On the Acceptability of Meta-Arguments

Guido Boella
University of Turin
Italy
Email: guido@di.unito.it

Leendert van der Torre
University of Luxembourg
Luxembourg
Email: leendert@vandertorre.com

Serena Villata
University of Turin
Italy
Email: villata@di.unito.it

Abstract

In this paper we introduce a theory of meta-argumentation, by using Dung's theory of abstract argumentation to reason about itself. Meta-arguments are generated from atomic arguments, and extensions of acceptable meta-arguments are based on Dung's argumentation semantics. To illustrate our theory, we show how to represent Toulmin schemes in this theory by introducing meta-arguments using the Caminada labeling, and meta-arguments for support.

1. Methodology

Usually, when people encounter that the abstract nature of Dung's argumentation theory makes it difficult to represent an example of argumentation, their first inclination is to *extend* Dung's theory, for example with preferences among arguments [1], second and higher-order attack relations [2], [3], support relations among arguments [4], or priorities among arguments. However, we propose to instantiate Dung's theory rather than to extend it.

In a sense, it may be argued that Dung and colleagues [5] propose already to *instantiate* his theory rather than to extend it, and abstract arguments have been instantiated by, for example, assumptions, default rules, or clauses from a logic program. However, Dung's framework is seen as an abstract reference model into which less abstract models can be mapped, but is not meant to be the "starting point" of a modeling activity. [5] refers to Dung's framework as an abstraction of logic programming semantics interpretation, and the assumption-based approach proposed there is not introduced as an instantiation of Dung's framework but rather as a sort of intermediate abstraction with respect to various non-monotonic logics.

We instantiate Dung's theory with meta-arguments, such that we use Dung's theory to reason about itself. E.g., one may argue whether 'don't throw rubbish on the floor!' counts as an argument or not, whether it counts as an attack on 'be free!', or whether it supports 'respect other people!', or which argumentation semantics should be used. Or, one may argue whether a coalition of the Left is preferred to one of the Right or whether a larger coalition is preferred for getting more votes despite being less stable.

In particular, for argument a, we introduce meta-arguments $a_\in$ and $a_\notin$ of the Jacobovits - Vermeir - Caminada labeling, representing whether the argument a is an element of the set of acceptable arguments or not, which are used to represent the Toulmin scheme [6]. Moreover, for arguments a, b and c, we can add meta-arguments like a supports b,' 'a attacks b', 'c attacks the attack from a to b', etc.

Our methodology is based on the idea that we take the acceptable meta-arguments using some argumentation semantics, filter out the atomic arguments like $a_\in$ by removing auxiliary arguments like $a_\notin$, and then the set of atomic arguments represents the acceptable arguments of the extended theory. In the trivial cases where no additional information is added, they will correspond to the acceptable arguments using the same argumentation semantics used for the meta-argumentation theory.

As an example of our methodology and its challenges, consider a meta-argumentation framework with meta-arguments $a_\in$ and $a_\notin$. Without further motivations, the most natural representation may seem that they attack each other, which means that neither of them will be in the grounded extension, and there will be two preferred or stable extensions, one containing $a_\in$ and one containing $a_\notin$. However, since there is no such symmetry in the Jacobovits - Vermeir - Caminada labeling, this intuition is flawed. More importantly, such a symmetric attack relation would not give us any accepted arguments in the grounded extension. The other symmetric alternative, of not imposing any constraints on the attack relation between the two arguments, would lead to the possibility where both would be accepted, which is flawed as well. Thus we have to demand that either all the *in* meta-arguments attack their *out* counterparts or vice versa; we adopt the former approach.

As a more involved example, consider the addition of attack arguments [7], [8],[9],[10]. In that case, the attack from argument a to argument b is represented by adding an attack meta-argument 'a attacks b' in between. This attack meta-argument itself attacks meta-argument $b_\in$, which represents that if the attack meta-argument is accepted, we cannot have that argument b is accepted. Moreover, if the attack meta-argument is not accepted, because it is itself attacked, then meta-argument $b_\in$ can be accepted, and thus argument b can be accepted. Now, how do we represent that if argument a (i.e. meta-argument $a_\in$) is not accepted, then we cannot have an attack from a to b either? We cannot represent this by an attack from meta-argument $a_\in$ to the attack meta-argument, but we represent it by an attack from the meta-argument $a_\notin$ to the attack meta-argument. This illustrates the essential role of *out* meta-arguments in the representation of second or higher order attacks.

