Balanced Semantics for Argumentation based on Heider's Socio-Psychological Balance Theory

Hajime SAWAMURA [a,1], Jacques RICHE [b], Yutaka OOMIDOU [c], and Takeshi HAGIWARA [a]

[a] *Institute of Science and Technology, Niigata University, Japan*
[b] *Department of Computer Science, KU Leuven, Belgium*
[c] *Graduate School of Science and Technology, Niigata University, Japan*

Abstract. Argumentation, whether philosophical or formal and mathematical, is a discipline of interdisciplinary nature, per se. The recent works on the computational argumentation formalism and their foundations, however, have rested only on logic or logical account. In this paper, we reconsider Dung's seminal argument acceptability notion in the context of Heider's socio-psychological balance theory, where there can be 4 balanced (stable) interaction rules of the form of a triad: (1) the friend of my friend is my friend, (2) the friend of my enemy is my enemy, (3) the enemy of my enemy is my friend, and (4) the enemy of my friend is my enemy. The third one may be a counterpart of Dung's argument acceptability. We propose an innovative argumentation semantics named balanced semantics, taking into account all of the four balanced triads. It naturally leads to an argumentation framework with both attack and support incorporated from the start.

Keywords. Argumentation, argumentation semantics, Heider's balance theory, balanced semantics, PIRIKA

1. Introduction

With the advent of Dung's [2] seminal paper, abstract argumentation semantics has received growing attention from the community of researchers in computational or mathematical argumentation as well as in agent-oriented computing. In the aftermath of its publication, and in various ways, it has provided a tremendous amount of leverage in argumentation research as witnessed by a large corpus of scientific literature. So far, most of these works have centered on Dung's abstract argumentation semantics, and they are mainly focused on the extension and improvement of the Dungean argumentation semantics [8]. However, argumentation per se is a social notion and phenomenon. We have thus felt the need for an alternative and supplemental approach to argumentation from a more sociological point of view, in the same way as agent-oriented computing used to be inspired in terms of a societal view of computation.

[1]Corresponding Author: Hajime Sawamura, 8050, 2-cho, Ikarashi, Nishiku, Niigata, Japan. Tel/Fax: +81 25 262 6753; E-mail: sawamura@ie.niigata-u.ac.jp

In this paper, we will have another look at Dung's argument acceptability by positioning it in the broader context of Heider's socio-psychological balance theory [3][4]. Heider studied a special triadic interaction rule stated in the following form [6]: (1) the friend of my friend is my friend, (2) the friend of my enemy is my enemy, (3) the enemy of my enemy is my friend, and (4) the enemy of my friend is my enemy. These are balanced (stable) interaction rules of the form of a triad. The third one can be seen as a counterpart of the principle of Dung's argument acceptability. It is often described as an old Arabic or Chinese proverb, and a doctrine commonly used in foreign policy. We give an augmented argument acceptability notion, taking into account all of the four balanced triads, and suggest a new direction to computational argumentation research.

The paper is organized as follows. In Section 2, we briefly describe Heider's balance theory. In Section 3, we give an intuitive idea on how to reinterpret Heider's balance theory in the context of computational argumentation. Section 4 is the main part of this paper, where we present an augmented acceptability framework for argumentation based on the considerations of Section 3. The final section includes reflections and implications of the balanced semantics.

2. Heider's Balance Theory

Balance Theory is a motivational theory of attitude change proposed by Fritz Heider [3][4], which conceptualizes the consistency (coherence) motive as a drive toward psychological balance. Heider proposed that sentiment or liking relationships are balanced if the affect valence in a system multiplies out to a positive result.

Let us have a quick look at the balance theory through an example. If a person P segregates waste X for recycling and is in love with a person O, what does P feel upon learning that O does not segregate waste X?

Figure 1. Relationship of P, O and X with a negative multiplicative product

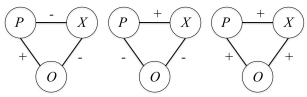

Figure 2. Changing attitude to the balanced triangles with a positive multiplicative product

The person P will perceive imbalance in this relationship. Such an imbalance is depicted as a triangular diagram with a negative multiplicative product in Fig. 1. Then he or she will be motivated to correct the imbalance. The person P can either:

- Decide that waste segregation X may be futile,
- Fall out of love for O, or
- Persuade O that waste segregation X is friendly to the Earth.

