Justifying Argument Acceptance with Collective Attacks: Discussions and Disputes

Giovanni Buraglio[1] , **Wolfgang Dvořák**[1] , **Matthias König**[1] and **Markus Ulbricht**[2]
[1]Institute of Logic and Computation, TU Wien
[2]ScaDS.AI Dresden/Leipzig, Leipzig University
{giovanni.buraglio, wolfgang.dvorak, matthias.koenig}@tuwien.ac.at,
mulbricht@informatik.uni-leipzig.de

Abstract

In formal argumentation one aims for intuitive and concise justifications for the acceptance of arguments. Discussion games and dispute trees are established methods to obtain such a justification. However, so far these techniques are based on instantiating the knowledge base into graph-based Dung style abstract argumentation frameworks (AFs). These instantiations are known to produce frameworks with a large number of arguments and thus also yield to long discussion games and large dispute trees.

To obtain more concise justifications for argument acceptance, we propose to instantiate the knowledge base as an argumentation framework with collective attacks (SETAFs). Remarkably, this approach yields smaller frameworks compared to traditional AF instantiation, while exhibiting increased expressive power. We then introduce discussion games and dispute trees tailored to SETAFs, show that they correspond to credulous acceptance w.r.t. the well-known preferred semantics, analyze and tune them w.r.t. the size, and compare the two notions. Finally, we illustrate how our findings apply to assumption-based argumentation.

1 Introduction

Computational models of argumentation in Artificial Intelligence (AI) [Gabbay *et al.*, 2021; Bench-Capon and Dunne, 2007] provide formal approaches to reason argumentatively, with a wide variety of application avenues, such as legal reasoning, medical sciences, and e-governmental issues [Atkinson *et al.*, 2017]. A main booster for this line of research was Dung's seminal paper [Dung, 1995] introducing *abstract argumentation frameworks* (AFs). In Dung-style AFs, arguments are atomic entities and their internal structure, e.g. the premises required to derive the respective conclusion, is abstracted away. Viewing the conflicts between arguments as a binary relation, Dung's AFs are directed graphs, interpreting nodes as arguments and edges as attacks between them.

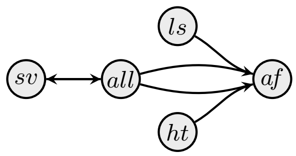
Established methods based on discussion games [Caminada, 2018] and dispute trees [Čyras *et al.*, 2018; Dung *et al.*, 2006; Modgil and Caminada, 2009] can be employed in order to justify whether an argument in the given AF can be accepted or not. Such techniques have been successfully applied to rule-base argumentation systems like Assumption-Based Argumentation (ABA) [Bondarenko *et al.*, 1997] (see also [Toni, 2014] for a gentle introduction). For all such systems, arguments and attacks are derived from a given knowledge base to specify argumentative workflows [Besnard and Hunter, 2008; Bondarenko *et al.*, 1997; García and Simari, 2004; Modgil and Prakken, 2013]. In this so-called *instantiation* procedure, different conclusions as well as their conflicts within the knowledge base are made explicit by constructing an associated Dung-style AF.

Hence, transforming a knowledge base D into an AF F_D provides a descriptive and human-understandable depiction of the conflicts arising in D. Researchers thus make use of these advantages by extracting justifications for reasoning in D from the constructed argumentation graph [Čyras *et al.*, 2021]. Thereby, discussion games or dispute trees oftentimes serve as the conceptual basis. They can be employed to engage the user in a dialogue that addresses all conceivable concerns regarding acceptance of a given argument.

However, when instantiating a knowledge base as an AF, in general a large number of arguments is constructed, several of which entailing the same conclusion. The underlying conceptual issue is that many structured argumentation formalisms give rise to situations where atoms attack certain arguments collectively, which cannot be captured by AFs. These issues impair the aforementioned advantages of constructing an argumentation graph. To overcome them we propose to utilize AFs with *collective attacks*. A popular abstract formalism to model this kind of situations are SETAFs [Nielsen and Parsons, 2006; Dvořák *et al.*, 2024], introduced as a generalization of Dung's Argumentation Frameworks by means of hypergraphs. Recently, SETAFs were used for instantiations of inconsistent knowledge bases [Yun *et al.*, 2020]. The additional expressive power stemming from the collective attacks [Dvořák *et al.*, 2019] can help to reduce the overall size of the instantiated argumentation framework, yielding more natural and concise representations of the knowledge base.

To see how turning to SETAFs yields a more succinct representation, consider the following simple example.

Example 1.1. Our protagonist Caroline would like to go on a trip for vacation. In particular, she intends to go for a long stay (ls) at a fancy hotel (ht). Among the available options she could either decide to visit a small village (sv) or book an all-inclusive type of vacation (all). However, the latter is only available during holiday seasons (s) for popular destinations (pd). For these reasons, neither a long stay nor a fancy hotel would be affordable (af) with the all-inclusive option. Therefore, she decides for the small village. We represent the described situation by means of a SETAF as follows.



It can easily be noted that this yields a more natural and succinct representation in contrast with the traditional AF instantiation (see Section 6). By capturing collective attacks through SETAFs, one is able to cut down disposable arguments deriving the same conclusion. Despite this evident advantage, no previous work has focused on implementing discussion games and dispute trees tailored for SETAFs. The present work aims at bridging this gap. As it turns out, these techniques exhibit further benefits in this setting, improving on the length of justifications for argument acceptance with respect to the corresponding AF disputes and discussions.

More precisely, the main contributions of the present paper can be summarized as follows.

- We generalize discussion games and dispute trees for SETAFs capturing credulous acceptance for preferred semantics.
- We demonstrate how our choice to instantiate knowledge bases with SETAFs leads to more concise justifications in the presence of collective attacks. In particular: (i) The instantiated framework is smaller than the one obtained in the usual AF instantiation; (ii) The length of SETAFs discussions is improved with respect to the corresponding AF discussions.
- We study computational aspects of discussion games.
- We compare AF and SETAF discussions and dispute trees on ABA, showcasing the benefits of our approach.

