Picking the Right Expert to Make a Debate Uncontroversial

Dionysios KONTARINIS [a,1], Elise BONZON [a,1], Nicolas MAUDET [b,2] and Pavlos MORAITIS [a,1]

[a] *Paris Descartes University, LIPADE, Paris, France*

[b] *UPMC, LIP6, Paris, France*

Abstract. Agents contributing to (online) debate systems often have different areas of expertise. This must be considered if we want to define a decision making process based on the output of such a system. Distinguishing agents on the basis of their areas of expertise also opens an interesting perspective: when a debate is deemed "controversial", calling an additional expert may be a natural way to make the decision easier. We introduce possible definitions that capture these notions and we provide a preliminary analysis with the objective to help a designer find the "right" expert.

Keywords. Argumentation, Multi-agent systems, Online debate systems

1. Introduction

One of the greatest difficulties in group decision-making is to achieve an agreement accepted by all the participating agents. In particular, agents may be reluctant to accept a decision if they have the feeling that the decision could have "easily" been different.

In recent years, there has been a considerable advance in the area of on-line, multiparty argumentation [18]. These works aim at building systems which let users engage in an asynchronous multiparty discussion. An example of application domain is the construction of more interactive forums on the Web like *DebateGraph*[3]. Some of these systems just provide a way to represent arguments, attacks and information about them, while others (like the Parmenides system [6]) include a reasoning machinery, usually from argumentation theory, which provides a formal way to decide on the acceptability of statements (arguments). Abstract argumentation is commonly used in this type of systems, as it is relatively easy to represent arguments in the form of abstract entities, or at least in the form of entities containing only some lines of text, or hyperlinks to resources on the Web [17]. These systems raise a number of challenges [19]. For instance, one of their characteristic features is that they allow users to also *vote* on arguments or attacks, in order to express their agreement or disagreement with respect to some information that was previously put forward in the debate. Reasons for disagreement may

[1] E-mail: dionysios.kontarinis@parisdescartes.fr; {elise.bonzon, pavlos}@mi.parisdescartes.fr

[2] E-mail: nicolas.maudet@lip6.fr

[3] http://www.debategraph.org

be the different underlying assumptions or preferences of the agents [1], or the different interpretations of the content of arguments. In this case, it is crucial to find a way to decide on the attacks which are to be considered, and voting is a natural way to do it. In this paper we suppose that only voting on attacks is permitted. Another challenging aspect is that agents may not be treated similarly, because some of them may be experts in some topics of the discussion. Now, once the debate is over, participating agents may not be entirely satisfied by the procedure's final outcome. First, of course, they may not be satisfied with the outcome itself [15,5]. In this paper we concentrate on a different issue though, namely the fact that the obtained result may in some way be *controversial*. In particular, in our context, this may result from two (distinct, but related) situations: (a) *argumentative controversy*: when argumentation theory does not provide a clean-cut decision: this is the case in particular when several (conflicting) acceptable outcomes are returned; (b) *voting controversy*: when voting does not offer a clear majority to support the fact that an attack should be taken into account (or not).

Our objective in this work is to set up a framework where these issues can be formally studied. Once a debate is obtained, we discuss how the choice of an additional expert should be made in order to make the result less controversial. We emphasize that our work is *not* dependent on a specific protocol. For what matters, the resulting debate may be the outcome of a multilateral protocol like the one proposed in [3], or of a merging process [7,8]. Instead we study how the different expertise of agents should be modeled, and how the additional expert may affect the current debate.

The fact that agents may be treated differently is also present in the work of [2] where a notion of trust is attached to agents, whereas we focus on the notion of expertise of the agents. The resulting object we deal with is a Weighted Argumentation System (WAS), as defined in [12,7]. It allows us to conveniently quantify the impact of the experts' opinions and aggregate them on a common WAS. Once this is done, we consider the non-weighted counterpart argumentation system and draw conclusions according to classical Dung's semantics. Some properties studied in [12] are connected to our work, they investigate in particular how a given argumentation framework (and its outcome) may be affected by removing attacks up to a fixed "inconsistency budget". Similar notions of dynamics are proposed here, but we study removal and addition of attacks that are insufficiently supported by the votes. Approaches not relying on Dung's semantics are also possible: in particular, the recent work of [14] builds a framework based on a different semantics to account for systems with arguments and votes. The issue of argumentative controversy has also connections with the work of [16]: they study how different subjective views on the same debate may be reconciled (or not) by means of voting, whereas we suppose that an additional agent can be called for, in the hope to make the debate "less controversial".

The remainder of this paper is as follows. In Section 2 we set up the basics of our framework, showing in particular how the notion of expertise can be taken into account. This allows to model a debate resulting from several experts. We next define how the conclusions should be drawn from the resulting (weighted) argumentation system (Section 3). Then comes the description of the actual procedure (Section 4): in the first phase experts put forward arguments and attacks and they vote on the attacks. The resulting debate is then analyzed: attacks are classified in three classes, depending on the support they obtained throughout the voting process. Finally, we discuss and investigate in Section 5 how an additional expert should be chosen with the aim of making the debate less controversial. Section 6 concludes.

2. Arguments, topics and expertise

As mentioned in the introduction, this paper deals with systems where arguments are put forward in a debate by users. We follow Dung [11] and define an argumentation system as a (finite) set of arguments together with the different (binary) conflicts among them.

Def. 1 *An* **argumentation system (AS)** *is a pair $\langle A, \mathcal{R} \rangle$ of a set A of arguments and a binary relation $\mathcal{R}$ on A called the* **attack relation**. *$\forall a,b \in A$, $a\mathcal{R}b$ (or $(a,b) \in \mathcal{R}$) means that a* **attacks** *b. An AS may be represented by a directed graph, called the* **argumentation graph**, *whose nodes are arguments and edges represent the attack relation.*

Observe that in this definition the structure of arguments is unspecified. In our context, there is no semantical analysis of the content of arguments, but nevertheless users can *tag* arguments with keywords specifying which topics the argument is about. It is common practice in such systems [19]. In this paper we assume that the set of potential *topics*, denoted T, is known and fixed a priori by the system. Attached to each argument is a set of topics which, in principle, can be empty or contain as many topics as wished.