Any of these will result in psychological balance (positive multiplicative products), thus resolving the dilemma and satisfying the drive, as depicted in Fig. 2. They actually correspond respectively to (3), (4), and (1) of Heider's verbal form stated in Introduction. For example, the relationship of P and X with a positive sign + in Fig. 1 changes to a negative sign - as seen in Fig. 2, by his or her attitude change from 'segregate waste X for recycling' to 'decide that waste segregation X may be futile'. This corresponds to (3) of the Introduction, the enemy of my enemy is my friend, resulting in a balanced (stable) triangle with a positive multiplicative product.

3. Four Triadic Interactions for Argumentation

As can be seen above, the directionality of the relationships of P, O and X is immaterial in Heider's balance theory. In this section, we reconsider the four triads of the theory by taking into consideration the directionality of the relationship, and we describe the basic ideas for reconstructing the argumentation semantics.

From now on, we use a general notation which no longer refers to P, O and X with specific meanings like person and object. The nodes i, j and k in Figures from 3 to 6 below may be agents, nations, arguments and so on, and the edge r_{ij} between the nodes i and j simply stands for friendly (positive) or hostile (negative) relationships (bonds) among them. In this paper, of course, the nodes stand for arguments.

There can be 4 balanced (stable) interaction rules (of the form of a triad) when we are looking at them through the eyes of the node i.

(1) The friend of my friend is my friend: $r_{ij} > 0 \wedge r_{jk} > 0 \rightarrow r_{ik} > 0$. We call this a Type 1 triad (see the left triad in Fig. 3). The rule says that if i and j are initially friends and the same is true of j and k, these two friendships, i. e., positive relationships, make i feel positive or friendly towards k.

We capture this as in the right triad for argumentation of Fig. 3, where the relationship between each edge is directed with reference to the initiator or the target, and to each node a valuation is associated in which k has values +1,-0, representing that it is a non-attacked argument which is given value +1 (advantage), j has values +1,-0, representing that it has one support from non-attacked argument k and no attack, and i has values +1,-0, representing that it has one support form k and no attack. This support represented in a dotted diagonal line is generated by the relationships of i, j and k in triad $\triangle ijk$. So we no longer need to count the support from j again. As a result, the extension of acceptable arguments can be considered as $\{i, j, k\}$ in Type 1 triad.

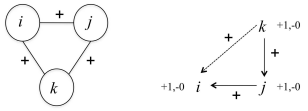

Figure 3. Type 1 triad

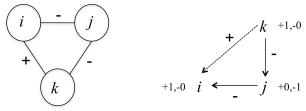

Figure 4. Type 2 triad

(2) The enemy of my enemy is my friend: $r_{ij} < 0 \wedge r_{jk} < 0 \rightarrow r_{ik} > 0$. We call this a Type 2 triad (see the left triad in Fig. 4). The rule says that if j is hostile towards i while k is hostile towards j, i and k are friendly. This case is a triadic paraphrase of the old saying (common wisdom), as stated by Dung [2]: *The one who has the last word laughs best*, which can be actually observed in our daily argumentation as well as in foreign policy, for example.

We capture this as in the right triad for argumentation of Fig. 4, where the relationship between each edge is directed with reference to the initiator or the target, and to each node a valuation is associated in which k has values +1,-0, representing that it is a non-attacked argument and hence has an advantage 1, j has values +0,-1, representing that it has no support and one attack from non-attacked argument k, and i has values +1,-0, representing that it has one support from k and no attack. This support represented in a dotted diagonal line is generated by the relationships of i, j and k in triad $\triangle ijk$. So, the extension of acceptable arguments can be considered as $\{i, k\}$ in Type 2 triad.

Type 2 triad is an empirical social truth or wisdom that has been evolved in various cultural spheres over generations and considered useful by people. Interestingly, such a wisdom often appears in other scientific disciplines such as ecology, sociology, political sciences, etc.

