Report on ongoing research

This document is an ongoing work regarding the exercise of realizing some of the abstract ideas and notions we discussed in the *BTC-collaboration* document.

1 Social Behavior

Ultimately we strive for our system to satisfy the following property:

Property 1. *A user of the system never has an incentive to change his/her vote.*

However that's not the case for the time being. Before going into formal definitions, with the following simple example we show that the property does not hold for ESAFs.

Example 1. *Assume the fictitious example of a framework with three arguments as depicted in Figure 1. Let $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ be an Abstract Argumentation Framework and $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ be a Social Argumentation Framework. Let $\mathcal{A} = \{a, b, c\}$, $\mathcal{R} = \{(b, a), (c, b)\}$. Suppose that there is a user utilizing the system, that has five votes to spare in her disposal, and her only objective is increasing argument a's value as much as possible*

As can be observed from the values in Table 1, it makes more sense for the user not to take the straight forward approach and use all of the votes s/he possess on argument a. But rather take an in direct approach and use them on argument c. Table 1 shows that a's social strength benefits greater from the increase in its attacker's attacker's social support.
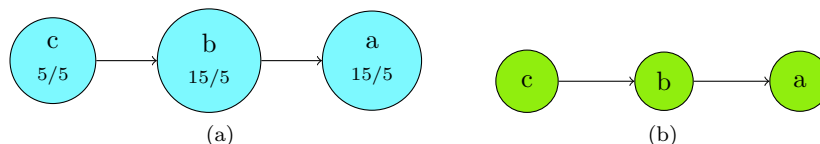

(a) (b)

Figure 1: Initial state with votes on arguments(left), Model evaluation under ESAF(right)

Thus the example shows in our system, in some settings users indeed have incentive to sway away from their original preferences and vote for other arguments(and/or attack relations). This makes it interesting to investigate and figure out potential behavioral patterns of the users.

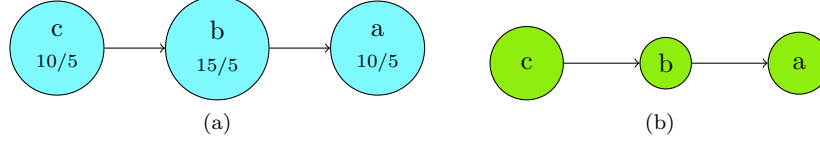(a)   (b)

Figure 2: Initial state with votes on arguments(left), Model evaluation under ESAF(right)

Table 1: Model of the Social Abstract Argumentation Framework considering: 1a) social support only; 1b) model under ESAF; 2a) social support only(where 5 positive votes transferred from a to c); 2b) model under ESAF(where 5 positive votes transferred from a to c)

Figure	Arg. a	Arg. b	Arg. c
1a	0.750	0.750	0.500
1b	0,469	0,375	0,500
2a	0,667	0,750	0,667
2b	0,492	0,255	0,667

In the light of this motivation, we first try to construct the formation. Again we try to define the model as generic as possible, so that later we have the flexibility of tailoring the system for distinctive settings. The model takes some inspiration from other work that mainly dwell in automated task planning and coalition formation.

1.1   Social Behavior System

Let us start by discussing the elements of the system. Firstly several users $\mathcal{U} = u_1, ..., u_n$ are involved in the system.

The system composes several moves $\mathcal{M} = m_1, ..., m_n$ that will be carried out by users. Foreseeable list of moves include introducing an argument to the system, introducing an attack relation to the system, voting on existing arguments(and/or attack relations).

Each user $u_i \in \mathcal{U}$ is supposed to have a vector of non-negative capabilities $\mathcal{B} = < b_1^i, ..., b_r^i >$. A capability b_j^i represents the capacity of the user u_i regarding the move m_j . One other way of imagining this vector is a collection of budgets. For instance m_2 might stand for the action of voting on an argument, and b_2^5 might correspond to the number of votes that user u_5 holds that could be used for voting on the existing arguments of the system (Henceforth with respect this modeling $|\mathcal{M}| = |\mathcal{B}^i|$ for any user $u_i \in \mathcal{U}$, but of course for some $b_j^i \in \mathcal{B}^i$, b_j^i might equal to 0).

Moreover each user $u_i \in \mathcal{U}$ possesses a vector called core, $\mathcal{C}^i = < c_1^i, ..., c_r^i >$. There are multiple possibilities in defining the core. Firstly it could be set

2

of desired entities from the framework, i.e. a subset of the argument set $\mathcal{A}$ (and/or the attack relations set $\mathcal{R}$). Moreover this set could be substituted by a preference vector. Thus the elements of the core would have a relative hierarchical order between themselves, and this would guide the choices of the user when selecting a move over the rest.

One more possibility, which a tad more complex is that $\mathcal{C}$ could stand for a desired state of the system, and each $c_i \in \mathcal{C}$ depicts a part of the state. For instance a user might yearn for a state where an argument $a_i \in \mathcal{A}$ enjoys a social strength higher than 0.7, and an attack relation $(a_m, a_n) \in \mathcal{R}$ to command a social support lower than 0.3, and an argument $a_j \in \mathcal{A}$ to be attacked by at least five arguments. Such a desired state can be defined via a core with three elements, $\mathcal{C} = < c_1, c_2, c_3 >$, where each element correspond to a distinctive one of the three concepts. Again just as stated in the previous discussion, this vector might be defined as preference vector(accompanied by the definition of the necessary operator).

We combine the notions discussed above to construct the formal definition of the envisioned system:

Definition 1. *[Social Behavior System]*
A Social Behavior System(SBS) is a triple $< \mathcal{F}, \mathcal{M}, \mathcal{U} >$ such that:

- *$\mathcal{F}$ is an ESAF.*

- *$\mathcal{M} = \{m_1, ..., m_k\}$ is set of moves allowed by the system.*

- *$\mathcal{U} = \{u_1, ..., u_l\}$ is a set of users, where $\forall u_i \in \mathcal{U}$ consists of:*

    - *a vector $\mathcal{C}^i = < c_1^i, ..., c_r^i >$ depicting its core.*
    - *a vector $\mathcal{B}^i = < b_1^i, ..., b_t^i >$ of its budgets corresponding each move.*

1.2 The concept of Coalition

In the light of the model, we believe that there are multiple reasons for the users to cooperate under certain settings.

- Firstly the shared elements in the cores of the distinct arguments points to an intrinsic motivation for cooperation.

- Moreover, recall that we had presented a candidate setting for the notion of core where the every $c_i \in \mathcal{C}$ corresponds to a desired state of some part of the framework. If a similar approach is adopted, some arbitrary sub-budget b_j^i of a user u_i might not prove to be enough for its corresponding element of the core. In such scenarios, we expect the user to look for other users in the hopes of forming a *coalition* to realize such goals which can't be achieved by their own.

