Toulmin Argumentation Logic

Floris BEX [a] Henry PRAKKEN [b]
[a] *Department of Information and Computing Sciences and School of Law, Utrecht University, The Netherlands*
[b] *Department of Information and Computing Sciences, Utrecht University, The Netherlands*

Abstract. In this paper we present a formal Toulmin Argumentation Logic (TAL) based on Toulmin's original argument scheme, and establish a formal correspondence between TAL and the *ASPIC*$^+$ framework for structured argumentation. Thus, we provide Toulmin's account with a precise notion of argument acceptability grounded in Dung's theory of abstract argumentation frameworks. Furthermore, the formalisation has direct practical relevance for recent work that uses Toulmin's scheme in the context of large language models.

Keywords. Toulmin, Argumentation, Logic

1. Introduction

Steven Toulmin's [21] account of the structure of arguments, captured in his famous Toulmin Argument Scheme (see Figure 1), has been extremely influential in many academic fields concerned with argumentation. Toulmin was one of the first who pointed at the defeasible nature of much ordinary, legal, ethical and scientific reasoning. It is therefore not surprising that Toulmin has influenced researchers in AI and computational argumentation (cf. [12,25]). For example, Toulmin's scheme has been used in various computer tools for argument diagramming and representation [13,18,17]; it has been discussed in the context of formal argumentation dialogue games [9,1]; and there have been several discussions of how aspects of Toulmin's argument scheme can be modelled in modern computational models of argument (e.g., [9,24]). Furthermore, in recent years, various researchers in natural language processing, including with large language models (LLMs), have experimented with Toulmin's original argument scheme [5,10,15].

Despite Toulmin's broad influence on AI and computational argumentation there is, as far as we are aware, no work that directly formally models Toulmin's original ideas.[1] Furthermore, the exact relation between Toulmin's scheme on the one hand and the various frameworks for constructing arguments and determining argument acceptability [2,6,14] has not yet been clarified. In this paper, we therefore provide a formal Toulmin Argumentation Logic (TAL), which stays as close as possible to Toulmin's original ideas, and establish a formal correspondence between TAL and the *ASPIC*$^+$ framework for structured argumentation [14].

[1] The formal interpretation of elements of Toulmin in [24] comes close, but does not provide a fully worked-out formalisation.

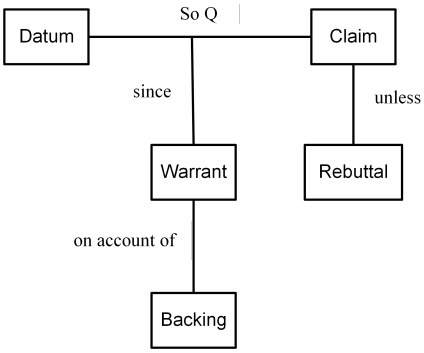

Figure 1. Toulmin's Argument Scheme

While Toulmin was clear on the structure of arguments, he did not have a precise notion of argument acceptability in mind. However, he stressed that his account of argument evaluation is procedural in that it is about whether an argument can be defended in a proper dispute. For our formalisation, this naturally points to choosing an approach in which arguments are first-class citizens, as opposed to, for instance, DefLog [23] or abstract dialectical frameworks [4], in which not arguments but statements are the primary objects of evaluation. Accordingly, we will present our formalisation in the context of [6]'s theory of abstract frameworks for argumentation. This also makes *ASPIC*$^+$ a natural choice, since it is a structured argumentation framework that directly generates Dung-style abstract argumentation frameworks.

Our formal account in terms of TAL and its correspondence to *ASPIC*$^+$ thus allows us to investigate to what extent Toulmin's original ideas were coherent and non-ambiguous. Furthermore, it provides Toulmin's account with a precise notion of argument acceptability. Finally, it has immediate practical relevance given the recent work on Toulmin and large language models [5,10,15]: our formalisation defines what a well-formed Toulmin argument is and what constitutes a legitimate attack, which can serve both as an evaluation standard for LLM-generated arguments and as a principled basis for prompting and argument construction.

This paper is organised as follows. In Section 2 we informally analyse Toulmin's argument scheme to motivate our formalisation. In Section 3 we present our formal definition of Toulmin arguments and the TAL argumentation framework, and illustrate the framework with an example. In Section 4 we show how TAL arguments can be translated into *ASPIC*$^+$ and prove that the translation preserves attack relations. In Section 5 we discuss related research, and Section 6 concludes.

2. Informal Analysis of Toulmin's Argument Scheme

The core elements of Toulmin's argument scheme are as follows (see Figure 1). The *Claim* can be obtained via a *Ground* [22] (called Datum in [21]), which is made into a reason for the claim by a *Warrant*. Claims are *qualified* and when their qualifier indicates a defeasible inference, they can be *rebutted* if exceptional circumstances apply. We will assume the list of qualifiers in [22, p. 86]:

> *necessarily, certainly, presumably, in all probability, so far as the evidence goes, for all that we can tell, very likely, very possibly, maybe, apparently, plausibly, or so it seems.*

The qualifiers *necessarily* and *certainly* indicate a deductive argument while all other qualifiers indicate a defeasible inference.