2 Background

We recall the definitions of SETAFs [Nielsen and Parsons, 2006; Bikakis *et al.*, 2021] and AF discussion games [Caminada, 2018]. A SETAF $SF = (A, R)$ consists of a finite set of arguments A and an attack relation $R \subseteq 2^A \setminus \{\emptyset\} \times A$ that contains attacks from sets of arguments towards a single argument. We extend the attack relation to sets of arguments as follows: A set $S \subseteq A$ attacks a set $T \subseteq A$ if there is $(S', t) \in R$ with $S' \subseteq S, t \in T$. We define the set of arguments attacked by S as $S^+_{SF} = \{a \in A \mid S \text{ attacks } \{a\}\}$ and denote by $S^\oplus_{SF} = S \cup S^+_{SF}$ the *range* of S. We drop the subscript SF whenever there is no risk of confusion.

A set $S \subseteq A$ is *conflict-free* in SF ($S \in cf(SF)$) if $S \cap S^+ = \emptyset$. A set $S \subseteq A$ *defends* some argument $a \in A$ if for any set T, $a \in T^+$ implies $S^+ \cap T \neq \emptyset$. In this work, we consider the SETAF semantics for admissible and preferred extensions. For a conflict-free set $S \in cf(SF)$: S is an admissible set ($S \in ad(SF)$), if S defends each $a \in S$; S is a preferred extension ($S \in pr(SF)$), if S is $\subseteq$-maximal in $ad(SF)$. If we restrict SETAFs to single arguments attacking other arguments then they amount to Dung-style AFs.

Discussion games for AFs [Caminada, 2018] show membership to a preferred extension for some specific argument, called *target* of the discussion. These take the shape of a dialogue among two agents: the proponent P of the target and its opponent O. In the spirit of a socratic dialogue, the proponent aims at showing that such an argument can be justified, while the opponent challenges P with the implications of P's own position. If the proponent is able to counter-attack all of the opponent moves, P yields a justification for the target. Formally, AF preferred discussions are defined as follows.

Definition 2.1. Let $F = (A, R)$ be an AF. A preferred discussion D is a sequence of moves $(m_1, m_2, \ldots, m_n)$ s.t.

- each move m_i ($1 \leq i \leq n$) of the form $\texttt{in}(a)$ is a P-move if i is odd and $a \in A$,
- each move m_i ($1 \leq i \leq n$) of the form $\texttt{out}(a)$ is an O-move if i is even and $a \in A$,
- for each O-move $m_i = \texttt{out}(a)$ ($2 \leq i \leq n$) there exists a P-move $m_j = \texttt{in}(b)$ ($j < i$) such that a attacks b,
- for each P-move $m_i = \texttt{in}(a)$ ($3 \leq i \leq n$) it holds that m_{i-1} is an O-move $m_j = \texttt{out}(b)$ such that a attacks b,
- for all O-moves m_i, m_j with $i \neq j$ we have $m_i \neq m_j$.

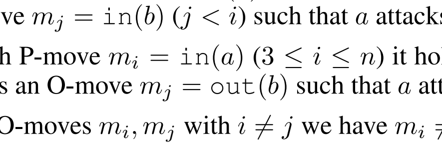
3 SETAF Discussion Games

In this section, we will introduce SETAF discussion games. The goal is to come up with a notion that intuitively captures the reason as to why a certain argument is acceptable. The definition of admissibility in SETAFs gives rise to the following aspects that need to be considered:

- in order for a to be accepted, *each* attack (T, a) directed towards a needs to be counter-attacked;
- in order to counter-attack some attack (T, a) it suffices to defeat *any* $t \in T$;
- each argument in the extension needs to be defended and our discussion shall guide the user through the graph step-by-step.

Let us illustrate the procedure in an example.

Example 3.1. Consider the following SETAF SF:

SF :


Suppose we want to argue for acceptance of a. The argument a is threatened by b and c collectively, so at least one argument in the set $\{b, c\}$ has to be defeated in order to defend a. We will thus consider a move $\texttt{out}_\exists(\{b, c\})$ formalizing this requirement. Attacking the set $\{b, c\}$ can be achieved via the attack $(\{d, e\}, b)$ which requires acceptance of both d and e. To this end we consider a move $\texttt{in}(\{d, e\}) : \texttt{out}(b)$. Since this move requires both d and e to be in, the opponent player can now argue against either of those, resulting in bringing forward f or g. Due to the presence h, both of these arguments can be countered. In summary, the admissible extension $E = \{a, d, e, h\}$ is rightfully spotted.

We are now ready to formalize the intuition of proponent- and opponent-moves given in Example 3.1 via the following definition of a preferred discussion. This generalizes AF discussions from [Caminada, 2018, Definition 13].

Definition 3.2. Let $SF = (A, R)$ be a SETAF. A preferred discussion D is a sequence of moves $(m_1, m_2, \ldots, m_n)$ s.t.

- move m_1 is a P-move of the form $\mathtt{in}(a)$ for some $a \in A$,
- if i with $3 \leq i \leq n$ is odd, then m_i is a P-move of the form $\mathtt{in}(S)\!:\!\mathtt{out}(t)$ where there is an attack $(S, t) \in R$,
- if i with $2 \leq i \leq n$ is even, then m_i is an O-move of the form $\mathtt{out}_\exists(S)$. Moreover, there is an attack $(S, t) \in R$ such that either (i) $t = a$ or (ii) there exists an earlier P-move $m_j = \mathtt{in}(T)\!:\!\mathtt{out}(x)$ with $t \in T$ (and $j < i$),
- for each P-move $m_i = \mathtt{in}(S)\!:\!\mathtt{out}(t)$ for $3 \leq i \leq n$ it holds that m_{i-1} is of the form $\mathtt{out}_\exists(T)$ where $(S, t) \in R$ with $t \in T$,
- for all O-moves m_i, m_j with $i \neq j$ we have $m_i \neq m_j$.

