired with an opposing assumption that captures the absence of the condition.
- Each positive effect on a condition is represented as a rule whose premise is the corresponding recommendation and whose conclusion is the contrary of the assumption representing the absence of the condition e.g. $\overline{\text{fever}}$.
- Each negative effect on a condition, e.g., increase fatigue, is represented as a rule whose premise is the assumption representing the condition and whose conclusion is the contrary of the corresponding recommendation.
- A conflicting interaction between two recommendations is represented by a rule in which a recommendation derives the contrary of the assumption corresponding to the recommendation it is in conflict with.

Recommendation	Effects	Cond.	Contrib.
Chemotherapy (r1)	Decrease	Tumor	+
	Increase	Fatigue	-
	Decrease	Fitness	-
Low Int. Exerc. (r3)	Decrease	Fatigue	+
	Increase	Fitness	+

Table 1: Recommendations Chemotherapy and Low Intensity Exercise, their effects on conditions and contribution.

In our view, clustering is a potentially useful tool for handling interacting medical recommendation guidelines. Guidelines and recommendation hierarchies, where recommendations are organized according to specific grouping criteria, can be interpreted as clusters. Then, given information about a specific patient, we can refine those clusters to determine whether a particular treatment is recommended.

The following example showcases how guidelines viewed as clusters can be refined to determine whether a treatment is permissible, without requiring access to full information. We use over-approximating semantics as a guide for determining which sets are worth refining, and then use under-approximating semantics on the refined versions. Combining over- and under-approximation, as in the following example, can enhance the refinement process by avoiding unnecessary refinements while ensuring sound reasoning.

Example 10. *Consider a clinical guideline G_1 that recommends the administration of the following medications: $\{$Adm Aspirin, Adm Ibuprofen, Adm Tramadol$\}$. Each of these actions contributes to reducing blood coagulation. Notably, both Aspirin and Ibuprofen belong to the broader category of NSAIDs (Non-Steroidal Anti-Inflammatory Drugs). Given this, we can group $\{$Adm Aspirin, Adm Ibuprofen$\}$ in one cluster, called Adm NSAID, thereby forming a clustered version of the guideline, denoted as G_1'. Then, we can cluster all recommendations within G_1 into one single entity, forming G_1''. Let us introduce a second guideline, G_2, stating that administering NSAID increases the risk of gastrointestinal bleeding and therefore advises against it.*

Using under-approximating semantics there is no set deemed stable. However, over-approximating semantics suggest that potentially there is a stable set containing some part of G_1'' and G_2. It is then worthwhile to better understand this interaction by refining the abstraction of G_1'' down to G_1', revealing that G_2 specifically conflicts with the clustered assumption Adm NSAID and not with the singleton Adm Tramadol. As a result, we can under-approximate a stable set of actions that addresses high blood coagulation. This stable set also follows guideline G_2 and recommends the administration of Tramadol, instead of NSAID. This conclusion was reached without fully refining the abstraction, demonstrating that we were able to resolve the conflict without accessing all the information.

Conclusions

We introduced under-approximation to ABA frameworks, with an underlying motivation that our newly proposed semantics have a corresponding set in the classical ABA semantics for any clustering. We provided foundational properties, showed relations to existing semantics of clusterings, and showed that the new abstract semantics exhibit the same complexity as the classical ABA semantics. Via the illustration on medical recommendations we shared our view on how clustering in ABA can be utilized. As future work, we aim to capture more TMR interactions in ABA. Additionally, it would be interesting to investigate ways to cluster more efficiently and to study connections to forgetting in answer set programming (Gonçalves, Knorr, and Leite 2023).

Acknowledgements

This research was funded in whole or in part by the Austrian Science Fund (FWF) grants P35632 and 10.55776/COE12. For open access purposes, the authors have applied a CC BY public copyright license to any author accepted manuscript version arising from this submission.

References

Amgoud, L.; Dimopoulos, Y.; and Moraitis, P. 2007. A unified and general framework for argumentation-based negotiation. In *Proc. AAMAS*, 967–974. IFAAMAS.

Apostolakis, I.; Saribatur, Z. G.; and Wallner, J. P. 2024a. Abstraction in Assumption-based Argumentation. In *Proc. KR*, 49–59. ijcai.org.

Apostolakis, I.; Saribatur, Z. G.; and Wallner, J. P. 2024b. A Semantical Approach to Abstraction in Answer Set Programming and Assumption-Based Argumentation. In *Proc. LPNMR*, volume 15245 of *LNCS*, 228–234. Springer.

Atkinson, K.; Baroni, P.; Giacomin, M.; Hunter, A.; Prakken, H.; Reed, C.; Simari, G. R.; Thimm, M.; and Villata, S. 2017. Towards Artificial Argumentation. *AI Mag.*, 38(3): 25–36.

Baroni, P.; Boella, G.; Cerutti, F.; Giacomin, M.; van der Torre, L. W. N.; and Villata, S. 2014. On the Input/Output behavior of argumentation frameworks. *Artif. Intell.*, 217: 144–197.