A warrant should be justified by a *backing*. In factual domains a backing can, for example, be observations of certain regularities of which the warrant is the generalisation by way of induction. In the legal domain an example of a backing is observations about the terms and dates of enactment of the relevant statutory provisions.

In [22] Toulmin et al. generalise the scheme from [21] in two ways. They allow that an argument has more than one ground and that that each ground can be the claim of another argument. Thus they in fact make the definition of an argument recursive.

A legal example argument from [21] is an argument for the claim that Harry presumably is a British subject on the ground that Harry was born in Bermuda, which is warranted by the fact that anyone born in Bermuda is a British subject. The backing here consists of a reference to the relevant statutes and other provisions. A factual example from [21] is an argument for the claim that Petersen is almost certainly not a Roman Catholic on the ground that Petersen is a Swede, which is warranted by the generalisation that scarcely any Swede is a Roman Catholic. This warrant is backed by a reference to statistics about the proportion of Swedes who are Roman Catholic.

We next discuss some elements separately to explain and motivate our design decisions in our formalisation of Toulmin's argument scheme.

Grounds Unlike modern formal argumentation systems, neither [21] nor [22] explicitly assume a knowledge base from which arguments take their grounds. Therefore, we will also not assume such a knowledge base. This will somewhat complicate our definition of arguments compared to most modern argumentation systems, such as Pollock's system [16], assumption-based argumentation [7] and *ASPIC*$^+$ [14].

Rebuttals As also noted by [24], Toulmin in [21] is somewhat ambiguous on whether a rebuttal just makes the warrant inapplicable or also licenses the opposite conclusion (in Pollock's system [16] and *ASPIC*$^+$ [14] this is undercutting versus rebutting attack). The description of rebuttals in [22, p. 96] points at the former interpretation:

> It [a rebuttal] registers the fact that the inference is warranted – that the claim is directly supported by the grounds – only *in the absence of some particular exceptional condition*, which would undercut (i.e., withdraw the authority of the warrant for) the inference. [...] An argument that would normally have been sound is invalidated as a result of the discovery of special circumstances.

On the other hand, in several examples the rebuttal also seems to allow to conclude the opposite. Nevertheless, since this is not generally the case, we have made the choice to let a rebuttal do no more than making the warrant inapplicable, so we interpret a rebuttal as something that would, if true, undercut in the sense of [16,14].

Note that a rebuttal only denotes, as Toulmin [21, p. 94] says, "the exceptional conditions which *might be* capable of defeating or rebutting the warranted conclusion" (emphasis added by B&P). That is, a rebuttal does not represent a direct counterclaim to the original argument, but only indicates what such a claim could be. Toulmin does not dis-

cuss where such an actual rebuttal would come from, so we assume it is based on another argument that has the exceptional condition (rebuttal) as its Claim.

Warrants We want to model arguments about warrants. At first sight, this would seem to deviate from Toulmin's account, but some of the discussions in [21,22] still point at the possibility of such arguments. The first quote below is from [21, p. 98].

> Indeed, if we demanded the credentials of all warrants at sight and never let one pass unchallenged, argument could scarcely begin. Jones puts forward an argument invoking warrant W1, and Smith challenges that warrant; Jones is obliged, as a lemma, to produce another argument in the hope of establishing the acceptability of the first warrant, but in the course of this lemma employs a second warrant W2; Smith challenges the credentials of this second warrant in turn; and so the game goes on. Some warrants must be accepted provisionally without further challenge, if argument is to be open to us in the field of question (...)

Although Toulmin here primarily argues that some warrants have to be accepted without discussion to have an argument started, this quote still leaves room for warrants for warrants. For example, on pp. 276-7 in [22] a distinction is made between *regular* or *rule-applying* arguments and *critical* or *rule-justifying* arguments (where rules are warrants). Likewise, [21, p. 111] distinguishes between 'warrant-using' and 'warrant-establishing' arguments. Of the latter Toulmin writes "In this type of argument the warrant, not the conclusion, is novel, and so on trial'. He gives an example of a warrant-establishing argument in which a scientist induces a statistical warrant from a series of experiments.

All these fragments indicate that Toulmin wanted to allow for arguments about warrants. We speculate that a reason why he did not explicitly model such arguments in his scheme may be that he was not aware that this could be done by applying his scheme recursively, as is usual in modern argumentation logics.