(3) The friend of my enemy is my enemy: $r_{ij} < 0 \wedge r_{jk} > 0 \rightarrow r_{ik} < 0$. We call this a Type 3 triad (see the left triad in Fig. 5). The rule says that if j is hostile towards i, and k, however, is friendly towards j, the enmity between i and j and the friendship between j and k makes i feel hostile towards k.

We capture this as in the right triad for argumentation of Fig. 5, where the relationship between each edge is directed with reference to the initiator or the target, and to each node a valuation is associated in which k has values +1,-0, representing that it is a non-attacked argument and hence has an advantage 1, j has values +1,-0, representing that it has one support from non-attacked argument k and no attack, and i has values +0,-1, representing that it has no support and one attack from k. This attack represented in a dotted diagonal line is generated by the relationships of i, j and k in triad $\triangle ijk$. So, the extension of acceptable arguments can be considered as $\{j, k\}$ in Type 3 triad.

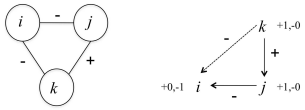

Figure 5. Type 3 triad

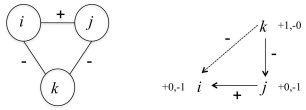

Figure 6. Type 4 triad

(4) The enemy of my friend is my enemy: $r_{ij} > 0 \wedge r_{jk} < 0 \rightarrow r_{ik} < 0$. We call this a Type 4 triad (see the left triad in Fig. 6). The rule says that the friendship between i and j and the enmity between j and k makes i feel hostile towards k.

We capture this as in the right triad for argumentation of Fig. 6, where the relationship between each edge is directed with reference to the initiator or the target, and to each node a valuation is associated in which k has values +1,-0, representing that it is a non-attacked argument and hence has an advantage 1, j has values +0,-1, representing that it has no support and one attack from non-attacked argument k, and i has values +0,-1, representing that it has no support and one attack from k. This attack represented in a dotted diagonal line is generated by the relationships of i, j and k in triad $\triangle ijk$. So, the extension of acceptable arguments can be considered as $\{k\}$ in Type 4 triad.

For actual argument graphs consisting of more attacks and supports, their extensions are to be calculated by the combination of these four types of triads.

4. Balanced Semantics for Argumentation

We have described an intuitive idea for a new argumentation semantics. In this section, we will describe a series of definitions to capture it formally. Heider's balance theory naturally leads to the balanced abstract argumentation framework with both attack and support relation among arguments from the start. So we first begin by extending Dung's abstract argumentation framework so that it incorporates the notion of argument support as follows.

Definition 1 (**Extended Abstract Argumentation Framework**) *The extended abstract argumentation framework $\mathcal{EAAF}$ is a triple $\langle AR, attack, support\rangle$, where AR is a set of arguments, $attack \subseteq AR \times AR$, and $support \subseteq AR \times AR$.*

It should be noted that we do not impose such an independent condition as $attack \cap support = \phi$ as in [1] since for the balanced semantics below, we may allow for *frenemy* that is a portmanteau of 'friend' and 'enemy' that can refer to either an enemy pretending to be a friend or someone who really is a friend but is also a rival.

Definition 2 (**Non-Attacked Arguments**) *Arguments in AR are called non-attacked arguments if and only if they are not attacked by any arguments in AR. NA denotes the set of non-attacked arguments.*

It should be noted that self-defeating arguments are not non-attacked arguments, and non-attacked arguments may be supported by other arguments.

Definition 3 (**Dyadic Relation**) *Let NA be the set of non-attacked arguments in $\mathcal{EAAF}$. The dyadic relation DR in $\mathcal{EAAF}$ is defined to be $\{(a,b) \mid a \in NA \text{ and } (a,b) \in attack \text{ or } support \text{ in } \mathcal{EAAF}\}$.*

That is, $(a,b) \in DR$ if and only if an argument b is attacked or supported by an argument $a \in NA$.

The balanced abstract argumentation framework is now defined by adding to $\mathcal{EAAF}$ the cognitive attack and support relations generated by Heider's balance theory.