Thus, the first move is made by the proponent player in favor of the query argument a, stating it should be "in" ($\mathtt{in}(a)$). The O-moves represent attacks towards our extension; $m_i = \mathtt{out}_\exists(S)$ is to be interpreted as "one argument in S needs to be out". In the same vein, P-moves represent means to defend our extension against the concerns raised by O-moves; $\mathtt{in}(S)\!:\!\mathtt{out}(t)$ is to be interpreted as "S is in and thus, t is out". Consequently, each O-move is directed towards a previous P-move and vice versa.

A more subtle observation is that discussion games are always finite (if SF is). The reason is a certain asymmetry in the O-moves compared to the P-moves, which stems from the condition that there exists no two O-moves m_i, m_j with $i \neq j$ and $m_i = m_j$. Consequently, the opponent is never allowed to make the same move twice. Otherwise, the discussion game could become infinite due to e.g. two mutually attacking arguments a and b.

Definition 3.3. A preferred discussion $D = (m_1, m_2, \ldots, m_n)$ is *finished* if (1) there exists no move m_{n+1} such that $(m_1, m_2, \ldots, m_n, m_{n+1})$ is a preferred discussion, or (2) there exist two P-moves $m_i = \mathtt{in}(S_i) : \mathtt{out}(t_i)$, $m_j = \mathtt{in}(S_j) : \mathtt{out}(t_j)$ such that $t_i \in S_j$, and no sub-sequence $(m_1, m_2, \ldots, m_k)$ with $k < n$ is finished. We say a preferred discussion is won by player O if it is finished and there are two P-moves $m_i = \mathtt{in}(S_i) : \mathtt{out}(t_i)$, $m_j = \mathtt{in}(S_j) : \mathtt{out}(t_j)$ such that $t_i \in S_j$. Otherwise, a finished discussion it is won by the player making the last move m_n.

The rationale behind finished discussion games D is that before a game is finished, both players still can make valid moves and we thus cannot decide yet what the outcome is.

Example 3.4. Recall Example 3.1. The sequence

$$\mathtt{in}(a),\ \mathtt{out}_\exists(\{b, c\}),\ \mathtt{in}(\{d, e\})\!:\!\mathtt{out}(b),$$
$$\mathtt{out}_\exists(\{f\}),\ \mathtt{in}(\{h\})\!:\!\mathtt{out}(f)$$

is a preferred discussion. Note that the proponent player challenges f via the O-move $\mathtt{out}(\{f\})$, but not g. Indeed, this preferred discussion is not finished, since it can be extended by the additional moves

$$\mathtt{out}_\exists(\{g\}),\ \mathtt{in}(\{h\})\!:\!\mathtt{out}(g)$$

which then corresponds to a finished discussion.

The intended meaning of a SETAF discussion game is that a finished discussion won by the proponent player corresponds to some admissible extension in the given SETAF. We formalize this as follows.

Definition 3.5. Let $SF = (A, R)$ be a SETAF and $D = (m_1, \ldots, m_n)$ a preferred discussion in SF. We call

$$\mathtt{IN}(D) = \{a \mid m_1 = \mathtt{in}(a)\} \cup \bigcup\{S \mid \mathtt{in}(S)\!:\!\mathtt{out}(b) \in D\}$$

the set of in-arguments, and we call

$$\mathtt{OUT}(D) = \bigcup\{b \mid \mathtt{in}(S)\!:\!\mathtt{out}(b) \in D\}$$

the set of out-arguments.

If P wins the discussion, then the $\mathtt{IN}$ arguments are admissible; moreover, all arguments used in the O-moves are out. This matches the intuition that they threaten our extension.

Proposition 3.6. *Let $SF = (A, R)$ be a SETAF and let D be a preferred discussion that is won by the proponent. Let $E = \mathtt{IN}(D)$. Then $E \in ad(SF)$ and $E^+ \supseteq \mathtt{OUT}(D)$.*

We consider proponent and opponent to play a *discussion game* where moves follow the rules of Definition 3.2, and each player aims to win the resulting discussion. We are particularly interested in *winning strategies* for the proponent. A winning strategy for an argument a is a function that for each O-move gives a P-move to play next such that, when starting with $\mathtt{in}(a)$, for each possible combination of O-moves the resulting discussion is won by the proponent. Admissible sets provide the blueprints for such winning strategies.

Theorem 3.7. *Let $SF = (A, R)$ be a SETAF and $a \in A$. There is an admissible set E containing a if and only if the proponent has a winning strategy for the preferred discussion game starting with $\mathtt{in}(a)$.*

Thus it is sufficient that the proponent wins a single discussion for an argument a in order to have a winning strategy. This is because, in a discussion won by the proponent, the opponent always plays all possible attacks on the P-moves.

Corollary 3.8. *Let $SF = (A, R)$ be a SETAF and $a \in A$. The proponent has a winning strategy for the preferred discussion game for a if and only if there is a preferred discussion for a that is won by the proponent.*

Example 3.9. Recall Example 3.1. The sequence from Example 3.4, say D_a, is a finished preferred discussion for a. The last move was made by P, so it is won by the proponent. We have $\mathtt{IN}(D_a) = E = \{a, d, e, h\}$ which is indeed an admissible extension in SF from Example 3.1. Moreover, $\mathtt{OUT}(D_a) = \{b, f, g\}$, indeed corresponding to E^+.

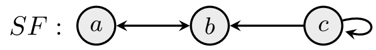
On the other hand, let us try to find a discussion D_b for b. One possibility is the sequence

$$\mathtt{in}(b),\ \mathtt{out}_\exists(\{d, e\}),\ \mathtt{in}(\{f\})\!:\!\mathtt{out}(d),\ \mathtt{out}_\exists(\{h\}).$$

We see that the opponent won. The same would have happened if we had made the P move $\texttt{in}(\{g\}) : \texttt{out}(e)$; this matches the fact that b does not occur in any admissible set.

In the previous example, D_a was won by P and D_b was won by O, corresponding exactly to the acceptance of a and non-acceptance of b, respectively. We want to mention, however, that the proponent P can lose although there is a winning strategy. So it is possible to make a "wrong" P-move.