Baroni, P.; Caminada, M.; and Giacomin, M. 2011. An introduction to argumentation semantics. *Knowl. Eng. Rev.*, 26(4): 365–410.

Baroni, P.; Gabbay, D.; Giacomin, M.; and van der Torre, L., eds. 2018. *Handbook of Formal Argumentation*. College Publications.

Besnard, P.; and Hunter, A. 2018. A Review of Argumentation Based on Deductive Arguments. In Baroni, P.; Gabbay, D.; Giacomin, M.; and van der Torre, L., eds., *Handbook of Formal Argumentation*, 437–484. College Publications.

Boella, G.; Kaci, S.; and van der Torre, L. W. N. 2009. Dynamics in Argumentation with Single Extensions: Abstraction Principles and the Grounded Extension. In *Proc. ECSQARU*, volume 5590 of *LNCS*, 107–118. Springer.

Bondarenko, A.; Dung, P. M.; Kowalski, R. A.; and Toni, F. 1997. An Abstract, Argumentation-Theoretic Approach to Default Reasoning. *Artif. Intell.*, 93: 63–101.

Čyras, K.; Fan, X.; Schulz, C.; and Toni, F. 2018. Assumption-Based Argumentation: Disputes, Explanations, Preferences. In Baroni, P.; Gabbay, D.; Giacomin, M.; and van der Torre, L., eds., *Handbook of Formal Argumentation*, 365–408. College Publications.

Čyras, K.; Oliveira, T.; Karamlou, A.; and Toni, F. 2021a. Assumption-based argumentation with preferences and goals for patient-centric reasoning with interacting clinical guidelines. *Argument Comput.*, 12(2): 149–189.

Čyras, K.; Rago, A.; Albini, E.; Baroni, P.; and Toni, F. 2021b. Argumentative XAI: A Survey. In *Proc. IJCAI*, 4392–4399. ijcai.org.

Dimopoulos, Y.; Mailly, J.; and Moraitis, P. 2019. Argumentation-based Negotiation with Incomplete Opponent Profiles. In *Proc. AAMAS*, 1252–1260. IFAAMAS.

Dimopoulos, Y.; Nebel, B.; and Toni, F. 2002. On the computational complexity of assumption-based argumentation for default reasoning. *Artif. Intell.*, 141(1/2): 57–78.

Dung, P. M. 1995. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. *Artif. Intell.*, 77(2): 321–358.

Fan, X.; and Toni, F. 2012. Agent Strategies for ABA-based Information-seeking and Inquiry Dialogues. In *Proc. ECAI*, volume 242 of *FAIA*, 324–329. IOS Press.

Fan, X.; and Toni, F. 2015. On Computing Explanations in Argumentation. In *Proc. AAAI*, 1496–1502. AAAI Press.

Gabbay, D.; Giacomin, M.; Simari, G. R.; and Thimm, M., eds. 2021. *Handbook of Formal Argumentation*, volume 2. College Publications.

Gao, Y.; Toni, F.; Wang, H.; and Xu, F. 2016. Argumentation-Based Multi-Agent Decision Making with Privacy Preserved. In *Proc. AAMAS*, 1153–1161. ACM.

García, A. J.; and Simari, G. R. 2018. Argumentation Based on Logic Programming. In Baroni, P.; Gabbay, D.; Giacomin, M.; and van der Torre, L., eds., *Handbook of Formal Argumentation*, 409–435. College Publications.

Gonçalves, R.; Knorr, M.; and Leite, J. 2023. Forgetting in Answer Set Programming - A Survey. *Theory Pract. Log. Program.*, 23(1): 111–156.

Gordon, T. F.; Prakken, H.; and Walton, D. 2007. The Carneades model of argument and burden of proof. *Artif. Intell.*, 171(10-15): 875–896.

Kakas, A. C.; Moraitis, P.; and Spanoudakis, N. I. 2019. *GORGIAS*: Applying argumentation. *Argument Comput.*, 10(1): 55–81.

Modgil, S.; and Prakken, H. 2018. Abstract Rule-Based Argumentation. In Baroni, P.; Gabbay, D.; Giacomin, M.; and van der Torre, L., eds., *Handbook of Formal Argumentation*, 287–364. College Publications.

Prakken, H.; and Sartor, G. 2015. Law and logic: A review from an argumentation perspective. *Artif. Intell.*, 227: 214–245.

Saribatur, Z. G.; and Wallner, J. P. 2021. Existential Abstraction on Argumentation Frameworks via Clustering. In *Proc. KR*, 549–559. ijcai.org.

Vassiliades, A.; Bassiliades, N.; and Patkos, T. 2021. Argumentation and explainable artificial intelligence: a survey. *Knowl. Eng. Rev.*, 36: e5.

Zamborlini, V.; Silveira, M. D.; Pruski, C.; ten Teije, A.; Geleijn, E.; van der Leeden, M.; Stuiver, M.; and van Harmelen, F. 2017. Analyzing interactions on combining multiple clinical guidelines. *Artif. Intell. Medicine*, 81: 78–93.