- Lastly some, negotiation procedures may be defined for the model in order to mimic possible strategic partnerships that may emerge amongst the human users. It may well be plausible for a user to sacrifice the most preferred strategy for the sake of initiating a coalition.

3

In the light of the listed incentives, we follow with the formal definition of the concept of coalition in our system:

Definition 2. *[Coalition]*
Let $< \mathcal{F}, \mathcal{M}, \mathcal{U} >$ be an SBS, a coalition is a tuple $< \mathcal{K}, \mathcal{G} >$ such that:

- $K \subseteq U$

- $\mathcal{G} = \{c_1, ..., c_n\} \subseteq \bigcup_{x=1}^{m} C^x$, *where $u_i \in K$ for $i \in [1..m]$*

- $\mathcal{K} \Vdash \mathcal{G}$

Where K is the coalition group composed of a set of users, and G is the shared goal which consists of a subset of the union of the cores of the users that belong to the coalition group. $K \Vdash G$ denotes that goal G might be realized via K.

We must note that there are loose ends in the current state of the model, including the formal definitions. Most cases the reason is that, as mentioned above in detail, there are many candidates to be adopted for each concept in the model and the definitions have to be revised for some.

For instance the meaning of the operator $\Vdash$ from Definition 2 relies on the choice of $\mathcal{C}$. As an example, if we are to follow the first interpretation of $\mathcal{C}$ from Section 1.1 (where roughly, c_j^i corresponds to an entity from the framework, for which the user u_i desires to see an increase in its value), it's probably better to simply drop the operator from the definition. It's hard to talk about *the satisfiability of a goal*, since even an episolon-increase results in a justified cooperation.

On the other hand if the last approach is adopted from the same section (where briefly, c_j^i a part of a concrete desired state of the frame such as *argument a's social strength being under 0.4*), then the operator indeed serves the intended purpose. This can be seen from its formal definition, which is as follows:

Conjecture 1. *[Achievement of a unified goal]*
Let $< \mathcal{F}, \mathcal{M}, \mathcal{U} >$ be an SBS, $\mathcal{K} \subseteq \mathcal{U}$ and $\mathcal{G} \subseteq \bigcup_{x=1}^{m} \mathcal{C}^x$, for $u_i \in K$, $i \in [1..m]$.
The unified goal $\mathcal{G}$ can be achieved by the group $\mathcal{K}$ of users, denoted by $\mathcal{K} \Vdash \mathcal{G}$, iff $\sum_{u_i \in \mathcal{K}} \mathcal{B}^i \geq \sum_{c_i \in \mathcal{G}} c_i$

In plain words, the unified goal is satisfiable if and only if the total budgets of the users in the coalition exceed the numerical value needed to satisfy all the respective cores in the goal. For the sake of simplicity, we suppose c_is are numerical values. Recall that they can be complex sub-states of the framework, so normally these numerical values have to be calculated with a process preceding this check. Moreover we have taken a single-typed budget for all users, where as in sophisticated setting we have to take into consideration of sub-budgets for each move(in other words the definition works adopts a setting of the framework where there is a budget for only one move i.e. solely voting on arguments).

1.3 Ongoing work

As stated earlier, in our system each user aims at maximizing its own utility, or wants to reach its desired state of the framework. When that is not achievable by its own means, each user would have an incentive to seek a coalition with some fellow user(s). With this motivation, in this section we have regarded the notion of Social Behavior in our system as a problem of coalition formation that can be defined as a finite set N of users who should achieve a finite set M of cores.

Ideally, a coalition formation process can be envisioned as a two steps procedure:

1. Generating the coalition structures. The idea here is to form the coalitions such that users within a coalition should coordinate to decide on a set of (shared)cores and then achieve those, but those in different coalitions do not.

2. Discussing these structures between the users in order to select the best ones which will be adopted.

1.3.1 Constructing coalition structures

In the current state of the proposed model, there is no elaborate methodology for constructing coalitions. We may create some simplistic procedures by introducing a few assumptions. For instance we may restrict the cardinality of core set C for each user to one, and state that the users with shared cores form coalitions. This indeed reminds the setting from Prof.Maudet's research, where each user can be easily attained to a party. However this time the number of parties is equal to the number of distinct cores in the system, rather than just two. However there are existing works which may help us in devising a more sophisticated framework in which users have the initiative of choosing the coalition to their liking amongst multiple possibilities.

Firstly in order to be able to make a choice between the candidate coalitions, users need some type of a preference mechanism. Perhaps this could be achieved by the introduction of a utility function for each user. Given a tuple consisting of a coalition k_i and a user u_j, the value function may return the benefit of attending to k_i for u_j.

Conjecture 2. *There exists a function* $\mu : \mathcal{U} \times \mathcal{K} \to \mathcal{P}$*, where* $(\mathcal{P}, \leq)$ *is a poset.*

So as discussed earlier, we want a function that would give us some idea on how preferable is a certain coalition for a particular user. We envision that this notion would roughly relate to the difference of the profit the user gets via the coalition and the cost of joining the coalition. The profit would perhaps rely on the gain with respect to the cores of the user. And in a similar sense, the cost would depend on the use of budget as a result of the constructed coalition.

Moreover we envisioned the co-domain of the function to form a poset together with the necessary relation, so that we would have a measure of comparison between distinct coalitions. A total-order or a lattice may prove to be

more fruitful, in terms of a more complete comparison measure. On the other hand accomplishing the task of defining such functions with stricter constraints perhaps may turn out to be difficult.

It's also important to note that with the help of this function, we may formalise Property 1 which was stated in a highly abstract sense:

Conjecture 3. $\forall u \in \mathcal{U}, \ \nexists k \in K \ s.t. \ \mu(u, k) > \mu(u, \{u\}) \ , \ where \ k \neq \{u\}.$

So what the conjecture conveys is that, in the optimal framework for all users there is no coalition setting which would return a higher benefit to the user then acting on its own. Or simpler words, there's no reason at all for cooperating with other users.

Another interesting notion is *conflicts* between coalitions. Since the budgets are finite values, it might make sense to restrict each user to a single coalition, or a more complex setting could be letting users participate in multiple coalitions as long as they have some remaining budget. However it's thought provoking as it leads to exciting notions like defeasibility. Briefly we may define a *defeat* relation between coalitions, where a coalition defeats another when they share the same cores but the former possess a lower price for formation.

If all these tasks are to be accomplished, then we may upgrade our coalition formation process to a more sophisticated level, consisting of the following four steps:

1. Constructing all candidate coalitions.

2. Defining preference and defensibility relations amongst the candidate coalitions.

3. Hierarchical ordering of the acceptable coalitions (for each user).

4. Actual participation to a certain coalition (for each user).

2   Vectors meet Pro-arguments

In this section we try to integrate two distinct research topics that were previously discussed in the *BTC-collaboration* document. We start with a brief recap of the two underlying ideas', possible approaches and their shortcomings.