3. Formalisation of the Toulmin Argumentation Scheme

We now define arguments of a Toulmin argumentation logic, sticking closely to the extended Toulmin argument scheme (TAS) as proposed in [22]. Then informally, elementary arguments have the following form: *Grounds, so (Qualifier) Claim since Warrant, on account of Backing, except when Rebuttal*. Arguments can be combined by regarding the claim of one argument as a ground of another argument. Since in Toulmin's account of his argument scheme the logical form of the statements in the scheme's elements does not matter, no logical language for these elements needs to be specified, except that we assume that warrants have a list of conditions and a consequent. Below we only give the rules for arguments of which the qualifier indicates a defeasible inference. These arguments all have a possible rebuttal. The rules for arguments with a qualifier indicating a deductive inference are the same except that such arguments do not have a rebuttal.

Definition 1 [**TAS-Argument**] A *TAS-argument* A is:

1. $G_1, \ldots, G_n$ *so* (Q) C *since* W *on account of* B *except when* R, if
   - $n \geq 1$,
   - $Q \notin \{Necessarily, Certainly\}$,

- W has conditions $G_1, \ldots, G_n$ and consequent C
where:
$\texttt{Grounds}(A) = \{G_1, \ldots, G_n\}$,
$\texttt{Claim}(A) = C$,
$\texttt{Warrant}(A) = W$,
$\texttt{Backing}(A) = B$,
$\texttt{Qualifier}(A) = \{Q\}$,
$\texttt{Rebuttal}(A) = R$,
$\texttt{Sub}(A) = \{A\}$.

2. $A_1, \ldots, A_i, G_j, \ldots G_n$ *so* (Q) C *since* W (*on account of* B) *except when* R, if
- $i \geq 1, n \geq 0$
- $Q \notin \{Necessarily, Certainly\}$,
- $A_1, \ldots, A_i$ are arguments,
- W has conditions $\texttt{Claim}(A_1), \ldots, \texttt{Claim}(A_i), G_j, \ldots G_n$ and consequent C
where:
$\texttt{Grounds}(A) = \texttt{Grounds}(A_1) \cup \ldots \cup \texttt{Grounds}(A_i) \cup \{G_j, \ldots, G_n\}$,
$\texttt{Claim}(A) = C$,
$\texttt{Warrant}(A) = W$,
$\texttt{Backing}(A) = B$,
$\texttt{Qualifier}(A) = \{Q\}$,
$\texttt{Rebuttal}(A) = R$,
$\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_i) \cup \{A\}$.

3. $A_0, \ldots, A_i, G_j, \ldots G_n$ *so* (Q) C *since* W *on account of* A_m *except when* R if -
- $i \geq 1$ or $n \geq 1$,
- $Q \notin \{Necessarily, Certainly\}$,
- $A_0, \ldots, A_i, A_m$ are arguments,
- $\texttt{Claim}(A_m) = W$
where:
$\texttt{Grounds}(A) = \texttt{Grounds}(A_1) \cup \ldots \cup \texttt{Grounds}(A_i) \cup \texttt{Grounds}(A_m) \cup \{G_j, \ldots, G_n\}$,
$\texttt{Claim}(A) = C$,
$\texttt{Warrant}(A) = W$,
$\texttt{Backing}(A) = A_m$,
$\texttt{Qualifier}(A) = \{Q\}$,
$\texttt{Rebuttal}(A) = \{R\}$,
$\texttt{Sub}(A) = \texttt{Sub}(A_0) \cup \ldots \cup \texttt{Sub}(A_i) \cup \texttt{Sub}(A_m) \cup \{A\}$.

Clause 1 is the base case for defeasible inferences, where arguments are defined as instantiations of Toulmin's argument scheme and have a qualifier that indicates a defeasible inference. Clause 2 recursively defines complex arguments that support some or all grounds with a further argument, and clause 3 recursively defines complex arguments that provide a further argument for a warrant. Note that since clause 3 refers to A_0, it allows arguments of which the grounds are not arguments themselves but in which the warrant is given in the form of an argument.

The inclusion of $G_j, \ldots G_n$ in clauses 2 and 3 is to respect that arguments do not have to 'bottom out' in a knowledge base but can state their own grounds. Note that because of this design choice there can be two versions of an argument, for instance, a

version with a ground G and a version with a subargument for G. We think this is not a problem since in actual disputes one person may state an argument with ground G after which another person supports it with an argument for G. These two arguments can be said to have different owners.

We next define the notion of an argumentation framework for the Toulmin Argumentation Logic, by only regarding an argument for a rebuttal of a warrant as an attack.

Definition 2 [**Attack**] Argument A *attacks* argument B on B' iff $\texttt{Claim}(A) = R$ for some $B' \in \texttt{Sub}(B)$ such that $\texttt{Rebuttal}(B') = R$.