2. An abstract theory of meta-argumentation

2.1. Abstract argumentation

We follow Baroni and Giacomin [11]. An underlying mechanism of argument generation defines a set of arguments, which is typically infinite, and which we call the universe of arguments and represent by $\mathcal{U}$. An acceptance function $\mathcal{E}$ is a function that associates with a set of arguments $\mathcal{A} \subseteq \mathcal{U}$, a set of arguments produced by a reasoner at a given instant of time, and a binary relation $\rightarrow \subseteq \mathcal{A} \times \mathcal{A}$, representing the dominance or attack relation among these arguments, the sets of acceptable arguments of $\mathcal{A}$.

Definition 1: Let $\mathcal{U}$ be the universe of arguments. An acceptance function $\mathcal{E} : 2^{\mathcal{U}} \times 2^{\mathcal{U} \times \mathcal{U}} \rightarrow 2^{2^{\mathcal{U}}}$ is a partial function which is defined for each argumentation framework $\langle \mathcal{A}, \rightarrow \rangle$ with finite $\mathcal{A} \subseteq \mathcal{U}$ and $\rightarrow \subseteq \mathcal{A} \times \mathcal{A}$, and which associates with argumentation framework $\langle \mathcal{A}, \rightarrow \rangle$ sets of subsets of $\mathcal{A}$: $\mathcal{E}(\langle \mathcal{A}, \rightarrow \rangle) \subseteq 2^{\mathcal{A}}$.

Baroni and Giacomin identify two fundamental principles underlying the definition of extension-based semantics in Dung's framework, the *language independent* principle and the *conflict free* principle. See [11] for a discussion on these principles.

The following definition summarizes the most widely used acceptability semantics, that satisfy these two principles. Which semantics is most appropriate depends on the application domain of the argumentation theory.

Definition 2 (Acceptability semantics): Let $AF = \langle \mathcal{A}, \rightarrow \rangle$ be an argumentation framework. Let $\mathcal{S} \subseteq \mathcal{A}$. $\mathcal{S}$ defends a if $\forall b \in \mathcal{A}$ such that $b \rightarrow a$, $\exists c \in \mathcal{S}$ such that $c \rightarrow b$. Let $D(\mathcal{S}) = \{a \mid \mathcal{S} \text{ defends } a\}$.

- $\mathcal{S} \in \mathcal{E}_{\text{admiss}}(AF)$ iff $cf(\mathcal{S})$ and $\mathcal{S} \subseteq D(\mathcal{S})$.
- $\mathcal{S} \in \mathcal{E}_{\text{compl}}(AF)$ iff $cf(\mathcal{S})$ and $\mathcal{S} = D(\mathcal{S})$.
- $\mathcal{S} \in \mathcal{E}_{\text{ground}}(AF)$ iff $\mathcal{S}$ is smallest in $\mathcal{E}_{\text{compl}}(AF)$.
- $\mathcal{S} \in \mathcal{E}_{\text{pref}}(AF)$ iff $\mathcal{S}$ is maximal in $\mathcal{E}_{\text{admiss}}(AF)$.
- $\mathcal{S} \in \mathcal{E}_{\text{skep-pref}}(AF)$ iff $\mathcal{S} = \cap \mathcal{E}_{\text{pref}}(AF)$.
- $\mathcal{S} \in \mathcal{E}_{\text{stable}}(AF)$ iff $cf(\mathcal{S})$ and $\forall b \in \mathcal{A} \backslash \mathcal{S} \exists a \in \mathcal{S} : a \rightarrow b$.

Jacobovits - Vermeir [12] and Caminada [13] show that these semantics can also be described by a three valued argument labeling, where the first two conditions represent conflict free and defense, and the third one represents the so-called reinstatement principle.