2.1   Pro-arguments

We have been searching for possible methods of including the concept of support into our framework. One candidate extension to the existing framework is the addition of the notion of Pro-arguments, which are envisioned to be structural parts of the arguments they defend rather than stand alone arguments.

Our first take is realizing them as solely additional reasons to make users vote on a particular argument, and leaving them out of the social voting process. This approach does not introduce any additional ways for exploitation of the

system (via the users), however it could be seen as merely an enhancement on the GUI side.

Second interpretation is adopting them with the same set of functionalities/privileges as the regular arguments in the framework. However this leaves the system vulnerable against harmful manipulation, i.e. the introduction of fallacious pro-arguments and associating them to existing arguments with strong social support.

One last realization method of the notion is taking a similar approach to value-based argumentation (as pointed out by Prof.Baroni in his last correspondence). Instead of implementing them as some stripped version of an argument, the idea is mimicking the value functions and making pro-arguments some value promoting relation amongst arguments. However this method again gives way to the exact same problem as in the second approach.

2.2 Multi-Dimensional Representation

2.2.1 Preliminaries

This section will present some useful definitions and notation regarding to (a slightly modified version of) ESAFs.

Definition 3 (Extended social argumentation frameworks). *An* extended social argumentation framework *is a 4-tuple $F = \langle \mathcal{A}, \mathcal{R}, V_{\mathcal{A}}, V_{\mathcal{R}} \rangle$, where*

- *$\mathcal{A}$ is the set of arguments,*

- *$\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ is a binary attack relation between arguments,*

- *$V_{\mathcal{A}} : \mathcal{A} \to \mathbb{N} \times \mathbb{N} \times \mathbb{N}$ stores the crowd's pro and con votes for each argument together with the maximum number of votes for an argument in the system.*

- *$V_{\mathcal{R}} : \mathcal{R} \to \mathbb{N} \times \mathbb{N}$ stores the crowd's pro and con votes for each attack.*

Definition 4 (Semantic Framework). *A semantic framework is a 6-tuple $\langle L, \curlywedge_1, \curlywedge_2, \curlyvee, \neg, \tau \rangle$ where:*

- *L is a totally ordered set with top and bottom elements $\top$, $\bot$, containing all possible valuations of an argument.*

- *$\curlywedge_{\mathcal{A}}, \curlywedge_{\mathcal{R}} : L \times L \to L$, are two binary algebraic operations used to restrict strengths to given values.*

- *$\curlyvee : L \times L \to L$, is a binary algebraic operation on argument valuations used to combine or aggregate valuations and strengths.*

- *$\neg : L \to L$ is a unary algebraic operation for computing a restricting value corresponding to a given valuation or strength.*

- *$\tau : \mathbb{N} \times \mathbb{N} \times \mathbb{N} \to L$ is a function that aggregates positive and negative votes and the maximum number of total votes into a* social support *value.*

7

Notation 1. *Let* $F = \langle \mathcal{A}, \mathcal{R}, V_{\mathcal{A}}, V_{\mathcal{R}} \rangle$ *be an ESAF,* $\mathcal{S} = \langle L, \curlywedge_1, \curlywedge_2, \curlyvee, \neg, \tau \rangle$ *a semantic framework and*

- $\mathcal{R}^{\text{-}}(a) \triangleq \{a_i \in \mathcal{A} : (a_i, a) \in \mathcal{R}\}$ *be the set of direct attackers of an argument* $a \in \mathcal{A}$,

- $V_{\mathcal{A}}^{+}(a) \triangleq x$ *denote the number of positive votes for argument* a,

- $V_{\mathcal{A}}^{-}(a) \triangleq y$ *denote the number of negative votes for argument* a,

- $V_{\mathcal{A}}^{max}(a) \triangleq z$ *denote the number of maximum votes for an argument in the ESAF that argument* a *belongs to*,

- $\tau(a) \triangleq \tau(V_{\mathcal{A}}(a)) = \tau(x, y, z)$ *denote the social support for an argument* a *via utilizing a vote aggregation function* τ,

- $v^r(a) = \frac{V_{\mathcal{A}}^{+}(a)}{V_{\mathcal{A}}^{-}(a)}$ *denote the ratio of positive votes to negative votes for argument* a,

- $v^t(a) = V_{\mathcal{A}}^{+}(a) + V_{\mathcal{A}}^{-}(a)$ *denote the total number of votes for argument* a,

- $V_{\mathcal{R}}^{+}((a, b)) \triangleq m$ *denote the number of positive votes for an attack relation between arguments* a *and* b,

- $V_{\mathcal{R}}^{-}((a, b)) \triangleq n$ *denote the number of negative votes for an attack relation between arguments* a *and* b,

- $\tau((a, b)) \triangleq \tau(V_{\mathcal{R}}(a, b)) = \tau(m, n)$ *denote the social support for an attack relation between arguments* a *and* b *by utilizing a vote aggregation function* τ,

-

$$\bigcurlyvee_{x \in R} x \triangleq (((x_1 \curlyvee x_2) \curlyvee ...) \curlyvee x_n)$$

$R = \{x_1, x_2, ..., x_n\}$ *denote the aggregation of a multiset of elements of* L.

Definition 5 (Model)**.** *Let* $F = \langle \mathcal{A}, \mathcal{R}, V_{\mathcal{A}}, V_{\mathcal{R}} \rangle$ *be a social argumentation framework,* $\mathcal{S} = \langle L, \curlywedge_1, \curlywedge_2, \curlyvee, \neg, \tau \rangle$ *be a semantic framework. A* $\mathcal{S}$*-model of* F *is a total mapping* $M : \mathcal{A} \to L$ *such that for all* $a \in \mathcal{A}$,

$$M(a) = \tau(a) \curlywedge_{\mathcal{A}} \neg \bigcurlyvee_{a_i \in \mathcal{R}^{\text{-}}(a)} (\tau((a_i, a)) \curlywedge_{\mathcal{R}} M(a_i))$$

2.2.2 Properties and Propositions

Again in the previous document, we had discussed the fact that even though we have functions for unifying the social such as the one depicted in Definition 4, accompanied by a set of desired properties(P1, P2, P3), there is substantial room for improvement.