An *abstract* argumentation framework is then a pair $AF = (\mathcal{A}, \mathcal{C})$ where $\mathcal{A}$ is a set of arguments closed under subarguments and $\mathcal{C}$ is the attack relation defined over $\mathcal{A}$. Given such an AF, the acceptability of TAS-arguments can then be determined under various semantics [6].

Figure 2 illustrates arguments about warrants with an extended and slightly modified version of an example from [21, p. 111] (which included no rebuttal of the main warrant). The top right part *Petersen is a Swede so he almost certainly is not a Roman Catholic since a Swede can be taken to be almost certainly not a Roman Catholic (unless he was born in a Roman Catholic country)* instantiates Toulmin's original scheme while the rest of the figure, with the boxes in grey colour, displays an argument for a warrant licensing this inference. This grey part corresponds to argument A_m in Definition 1(3).

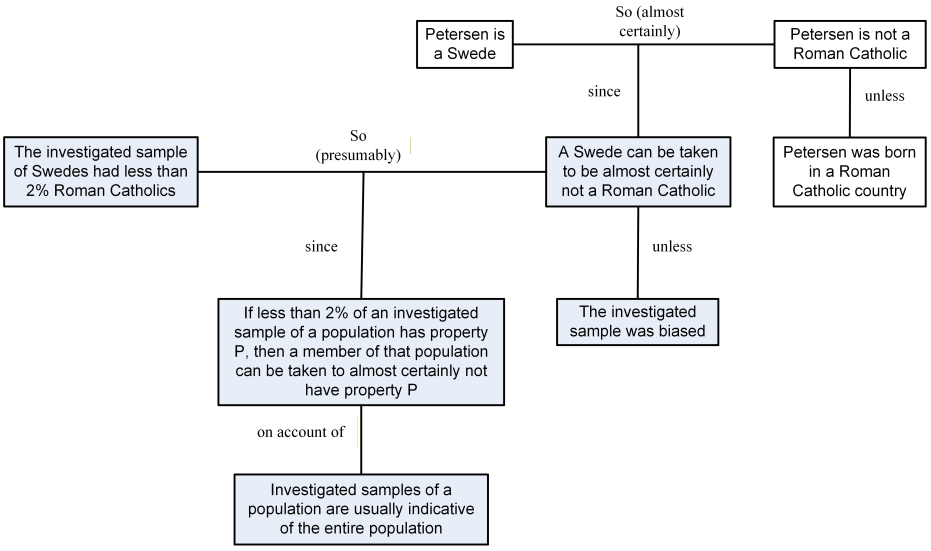

Figure 2. An argument about a warrant

4. Translating into current argumentation formalisms

We next briefly sketch how our Toulmin Argument Logic can be reconstructed in a state-of-the-art argumentation formalism, for which we choose *ASPIC*$^+$ (readers familiar with

ASPIC$^+$ will have noted that the form of Definition 1 is inspired by *ASPIC*$^+$'s definition of an argument). One reason for choosing *ASPIC*$^+$ is that, like TAL (Definition 2), it directly instantiates [6]'s Argumentation Frameworks. Furthermore, Toulmin's warrants seem to naturally correspond to the inference rules of *ASPIC*$^+$. Note that our account can be translated into assumption-based argumentation (ABA) by applying [8]'s translation of *ASPIC*$^+$ rules into ABA rules, although, unlike with *ASPIC*$^+$, extending the ABA modelling with preferences would not be straightforward since 'standard' ABA has no preferences.

The *ASPIC*$^+$ framework [14] defines notions of argument, attack and defeat and thus generates abstract argumentation frameworks in which the *ASPIC*$^+$ defeat relation instantiates Dung's attack relation. For present purposes an informal summary suffices. Each step in an argument is of the form $P_1, \ldots, P_n$, *therefore* Q where

- each P_i and Q is a statement in some language $\mathcal{L}$,
- The inference step is supposed to be based on an inference rule of the form $P_1, \ldots, P_n \rightarrow/\Rightarrow Q$, where a rule with $\rightarrow$ is *strict* and a rule with $\Rightarrow$ is *defeasible*. Strict rules guarantee that their consequent Q is acceptable if their antecedents $P_1, \ldots, P_n$ are acceptable, while defeasible rules in that case only create a presumption for the acceptability of their conclusion.
- Each P_i is either given as an element of the knowledge base $\mathcal{K}$ (divided into two disjoint subsets $\mathcal{K}_n$ of necessary and $\mathcal{K}_p$ of ordinary premises) or a conclusion from a preceding inference in the argument,
- Q is the conclusion of the inference step.

Arguments are then trees (or more generally directed acyclic graphs) of such inference steps, where the leaves are its *premises*, the links are inference steps, the root is the *final conclusion* and all other nodes are *intermediate conclusions*. Subtrees of an argument are its *subarguments*.