Def. 2 *Let T be the set of topics. The set of* **topics of an argument** *$a \in A$ is given by function $top(a) \subseteq T$.*

Ex. 1 *Consider a debate system designed to support discussion among PC members about papers to accept or reject for a conference. There is a list of keywords that PC members can choose to indicate their area of expertise,* e.g. *T = {comp, kr, ml, cog}, where comp stands for "complexity", kr stands for "knowledge representation", ml stands for "machine learning", and cog stands for "cognitive science". A first reviewer ($PC1$) argues that the paper is good because it presents an interesting representation formalism, very elegant, and very much plausible from the cognitive point of view (argument* a, *with $top(a) = \{kr, cog\}$). A second reviewer ($PC2$) challenges this on the basis that the formalism is too expressive, so the reasoning tasks would be intractable (argument* b, *with $top(b) = \{comp\}$), and that the formalism contains some imperfections that should be worked on before publication (argument* d, *with $top(d) = \{kr\}$). A third reviewer ($PC3$) challenges argument* b *by saying that he is aware of related formalisms and problems in machine learning for which very good approximation algorithms work in practice, so it is not unlikely that the same could happen with this one (argument* c, *with $top(c) = \{comp, ml\}$).*

In the following, we will denote by $R = A \times A$ the set of *potential attacks*[4]. Now, from topics attached to arguments, we deduce how topics are attached to potential attacks.

Def. 3 *Let T be the set of topics. The set of* **topics of a (potential) attack** *$(a,b) \in R$ is given by the function*[5] *$top(a,b) = top(a) \uplus top(b) \subseteq T \uplus T$.*

As attacks are binary, this is simply a multiset where topics appearing in the attacking and attacked arguments appear twice. For an attack, it is thus possible to distinguish three levels of "relevance" for topics: *prominent* topics (attached to both arguments) denoted

[4] For simplicity reasons, we will just call them *attacks* in the remainder of the paper.

[5] $\uplus$ indicates the multiset union

$prom(a,b) \subseteq T$, *relevant* topics (attached to one argument) denoted $rel(a,b) \subseteq T$, and *irrelevant* topics (not attached to either argument) denoted $irr(a,b) \subseteq T$.

Ex. 1, cont. *We have* $top(c,b) = \{comp, comp, ml\}$*, thus* $prom(c,b) = \{comp\}$*,* $rel(c,b) = \{ml\}$*, and* $irr(c,b) = \{cog, kr\}$*. We also have* $top(b,a) = \{comp, kr, cog\}$*, so* $prom(b,a) = \{\}$*,* $rel(b,a) = \{comp, kr, cog\}$*, and* $irr(b,a) = \{ml\}$*.*

When the agents express some opinion regarding an attack (in our context by voting, or initially stating the attack), a weight will be attached to that vote. Note that weights are not assigned to agents but to pairs agents-attacks depending on their expertise.

Def. 4 *The* **expertise of agent** i *is given by a function* $exp(i) \subseteq T$*.*

Experts express their opinions on attacks by casting positive or negative votes.

Def. 5 *A* **vote** *is a tuple* $\langle (a,b), s, i \rangle$ *where* $(a,b) \in A \times A$ *is the attack concerned by the vote,* $s \in \{-1, +1\}$ *is the polarity (sign) of the vote, and i is the voter.*

The impact of the vote of an expert *i* for or against an attack depends on her expertise over the topics of this attack. Intuitively, the opinion of an expert on the topics of an attack should have more importance than the opinion of a non-expert on the same attack. However, this general principle needs to be made much more precise, as illustrated in the next example.

Ex. 1, cont. *Suppose reviewers have to choose two keywords from a list. PC1 has expertise in* $\{kr, cog\}$*, PC2 has expertise in* $\{kr, comp\}$*, and PC3 has expertise in* $\{comp, ml\}$*. We focus on attack* (c,b)*. As PC3 put forward argument* c *and attack* (c,b)*, she voted by default positively on the attack. PC1 voted against this attack because she thinks it is not valid, while PC2 voted in favour of it. As reviewers have different topics of expertise, we can summarize their votes, as well as their expertise in the prominent, relevant and irrelevant topics, by means of vectors as follows:*

$$\begin{array}{l}
\textit{Vote of PC3 on } (c,b) : \langle \langle comp : +1 \rangle, \langle ml : +1 \rangle, \langle kr : 0, cog : 0 \rangle \rangle \\
\textit{Vote of PC2 on } (c,b) : \langle \langle comp : +1 \rangle, \langle ml : 0 \rangle, \quad \langle kr : +1, cog : 0 \rangle \rangle \\
\textit{Vote of PC1 on } (c,b) : \langle \langle comp : 0 \rangle, \quad \langle ml : 0 \rangle, \quad \langle kr : -1, cog : -1 \rangle \rangle
\end{array}$$

How should we aggregate these different votes? The question is difficult because we need to aggregate different topics within one vote, but also different votes. There are key assumptions that need to be made explicit: (i) the *independence of expertise* (Suppose an expert in *kr* votes for (b,a), then another expert in *comp* votes the same way. Does it have the same impact as one expert in both topics voting once?); and (ii) *compensation among topics* (Should we allow compensation among levels of topics? For instance, should two votes on a relevant topic be as important as one vote on a prominent one? Should irrelevant topics be considered in the first place?). There are many ways to aggregate the votes. Some interesting ideas can be found in [10]. In this paper we make the following simple choices: we suppose that independence of expertise holds, we allow compensation among topics (considering prominent topics to be twice as important as relevant topics) and we disregard votes on irrelevant topics.