Proposition 1: Let $L^{\langle \mathcal{A}, \rightarrow \rangle}_{\text{admiss}} : \mathcal{A} \rightarrow \{\mathit{in}, \mathit{out}, \text{undecided}\}$ be a complete labeling function for semantics satisfying the conflict free principle such that:

1) $L^{\langle \mathcal{A}, \rightarrow \rangle}_{\text{admiss}}(b) = \mathit{out}$ iff $\exists a \rightarrow b : L^{\langle \mathcal{A}, \rightarrow \rangle}_{\text{admiss}}(a) = \mathit{in}$
2) if $L^{\langle \mathcal{A}, \rightarrow \rangle}_{\text{admiss}}(b) = \mathit{in}$ then $\forall a \rightarrow b : L^{\langle \mathcal{A}, \rightarrow \rangle}_{\text{admiss}}(a) = \mathit{out}$

and let $L^{\langle \mathcal{A}, \rightarrow \rangle}_{\text{compl}} : \mathcal{A} \rightarrow \{\mathit{in}, \mathit{out}, \text{undecided}\}$ be a complete labeling function such that in addition:

3) if $\forall a \rightarrow b : L^{\langle \mathcal{A}, \rightarrow \rangle}_{\text{compl}}(a) = \mathit{out}$ then $L^{\langle \mathcal{A}, \rightarrow \rangle}_{\text{compl}}(b) = \mathit{in}$.

Then we have the following results.
$\mathcal{E}_{\text{admiss}}(AF) = \{\{a \mid L^{\langle \mathcal{A}, \rightarrow \rangle}_{\text{admiss}}(a) = \mathit{in}\} \mid \exists L^{\langle \mathcal{A}, \rightarrow \rangle}_{\text{admiss}}\}$
$\mathcal{E}_{\text{compl}}(AF) = \{\{a \mid L^{\langle \mathcal{A}, \rightarrow \rangle}_{\text{compl}}(a) = \mathit{in}\} \mid \exists L^{\langle \mathcal{A}, \rightarrow \rangle}_{\text{compl}}\}$
They show how the other semantics can be defined in terms of these labelings too. Actually, our way of defining admissible labellings is mix of the work of Jacobovits & Vermeir and Caminada's work. The differences:

- In the work of Jacobovits & Vermeir, there are only two labels: in ("+") and out ("-"), but an argument can get 0, 1 or 2 labels. This basically destroys the correspondence with Dung's extension based approach.
- In Caminada's work [13] a weaker condition is used to describe admissible labellings (condition 1 is "if ... then ..." instead of "iff"). The stronger condition in our work causes each admissible set to have exactly one admissible labelling, whereas in Caminada's description, an admissible set can have one or more associated admissible labellings (so there's a 1-n relation instead of a 1-1 relation like in our work).

Clearly, one of the advantages is the fact that there is now an 1-1 relation with admissible sets. The disadvantage is that the stronger condition is perhaps too strong for some purposes, and that for instance the discussion game for preferred semantics does yield "weakly" (Caminada version) but not "strongly" (our version) admissible labellings. Also the preferred/stable/semi-stable algorithm of [13] yields labellings that are weakly admissible but not strongly admissible, so they could not directly be applied to our version of admissible labellings.

2.2. Representation theory

Our theory of meta-argumentation can be seen as a special kind of representation, namely as representation by an argumentation theory where the set of arguments is extended. It is based on the following three steps.

1) Extend the set of arguments with auxiliary arguments; we call the extended set 'meta-arguments'.
2) Calculate the extensions of the extended theory using one of Dung's semantics; we call them 'meta-extensions'.
3) For each meta-extension, filter out the auxiliary arguments; the resulting sets of arguments are the extensions of the theory.

The three-step process terminates by filtering out auxiliary arguments and producing a set of extensions. The difference with respect to the set of extensions one would produce without using the three-step process involving meta-arguments is that we can represent more complex structures formerly introduced in the literature as extensions of Dung's framework. To define a particular meta-argumentation framework, we have to define how the set of meta-arguments is generated from the set of atomic arguments, and which conditions argumentation frameworks have to satisfy, i.e., for which argumentation frameworks the acceptance function is defined.