Arguments can be *attacked* in three ways: on an ordinary premise (*undermining attack*), on their defeasible inference steps (*undercutting attack*) and on their defeasibly derived conclusions (*rebutting attack*). Attacks can be *direct* (on an argument's final inference step of conclusion) or *indirect* (on an intermediate inference step or conclusion). In general, *ASPIC*$^+$ distinguishes between attack and defeat relations between arguments, where the defeats are the attacks that are successful according to some preference relation on the set of all arguments. However, in this paper we assume that there are no preferences, so all attacks succeed as defeats. If A defeats B but B does not defeat A we say that A *strictly defeats* B.

We next sketch how the formalisation of Section 3 can be translated into *ASPIC*$^+$. The *ASPIC*$^+$ knowledge base and sets of inference rules consist of all and only the elements specified below.

Grounds First, since *ASPIC*$^+$ assumes that all premises are taken from $\mathcal{K}$, we assume that all grounds of a TAS argument are in $\mathcal{K}_n$ and form the base cases of the inductive definition of an argument. Grounds are in $\mathcal{K}_n$ since Toulmin does not consider attacks on grounds.

Warrants Warrants are *ASPIC*$^+$-style inference rules. They are strict if their qualifier is *Necessarily* or *Certainly* and defeasible otherwise (recall the list of qualifiers from [22] cited above).

Backings Because backings are about the question whether a warrant is valid, we include an additional antecedent $Backed(w)$ to a warrant/rule, which expresses that the warrant is backed, or valid. This idea is inspired by early work in AI & Law, such as [20,9,11]). In particular, Hage [11] explains that invalidity is not the same as inapplicability due to an exception (Toulmin's rebuttal, undercutter in *ASPIC*$^+$). For instance, a legal rule can be a valid rule of law while yet inapplicable to a particular case because of an exception, or a generalisation can be justified in general while yet being inapplicable to a specific case. A backing B for a warrant can then be modelled as a simple argument $B \Rightarrow Backed(w)$, where B is added to $\mathcal{K}_n$, since Toulmin does not consider attacks on backings. More complex arguments for a warrant (cf. Clause 3 of Definition 1) can also be modelled as arguments for this additional backing condition $Backed(w)$.

Rebuttals Rebuttals R are modelled by adding the weak negation $\sim R$ as an additional antecedent of the warrant and as an item in $\mathcal{K}_p$. In *ASPIC*$^+$ an argument with conclusion R asymmetrically attacks a premise $\sim R$.

Warrants Warrants w then have the form

$$w : G_1, \ldots G_n, Backed(w), \sim R \Rightarrow C$$

where $G_1, \ldots G_n$ are the conditions of the Toulmin warrant, $\texttt{Backed}(w)$ is the validity condition corresponding to the backing for a warrant, and $\sim R$ expresses that the defeasible assumption that the rebuttal does not apply. Arguments for the rebuttal then asymmetrically undermine arguments using the warrant to which the rebuttal is attached.

Finally, apart from the use of undercutters or weak negation to model rebuttals, the object language $\mathcal{L}$ does not contain a negation symbol, to model that rebuttals are the only form of attack considered by Toulmin. It is then easy to see that the only possible attacks are those corresponding to an attack according to Definition 2. Thus warrants effectively correspond to defaults of default logic [19] or rules of answer set programming [3].

We now inductively define a formal translation of TAL arguments to *ASPIC*$^+$ arguments. Like for TAL arguments we only give the rules for defeasible inferences. The rules for deductive inferences are the same except that they do not contain the element $\sim R$ and the indicated *ASPIC*$^+$ rule is strict ($\rightarrow$). Note that while in TAL as defined above grounds are introduced in arguments, *ASPIC*$^+$ arguments take premise from their knowledge base and premises are defined as basic arguments. Below this is left implicit. Finally, $\rightsquigarrow$ is a variable ranging over $\{\Rightarrow, \rightarrow\}$.

Definition 3 For every TAL argument A the *ASPIC*$^+$ translation $t(A)$ is defined as follows.

1. If $A = G_1, \ldots, G_n$ *so* (Q) C *since* W *on account of* B (*except when* R) then
   $t(A) = G_1, \ldots, G_n, A_B \ \rightsquigarrow \ C$ where $\rightsquigarrow = \Rightarrow$ if A has a rebuttal and $\rightsquigarrow = \rightarrow$ otherwise and
   $A_B = B \Rightarrow Backed(w)$;
   $G_1, \ldots, G_n, B \in \mathcal{K}_n$ and $\sim R \in \mathcal{K}_p$;
   $\texttt{Prem}(A) = \{G_1, \ldots, G_n, B, \sim R\}$;
   $\texttt{Conc}(A) = C$;
   $\texttt{Sub}(A) = \{G_1, \ldots, G_n, B, A_B, \sim R, A\}$;
   $\texttt{Rules}(A) = \{G_1, \ldots, G_n, Backed(w), \sim R \rightsquigarrow C\} \ \cup \ \{B \Rightarrow Backed(w)\}$.