Following this discussion, we propose the following definition of impact, which has two advantages. First, the more topics of an attack an agent is expert in, the greater her impact is on the attack. Second, expertise in prominent topics of an attack leads to a greater impact than expertise in its relevant topics.

Def. 6 *Let i be an agent. The* **impact of** i **on** $(a,b) \in A \times A$ *is denoted* $imp_i(a,b)$ *and defined by* $imp_i(a,b) = 2 \times |exp(i) \cap prom(a,b)| + |exp(i) \cap rel(a,b)|$. *If* $imp_i(a,b) = 0$ *we say that i is a* **dummy voter** *on* (a,b).

Ex. 1, cont. *Since* $exp(PC2) = \{kr, comp\}$ *and* $top(c,b) = \{comp, comp, ml\}$*, it holds that* $imp_{PC2}(c,b) = 2 \times 1 + 0 = 2$*, as PC2 is expert in one prominent topic of the attack (and in no relevant topics).*

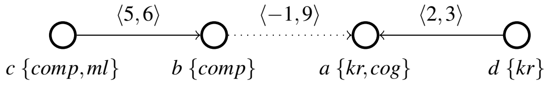
Def. 7 *The* **evaluation vector of** $(a,b) \in A \times A$ *is denoted* $v(a,b) = \langle w(a,b), mw(a,b) \rangle$*, where* $w(a,b)$ *(called the* **weight** *of* (a,b)*) is the aggregated impact of all the experts who have voted for or against* (a,b)*, and* $mw(a,b)$ *(called the* **max-weight** *of* (a,b)*) is the aggregated impact of all the voters on* (a,b)*, assuming they had all voted in favour of it.*

Before considering any votes on (a,b), we have, $\forall (a,b) \in A \times A$, $v(a,b) = \langle 0,0 \rangle$. We can now iteratively define how the evaluation vector $v(a,b) = \langle w(a,b), mw(a,b) \rangle$ is updated after a vote $\langle (c,d), s, i \rangle$ of an expert i:

$$upd(v(a,b), \langle (c,d), s, i \rangle) = \begin{cases} v(a,b), \text{ if } (a,b) \neq (c,d) \text{ or } imp_i(a,b) = 0 \\ \langle w(a,b) + (s \times imp_i(a,b)), mw(a,b) + |top(a,b)| \rangle \text{ otherwise} \end{cases}$$

A vote on attack (a,b) can only change the evaluation vector of this specific attack, and only if the voter has some relevant expertise on its topics. The weight $w(a,b)$ aggregates the positive and negative votes on (a,b) by using a sum, though other possibilities exist. Finally, the value of $mw(a,b)$ is always equal to the product of the number of (non-dummy) voters on (a,b) and the cardinality of $top(a,b)$.

Ex. 1, cont. *The evaluation vector of* (c,b) *after the vote of PC3 is* $v(c,b) = \langle 3,3 \rangle$*. It becomes* $v(c,b) = \langle 3+2, 3+3 \rangle = \langle 5,6 \rangle$ *after the vote of PC2, and remains unchanged after the vote of PC1 who has no expertise in the topics of this attack (and thus is a dummy voter on* (c,b)*). Now, if we assume that PC2 votes for the attack* (b,a) *whereas PC1 and PC3 vote against it; and that PC2 is the only expert who expresses her opinion on* (d,a)*, we obtain the following AS, with* $v(b,a) = \langle -1,9 \rangle$*,* $v(c,b) = \langle 5,6 \rangle$ *and* $v(d,a) = \langle 2,3 \rangle$*. Note that the dotted edge denotes an attack with negative weight.*



3. Reasoning with weighted argumentation systems

Once the experts have expressed their points of view on a subset of attack relations, we obtain an aggregated argumentation system with weighted attacks.

Def. 8 *A* **weighted argumentation system (WAS)** *is a triplet* $\mathbb{W} = \langle A, R, v \rangle$ *where A is a set of arguments,* $R = A \times A$ *is the set of potential attacks between pairs of arguments, and* $v \colon R \rightarrow \langle \mathbb{Z}, \mathbb{N} \rangle$ *is a function which returns the evaluation vector of each attack in R.*

Let a WAS $\mathbb{W} = \langle A, R, v \rangle$ and $(a,b) \in R$. Let $v(a,b) = \langle w(a,b), mw(a,b) \rangle$. If $w(a,b) > 0$, we say that the attack (a,b) holds, whereas if $w(a,b) \leq 0$, we say that the attack (a,b) does not hold.

Def. 9 *Given a WAS $\mathbb{W} = \langle A, R, v \rangle$ and an agent i, a* **move by agent** i **on** $\mathbb{W}$ *is a vector of votes $m = \langle \langle (a_1,b_1), s_1, i \rangle, \ldots, \langle (a_n,b_n), s_n, i \rangle \rangle$ such that $\forall k,l \in \{1,\ldots,n\}$, if $k \neq l$, it holds that $(a_k,b_k) \neq (a_l,b_l)$. When the move m is played on $\mathbb{W}$ we obtain a modified WAS denoted $\mathbb{W} \oplus m = \mathbb{W}' = \langle A, R, v' \rangle$ where $\forall k \in \{1,\ldots,n\}$ it holds that $v'(a_k,b_k) = upd(v(a_k,b_k), \langle (a_k,b_k), s_k, i \rangle)$.*

In order to use acceptability semantics of abstract argumentation [11], we need to define the notion of *non-weighted counterpart AS of a WAS*. To do so, we simply chose to remove all attacks with non-positive weights. We note that more sophisticated alternatives could have been used, based for example on the notion of “inconsistency budget” of [12], which could give rise to new ways to define extensions [9].