Definition 6. *[Enhanced Vote Aggregation]*
An enhanced vote aggregation function is any function $\tau : \mathbb{N} \times \mathbb{N} \times \mathbb{N} \rightarrow [0,1]$
such that $v_{max} \geq 0$

$$\tau\left(v^+, v^-, v_{max}\right) = \begin{cases} 0 & v_{max} = 0 \\ \frac{v^+}{v^+ + v^- + \frac{1}{v_{max}}} & otherwise \end{cases}$$

Property 2. *[Perfect argument freeness]*
Let $F = \langle \mathcal{A}, \mathcal{R}, V_{\mathcal{A}}, V_{\mathcal{R}} \rangle$ *be an ESAF and* $\mathcal{S} = \langle L, \curlywedge_1, \curlywedge_2, \curlyvee, \neg, \tau \rangle$ *a semantic framework. A vote aggregation function* $\tau : \mathbb{N} \times \mathbb{N} \times \mathbb{N} \rightarrow L$ *is said to be "perfect argument free" if*

$$\forall a \in \mathcal{A}, \ \tau(a) < 1.$$

The gist of the property can be captured in a verbal context with the following sentence: *No argument enjoys perfect Social Support.*

Proposition 1. *Enhanced Vote Aggregation function enjoys Property 2.*

Proof. Trivial.
Suppose an arbitrary $a \in \mathcal{A}$ of an extended social argumentation framework $\mathcal{F}$ with some well-behaved semantics $\mathcal{S}$:
If $v^+ = 0$ then $\tau(a) = \tau(v^+, v^-, v_{max}) = 0 \leq 1$.
Else if $v^+ \neq 0$, then $v_{max} \neq 0$ and since $v_{max} \in \mathcal{Z}^+$, then $\frac{1}{v_{max}} > 0$. Since denominator equals to the addition of the numerator and some $r \in R^+$, denominator is bigger than the numerator and thus
$\frac{v^+}{v^+ + v^- + \frac{1}{v_{max}}} = \tau(v^+, v^-, v_{max}) \leq 1$.
□

Property 3. *[Precedence of the vote ratio]*
Let $F = \langle \mathcal{A}, \mathcal{R}, V_{\mathcal{A}}, V_{\mathcal{R}} \rangle$ *be an ESAF and* $\mathcal{S} = \langle L, \curlywedge_1, \curlywedge_2, \curlyvee, \neg, \tau \rangle$ *a semantic framework. For* $a_1, a_2 \in \mathcal{A}$, *a vote aggregation function* $\tau : \mathbb{N} \times \mathbb{N} \times \mathbb{N} \rightarrow L$ *is said to be "vote ratio precedent" if*

$$(v^r(a_1) \geq v^r(a_2)) \implies (\tau(a_1) \geq \tau(a_2)).$$

The gist of the property can be captured in a verbal context with the following sentence: *If the value of the ratio of positive votes to the total amount of votes is higher for an argument a than an argument b; a's social support value exceeds the social support of b.*

Conjecture 4. *Vote Aggregation function enjoys Property 3.*

Sketch of proof. It should suffice to show that:
For two arbitrary tuples (v_1^+, v_1^-) and (v_2^+, v_2^-), if $\frac{v_1^+}{v_1^-} \geq \frac{v_2^+}{v_2^-}$ then $\tau(v_1^+, v_1^-, v_{max}) \geq \tau(v_2^+, v_2^-, v_{max})$.
□

As an example for $(v_1^+, v_1^-) = (10, 0)$ and $(v_2^+, v_2^-) = (999, 1)$, $v_{max} = 1000$ and $\tau(v_1^+, v_1^-, v_{max}) = 0.9999$ and $\tau(v_2^+, v_2^-, v_{max}) = 0.9989$.

Property 4. *[Precedence of the total number of votes]*
Let $F = \langle \mathcal{A}, \mathcal{R}, V_{\mathcal{A}}, V_{\mathcal{R}} \rangle$ be an ESAF and $\mathcal{S} = \langle L, \curlywedge_1, \curlywedge_2, \curlyvee, \neg, \tau \rangle$ a semantic framework. For $a_1, a_2 \in \mathcal{A}$, a vote aggregation function $\tau : \mathbb{N} \times \mathbb{N} \times \mathbb{N} \to L$ is said to be "total votes precedent" if

$$((v^r(a_1) = v^r(a_2)) \wedge (v^t(a_1) \geq (v^t(a_2))) \implies (\tau(a_1) \geq \tau(a_2)).$$

The gist of the property can be captured in a verbal context with the following sentence: *When the ratios are equal, the function should return a higher social support value for the one with the higher number of total votes.*

Conjecture 5. *Enhanced Vote Aggregation function enjoys Property 4.*

Sketch of proof. It should suffice to show that:
For two tuples (v_1^+, v_1^-) and (v_2^+, v_2^-), if $\frac{v_1^+}{v_1^-} = \frac{v_2^+}{v_2^-}$ and $(v_1^+ + v_1^-) \geq (v_2^+ + v_2^-)$ then $\tau(v_1^+, v_1^-, v_{max}) \geq \tau(v_2^+, v_2^-, v_{max})$. $\qquad\square$

As an example for $(v_1^+, v_1^-) = (45, 5)$ and $(v_2^+, v_2^-) = (9, 1)$, $v_{max} = 50$ and $\tau(v_1^+, v_1^-, v_{max}) = 0.8996$ and $\tau(v_2^+, v_2^-, v_{max}) = 0.8982$

However firstly, the existing function's result does not reflect the total number of votes casted on a particular argument to a decent extent. Moreover since the function boils all the inputs to a single value prior to the model evaluation, we don't have much flexibility in the upcoming stages of the computation.

With ESAFs, when two distinct arguments have the same ratio of positive and negative votes, even if they differ immensely with respect to the total number of votes they've received, this fact is not reflected in their model evaluation values.

We may capture the idea with the following property:

Property 5. *"For two arguments sharing the same ratio of positive and negative votes and the same set of attackers, if the total number of votes differ from each other, so should the model evaluations."*
Let $F = \langle \mathcal{A}, \mathcal{R}, V_{\mathcal{A}}, V_{\mathcal{R}} \rangle$ be an extended social argumentation framework, $\mathcal{S} = \langle L, \curlywedge_1, \curlywedge_2, \curlyvee, \neg, \tau \rangle$ a well behaved semantics and $\mathcal{M}_{\mathcal{S}}^F$ the set of all $\mathcal{S}$-models of $\mathcal{F}$. For $a_1, a_2 \in \mathcal{A}$, where $a_1 = (v_1^+, v_1^-)$, $a_2 = (v_2^+, v_2^-)$, and $\mathcal{R}^-(a_1) = \mathcal{R}^-(a_2)$, and $v_1^+ \neq 0$, $v_2^+ \neq 0$,

$$\forall M \in \mathcal{M}_{\mathcal{S}}^F, \left(\left(\frac{v_1^+}{v_1^-} = \frac{v_2^+}{v_2^-} \right) \wedge \left((v_1^+ + v_1^-) \neq (v_2^+ + v_2^-) \right) \right) \implies$$
$$\left(M(v_1^+, v_1^-) \neq M(v_2^+, v_2^-) \right).$$

We show that the original proposal does not possess to this property.