2. If $A = A_1, \ldots, A_i$ *so* (Q) C *since* W *on account of* B (*except when* R), then
   $t(A) = t(A_1), \ldots, t(A_i), A_B \rightsquigarrow C$ where $\rightsquigarrow = \Rightarrow$ if A has a rebuttal and $\rightsquigarrow = \rightarrow$ otherwise and
   $A_B = B \Rightarrow Backed(w)$;
   $\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_i) \cup \{B\}$;
   $\texttt{Conc}(A) = C$;
   $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_i) \cup \{B, A_B, A\}$;
   $\texttt{Rules}(A) = \texttt{Rules}(A_1) \cup \ldots \cup \texttt{Rules}(A_i) \cup$
   $\{\texttt{Conc}(A_m), \texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_i), Backed(w), \sim R \rightsquigarrow C\} \cup$
   $\{B \Rightarrow Backed(w)\}$.
3. If $A = A_1, \ldots, A_i$ *so* (Q) C *since* W *on account of* A_m (*except when* R), then
   $t(A) = t(A_m), t(A_1), \ldots, t(A_i) \rightsquigarrow C$ where $\rightsquigarrow = \Rightarrow$ if A has a rebuttal and $\rightsquigarrow = \rightarrow$ otherwise, and
   $\texttt{Conc}(A_m) = Backed(w)$ where w is the top rule of $t(A)$;
   $\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_i) \cup \texttt{Prem}(A_m)$;
   $\texttt{Conc}(A) = C$;
   $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_i) \cup \{A\} \cup \texttt{Sub}(A_m)$;
   $\texttt{Rules}(A) = \texttt{Rules}(A_1) \cup \ldots \cup \texttt{Rules}(A_i) \cup \texttt{Rules}(A_m) \cup$
   $\{\texttt{Conc}(A_m), \texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_i), \sim R \rightsquigarrow C\}$.

For any set S of TAL arguments we denote the set of all their *ASPIC*$^+$ translations as $t(S)$. Then for a TAL $AF = (\mathcal{A}, \mathcal{C})$ we write $t(AF)$ for $(t(\mathcal{A}) \cup \texttt{Prem}(t(\mathcal{A})), \mathcal{C}')$ where $\texttt{Prem}(t(\mathcal{A}))$ is the set of all premises of any argument in $\mathcal{A}$ and $\mathcal{C}'$ is the *ASPIC*$^+$ attack relation on $t(\mathcal{A}) \cup \texttt{Prem}(t(\mathcal{A}))$.

Figure 3 displays the *ASPIC*$^+$ version of the argument from Figure 2. The grey part corresponds to the grey part in Figure 2. The argument uses the following warrants, in which rebuttals are weakly negated in the warrants' antecedents.

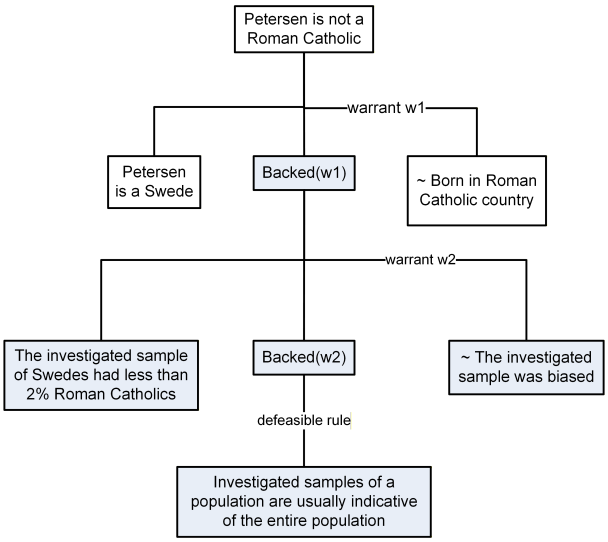

Figure 3. An *ASPIC*$^+$ argument with a subargument about a warrant

w_1: *Swede*, $Backed(w_1)$, $\sim$ *Born in R.C. country* $\Rightarrow$ *Not a Roman Catholic*
w_2: *Investigated sample* $< 2\%$ *R.C.*, $Backed(w_2)$, $\sim$ *Biased sample* $\Rightarrow$ $Backed(w_1)$

Given the translation from Definition 3, we can prove that attacks between arguments in the original TAL-AF correctly transfer to the translated AF and vice versa.

Proposition 1 For any TAL arguments $A, B \in AF$ and *ASPIC*$^+$ arguments $t(A), t(B) \in t(AF)$ it holds that B attacks A in AF iff $t(B)$ attacks $t(A)$ in $t(AF)$.