Def. 10 *Given a WAS $\mathbb{W} = \langle A, R, v \rangle$, we define its* **non-weighted counterpart argumentation system** $\mathbb{W}_{cp} = \langle A_{cp}, R_{cp} \rangle$ *as follows: $A_{cp} = A$ and $(a,b) \in R_{cp}$ iff $w(a,b) > 0$.*

The different concepts of admissibility are originally stated in terms of sets of arguments (see [11]). However, Jakobovits and Vermeir [13] and later Caminada [4] have shown that we can express these concepts using *argument labelling*. This labelling specifies the accepted arguments (labelled IN), the rejected ones (labelled OUT), and the ones' whose status cannot be decided (labelled UND). More specifically, a *reinstatement labelling* satisfies the condition that an argument is IN, if and only if all of its attackers are OUT; and that an argument is OUT, if and only if at least one of its attackers is IN. Otherwise the label is UND. Given a WAS $\mathbb{W}$, in order to find its reinstatement labelling, we must first consider its non-weighted counterpart AS $\mathbb{W}_{cp}$. We denote its labelling by $\mathbb{L}^{\mathbb{W}}$. The label of a specific argument $a \in A$ will be denoted by $\mathbb{L}^{\mathbb{W}}(a)$. The set of arguments labelled IN (resp. OUT, UND) will be denoted $\mathbb{L}^{\mathbb{W}}_{IN}$ (resp. $\mathbb{L}^{\mathbb{W}}_{OUT}$, $\mathbb{L}^{\mathbb{W}}_{UND}$).

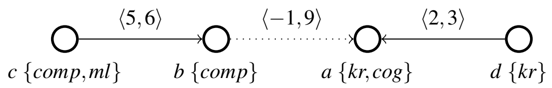
4. The process: collecting arguments, evaluating the system, and selecting an expert

The procedure we propose here proceeds in three phases. The first phase consists in the aggregation of the different opinions of the agents, and allows to obtain an aggregated WAS. Recall that we do not commit to any specific protocol here. Then comes an evaluation phase which allows to determine how controversial the aggregated WAS is. If required, this second phase leads to a third phase where we chose an expert to make the aggregated WAS less controversial.

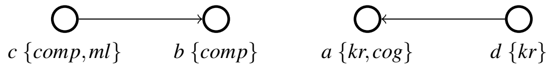
4.1. Phase 1: Experts express their opinions

In this first phase, the agents place their arguments on the board, and express their opinions by voting on the attack relations. This phase has been described in previous sections.

Ex. 1, cont. *The WAS $\mathbb{W}$ obtained after the vote of the three PC members is the following:*



The counterpart AS will be as follows, with $\mathbb{L}^{\mathbb{W}}_{IN} = \{c,d\}$, $\mathbb{L}^{\mathbb{W}}_{OUT} = \{a,b\}$ *and* $\mathbb{L}^{\mathbb{W}}_{UND} = \{\}$.



4.2. Phase 2: Evaluation of the aggregated WAS

Once the agents have expressed their opinion, we obtain an aggregated WAS. The question is then if this WAS is controversial, and to what extent. Several criteria can assess how controversial a WAS is. In the following, we focus on two such criteria: the first one is the *stability of attacks*. An attack relation can be seen as *stable* if it is difficult to question it, more specifically, if a single expert cannot change its sign. The second one is the *persistence of arguments' labels*. An argument can be seen as *persistent* if its label does not depend on unstable (controversial) attacks, that is if a single expert cannot change its label if she changes some signs of these attacks.

For the first criterion, we consider a natural qualitative scale and introduce three types of attacks. The *beyond any doubt* attacks ($R^{\mathbb{W}}_{bd}$) are the ones on which sufficiently many agents agree (or, as a particular case, no agent has stated them). The *strong* attacks ($R^{\mathbb{W}}_{str}$) are defined as the attacks which are not beyond any doubt, but such that a single expert cannot change the sign of their weights. Finally, the *weak* attacks ($R^{\mathbb{W}}_{wk}$) are neither beyond any doubt nor strong, thus they are the most controversial ones.

Def. 11 *Let* $\mathbb{W} = \langle A,R,v\rangle$ *be a WAS. Recall that* $\forall(a,b) \in R$, $v(a,b) = \langle w(a,b), mw(a,b)\rangle$. *The set of attacks R is partitioned into the three following sets:*

- *An attack* (a,b) *is beyond any doubt if either* $v(a,b) = \langle 0,0\rangle$, *or if its weight is "very significant". The latter is true when two conditions hold: The number of voters on* (a,b) *is greater than a threshold* $\delta \in \mathbb{N}$, *and* $\frac{|w(a,b)|}{mw(a,b)}$ *is greater than a threshold* $\varepsilon \in]0,1[$. *Thus, the set of* **beyond any doubt attacks** *is defined as:* $R^{\mathbb{W}}_{bd} = \{(a,b) \in R \mid \text{either } v(a,b) = \langle 0,0\rangle, \text{ or } \frac{mw(a,b)}{|top(a,b)|} > \delta \text{ and } \frac{|w(a,b)|}{mw(a,b)} > \varepsilon\}$
- *The set of* **strong attacks** *is defined as:* $R^{\mathbb{W}}_{str} = R^{\mathbb{W},+}_{str} \cup R^{\mathbb{W},-}_{str}$, *where* $R^{\mathbb{W},+}_{str} = \{(a,b) \in R \mid (a,b) \notin R^{\mathbb{W}}_{bd}, w(a,b) > 0 \text{ and } w(a,b) - |top(a,b)| > 0\}$ *and* $R^{\mathbb{W},-}_{str} = \{(a,b) \in R \mid (a,b) \notin R^{\mathbb{W}}_{bd}, w(a,b) \leq 0 \text{ and } |w(a,b)| - |top(a,b)| \geq 0\}$
- *The set of* **weak attacks** *is defined as:* $R^{\mathbb{W}}_{wk} = R^{\mathbb{W},+}_{wk} \cup R^{\mathbb{W},-}_{wk}$, *where* $R^{\mathbb{W},+}_{wk} = \{(a,b) \in R \mid (a,b) \notin R^{\mathbb{W}}_{bd}, w(a,b) > 0 \text{ and } w(a,b) - |top(a,b)| \leq 0\}$ *and* $R^{\mathbb{W},-}_{wk} = \{(a,b) \in R \mid (a,b) \notin R^{\mathbb{W}}_{bd}, w(a,b) \leq 0 \text{ and } |w(a,b)| - |top(a,b)| < 0\}$

Note that the designer can set the values of δ and ε as wished for a given debate. Intuitively, the number of non-dummy voters on a beyond any doubt attack is greater than the value of δ, and ε is the threshold used to assess how unanimous the votes have been.