Example 3.10. Suppose SF is given as follows.



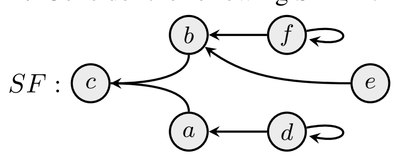
Consider the following finished game:

$$\texttt{in}(a),\ \texttt{out}_\exists(\{b\}),\ \texttt{in}(\{c\}){:}\texttt{out}(b),$$
$$\texttt{out}_\exists(\{c\}),\ \texttt{in}(\{c\}){:}\texttt{out}(c).$$

Clearly it was the wrong move to choose $\texttt{in}(\{c\}){:}\texttt{out}(b)$, if the proponent had instead chosen $\texttt{in}(\{a\}){:}\texttt{out}(b)$ the game would have been won for P.

We want to mention that compared to AF discussion games, there is an additional layer of complexity from the point of view of the proponent: in an AF, each argument a attacking our extension E needs to be countered. Hence the only way to make such a "wrong" move is by choosing the wrong defender. In SETAFs, it is not clear which argument in the tail T of an attack one should counter. So there is an additional source for "wrong" moves.

Example 3.11. Consider the following SETAF.



Consider the finished game:

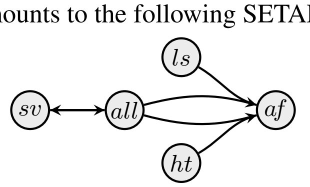
$$\texttt{in}(c),\ \texttt{out}_\exists(\{a,b\}),\ \texttt{in}(\{d\}){:}\texttt{out}(a),$$
$$\texttt{out}_\exists(\{d\}),\ \texttt{in}(\{d\}){:}\texttt{out}(d).$$

Here the P-move $\texttt{in}(\{d\}){:}\texttt{out}(a)$ was "wrong" in the sense that it caused O to win. In order to successfully defend c, the P-move must be directed towards b. However, even after this has been spotted there is still a possible "wrong" move ($\texttt{in}(\{f\}){:}\texttt{out}(b)$) which would again result in a win for O.

We want to mention that, in contrast to the proponent, the opponent player has no notion of "right" or "wrong" moves. This is due to the fact that while P is responsible for finding an extension, O merely raises any conceivable concern. Formally, we note the following corollary of Theorem 3.7.

Corollary 3.12. *Let $SF = (A, R)$ be a SETAF and $a \in A$. If there is no $E \in ad(SF)$ with $a \in E$, then any finished preferred discussion for a is won by O.*

Now let us head back to the introductory example to see how it is handled by our discussion game notion. Recall that this example amounts to the following SETAF:



Striving to argue for an affordable trip, the following discussion is induced.

$$\texttt{in}(af),\ \texttt{out}_\exists(\{ls, all\}),\ \texttt{in}(\{sv\}){:}\texttt{out}(all),$$
$$\texttt{out}_\exists(\{ht, all\}),\ \texttt{in}(\{sv\}){:}\texttt{out}(all),$$
$$\texttt{out}_\exists(\{all\}),\ \texttt{in}(\{sv\}){:}\texttt{out}(all).$$

The discussion rightfully spots that an affordable trip is only possible if the agent visits a small village. Moreover, it is clearly visible that the all-inclusive trip is the culprit here. While this is already shorter than a corresponding AF discussion (see Section 6 below), there are still some redundancies in this examples: the P-move in favor of sv is made multiple times. The next section is devoted to handling such issues.

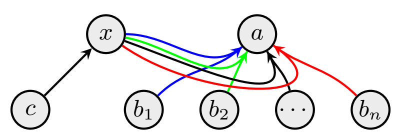
4 Concise Discussions

For regular AFs, the size of a preferred discussion can be equally measured as the number of moves as well as by the number of $\texttt{in}$, $\texttt{out}$ labeled arguments in the discussion. However, this is not the case for SETAFs: a single P-move can label several arguments $\texttt{in}$, which might result in quite different sizes for these measures. Here we hold the number of moves to be the most relevant aspect. In the context of human-computer interaction, for example, it is safe to assume that the user does not care about the number of the labeled arguments, but rather on the number of interactions in the discussion they engage in. Therefore, we define the size of a discussion game as the number of moves it contains. In what follows we investigate how the size of an admissible set relates to that of the discussion game induced by it.

4.1 Size of Discussion Games

As the following example illustrates, small admissible sets do not automatically induce short discussion games.

Example 4.1. In the following SETAF we have n attacks $(\{b_i, x\}, a)$ for $1 \leq i \leq n$ towards the argument a.



$E = \{a, c\}$ is admissible with $|E| = 2$, but the size of the smallest preferred discussion for a is $2n + 1$.

However, the previous example already intuitively illustrates that the size of a discussion game correlates with the number of *attacks towards* the accepted arguments rather than the number of accepted arguments. Indeed we show that the size of a discussion game is bounded by the number of incoming attacks of an admissible set E. This resembles a result for regular AFs in [Caminada *et al.*, 2016].

Theorem 4.2. *Let $SF = (A, R)$ be a SETAF and let $E \in ad(SF)$. Then for each $a \in E$ there exists a winning strategy for the preferred discussion game such that for all the possibly resulting discussions $D_a = (m_1, \ldots, m_{n_a})$ for a the length n_a is at most $2 \cdot |\{T \mid (T, b) \in R, b \in E\}| + 1$.*

4.2 Towards More Concise Games

Let us now go one step further and make our discussions even more concise, by employing an additional rule that allows us to avoid certain redundancies in discussions. To this end recall Example 4.1. We see a pattern in the proponent's moves:

$$\texttt{in}(a),\ \texttt{out}_\exists(\{b_1, x\}),\ \texttt{in}(\{c\})\texttt{:out}(x),$$
$$\texttt{out}_\exists(\{b_2, x\}),\ \texttt{in}(\{c\})\texttt{:out}(x),\ \ldots$$

The proponent keeps playing the same move—in particular, the argument that is set *out* (namely x) is repeated over and over. It is clear, however, that any move of the opponent containing x can be countered by the proponent in this way, which means the repetition is unnecessary. We can therefore improve the rule in question in the following way to prohibit the opponent from playing a move that contains x once we know it is attacked by the admissible set:

Definition 4.3. Let $SF = (A, R)$ be a SETAF. A *concise preferred discussion* D is a preferred discussion where in addition, the following rule holds:

- Let $D_i = \{t \mid \texttt{in}(S)\texttt{:out}(t) = m_j \in D, j < i\}$ be the set of defeated arguments before move i. There exists no O-move $m_i = \texttt{out}_\exists(T_i)$ such that $T_i \cap D_i \neq \emptyset$.