3. Toulmin scheme in abstract argumentation

3.1. Toulmin scheme

Loui [14] finds that Toulmin is ninth in total number of citations for philosophers of science and logic between 1988 and 2004. He concludes that after paradigm shifts and methods (Kuhn, Lakatos, Feyerabend), fuzzy logic (Zadeh), illocutionary force (Austin), the analytic-synthetic distinction (Quine), supervenience (Putnam), deductive-nomological explanation (Hempel), Toulmin's scheme must be mentioned next, before, for example, Carnap, Church, Tarski and Russell-Whitehead. Hitchcock and Verhey [15] explain Toulmin's scheme in Figure 1 as follows.

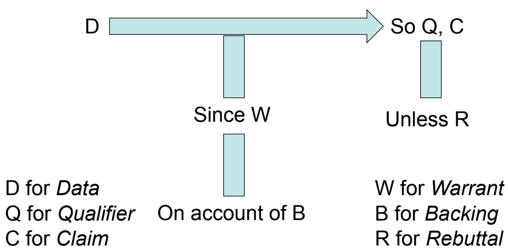

Fig. 1. Toulmin scheme [Toulmin, 1958, p.108].

"First we assert something, and thus make a *claim*. Challenged to defend out claim by a questioner who asks, "What have you got to go on?", we appeal to the relevant facts at our disposal, which Toulmin calls our *data* (D). It may turn out to be necessary to establish the correctness of these facts by a preliminary argument. But their acceptance by the challenger, whether immediate or indirect, does not necessarily end the defense. For the challenger may ask about the bearing of our data on our claim: "How did you get there?" Our response will be at our most perspicuous take the form: "Data such as D entitle one to draw conclusions, or make claims, such as C." [6, p.98]. A proposition of this form Toulmin calls a *warrant* (W). Warrants, he notes, confer different degrees of force on the conclusions that justify, which may be signaled by qualifying our conclusion with a *qualifier* (Q) such as "necessarily", "probably" or "presumably". In the latter case, we may need to mention conditions of *rebuttal* (R) "indicating circumstances in which the authority of the warrant would have to be set aside" [6, p.110]. Our task, however, is still not necessarily finished. For our challenger may question the general acceptability of our warrant: "Why do you think that?" Toulmin calls our answer to this question our *backing* (B)." He emphasizes the great differences in kind between backings in different fields. Warrants can be defended by appeal to a system of taxonomic classification, to a statute, to statistics from census, and so forth. It is this difference in backing that constitutes the field-dependence of our standards of argument. Ultimately, all micro-arguments depend of the combination of data and backing. In rare cases, checking the backing will involve checking the claim; Toulmin calls such arguments "analytic arguments". Most arguments are not of this sort, so that purely formal criteria do not suffice for their assessment; Toulmin calls them "substantial arguments". The sort of backing that is acceptable for a given substantial argument will depend on the field to which it belongs."

There are various challenges to formalize Toulmin's field-independent scheme. First, the scheme, like Dung's argumentation theory, differs radically from the traditional logical analysis of a micro-argument into premises and conclusion, and it was precisely this difference that has made the scheme so popular and influential. However, Toulmin's argument against formalization of his scheme can be countered by the argument that over the past five decades, many new kinds of formalisms have been developed. The second challenge is that there are great differences between the kind of backings in different fields, as emphasized by Toulmin, and thus backing B is abstract like arguments in Dung's theory. The third challenge, which is the challenge of this paper, is to represent the defense of C by D. Extensions of Dung's theory with a binary support relation among arguments [4] do not allow for the support itself to be attacked [3], which is the core of Toulmin's scheme.