Ex. 1, cont. *If we consider that* $\delta = 4$ *and* $\varepsilon = 0.5$*, we have* $R^{\mathbb{W}}_{str} = \{(c,b)\}$*,* $R^{\mathbb{W}}_{wk} = \{(b,a),(d,a)\}$*, and* $R^{\mathbb{W}}_{bd} = R \setminus (R^{\mathbb{W}}_{str} \cup R^{\mathbb{W}}_{wk})$*.*

The set of stable attacks are the attacks which cannot be changed easily, that is the strong and beyond any doubt attacks.

Def. 12 *Let* $\mathbb{W} = \langle A, R, v\rangle$ *be a WAS. The set of* **stable attacks** *of* $\mathbb{W}$ *is* $R^{\mathbb{W}}_{stab} = R^{\mathbb{W}}_{bd} \cup R^{\mathbb{W}}_{str}$*. The set of* **unstable attacks** *of* $\mathbb{W}$ *is* $R^{\mathbb{W}}_{\overline{stab}} = R^{\mathbb{W}}_{wk}$*.*

We now focus on the second criterion. Intuitively, a *persistent argument* is an argument whose label remains unchanged, regardless of any changes in the weak attacks. Thus, a persistent argument is an argument whose label cannot be changed by the vote of a single expert. To compute the set of persistent arguments, we need to consider all the possible changes over the set of unstable (weak) attacks.

Def. 13 *Let* $\mathbb{W} = \langle A, R, v\rangle$ *be a WAS, and* $R^{\mathbb{W}}_{\overline{stab}}$ *its set of unstable attacks. For all* $R^{\mathbb{W}}_{\overline{stab},i} \subseteq R^{\mathbb{W}}_{\overline{stab}}$ *let* $\mathbb{W}^{alt}_i = \langle A, R, v^{alt}_i\rangle$ *be the* **alternative WAS** *such that* $\forall r \in R^{\mathbb{W}}_{\overline{stab},i}$ *with* $w(r) \neq 0$*, we have* $w^{alt}_i(r) = -w(r)$*;* $\forall r \in R^{\mathbb{W}}_{\overline{stab},i}$ *with* $w(r) = 0$*, we have* $w^{alt}_i(r) = +1$*; and* $\forall r \in (R \setminus R^{\mathbb{W}}_{\overline{stab},i})$ *we have* $w^{alt}_i(r) = w(r)$*. We denote by* $\mathrm{Alt}(\mathbb{W})$ *the* **set of alternative WAS** *of* $\mathbb{W}$*. It holds that* $|\mathrm{Alt}(\mathbb{W})| = 2^{|R^{\mathbb{W}}_{\overline{stab}}|}$*.*

We can now define the sets of persistent and non-persistent arguments.

Def. 14 *Given a WAS* $\mathbb{W} = \langle A, R, v\rangle$ *and* $\mathrm{Alt}(\mathbb{W})$ *its set of alternative WAS, we define the* **set of persistent arguments** *of* $\mathbb{W}$*, denoted* $A^{\mathbb{W}}_{pers}$*, as follows: an argument* $a \in A^{\mathbb{W}}_{pers}$ *iff* $\forall \mathbb{W}^{alt}_i \in \mathrm{Alt}(\mathbb{W})$*, it holds that* $\mathbb{L}^{\mathbb{W}}(a) = \mathbb{L}^{\mathbb{W}^{alt}_i}(a)$*. The* **set of non-persistent arguments** *of* $\mathbb{W}$ *is defined as* $A^{\mathbb{W}}_{\overline{pers}} = A \setminus A^{\mathbb{W}}_{pers}$*.*

Ex. 1, cont. *It holds that* $A^{\mathbb{W}}_{pers} = \{b,c,d\}$ *and* $A^{\mathbb{W}}_{\overline{pers}} = \{a\}$*.*

These two (related) notions allow us to determine to what extent the aggregated WAS is controversial. The next phase of the procedure depends on the result of this analysis. If the debate is too controversial at this point, it could be useful to know how to choose an expert in order to stabilize the aggregated WAS.

4.3. Phase 3: Asking the opinion of an expert

The main difficulty at this phase is that the choice of an expert depends on her expertise, but the decision-maker cannot know the expert's *opinion*. So, we consider an expert *i* who has not taken part to the discussion so far, and ask for her opinion on $\mathbb{W}$. We assume that we know *i*'s topics of expertise (so we can calculate the effects of her possible moves on $\mathbb{W}$), but we do not know *a priori* her opinion on the attacks. We will not ask *i*'s opinion on beyond any doubt attacks ($R^{\mathbb{W}}_{bd}$), as we are certain about them, but only on strong and weak attacks ($R^{\mathbb{W}}_{str} \cup R^{\mathbb{W}}_{wk}$) (a strong attack can become weak after the expert's vote). So, by asking the opinion of expert *i*, we will face up to $2^{|R^{\mathbb{W}}_{str} \cup R^{\mathbb{W}}_{wk}|}$ possible WAS, without being able to know which one we will end up with.