All the notions for preferred discussion are extended to concise preferred discussion in the natural way.

Example 4.4. We observe that Example 4.1 now yields a concise game for a:

$$\texttt{in}(a),\ \texttt{out}_\exists(\{b_1, x\}),\ \texttt{in}(\{c\})\texttt{:out}(x).$$

The opponent cannot play another move, as c is unattacked and every attack towards a contains the already defeated argument x. Consequently, the game is finished.

In fact, we can now see that the length of discussion games in the improved form is also bound by the number of *defeated* arguments. As illustrated in Example 4.1, this number can be arbitrarily smaller then the number of incoming attacks.

Theorem 4.5. *Let $SF = (A, R)$ be a SETAF and let $E \in ad(SF)$. Then for each $a \in E$ there exists a winning strategy for the concise preferred discussion game such that for all the possibly resulting discussions $D_a = (m_1, \ldots, m_{n_a})$ for a the length n_a is bounded by $2 \cdot \min(|E^+|, |\{T \mid (T, b) \in R, b \in E\}|) + 1$.*

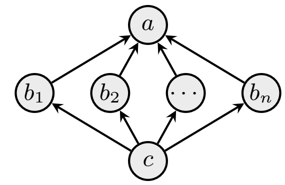
We refine this upper bound using so-called hitting sets [1].

Theorem 4.6. *Let $SF = (A, R)$ be a SETAF, $E \in ad(SF)$ and let $H \subseteq E^+$ be a hitting set of $\{T \mid (T, b) \in R, b \in E\}$. Then for each $a \in E$ there exists a winning strategy for the concise preferred discussion game such that for all the possibly resulting discussions $D_a = (m_1, \ldots, m_{n_a})$ the length n_a is bounded by $2 \cdot |H| + 1$.*

We will now illustrate that this upper bound cannot be improved in the general case.

Example 4.7. In the following SETAF we can see that we have to account for each incoming attack, i.e., for the only admissible set containing a, namely $E = \{a, c\}$, we have $|E^+| = |\{(T, a) \in R \mid a \in \texttt{IN}(D)\}|$.

[1] A hitting set H of a set of sets $\mathcal{S} = \{S_1, \ldots, S_n\}$ is a set that satisfies $H \cap S_i \neq \emptyset$ for all $1 \leq i \leq n$.



First note that the improved rule inducing concise discussions is still a faithful generalization of the rules for AFs, as in this case it coincides with the rule that the opponent must not repeat moves.

Example 4.8. Let us see how this improved rule impacts our introductory example. Since the opponent is not allowed to make use of attacks containing all multiple times anymore, we end up with the discussion

$$\texttt{in}(af),\ \texttt{out}_\exists(\{ls, all\}),\ \texttt{in}(\{sv\})\texttt{:out}(all)$$

capturing the intuition that once we decide to go for a small village (sv), we do not have any issue with affordability anymore; thus the discussion is finished at this point.

Let us briefly discuss the computational aspects of finding a shortest discussion for a given argument a. The canonical decision problem that corresponds to this optimization problem is: given a SETAF F, argument a and integer k, is there a winning strategy that ensures a discussion of length $\leq k$? For regular AFs this is known to be NP-complete [Caminada *et al.*, 2016]. This also holds for our case.

Proposition 4.9. *Deciding whether for an input SETAF F, argument a, and integer k, there is a winning strategy ensuring that each finished preferred discussion D is of length $\leq k$ is* NP*-complete.*

5 SETAF Dispute Trees

Another well-established concept in the literature is the notion of dispute trees [Čyras *et al.*, 2018; Dung *et al.*, 2006; Modgil and Caminada, 2009]. They are similar in spirit to discussion games, but represent the debate in a tree. In this section, we want to generalize dispute trees to SETAFs as well, and compare them to our SETAF discussion games.

5.1 Dispute Trees

We follow the usual terminology for dispute trees, but we note that many concepts resemble previous notions from discussion games. Given a SETAF $SF = (A, R)$, a dispute tree $\texttt{t}$ for $a \in A$ is a (possibly infinite) tree s.t.

1. each node of t is either of the form $[L : S]$ where $S \subseteq A$ is called the *label* of the node and $L \in \{P, O\}$ the *status* of either opponent (O) or proponent (P),
2. the root is of the form $[P : \{a\}]$ for some $a \in A$,
3. every node of the form $[P : S]$ has children of the form $[O : T]$ for each $(T, s) \in R$ with $s \in S$, which are connected via edges labeled s for each such $(T, s) \in R$,
4. every node of the form $[O : T]$ has either no child node, or exactly one child node $[P : S]$ for some $(S, t) \in R$ with $t \in T$ which is connected via an edge labeled t,
5. no other node or edge occurs in $\texttt{t}$.

The root label a is called the *topic* of the tree. The union of sets of all arguments labeling P nodes in $\mathtt{t}$ is called the defense set of $\mathtt{t}$, denoted by $D(\mathtt{t})$; the set of all arguments labeling attack-edges in $\mathtt{t}$ is called the attack set of $\mathtt{t}$, denoted by $Att(\mathtt{t})$. Edges below P-nodes correspond to attacks towards $D(\mathtt{t})$, and we call them *attack-edges*; edges below O-nodes correspond to defending attacks, we call them *defense-edges*.

A dispute tree is *admissible* if i) every O node has a child and ii) $D(\mathtt{t}) \cap Att(\mathtt{t}) = \emptyset$.