3.2. Jakobovits - Vermeir - Caminada labeling

Figure 2 visualizes our representation of Toulmin's scheme in abstract argumentation. Each square is a meta-argument, stating that the argument inside the square is *in*, and each circle is a meta-argument stating that the argument written inside it is *out* (neither is undecided). The qualifier Q is not represented, and rebuttal is represented by an optional counterargument R to C. If we have $D_{\in}$ and $B_{\in}$ then we have $W_{\in}$ and accordingly $C_{\in}$ for any of Dung's argumentation semantics. If we don't have $B_{\in}$, then we don't have $W_{\in}$, and consequently we don't have $C_{\in}$. In the bottom left corner of the figure, we suggest a more convenient visualization. We add $C_{\notin}$ and $R_{\notin}$ for symmetry and to combine micro-arguments, but for a single micro-argument we could have left them out.

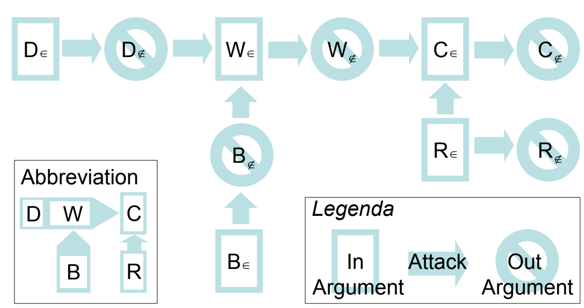

Fig. 2. Toulmin scheme in abstract argumentation.

One may wonder whether there are other representations of Toulmin's scheme in our meta argumentation framework. For example, at first sight it may seem that if there is an attack from $D_{\in}$ to $D_{\notin}$, then there might also be an attack the other way around. However, this would not represent the defense of

C by D, but a conditional defense: if D would be acceptable, then C would be acceptable too. However, we do not claim that our representation is the only one which can be used, and a more systematic exploration of the kind of schemes which can be represented in our meta-argumentation theory is a topic of further investigation.

Figure 3 visualizes some examples of combining micro-arguments into larger arguments [16]. This is one contribution of Dung's theory since it talks about larger networks, and offers the distinction between various kinds of semantics to deal with cycles, and it is well known that without cycles all Dung's semantics coincide. An example of cycle is provided on the bottom left of Figure 3. For example, there can be attacks on data D too as in the example provided on the bottom right of Figure 3 where the two data D and E attack each other.

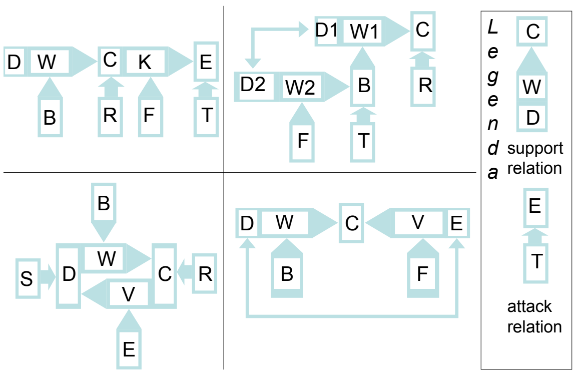

Fig. 3. Combining micro-arguments.

The generation of meta-arguments and the condition on meta-argumentation frameworks are thus very simple, and formalized as follows.

Definition 3: Let A_0 be a set of atomic arguments. Let the universe of arguments U of a Toulmin argumentation framework be the minimal set of arguments such that if a in A_0, then $a_\in$ and $a_\notin$ in U. A Toulmin argumentation framework is an argumentation framework $\langle \mathcal{A}, \rightarrow \rangle$, where $a_\in$ in $\mathcal{A}$ iff $a_\notin$ in $\mathcal{A}$, and if $a_\in$ in $\mathcal{A}$, then $a_\in \rightarrow a_\notin$, and this is the only attack on $a_\notin$.

Toulmin does not consider examples with cycles, so the formalization of his examples is straightforward (and all semantics coincide). A topic for further exploration concerns the effect of adding the meta-arguments is: how are the original semantics related to the semantics of the extended framework? How to define the influence from meta-arguments to original arguments, e.g. argument b should be out if meta-argument $b_\notin$ is in?