Def. 15 *Given a WAS* $\mathbb{W} = \langle A, R, v \rangle$ *and an expert i, we will denote by* $\mathrm{Poss}_i(\mathbb{W}) = \{\mathbb{W}_{i,1}, \ldots \mathbb{W}_{i,n}\}$ *the set of* **possible WAS** *that could be reached if i was chosen. Note that* $|\mathrm{Poss}_i(\mathbb{W})| \leq 2^{|R_{str}^{\mathbb{W}} \cup R_{wk}^{\mathbb{W}}|}$.

The main difficulty now lies in the comparison of the available experts, in order to choose the one who can make the WAS as uncontroversial as possible. In particular, we observe that it may not be a good heuristic to select the expert with the highest number of topics of expertise, because these topics may not be the most relevant ones. More surprisingly, we also observe that it may not be appropriate to always prefer an expert who declares a strict superset of topics over another expert, because the additional impact provided by the extra topics may actually jeopardize an attack which was considered "strong" before. This requires a careful study that we initiate in the next section.

5. Choosing an expert

The objective of this section is to compare available experts in order to choose the one who is the most able to make the debate uncontroversial. To do so, we focus on the two notions presented in Section 4. First, we study the stability of attacks and we define a relation of dominance among agents depending on their ability to "reinforce" and "weaken" some attacks. Then, we focus on the persistence of the arguments' labels, and we define a relation of dominance among experts depending on their ability to turn the arguments' labels more persistent. It is important at this point to observe that a difficulty we face here is that, when comparing experts, we do not compare two WAS, but two *sets* of possible WAS (those that can be obtained when questioning the experts). This leads to various natural definitions of (strict, easily adapted to weak) dominance:

- *i necessarily dominates j* if any WAS that can be reached by *i* is "better" than any WAS that can be reached by *j*.
- *i possibly dominates j* if there exists a WAS that can be reached by *i* which is "better" than a WAS that can be reached by *j*.
- *i optimistically dominates j* if the best WAS that can be reached by *i* is "better" than the best WAS that can be reached by *j*.
- *i pessimistically dominates j* if the worst WAS that can be reached by *i* is "better" than the worst WAS that can be reached by *j*.

Observe that while the necessary dominance guarantees that the WAS obtained will be better, the optimistic and pessimistic dominance do not. However, they provide good reasons to prefer an expert over another one. By "better" we essentially mean in the sense of Pareto. But what is compared precisely? In what follows we instantiate this, and we provide first properties by focusing on optimistic and pessimistic dominance.

5.1. Stability of attacks

We start by focusing on the experts who can increase (resp. decrease) the weights of some weak (resp. strong) attacks and turn them into strong (resp. weak) attacks.

Def. 16 *Let a WAS* $\mathbb{W} = \langle A, R, v \rangle$*, an expert i and* $\mathbb{W}_i \in \mathrm{Poss}_i(\mathbb{W})$ *a possible WAS among those i can reach. The set of attacks* $R_i \subseteq R_{wk}^{\mathbb{W}}$ **are reinforced** *iff* $\forall r \in R_i$ *it holds that* $r \in R_{str}^{\mathbb{W}_i}$*. The set of attacks* $R'_i \subseteq R_{str}^{\mathbb{W}}$ **are weakened** *iff* $\forall r \in R'_i$ *it holds that* $r \in R_{wk}^{\mathbb{W}_i}$.

An expert i optimistically reinforce-dominates an expert j on a WAS $\mathbb{W}$ iff j can reinforce only a subset of the attacks that i can reinforce, and i pessimistically reinforce-dominates j iff i can weaken only a subset of the attacks that j can weaken.

Def. 17 *Let a WAS $\mathbb{W} = \langle A, R, v\rangle$, and experts i and j. We say that i* **optimistically reinforce-dominates** *j on $\mathbb{W}$ iff: given that $\mathbb{W}_i \in \mathrm{Poss}_i(\mathbb{W})$ (resp. $\mathbb{W}_j \in \mathrm{Poss}_j(\mathbb{W})$) is the possible WAS which contains the $\subseteq$-maximal set of stable attacks, denoted R_i (resp. R_j), it holds that $R_j \subseteq R_i$. i* **pessimistically reinforce-dominates** *j iff given that $\mathbb{W}'_i \in \mathrm{Poss}_i(\mathbb{W})$ (resp. $\mathbb{W}'_j \in \mathrm{Poss}_j(\mathbb{W})$) is the possible WAS which contains the $\subseteq$-maximal set of unstable attacks, denoted R'_i (resp. R'_j), it holds that $R'_i \subseteq R'_j$. We say that i* **reinforce-dominates** *j if i optimistically and pessimistically reinforce-dominates j.*

5.2. Persistence of arguments' labels

We now turn our attention to the arguments' labels.

Def. 18 *Let a WAS $\mathbb{W} = \langle A, R, v\rangle$, an expert i and $\mathbb{W}_i \in \mathrm{Poss}_i(\mathbb{W})$. The set of arguments $A_i \subseteq A^{\mathbb{W}}_{\overline{pers}}$* **are turned persistent** *iff $\forall a \in A_i$ it holds that $a \in A^{\mathbb{W}_i}_{pers}$. The set of arguments $A'_i \subseteq A^{\mathbb{W}}_{pers}$* **are turned non-persistent** *iff $\forall a \in A'_i$ it holds that $a \in A^{\mathbb{W}_i}_{\overline{pers}}$.*

An expert i optimistically persist-dominates an expert j on a WAS $\mathbb{W}$ if and only if j can turn persistent only a subset of the arguments that i can turn persistent, and i pessimistically persist-dominates j iff i can turn non-persistent only a subset of the arguments that j can turn non-persistent.