Let us familiarize with SETAF dispute trees by recalling our Example 3.1.

Example 5.1. Recall that in the SETAF from Example 3.1, argument a occurs in an admissible extension, while b does not. Consider the two dispute trees $\mathtt{t}$ and $\mathtt{t}'$ with topics a and b, respectively.

$$\begin{tikzcd}[column sep=small]
& \mathtt{t} : [P:\{a\}] \arrow[d, dash, "a"] & \\
& {[O:\{b,c\}]} \arrow[d, dash, "b"] & \\
& {[P:\{d,e\}]} \arrow[dl, dash, "d"'] \arrow[dr, dash, "e"] & \\
{[O:\{f\}]} \arrow[d, dash, "f"'] & & {[O:\{g\}]} \arrow[d, dash, "g"] \\
{[P:\{h\}]} & & {[P:\{h\}]}
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathtt{t}' : [P:\{b\}] \arrow[d, dash, "b"] \\
{[O:\{d,e\}]} \arrow[d, dash, "d"] \\
{[P:\{f\}]} \arrow[d, dash, "f"] \\
{[O:\{h\}]}
\end{tikzcd}$$

We see that $\mathtt{t}$ is an admissible dispute tree since we have $D(\mathtt{t}) \cap Att(\mathtt{t}) = \emptyset$ and each O node has a child. On the other hand, $\mathtt{t}'$ is not admissible since $[O : \{h\}]$ has no child.

As in the case of discussion games, dispute trees characterize credulous acceptance under admissible semantics.

Theorem 5.2. *If $\mathtt{t}$ is an admissible dispute tree, then $D(\mathtt{t})$ is admissible. If $a \in E \in ad(SF)$, then there is some admissible dispute tree t with topic a s.t. $D(t) \subseteq E$.*

When closely inspecting the branches in the trees from Example 5.1, it becomes evident that there is a close relation between dispute trees and discussion games. For instance, the opponent edges in $\mathtt{t}$ stem from $\{b, c\}$, $\{f\}$, and $\{g\}$. These coincide with the opponent moves from Example 3.4. Moreover, the dispute tree displays a winning strategy because each node of the form $O : S$ has a child. In our example, $O : \{b, c, \}$ has $P : \{d, e\}$ as child and indeed, answering the O-move $\mathtt{out}_\exists(\{b, c\})$ with the P-move $\mathtt{in}(\{d, e\}) : \mathtt{out}(b)$ is part a winning strategy for the proponent.

Indeed, this is no coincidence: from an admissible dispute tree, we can deduce a winning strategy and vice versa. To see this, suppose $\mathtt{t}$ is an admissible dispute tree for a. Consequently, the root node is $[P : \{a\}]$. Now the children of this root node depict all possible opponent moves directed towards $\{a\}$, say $O_1, \dots, O_n$. For each of those O_i, the dispute tree depicts a corresponding P-move which we utilize to construct our winning strategy (that is, for each possible O-move we need some corresponding P-move). Inductively, we find a P-move for each possible O-move. In a dispute tree, O-moves might occur multiple times over different branches, and the corresponding P-move is not uniquely determined; however, since each branch of an admissible dispute tree is won by the proponent, we can take any possible P-move. Vice versa, it is straightforward to turn a winning strategy in an admissible dispute tree: for each O-node, we have a blueprint for the P-node we have to append. This yields the following.

Theorem 5.3. *An admissible dispute tree for a depicts a winning strategy for a preferred discussion starting with $\mathtt{in}(a)$ and vice versa.*

Having introduced SETAF dispute trees and compared them to discussion games, let us now see how they handle our introductory example.

Example 5.4. For our Example 1.1, the following dispute tree $\mathtt{t}$ yields a justification for acceptance of af.

$$\begin{tikzcd}[column sep=small]
& {[P:\{af\}]} \arrow[dl, dash, "af"'] \arrow[dr, dash, "af"] & \\
{[O:\{ls, all\}]} \arrow[d, dash, "all"'] & & {[O:\{ht, all\}]} \arrow[d, dash, "all"] \\
{[P:\{sv\}]} & & {[P:\{sv\}]} \\
\vdots & & \vdots
\end{tikzcd}$$

Note that dispute trees do not prevent the opponent player to make certain moves twice. Thus both branches have infinite length and alternate between $[P : sv]$ and $[O : all]$.

5.2 Strict Dispute Trees

As our introductory example illustrates, dispute trees are infinite in general. Inspired by the handling of dialogue games [Modgil and Caminada, 2009] we deal with infinite branches in $\mathtt{t}$ by preventing the opponent from using any label more than once, i.e. we forbid to bring forward the same counter-argument multiple times.

Definition 5.5. A strict dispute tree follows the requirements 1, 2, 4, and 5 as above, but adapts requirement 3 as follows:

3' every node of the form $[P : S]$ has children of the form $[O : T]$ for each $(T, s) \in R$ with $s \in S$ s.t. no $t \in T$ labels a defense edge on the path from $[P : S]$ to the root node, which are connected via edges labeled s for each such $(T, s) \in R$.

All the other notions like e.g. a *strict admissible dispute tree* are defined in the natural way.

It is clear that strict dispute trees are always finite. Moreover, moving to strict dispute trees does not surrender any expressiveness, as the following result establishes.

Proposition 5.6. *Let $F = (A, R)$ be an AF and $a \in A$. There is an admissible dispute tree $\mathtt{t}$ with topic a iff there is a strict admissible dispute tree s with topic a.*

Example 5.7. A strict dispute tree for our Example 1.1 is given as follows.

$$\begin{tikzcd}[column sep=small]
& {[P:\{af\}]} \arrow[dl, dash, "af"'] \arrow[dr, dash, "af"] & \\
{[O:\{ls, all\}]} \arrow[d, dash, "all"'] & & {[O:\{ht, all\}]} \arrow[d, dash, "all"] \\
{[P:\{sv\}]} & & {[P:\{sv\}]}
\end{tikzcd}$$

Now in both branches, the opponent is not allowed to use all anymore, which finishes the tree.