PROOF. The proof is with induction on the structure of arguments. For reasons of space, we only give the proof of the only-if part. The proof of the if-part is similar. Furthermore, we only give the proof for TAL arguments satisfying Definition 1(2); the proof for arguments satisfying Definition 1(3) is similar. Finally, we ignore deductive inferences since these cannot be attacked.

Base case: $A = G_1, \ldots, G_n$ *so* (Q) C *since* W *on account of* B *except when* R. Let $\texttt{Conc}(B) = R$. Then B attacks A. But then $\sim R \in \texttt{Prem}(t(A))$ and $\texttt{Conc}(t(B)) = R$, so $t(B)$ attacks $t(A)$.

Induction hypothesis: If $A = A_1, \ldots, A_i$ *so* (Q) C *since* W *on account of* B *except when* R, then for all $A_j (1 \leq j \leq i)$: if B attacks A_j then $t(B)$ attacks $t(A_j)$.

Induction step: Consider $A = A_1, \ldots, A_i$ *so* (Q) C *since* W *on account of* B *except when* R, and let $\texttt{Conc}(B) = R$. Then B (directly) attacks A. But then $\sim R \in \texttt{Prem}(t(A))$ and $\texttt{Conc}(t(B)) = R$, so $t(B)$ (directly) attacks $t(A)$. Moreover, by the induction hypothesis, for any argument B that indirectly attacks A on $A_j (1 \leq j \leq i)$ it holds that $t(B)$ indirectly attacks $t(A)$ on $t(A_j)$. QED

Since for any TAL AF the corresponding $t(AF)$ further only contains premise arguments between which no odd defeat cycles are possible, Proposition 1 implies that in any of [6]'s semantics, every TAL argument A has the same acceptance status in AF as $t(A)$ has in AF'.

5. Related research

While Toulmin is often mentioned as an inspiration by researchers working on formal argumentation [1,9], there is not much work that explicitly connects formal accounts of argumentation – that is, frameworks for constructing arguments and determining argument acceptability [2,6,14] – to Toulmin's scheme. The only such work we are aware of is [24]. Verheij directly interprets and discusses Toulmin's scheme in the context of his formal DefLog system [23], which focuses on support and attack between statements. The main focus is on single-step Toulmin arguments, where Datum supports a Claim (denoted as a conditional D $\sim>$ C in DefLog). Warrants are effectively 'duplicated' by interpreting them as warranting such a conditional (i.e., W $\sim>$ (D $\sim>$ C), so both W and the conditional D $\sim>$ C represent the Warrant). Thus, Verheij's Warrants are not, like in our approach, rules that can be directly used to infer the claim when its antecedents are satisfied. Backings are modelled as reasons for Warrants, and Qualifiers are included as part of the Claim statement. Thus, multi-step arguments, arguments with multiple Grounds/Data and arguments for Warrants are not considered, even though DefLog allows for such extensions. The definition of Rebuttals is broadened into five at-

tack types targeting different statements or conditionals in the core Toulmin scheme. Finally, Verheij demonstrates how statement acceptance (i.e., justified or defeated) can be determined using DefLog's idea of dialectical interpretation. While Verheij thus adds a certain level of formal precision and shows how the acceptability of statements can be determined, he stops short of providing a full formalisation. Another difference with our approach is one of fidelity: where Verheij interprets the individual elements of Toulmin's scheme as (connected) statements in DefLog, we treat Toulmin's original arguments as the primary object of formalisation and aim to preserve them as directly as possible in a Toulmin argumentation logic.

In recent years, there have been quite a few papers that use Toulmin's scheme as a basis for reasoning with Large Language Models. We believe the current formalisation TAL and its translation into *ASPIC*$^+$ is of use to this work, as follows. Gupta et al. [10] decompose natural language arguments into (claim, datum, warrant) triples using zero-shot LLM prompting, leaving out backing, rebuttal and qualifier. TAL makes explicit the formal commitments implicit in this setup and clarifies what is left out. Castagna et al. [5] use Toulmin's six components as the basis for critical questions that probe the validity of an LLM's reasoning steps, targeting data, warrant, backing, claim, qualifier and rebuttal. While TAL itself only models rebuttals as undercutters, the translation to *ASPIC*$^+$ opens up the full range of attack types – undermining, rebutting and undercutting – providing a principled formal basis for the kind of critical questioning their pipeline performs. Finally, Park et al. [15] go furthest in engaging with formal argumentation, building networks of datum-warrant-claim units in a special version of *ASPIC*$^+$ to generate and evaluate legal counterarguments. However, their formalisation of Toulmin-like notions is informal and incomplete: backings are absent, nesting of arguments is not defined, and the conditions under which one argument may attack another are not always clear. The present paper hence provides the formal foundation their system requires.