Def. 19 *Let a WAS $\mathbb{W}$, and two experts i and j. We say that i* **optimistically persist-dominates** *j on $\mathbb{W}$ iff given that $\mathbb{W}_i \in \mathrm{Poss}_i(\mathbb{W})$ (resp. $\mathbb{W}_j \in \mathrm{Poss}_j(\mathbb{W})$) is the possible WAS which contains the $\subseteq$-maximal set of persistent arguments, denoted A_i (resp. A_j), we have $A_j \subseteq A_i$. i* **pessimistically persist-dominates** *j iff given that $\mathbb{W}'_i \in \mathrm{Poss}_i(\mathbb{W})$ (resp. $\mathbb{W}'_j \in \mathrm{Poss}_j(\mathbb{W})$) is the possible WAS which contains the $\subseteq$-maximal set of non-persistent arguments, denoted A'_i (resp. A'_j), we have $A'_i \subseteq A'_j$. We say that i* **persist-dominates** *j if i optimistically and pessimistically persist-dominates j.*

The next properties study the relation between these notions of dominance.

Prop. 1 *If an expert i reinforce-dominates an expert j on a WAS $\mathbb{W}$, then i persist-dominates j on $\mathbb{W}$. The inverse does not always hold.*

Proof 1 *($\rightarrow$) Expert i reinforce-dominates expert j on $\mathbb{W} = \langle A, R, v\rangle$. (1) Assume that j can turn $a \in A$ persistent. To do so, j has to reinforce a set of attacks R_j. As i reinforce-dominates j, i can also reinforce R_j, thus i can also turn a persistent. (2) Assume that i can turn $a \in A$ non-persistent. To do so, i has to weaken a set of attacks R_i. As i reinforce-dominates j, j can also weaken R_i, thus j can also turn a non-persistent. From (1) and (2), we obtain that i persist-dominates j. ($\leftarrow$) Consider the following example: $\mathbb{W} = \langle A, R, v\rangle$, with $A = \{a, b, c\}$, $R^{\mathbb{W},+}_{wk} = \{(b, a), (c, a)\}$ and also $R^{\mathbb{W}}_{bd} = R \setminus R^{\mathbb{W},+}_{wk}$. It holds that $\mathbb{L}^{\mathbb{W}}(a) = OUT$, and $a \in A^{\mathbb{W}}_{\overline{pers}}$ (because if the weights of both weak attacks become negative, we have $\mathbb{L}^{\mathbb{W}}(a) = IN$). Assume that i can reinforce (b, a) while j can reinforce (c, a). Then, i persist-dominates j, but i does not reinforce-dominate j.*

Prop. 2 *Let a WAS* $\mathbb{W}$*. (1) If a subset of weak attacks* $R_1 \subseteq R_{wk}^{\mathbb{W}}$ *is reinforced, while the weights of the other attacks do not change, the number of persistent arguments will not decrease. (2) If a subset of strong attacks* $R_2 \subseteq R_{str}^{\mathbb{W}}$ *is weakened, while the weights of the other attacks do not change, the number of non-persistent arguments will not decrease.*

Proof 2 *(1) Let a subset of weak attacks* $R_1 \subseteq R_{wk}^{\mathbb{W}}$ *beeing reinforced, whereas the weights of the other attacks are unchanged. Let* $\mathbb{W}'$ *the WAS obtained. Let* $a \in A_{pers}^{\mathbb{W}}$*. So*[6] $\forall \mathbb{W}^{alt} \in \mathrm{Alt}(\mathbb{W})$, $\mathbb{L}^{\mathbb{W}}(a) = \mathbb{L}^{\mathbb{W}^{alt}}(a)$*. As* $R_{wk}^{\mathbb{W}'} \subseteq R_{wk}^{\mathbb{W}}$*, it holds that* $\mathrm{Alt}(\mathbb{W}') \subseteq \mathrm{Alt}(W)$*. Thus* $\forall \mathbb{W}'^{alt} \in \mathrm{Alt}(\mathbb{W}')$, $\mathbb{L}^{\mathbb{W}'}(a) = \mathbb{L}^{\mathbb{W}'^{alt}}(a)$*. So* $a \in A_{pers}^{\mathbb{W}'}$*. (2) Similar proof.*

Ex. 1, cont. *The PC chair is worried that the authors of the paper will not be convinced by the current decision, as two of the three attacks are weak, and the argument proposing the acceptance of the paper (a) is non-persistent. So, the question is which expert to choose in order to make the decision uncontroversial. Here are some available experts (strong attacks are in bold), together with the consequences of their (potential) votes.*

Expert	(c,b): $\langle$ **5,6** $\rangle$ *(strong)*		(b,a): $\langle -1,9 \rangle$ *(weak)*		(d,a): $\langle 2,3 \rangle$ *(weak)*	
	$s=+1$	$s=-1$	$s=+1$	$s=-1$	$s=+1$	$s=-1$
1: $\{comp, ml\}$	$\langle$ **8,9** $\rangle$	$\langle 2,9 \rangle$	$\langle 0,12 \rangle$	$\langle -2,12 \rangle$	$\langle 2,3 \rangle$	$\langle 2,3 \rangle$
2: $\{comp, kr\}$	$\langle$ **7,9** $\rangle$	$\langle 3,9 \rangle$	$\langle 1,12 \rangle$	$\langle$ **-3,12** $\rangle$	$\langle$ **4,6** $\rangle$	$\langle 0,6 \rangle$
3: $\{comp, cog\}$	$\langle$ **7,9** $\rangle$	$\langle 3,9 \rangle$	$\langle 1,12 \rangle$	$\langle$ **-3,12** $\rangle$	$\langle 3,6 \rangle$	$\langle 1,6 \rangle$
4: $\{ml, kr\}$	$\langle$ **6,9** $\rangle$	$\langle$ **4,9** $\rangle$	$\langle 0,12 \rangle$	$\langle -2,12 \rangle$	$\langle$ **4,6** $\rangle$	$\langle 0,6 \rangle$
5: $\{ml, cog\}$	$\langle$ **6,9** $\rangle$	$\langle$ **4,9** $\rangle$	$\langle 0,12 \rangle$	$\langle -2,12 \rangle$	$\langle 3,6 \rangle$	$\langle 1,6 \rangle$
6: $\{cog, kr\}$	$\langle$ **5,6** $\rangle$	$\langle$ **5,6** $\rangle$	$\langle 1,12 \rangle$	$\langle$ **-3,12** $\rangle$	$\langle$ **5,6** $\rangle$	$\langle -1,6 \rangle$