While dispute trees are generally much larger than discussion games (due to possibly repeated moves over different branches), by providing a winning strategy they are more informative than discussion games. Since we are interested in concise justifications, we put our focus on discussion games, whose linear structure is more natural for non-expert users.

6 SETAF Discussion Games for Instantiations

We showcase the advantages of SETAF instantiations in the context of ABA [Čyras *et al.*, 2018]. A flat ABA framework (ABAF) is a tuple $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^{-})$, where $\mathcal{L}$ is a set of atoms, $\mathcal{R}$ is a set of inference rules of form $a_0 \leftarrow a_1, \dots, a_n$ with $a_i \in \mathcal{L}$ and $a_0 \notin \mathcal{A}$, $\mathcal{A} \subseteq \mathcal{L}$ a set of assumptions, and $^{-} : \mathcal{A} \to \mathcal{L}$ is the contrary function.

In an ABAF, from a set S of assumptions we can construct an argument for an atom p, by iteratively applying the inference rules in the natural way. Then, an argument $S \vdash p$ *attacks* another argument $S' \vdash p'$ iff p is the contrary of some assumption in S'. This way, an ABAF D induces an AF F_D and reasoning in D can be captured by applying the usual semantics to F_D [Čyras *et al.*, 2018].

As noted in [König *et al.*, 2022], an ABAF D can also be captured by a SETAF $SF_D = (A, R)$ where $A = \mathcal{A}$ is the set of assumptions and $(T, a) \in R$ iff there is a tree-based argument from T deriving $\overline{a}$.

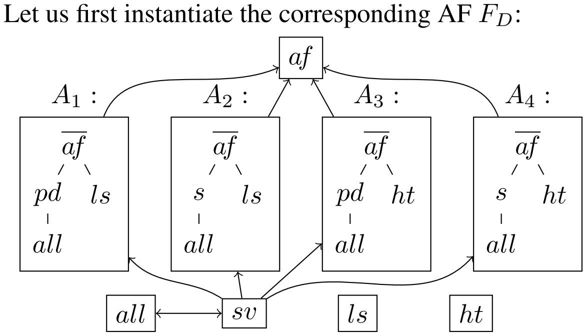
We model our introductory example as an ABA knowledge base. We use the assumptions all ("all inclusive"), sv ("small village"), ls ("long stay"), and ht ("hotel") to represent the trip specifications, i.e. if all is accepted, then our agents decides for an all inclusive vacation. Moreover, af ("affordable") indicates whether or not the trip is affordable. In summary, $\mathcal{A} = \{all, sv, ls, ht, af\}$. All inclusive and small village are mutually exclusive, so $\overline{all} = sv$ and $\overline{sv} = all$. Moreover, the all-inclusive option is only available in popular destinations (pd) during the holiday season (s), meaning that the destination is expensive. This yields the set $\mathcal{L} = \mathcal{A} \cup \{pd, s\}$ of atoms overall. Our agent wants to have a long stay in a hotel, but can afford this only if the destination is not expensive. This yields the following rules $\mathcal{R}$:

$$pd \leftarrow all \qquad s \leftarrow all$$

$$\overline{af} \leftarrow pd, ls \qquad \overline{af} \leftarrow s, ls \qquad \overline{af} \leftarrow pd, ht \qquad \overline{af} \leftarrow s, ht$$

We observe that this is a simple situation: since a long stay in a hotel is desired, the all-inclusive option is not affordable and our agent should thus visit a small village. Consequently, we would expect to obtain a simple justification which spots this in a crisp way.

Let us first instantiate the corresponding AF F_D:



For simplicity, we instantiate only the required arguments entailing $\overline{af}$ and not their (irrelevant) sub-arguments.

Now suppose our agent wants to know whether there is an acceptable set of arguments entailing ls, ht, and af representing an affordable long stay in some hotel. As one can easily see from the AF, this is possible by accepting the trip to a small village (sv). However, the AF discussion game for establishing this requires several steps:

$\texttt{in}(af)$, $\texttt{out}(A_1)$, $\texttt{in}(sv)$, $\texttt{out}(all)$, $\texttt{in}(sv)$, $\texttt{out}(A_2)$, $\texttt{in}(sv)$, $\texttt{out}(A_3)$, $\texttt{in}(sv)$, $\texttt{out}(A_4)$, $\texttt{in}(sv)$.

Note that the flow of the discussion indicates that sv is capable of defending our target set of arguments, but the AF F_D does not provide the tools to exploit this information. In particular, sv is called to counter-attack each and every one of the A_i arguments, making the discussion tedious and unnatural.

Meanwhile, the SETAF corresponding to the above ABAF D is actually the one we already saw in Example 1.1. Recall that at the end of Section 4 we found that the concise discussion yields the justification

$$\texttt{in}(af),\ \texttt{out}_{\exists}(\{ls, all\}),\ \texttt{in}(\{sv\})\texttt{:out}(all)$$

which is undoubtedly more efficient compared to the above AF discussion. Moreover, instead of iterating between sv and several arguments making use of all in a seemingly unorganized fashion, the above discussion neatly spots that all is the root of our agent's issues.

7 Conclusion

In this paper, we studied discussion games with a special emphasis on the existence of collective attacks. We advocated for the use of SETAFs as an abstract argumentation formalism to represent conflicts arising from a given knowledge base. We introduced discussion games for SETAFs, generalizing their AF counterparts by characterizing credulous acceptance for preferred semantics (as a base case for many other semantics). Due to the increased expressive power of SETAFs, they yield more concise justifications in the presence of collective attacks. In particular, we investigated how the discussion game's length relates to the corresponding admissible set. Further, we introduced the notion of *concise preferred discussions*, a notion that is genuine for SETAFs and finds no counter-part in AFs. We show that this notion can be employed to reduce the size of discussions even further. To round up our investigation, we presented SETAF dispute trees in Section 5 and compared them to SETAF discussion games. Applying these notions to ABA demonstrated the improvements our proposal achieves. Our research contributes to computing justifications for reasoning in rule-based systems [Čyras *et al.*, 2021]. As a future work direction, it would be exciting to see how utilizing SETAFs as means to instantiate a knowledge base can improve the quality of computed justifications. To broaden the range of applications, another natural step is to expand our study to further well-established argumentation semantics [Baroni *et al.*, 2018]. Moreover, it would be interesting to see how our strategy can be applied to other rule-based systems like ASPIC [Modgil and Prakken, 2013], logic-based argumentation [Besnard and Hunter, 2001] and DeLP [García and Simari, 2004].