Proposition 2. *ESAFs do not satisfy Property 5.*

Proof. It's enough to show a counter-example. Suppose two arguments $a_1, a_2 \in \mathcal{A}$ of an extended social argumentation framework $\mathcal{F}$ with some well-behaved semantics $\mathcal{S}$ such that

- the arguments have some positive votes: $v_1^+ \neq 0$, $v_2^+ \neq 0$,

- the arguments are unattacked: $\mathcal{R}^{\text{-}}(a_1) = \emptyset$, $\mathcal{R}^{\text{-}}(a_2) = \emptyset$

- and the ratios of positive to negative votes are the same: $\left(\frac{v_1^+}{v_1^-} = \frac{v_2^+}{v_2^-} \right)$,

- and the total number of casted votes are distinct: $\left((v_1^+ + v_1^-) \neq (v_2^+ + v_2^-) \right)$,

- however the model evaluations are the same: $M(v_1^+, v_1^-) = M(v_2^+, v_2^-)$.

Since $(\frac{v_1^+}{v_1^-} = \frac{v_2^+}{v_2^-})$, $\exists r \in \mathcal{R}^+$ s.t. $v_1^+ \cdot r = v_2^+$ and $v_1^- \cdot r = v_2^-$.

$$
\begin{aligned}
\tau(a_1) &= \frac{v_1^+}{v_1^+ + v_1^- + \varepsilon} \cong \frac{v_1^+}{v_1^+ + v_1^-} \\
&= \frac{v_1^+ \cdot r}{(v_1^+ + v_1^-) \cdot r} = \frac{v_1^+ \cdot r}{v_1^+ \cdot r + v_1^- \cdot r} \\
&= \frac{v_2^+}{v_2^+ + v_2^-} \cong \frac{v_2^+}{v_2^+ + v_2^- + \varepsilon} \\
&= \tau(a_2)
\end{aligned}
$$

We know that in ESAFS for any unattacked argument a, $\tau(a) = M(a)$ for $M \in \mathcal{M}_{\mathcal{S}}^F$. Thus

$$M(a_1) = \tau(a_1) = \tau(a_2) = M(a_2).$$
$$\text{CONTRADICTION!}$$

$\square$

A multi-dimensional vector representation of an argument lets one tailor operations for each distinct dimension, and then merge them at a phase up to the designer's liking. Since ESAFs assume is no restriction on the labeling set where arguments take their values from, arguments can be maintained as vectors all the way including the model evaluation phase, and the final result might be displayed by some GUI choice depicting all arguments' vector representations.

2.3 Marrying the two concepts

In the light of the aforementioned motivations, we try to propose a candidate vector representation.

The first step is defining the dimensions. In one of the earlier *BTC-collaboration* document's iterations we had briefly discussed the possible pitfalls of choosing hybrid dimensions - dimensions consisting of multiple variables, and some of these variables appearing in more than one dimension. This problem is inspected in more detail in the last subsection.

The obvious candidates are the positive and the negative number of votes an argument has received. Moreover in Section 2.1 we have motivated the extension of our system via pro-arguments, but the discussed methods all had their share bit of shortcomings. Here we try to incorporate them to the system by defining them as the third dimension.

Firstly we extend the framework as follows:

Definition 7 (Tridimensional Social argumentation frameworks). *A tridimensional social argumentation framework is a 6-tuple* $\langle \mathcal{A}, \mathcal{P}, \mathcal{R}, \mathcal{M}_P, V_{\mathcal{A}}, \mathcal{V}_P \rangle$, *where*

- $\mathcal{A}$ *is the set of arguments,*

- $\mathcal{P}$ *is the set of pro-arguments,*

- $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ *is a binary attack relation between arguments,*

- $\mathcal{M}_P : 2^{\mathcal{P}} \to \mathcal{A}$ *is a mapping from a set of pro-argument to an argument*

- $V_{\mathcal{A}} : \mathcal{A} \to \mathbb{N} \times \mathbb{N}$ *stores the crowd's pro and con votes for each argument.*

- $\mathcal{V}_P : \mathcal{P} \to \mathbb{N} \times \mathbb{N}$ *stores the crowd's pro and con votes for each pro-argument.*

   (We omit the social voting on attack relations for simplicity.)

Moreover assume we have an aggregation function that given a set of pro-arguments, return a reel value via aggregating each pro-argument's positive and negative votes by some method: $\psi : 2^{\mathcal{P}} \to \mathbb{N}$

We may also chose to normalise the pro and con votes of an argument for determining the associated dimension values with some function $\tau : \mathbb{N} \to \mathbb{N}$

Then we may represent each argument $a_i \in \mathcal{A}$ as a vector:
$\vec{a_i} = < \tau(v_i^+), \tau(v_i^-), \psi(\mathcal{M}_p^-(a_i)) >$.

Under this topic, current efforts are invested in finding a concrete complete framework utilizing vector operations and leaving the end result (after model evaluation) in vector form.

A very simplistic method τ can be defined for pro and con votes of the argument as such as:

$$\tau(x) = \frac{x}{v_{max} + \epsilon}, \text{ where } v_{max} \text{ stands for the maximum number of total votes for an argument in the system.}$$

Thus the values are restricted between the interval of $[0,1] \in \mathbb{R}$, no argument can take the perfect value(i.e. 1) for either dimension(as a result of epsilon addition), and only an argument possessing solely positive votes and the maximum number of votes in the system would enjoy near perfect value .

On the other hand, the following function could perhaps be a suitable choice for the sake of aggregating pro-arguments:

$$\psi(\{p_1, ..., p_n\}) = \frac{\sum\limits_{i=1}^{n} \frac{v_i^+}{v_i^+ + v_i^- + \frac{1}{(v_i^+ + v_i^- + \frac{1}{v_{max}})}}}{n},$$

where $p_i \in \mathcal{P}$ for $i \in [1..n]$, $p_i \doteq (v_i^+, v_i^-)$ with small abuse of notation.

So as can be observed the function first evaluates a single value for each pro-argument's positive and negative votes via utilizing function 6 which contained multiple desirable properties. Then process simply follows taking the average of these numbers.

So in the envisioned system one dimension of the vector representation of the arguments is saved for the value obtained by pro-arguments. We believe that this value should be acquired in a fair way from the complete set of pro-arguments, and so none of the pro-arguments should dominate the values of the rest by itself. This idea maybe captured with the following property:

Property 6. *"No single pro-argument should be an authority by itself."*
Let $\mathcal{F} = \langle \mathcal{A}, \mathcal{P}, \mathcal{R}, \mathcal{M}_P, \mathcal{V}_\mathcal{A}, \mathcal{V}_P \rangle$ be a tridimensional social argumentation framework and $\mathcal{S} = \langle L, \tau, \psi, \curlywedge, \curlyvee, \neg \rangle$ a well behaved semantics.

$$\nexists p \in \mathcal{P} \text{ s.t. } \forall S \subseteq \mathcal{P}, \ \psi(S \cup \{p\}) = \psi(\{p\}).$$