First, the PC chair oberves that expert 1 is necessarily reinforce-dominated by experts 4, 5, and 6. No other expert is necessarily reinforce (strictly) dominated in this example. For instance, expert 3 is not necessarily reinforce-dominated by expert 6, because if 3 votes negatively on (b,a) *while 6 votes positively on this attack, the WAS reached by expert 6 is not strictly better, in the sense of Pareto. Next, expert 6 reinforce-dominates all the other experts, as she can reinforce both* (b,a) *and* (d,a)*, and she cannot weaken* (c,b)*. No expert reinforce-dominates expert 6, for instance, expert 2 can weaken* (c,b)*, and expert 4 cannot reinforce* (b,a)*. Interestingly, expert 2 optimistically reinforce-dominates expert 4, but is pessimistically reinforce-dominated by the same expert. Finally, both expert 4 and expert 6 persist-dominate all the other experts (as they can turn a persistent, and they cannot turn b non-persistent).*

6. Conclusion

The first contribution of this paper is to set up a model where expertise can be meaningfully integrated in an argumentation framework, assuming that arguments are tagged with the topics they refer to. This is an important problem in online systems where several users are asked to vote: their different expertise may motivate us to weight their opinions accordingly. The second contribution of the paper is proposing a solution to the following problem: sometimes the resulting debate is controversial because users may

[6] For the sake of simplicity, and without loss of generality, we do not mention here the agent modifying $\mathbb{W}$.

have the feeling that the decision might have “easily” been different. This may result from a voting controversy, or an argumentative one. Having introduced notions to assess this controversy, we discuss how to choose an additional expert to make the debate as uncontroversial as possible. The problem is difficult, in particular because when we call an expert we only know her domain of expertise, but not the exact way she will contribute to the debate. So, we have to reason about the potential systems that may be reached after the contribution of the expert. In a preliminary analysis, we have provided initial results based on possible dominance relations among experts. Of course, much more work needs to be done in this respect. Another natural (probabilistic) approach would be for instance to consider an expected dominance, by quantifying over all the potential systems that can be reached (either by assuming equiprobable occurrence of these systems, or by injecting a prior probability based on further information we might have about the experts).

References

[1] T. Bench-Capon. Value-based argumentation frameworks. In *Proc. of NMR'02*, pages 443–454, 2002.

[2] G. Boella, D. M. Gabbay, L. van der Torre, and S. Villata. Arguing about trust in multiagent systems. In *Proc. of AI*IA*, 2010.

[3] E. Bonzon and N. Maudet. On the outcomes of multiparty persuasion. In *Proc. of AAMAS'11*, pages 47–54, May 2011.

[4] M. Caminada. On the issue of reinstatement in argumentation. In *Proc. of JELIA'06*, pages 111–123, 2006.

[5] M. Caminada and G. Pigozzi. On judgment aggregation in abstract argumentation. *Autonomous Agents and Multi-Agent Systems*, 22(1):64–102, 2011.

[6] D. Cartwright and K. Atkinson. Using computational argumentation to support e-participation. *IEEE Intelligent Systems*, 24(5):42–52, 2009.

[7] C. Cayrol and M.-C. Lagasquie-Schiex. Weighted argumentation systems: A tool for merging argumentation systems. In *Proc. of ICTAI'11*, pages 629–632, 2011.

[8] S. Coste-Marquis, C. Devred, S. Konieczny, M.-C. Lagasquie-Schiex, and P. Marquis. On the Merging of Dung's Argumentation Systems. *Artificial Intelligence*, 171:740–753, 2007.

[9] S. Coste-Marquis, S. Konieczny, P. Marquis, and M. A. Ouali. Weighted attacks in argumentation frameworks. In *Thirteenth International Conference on Principles of Knowledge Representation and Reasoning (KR'12)*, 2012.

[10] D. Dubois, M. Grabish, H. Prade, and P. Smets. Using the transferable belief model and a qualitative possibility theory approach on an illustrative example: The assessment of the value of a candidate. *International journal of intelligent systems*, 16(11):1245–1272, 2001.

[11] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[12] P. E. Dunne, A. Hunter, P. McBurney, S. Parsons, and M. Wooldridge. Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artif. Intell.*, 175(2):457–486, 2011.

[13] H. Jakobovits and D. Vermeir. Robust semantics for argumentation frameworks. *Journal of Logic and Computation*, 9(2):215–261, 1999.

[14] J. Leite and J. Martins. Social abstract argumentation. In *Proc. of IJCAI'11*, pages 2287–2292, 2011.

[15] I. Rahwan and K. Larson. Pareto optimality in abstract argumentation. In *Proc. of AAAI'08*, pages 150–155, 2008.

[16] I. Rahwan and F. Tohmé. Collective argument evaluation as judgement aggregation. In *Proc. of AAMAS'10*, pages 417–424, 2010.

[17] C. Reed and G. Rowe. Araucaria: Software for argument analysis, diagramming and representation. *International Journal of AI Tools*, 14:961–980, 2004.

[18] O. Scheuer, F. Loll, N. Pinkwart, and B. McLaren. Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning*, 5:43–102, 2010.

[19] F. Toni and P. Torroni. Bottom-up argumentation. In S. Modgil, N. Oren, and F. Toni, editors, *TAFA*, volume 7132 of *Lecture Notes in Computer Science*, pages 249–262. Springer, 2011.