Acknowledgments

The authors thank the reviewers for their helpful comments to improve the original version of this paper. This work has been supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 101034440). Moreover, this research has been supported by the Vienna Science and Technology Fund (WWTF) through project ICT19-065, the Austrian Science Fund (FWF) through project P32830, and by the Federal Ministry of Education and Research of Germany and by Sächsische Staatsministerium für Wissenschaft, Kultur und Tourismus in the programme Center of Excellence for AI-research "Center for Scalable Data Analytics and Artificial IntelligenceDresden/Leipzig", project identification number: ScaDS.AI.

References

[Atkinson *et al.*, 2017] Katie Atkinson, Pietro Baroni, Massimiliano Giacomin, Anthony Hunter, Henry Prakken, Chris Reed, Guillermo Ricardo Simari, Matthias Thimm, and Serena Villata. Towards artificial argumentation. *AI Magazine*, 38(3):25–36, 2017.

[Baroni *et al.*, 2018] Pietro Baroni, Martin Caminada, and Massimiliano Giacomin. Abstract argumentation frameworks and their semantics. In *Handbook of Formal Argumentation*. College Publications, February 2018.

[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.

[Besnard and Hunter, 2001] Philippe Besnard and Anthony Hunter. A logic-based theory of deductive arguments. *Artif. Intell.*, 128(1-2):203–235, 2001.

[Besnard and Hunter, 2008] Philippe Besnard and Anthony Hunter. *Elements of Argumentation*. MIT Press, 2008.

[Bikakis *et al.*, 2021] Antonis Bikakis, Andrea Cohen, Wolfgang Dvořák, Giorgos Flouris, and Simon Parsons. Joint attacks and accrual in argumentation frameworks. In *Handbook of Formal Argumentation*, volume 2, chapter 2. College Publications, 2021.

[Bondarenko *et al.*, 1997] Andrei Bondarenko, Phan Minh Dung, Robert A. Kowalski, and Francesca Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artif. Intell.*, 93:63–101, 1997.

[Caminada *et al.*, 2016] Martin W. A. Caminada, Wolfgang Dvořák, and Srdjan Vesic. Preferred semantics as socratic discussion. *J. Log. Comput.*, 26(4):1257–1292, 2016.

[Caminada, 2018] Martin Caminada. Argumentation semantics as formal discussions. In *Handbook of Formal Argumentation*, volume 1, chapter 10, pages 487–518. College Publications, 2018.

[Čyras *et al.*, 2018] Kristijonas Čyras, Xiuyi Fan, Claudia Schulz, and Francesca Toni. Assumption-based argumentation: Disputes, explanations, preferences. In *Handbook of Formal Argumentation*, volume 1, chapter 7, pages 365–408. College Publications, 2018.

[Čyras *et al.*, 2021] Kristijonas Čyras, Antonio Rago, Emanuele Albini, Pietro Baroni, and Francesca Toni. Argumentative XAI: A survey. In *Proc. IJCAI*, pages 4392–4399. ijcai.org, 2021.

[Dung *et al.*, 2006] Phan Minh Dung, Robert A. Kowalski, and Francesca Toni. Dialectic proof procedures for assumption-based, admissible argumentation. *Artif. Intell.*, 170(2):114–159, 2006.

[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):321–357, 1995.

[Dvořák *et al.*, 2019] Wolfgang Dvořák, Jorge Fandinno, and Stefan Woltran. On the expressive power of collective attacks. *Argument Comput.*, 10(2):191–230, 2019.

[Dvořák *et al.*, 2024] Wolfgang Dvořák, Matthias König, Markus Ulbricht, and Stefan Woltran. Principles and their computational consequences for argumentation frameworks with collective attacks. *J. Artif. Intell. Res.*, 79:69–136, 2024.

[Gabbay *et al.*, 2021] Dov Gabbay, Massimiliano Giacomin, Guillermo R. Simari, and Matthias Thimm, editors. *Handbook of Formal Argumentation*, volume 2. College Publications, 2021.

[García and Simari, 2004] Alejandro Javier García and Guillermo Ricardo Simari. Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming*, 4(1-2):95–138, 2004.

[König *et al.*, 2022] Matthias König, Anna Rapberger, and Markus Ulbricht. Just a matter of perspective – intertranslating expressive argumentation formalisms. In *Proc. COMMA 2022*, pages 212–223, 2022.

[Modgil and Caminada, 2009] Sanjay Modgil and Martin Caminada. Proof theories and algorithms for abstract argumentation frameworks. In *Argumentation in Artificial Intelligence*, pages 105–129. Springer, 2009.

[Modgil and Prakken, 2013] Sanjay Modgil and Henry Prakken. A general account of argumentation with preferences. *Artif. Intell.*, 195:361–397, 2013.

[Nielsen and Parsons, 2006] Søren Holbech Nielsen and Simon Parsons. A generalization of dung's abstract framework for argumentation: Arguing with sets of attacking arguments. In *Proc. ArgMAS*, volume 4766 of *Lecture Notes in Computer Science*, pages 54–73. Springer, 2006.

[Toni, 2014] Francesca Toni. A tutorial on assumption-based argumentation. *Argument Comput.*, 5(1):89–117, 2014.

[Yun *et al.*, 2020] Bruno Yun, Srdjan Vesic, and Madalina Croitoru. Sets of attacking arguments for inconsistent datalog knowledge bases. In *Proc. COMMA*, volume 326 of *Frontiers in Artificial Intelligence and Applications*, pages 419–430. IOS Press, 2020.