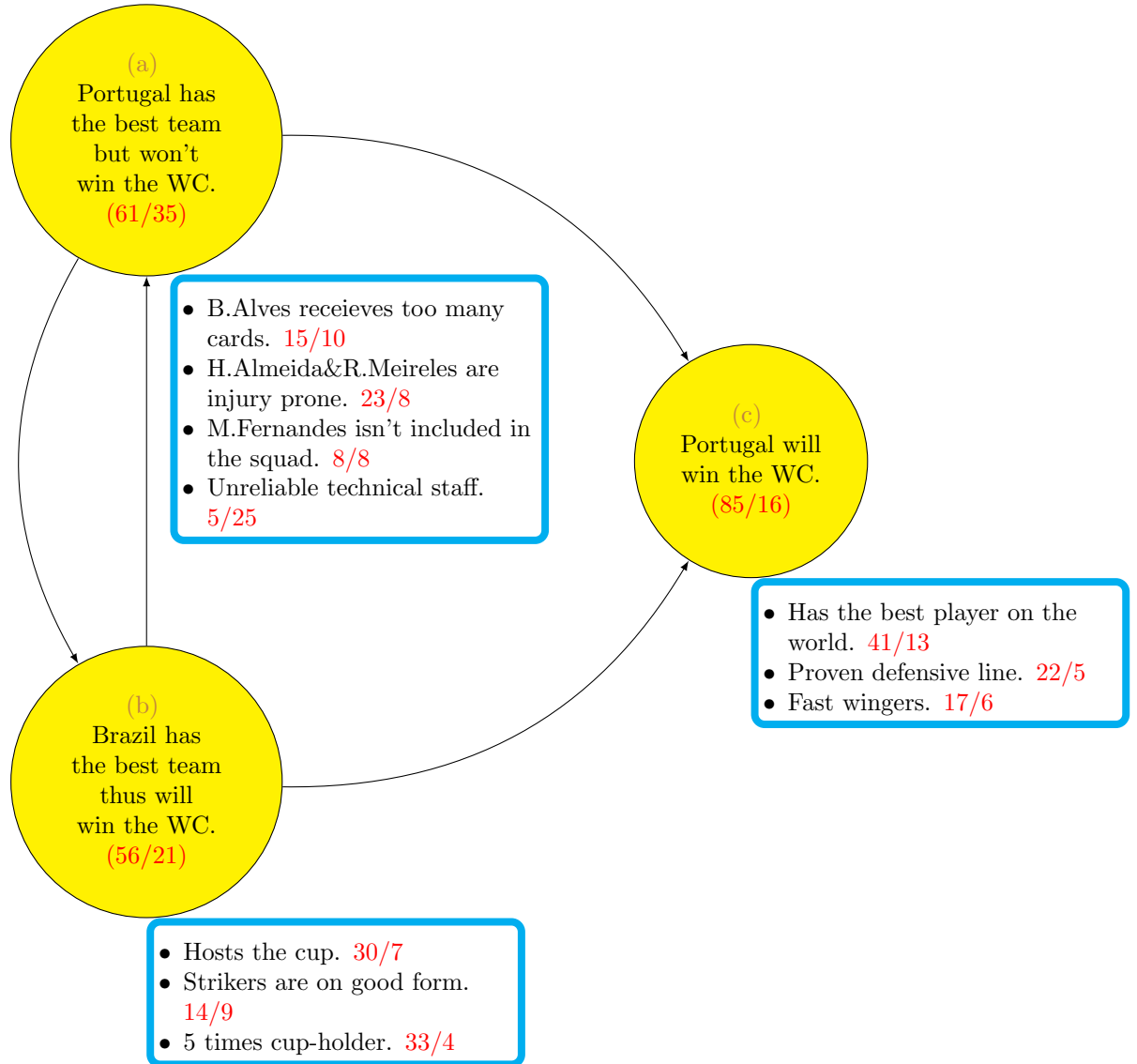**Proposition 3.** *Choice of T-norms(and respective co-norms) as semantics violate Property 6.*

Proof. For the proof it suffices to show that any t-conorm $\perp$ has a zero-element.

$$
\begin{aligned}
\perp(a,b) &= 1 - \top(1-a, 1-b) && \text{, Definition of t-conorm} \\
\perp(a,1) &= 1 - \top(1-a, 1-1) \\
&= 1 - \top(1-a, 0) && \text{, } \top(\text{x}, 0) = 0 \\
&= 1 - 0 \\
&= 1
\end{aligned}
$$

$\square$

Thus as mentioned above, since for any t-conorm there is a zero element(specifically 1), aforementioned property does not hold. If t-conorms where utilized for the aggregation of pro-arguments, one pro-argument with perfect score, or in other words a pro-argument free of negative votes could surpass the rest.

2.3.1 A complete example using 3D representation with pro-arguments

(a)
Portugal has the best team but won't win the WC. (61/35)

- B.Alves receieves too many cards. 15/10
- H.Almeida&R.Meireles are injury prone. 23/8
- M.Fernandes isn't included in the squad. 8/8
- Unreliable technical staff. 5/25

(c)
Portugal will win the WC. (85/16)

- Has the best player on the world. 41/13
- Proven defensive line. 22/5
- Fast wingers. 17/6

(b)
Brazil has the best team thus will win the WC. (56/21)

- Hosts the cup. 30/7
- Strikers are on good form. 14/9
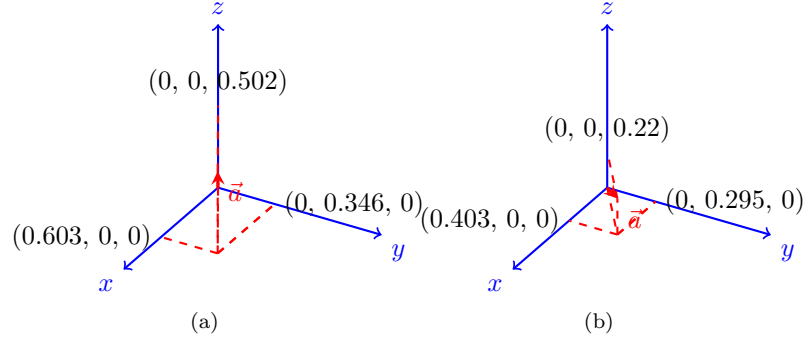- 5 times cup-holder. 33/4

Figure 3: Social support values (left), and social strength values of the dimensions for argument a (right)
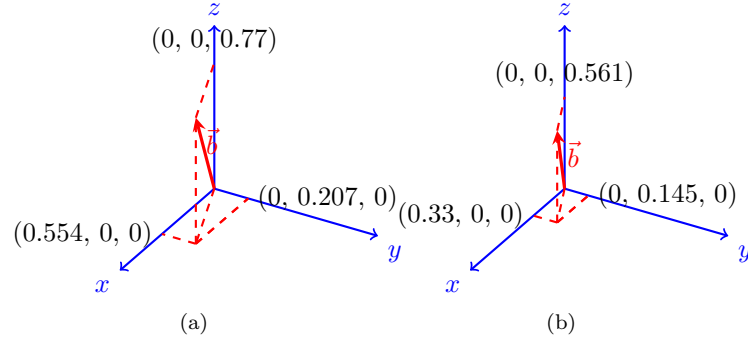