6. Conclusion

Toulmin's argument scheme [21,22] remains one of the most influential and widely used frameworks for the analysis of arguments, both in theoretical work on argumentation and in practical applications ranging from legal reasoning to natural language processing. Despite this broad influence also on computational argumentation, a full formal model of Toulmin's original ideas has so far been lacking. In this paper we have therefore presented a formal Toulmin Argumentation Logic (TAL) based on Toulmin's original work, including recursive arguments with multiple grounds, arguments for warrants, and rebuttals as exceptions to warrants. We have also established a formal correspondence between TAL and *ASPIC*$^+$ [14], proving that attacks are preserved under translation, which allows the acceptability of TAL arguments to be determined under Dung's standard argumentation semantics [6]. Future work includes extending TAL with other types of rebuttals, investigating the relation between Toulmin's qualifiers and argument strength and a preference relation on arguments, and exploring how the formalisation can be put to use as an evaluation standard in LLM-based argumentation systems.

References

[1] T.J.M. Bench-Capon. Specification and implementation of Toulmin dialogue game. In *Legal Knowledge-Based Systems. JURIX: The Eleventh Conference*, pages 5–19, Nijmegen, 1998. Gerard Noodt Instituut.

[2] T.J.M. Bench-Capon and P.E. Dunne. Argumentation in Artificial Intelligence. *Artificial intelligence*, 171(10-15):619–641, 2007.

[3] G. Brewka, T. Eiter, and M. Truszyński. Answer set programming at a glance. *Communications of the ACM*, 54:92–103, 2011.

[4] G. Brewka and S. Woltran. Abstract dialectical frameworks. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference*, pages 102–111. AAAI Press, 2010.

[5] F. Castagna, I. Sassoon, and Simon Parsons. Critical-questions-of-thought: Steering LLM reasoning with argumentative querying, 2024. arXiv:2412.15177.

[6] P.M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and *n*–person games. *Artificial Intelligence*, 77:321–357, 1995.

[7] P.M. Dung, R.A. Kowalski, and F. Toni. Assumption-based argumentation. In I. Rahwan and G.R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 199–218. Springer, Berlin, 2009.

[8] P.M. Dung and P.M. Thang. Closure and consistency in logic-associated argumentation. *Journal of Artificial Intelligence Research*, 49:79–109, 2014.

[9] T.F. Gordon. The Pleadings Game: an exercise in computational dialectics. *Artificial Intelligence and Law*, 2:239–292, 1993.

[10] A. Gupta, E. Zuckerman, and B. O'Connor. Harnessing Toulmin's theory for zero-shot argument explication. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10259–10276, 2024.

[11] J.C. Hage. A theory of legal reasoning and a logic to match. *Artificial Intelligence and Law*, 4:199–273, 1996.

[12] D. Hitchcock and B. Verheij, editors. *Arguing on the Toulmin Model. New Essays in Argument Analysis and Evaluation*. Argumentation Library (ARGA, volume 10). Springer, 2006.

[13] C. C. Marshall. Representing the structure of a legal argument. In *Proceedings of the 2nd International Conference on Artificial Intelligence and Law*, ICAIL '89, page 121–127, New York, NY, USA, 1989. Association for Computing Machinery.

[14] S. Modgil and H. Prakken. The ASPIC+ framework for structured argumentation: a tutorial. *Argument and Computation*, 5:31–62, 2014.

[15] S. Park, A. Choi, and R. Park. Objection, your honour!: an LLM-driven approach for generating Korean criminal case counterarguments. *Artificial Intelligence and Law*, 2025. https://doi.org/10.1007/s10506-025-09432-2.

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

[17] I. Rahwan and P.V. Sakeer. Towards Representing and Querying arguments on the Semantic Web. In *Proceedings of the 2006 conference on Computational Models of Argument: Proceedings of COMMA 2006*, pages 3–14. IOS Press, 2006.

[18] C. Reed and G. Rowe. Translating Toulmin diagrams: Theory neutrality in argument representation. *Argumentation*, 19(3):267–286, 2005.

[19] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[20] G. Sartor. A formal model of legal argumentation. *Ratio Juris*, 7:212–226, 1994.

[21] S.E. Toulmin. *The Uses of Argument – Updated Edition*. Cambridge University Press, Cambridge, 2003. Originally published in 1958.

[22] S.E. Toulmin, R. Rieke, and A. Janik. *An Introduction to Reasoning*. Macmillan Publishing, New York, NY, second edition, 1984.

[23] B. Verheij. DefLog: on the logical interpretation of prima facie justified assumptions. *Journal of Logic and Computation*, 13:319–346, 2003.

[24] B. Verheij. Evaluating arguments based on Toulmin's scheme. *Argumentation*, 19:347–371, 2005.

[25] B. Verheij. The Toulmin Argument Model in Artificial Intelligence: Or: how semi-formal, defeasible argumentation schemes creep into logic. In *Argumentation in artificial intelligence*, pages 219–238. Springer, 2009.