4. Concluding remarks

We show how Dung's abstract argumentation theory can reason about itself, using the Jakobovits - Vermeir - Caminada labeling to say whether an argument is acceptable or not; we call it a theory of meta-argumentation. In particular, we show how to represent the Toulmin scheme from informal argumentation by representing the defense of a claim by its data as a warrant argument, which can be attacked, and defended itself by a backing argument. Whereas many generalizations of Dung's abstract argumentation theory are now being proposed, leading to a fragmentation of the area, our aim in this paper is to represent such extensions as instances of Dung's theory. A further validation of our meta-argumentation approach is therefore the representation in our theory of other proposed extensions as well as existing instances of Dung's theory, such as his assumption based argumentation. A particular promising topic of further research is the use of collective meta-arguments. For example, whereas for Toulmin a claim C is defended by data D, for Dung an argument is defended by a *set of* arguments, such that we need collective arguments to represent meta-arguments about Dung's notion of defense. Moreover, whereas in the approach advocated in this paper each extension of meta-arguments corresponds to an extension of arguments, with collective meta-arguments we can represent multiple extensions into a single meta-extenson.

References

[1] L. Amgoud and C. Cayrol, "Inferring from inconsistency in preference-based argumentation frameworks," *J. Autom. Reasoning*, vol. 29, no. 2, pp. 125–169, 2002.

[2] S. Modgil, "An abstract theory of argumentation that accommodates defeasible reasoning about preferences," in *ECSQARU*, 2007, pp. 648–659.

[3] H. Barringer, D. M. Gabbay, and J. Woods, "Temporal dynamics of support and attack networks: From argumentation to zoology," *Mechanizing Mathematical Reasoning*, pp. 59–98, 2005.

[4] C. Cayrol and M.-C. Lagasquie-Schiex, "On the acceptability of arguments in bipolar argumentation frameworks," in *ECSQARU*, 2005, pp. 378–389.

[5] A. Bondarenko, P. M. Dung, R. A. Kowalski, and F. Toni, "An abstract, argumentation-theoretic approach to default reasoning," *Artif. Intell.*, vol. 93, pp. 63–101, 1997.

[6] S. Toulmin, *The Uses of Argument*. Cambridge University Press, 1958.

[7] G. Boella, L. van der Torre, and S. Villata, "Attack relations among dynamic coalitions," in *BNAIC 2008*.

[8] ——, "Social viewpoints for arguing about coalitions," in *PRIMA*, ser. Lecture Notes in Computer Science, T. D. Bui, T. V. Ho, and Q.-T. Ha, Eds., vol. 5357. Springer, 2008, pp. 66–77.

[9] S. Modgil, "Reasoning about preferences in argumentation frameworks," *Artif. Intell.*, vol. 173, no. 9-10, pp. 901–934, 2009.

[10] P. Baroni, F. Cerutti, M. Giacomin, and G. Guida, "Encompassing attacks to attacks in abstract argumentation frameworks," in *ECSQARU*, ser. Lecture Notes in Computer Science, C. Sossai and G. Chemello, Eds., vol. 5590. Springer, 2009, pp. 83–94.

[11] P. Baroni and M. Giacomin, "On principle-based evaluation of extension-based argumentation semantics," *Artif. Intell.*, vol. 171, no. 10-15, pp. 675–700, 2007.

[12] H. Jakobovits and D. Vermeir, "Robust semantics for argumentation frameworks," *J. Log. Comput.*, vol. 9, no. 2, pp. 215–261, 1999.

[13] M. Caminada, "On the issue of reinstatement in argumentation," in *JELIA*, 2006, pp. 111–123.

[14] R. Loui, "A citation based reflection on Toulmin and argument," in *Arguing On The Toulmin Model: New Essays In Argument Analysis And Evaluation*, D. Hitchcock and B. Verheij, Eds., 2007, pp. 31–38.

[15] D. Hitchcock and B. Verheij, "Introduction," in *Arguing On The Toulmin Model: New Essays In Argument Analysis And Evaluation*, 2007, pp. 1–23.

[16] T. J. Bench-Capon, "Specification and implementation of Toulmin dialogue game," in *JURIX*, 1998, pp. 5–20.