Figure 4: Social support values (left), and social strength values of the dimensions for argument b (right)
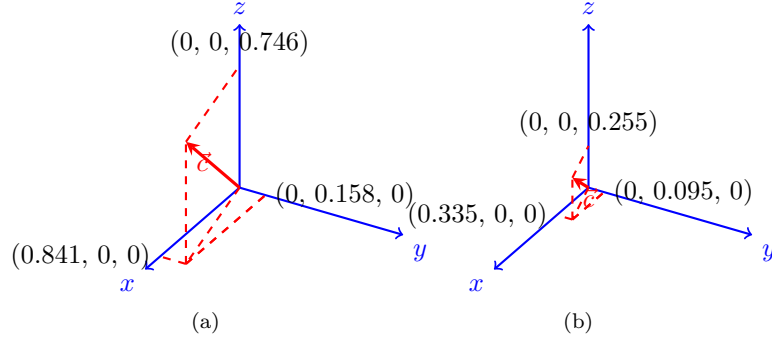

Figure 5: Social support values (left), and social strength values of the dimensions for argument c (right)

As can be seen from the previous example arguments free from attacks main-

tain their initial values for each dimension. This is indeed the desired pattern of behavior since in the absence of a challenger, there is no cause for the values to be diminished.

Property 7. *"An attack-free argument will always be assigned the same outcome as the initial value at every dimension."*
Let $\mathcal{F} = \langle \mathcal{A}, \mathcal{P}, \mathcal{R}, \mathcal{M}_P, V_{\mathcal{A}}, V_P \rangle$ be a tridimensional social argumentation framework, $\mathcal{S} = \langle L, \tau, \psi, \curlywedge, \curlyvee, \neg \rangle$ a well behaved semantics, $\mathcal{M}_{\mathcal{S}}^F$ the set of all $\mathcal{S}$-models of $\mathcal{F}$ and an argument $a \in \mathcal{A}$. $\forall M \in \mathcal{M}_{\mathcal{S}}^F$,

$$(\mathcal{R}^\cdot(a) = \emptyset) \implies \left(M(a) = < \tau(v^+), \tau(v^-), \psi(\mathcal{M}_p^-(a)) >\right)$$

We pursue via depicting that our tridimensional behaves accordingly with respect to the prior property.

Proposition 4. *Tridimensional SAFs obey property 7.*

Proof. It suffices to show that for an arbitrary dimension x of an attack-free argument a, $M(a_x) = a_x$.

Let $a \in \mathcal{A}$ such that $\mathcal{R}^\cdot(a) = \emptyset$ and $M \in \mathcal{M}_{\mathcal{S}}^F$. Then,

$$\begin{aligned}
M(a_x) &= a_x \curlywedge \neg \curlyvee_{b \in \mathcal{R}^\cdot(a)} \left((\tau(b,a)) \curlywedge M(b_x)\right) \\
&= a_x \curlywedge \neg \bot \\
&= a_x \curlywedge \top \\
&= a_x
\end{aligned}$$

$\square$

2.4   Hybrid dimensions

As discussed above, we may also define dimensions as the composition of variables. Here the idea is trying to conceptualize the generic steps that would guide us to a concrete methodology[1], and meanwhile eliminate the shortcoming of the existing aggregation function in taking into account the total number of received votes on an argument and lastly demonstrating on a numerical example.

We know that in ESAFs, the process of evaluating social support already gives us a great taste for the relative measure of of positive(v^+) and negative(v^-) votes of a. So here for simplicity, we choose a bi-dimensional vector representation in the form of $a = [a_x, a_y]$ for each argument, where a_x stands for the magnitude of total votes casted on a, and a_y is computed by boiling down two separate dimensions (i.e. the positive&the negative votes) via utilizing the help of ESAF.

We take a different step from the original work and define the social support as a local valuation, somewhat close to the idea behind the work of Gradual

---

[1]For the sake of brevity, we do quite a bit of "hand-waving" in our formal definitions, where the skipped notions are expected to be sufficiently clear

Valuation. So the social support value of an argument a is affected by the initial strength of a's direct attackers, namely an attacking argument b's own dimension values b_x and b_y. So all in all, to compute the social support of an argument a, we have to evaluated the values of it's dimensions separately via utilizing the votes casted on a, and then diminishing these values with the combined effect of a's direct attackers on each dimensions. In a very closely related manner to the way the model equation is defined in the original work, at this point we may also come up with a generic function by utilizing similar algebraic operators:

Definition 8 (Social support). *Social support of an argument $a \in \mathcal{A}$ is loosely $\tau(a) \triangleq \tau([a_x, a_y]) \triangleq [(\tau(a_x)), (\tau(a_y))]$ where $\tau(a_x)$ (respectively $\tau(a_y)$):*

$$\tau(a_x) = a_x \curlywedge \neg \bigvee_{a_i \in \mathcal{R}^-(a)} (a_i)_x$$

Please note that unlike the model function of the original work, the above function is not a recursive one and thus it does not beget a fixed-point calculation.

Next we pursue by concretely instantiating our operators. The simplest idea would be summing the distinct dimensions of attacking arguments and then subtracting(i.e. adding the negated value of) the aggregated value from the argument of focus. Since we would not expect a_x (the total size dimension) to attain negative values, we will take the maximum of zero and the resulting value when adding the two values (We will follow a similar path for a_y as well. At the first glance the motivation for preventing the value going below zero is not as intuitive as the other case, however as mentioned below this is the focus of the final discussion at the end of the chapter). This simple idea is captured by the following semantics:

Definition 9 (Simple Semantics). *1) $x_1 \curlywedge x_2 = max(0, x_1 +_2)$, 2) $x_1 \curlyvee x_2 = x_1 + x_2$, 3) $\neg x_1 = 1 - x_1$.*

And the resulting social support function(for single dimension):

$$\tau(a_x) = max(0, a_x - \sum_{a_i \in \mathcal{R}^-(a)} (a_i)_x)$$

Obviously this is a very naive approach, which simply equates adding the attacking vectors and then subtracting the combined value from the original argument with respect to the vector algebra. Before choosing the proper semantics, a very diligent investigation process should precede selection regarding figuring the desired properties of the operators. In our original work we choose to work with one specific triangular norm(product t-norm) and its dual t-conorm since they obey our desiderata. For instance T-norm conjunction between $\tau(a)$ and its attack should be a reduced value of $\tau(a)$, which is perfectly in accordance with argumentation intuitions: an attacked argument sees its strength reduced.

Since every t-norm is a function in the form of $T : [0,1] \times [0,1] \to [0,1]$, to be able to utilize them we have to normalize the dimensions. However, that's very easy to achieve in our setting by simply dividing both of the dimensions for each argument via the term $(v_{max}^+ + v_{max}^-)$, which stands for the size magnitude of the argument with the maximum number of total votes in the system. Thus the only case of the existence of an argument enjoying *perfect* social support is when an un-attacked argument has received the maximum number of total votes in the whole system, and all of those votes also have to be positive. As expected, such an argument's polar coordinates would be $(\sqrt{2}, 45°)$ wrt. our definition.

Anyhow, for the sake of a more comprehensive semantics, here we opt to work with another triangular norm, the Hamacher product t-norm for the conjunctive binary operation(and respectively with its t-conorm with respect to the disjunctive operator). The newly defined semantics is as follows:

Definition 10 (Hamacher Semantics)**.**

$$x_1 \curlywedge x_2 = \begin{cases} 0 & (x_1, x_2) = (0,0) \\ \frac{x_1 \cdot x_2}{x_1 + x_2 - x_1 \cdot x_2} & (x_1, x_2) \neq (0,0) \end{cases}$$

$$x_1 \curlyvee x_2 = \begin{cases} 1 & (x_1, x_2) = (1,1) \\ \frac{x_1 + x_2 - 2 \cdot x_1 \cdot x_2}{1 - x_1 \cdot x_2} & (x_1, x_2) \neq (1,1) \end{cases}$$

$$\neg x_1 = 1 - x_1$$

Hamacher product is a strict Archimedean t-norm with some really interesting properties for our application. For instance, very crudely in order to compute a high outcome it requires both of the parameters to be relatively high(inherits from archimedean property). As mentioned in the beginning, the whole motivation behind this line of research is being able to assure that total vote numbers also play an important role in the final outcome in addition to the relative ratio. Thus this kind of operator family, roughly spoken where one parameter does not cover for the other terms in conjunction, can be highly appealing. Moreover it differentiates small changes near zero much better than the Product T-norm, as mentioned earlier which is the choice of binary operator in our original work, which appears to normalise values in that area of its domain.

Property 8. *Chosen semantics should be smooth for each dimension throughout the interval of [0,1].*

Conjecture 6. *Hamacher semantics is a better choice then Product semantics with respect to Property .*

Sketch of proof. Hamacher T-norm is much better suited for situations where arguments have low crowd support. It differentiates small changes near the origin much better than the *Product T-norm* which appears to normalise values in that area of its domain. □
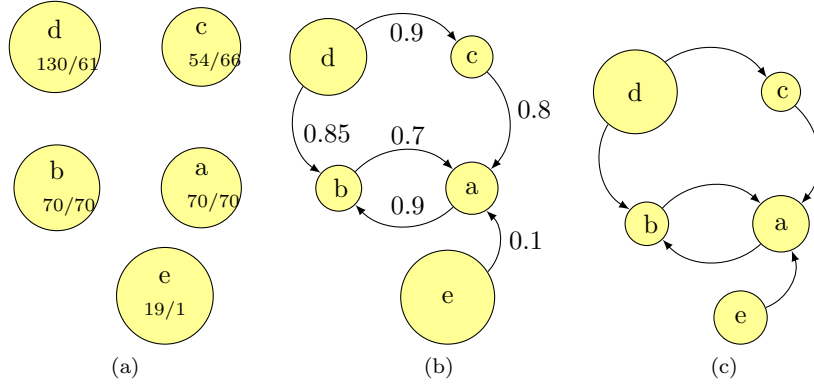
Figure 6: Framework with arguments scaled according to social support (a), ESAF-model of the debate(b), Model of the vector representation(c)

Table 2: Model of the Social Abstract Argumentation Framework considering: a) social support only; b) ESAF; c) Vector representation;

Argument	Fig. a	Fig. b	Fig. c
a	0.50	0.35	0.41
b	0,50	0,14	0.09
c	0,45	0,17	0.11
d	0,68	0,68	0.77
e	0,95	0,95	0.36

Let us see investigate a numerical example, as promised in the beginning, in order to demonstrate the applicability of the method.

At the first glance it seems as once the theoretic foundation is defined, this class of methodologies should not impose a computational overhead to the existing system; since computing the solutions will still be the problem of finding the fix-point of the model equations which is only preceded by a different method for computing social support. That being said, there are some significant pitfalls, sadly which oppose to the very nature of the incentives for using bi-dimensional representation:

1. Firstly, the choice of a bi-dimensional representation makes sense the most for a particular context, where the domains of the vector components are not too restricted in their respective dimensions. So that potentially vector directions can point to every angle in the four quadrants of the cartesian space. In our example, we defined the system in a way that it assures the dimension regarding the size magnitude (x-axis) is alway non-negative. Assume the contrary; so that when an argument a is attacked by a vector-force with a strong component in the x-axis, it may result to a

negative $\tau(a_x)$ value. Users with bad intentions might exploit this notion in the following way: Suppose there is an argument x that you want to defeat(see it's social support diminished). There are mainly two things that you will want to achieve: Firstly, trying to define as many attacks as you can. Followed by attempting to vote on these attacks as many times as possible. For our current topic of interest, it doesn't matter whether you can cast a high number of positive votes on the attack you defined. You are content as long as the total vote number increases. Because every vote casted on its attackers imposes a diminishing effect on $\tau(a_x)$, and since we allowed the component to obtain negative values, eventually the social support value, $\tau(a)$ might be decreased to a value point beyond recovery. On the other hand by restricting the domain of the component only to positive values, we are restricting the vector direction to two quadrants.

2. We have stated that the component in the y-axis of the vectors stand for the relative measure of positive and negative votes; and indeed utilized the original social support definition. In Social Argumentation Frameworks, we prohibit arguments obtaining negative social support values, and behaving as direct supporters to the arguments that they attack as it gives a way to the exploitation of the system. In order to strengthen an argument a user has defined, she may simply introduce a set of illogical arguments and attack relations from these argument to her initial argument. The basic idea is provoking the community and gathering as many negative votes as possible, and thus creating a set of *artificial* supporters for the desired argument. Preventing this type of behavior was exactly our motivation in defining the component of the y-axis as non-negative as well. That being said, by restraining the domain values from acquiring negative values, we also restrict the direction of vectors to a single quadrant, mainly the first quadrant of the 2D cartesian-space.

3. Lastly, we should perhaps consider actions in 2D-space by their very nature. In very simple words, taking a step in the x-axis, and taking a step in y-axis are totally independent, orthogonal actions, whose results then can be interpreted in an aggregated way. However in our setting, even that they are processed in distinct way, both parameters that constitute the values of two dimensions are the same (i.e. v^+ and v^- for a particular argument). Thus in our scenario, it does not seem as possible to claim the same orthogonal fundamentals with respect to definitions of our dimensions.

All in all, this technical exercise has been beneficial in providing us one concrete method of model computation. Even if we decide on abandoning hybrid dimensions for good, a similar methodology may be adopted for other approaches such as the one depicted in Section 2.3. It also seems to do a good job in incorporating the magnitude of total number of votes in the final evaluation.