Definition 4 (**Balanced Abstract Argumentation Framework**) *The balanced abstract argumentation framework $\mathcal{BAAF}$ is a triple $\langle AR, OR, CR\rangle$, where the original relation OR=$\langle OA, OS\rangle$ with OA (original attack) $\subseteq AR \times AR$ and OS (original support) $\subseteq AR \times AR$. The cognitive relation CR=$\langle CA, CS\rangle$ with CA (cognitive attack) $\subseteq AR \times AR$ and CS (cognitive support) $\subseteq AR \times AR$.*

OR corresponds to a tuple of attack and support in $\mathcal{EAAF}$. CR consists of cognitive attack and support newly generated by the method described in what follows. In the definition, we used the term 'cognitive' to signify a tacit attack and support in human cognition or the intentionality of attack and support. The balanced abstract argumentation framework is represented as a directed graph in an obvious way, similarly to the standard abstract argumentation framework.

4.1. Cognitive relation generated by four triadic interactions

In this subsection, we describe how to derive the cognitive attack and support relation, taking into account Heider's socio-psychological balance theory. Although there is no notion of directionality in the original Heider's balance theory, for argumentation, we need to consider directionality of the attack and support relations. Thus, we deal with four triadic interactions: the Type 1-4 triads described in Section 3 in a directional form as follows.

Definition 5 (**Four triads with directionality**) *For a given argument A, there can be four possible cognitive attacks or supports stipulated in terms of Type 1 triad, Type 2 triad, Type 3 triad and Type 4 triad in a way such as described in Fig. 7, where at least one of two sides of each right triangle which is drawn with the solid line is the original attack or support, and the dotted diagonal line represents the cognitive attack or support to be generated.*

Note that two sides of each right triangle have the directionality of the attack or support relation, in other words, they represent a flow of arguments.

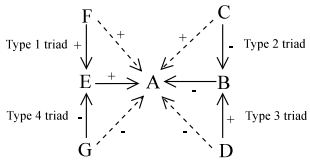

Figure 7. Four triads

4.1.1. Duplication of cognitive attacks and supports

Definition 6 (**Reduction of duplication**) *When cognitive attacks and supports with the same direction and sign are generated between two arguments, they are reduced to one. Cognitive attacks and supports with different directions or signs are left as they stand.*

4.1.2. Deriving cognitive relation

Definition 7 (**Derivative cognitive relation**) *Let* $\mathcal{BAAF}$=$\langle AR, OR, CR\rangle$,

- CR^0 *= the set of cognitive attacks or supports generated by the application of Definition 5 to OR in* $\mathcal{BAAF}$
- CR^{i+1}*= the union of* CR^i *with the set of cognitive attacks or supports generated by the application of Definition 5 to the combination of the elements of* CR^i *with the elements of OR in* $\mathcal{BAAF}$*, for* $i \geq 0$*.*

Then, CR *in* $\mathcal{BAAF}$ *is defined as* $CR = \bigcup_{i\geq 0} CR^i$.

It should be noted that Definition 5 is not allowed to apply to any two elements in CR^i $(i > 0)$ since we think that such an application turns out to weaken the socio-psychological or semantical relationship of attack and support among arguments.

4.2. Argument acceptability in $\mathcal{BAAF}$

We use the notions of NA and DR in $\mathcal{BAAF}$ as well as in $\mathcal{EAAF}$.

Definition 8 (**Strength of arguments**) *The strength of an argument* A *is defined to be the sum (=-l+m+-n+o) of the following values:*

- *-l if the number of the cognitive attacks for the argument* A *is l,*
- *m if the number of the cognitive supports for the argument* A *is* m*,*
- *-n if the number of the attacks from* NA *for the argument* A *is n,*
- *o if the number of the supports from* NA *for the argument* A *is o,*

provided that the arguments in NA are given an advantage 1 in advance.

In this definition, we have not explicitly taken into account the original relation OR since CR has been generated by taking in the original information and effect that OR had, as described in Definition 7. On the other hand, NA is obviously weighty for the strength of an argument since it has no attacks from other arguments, or rather receives supports as defined in Definition 2. This is a reason why we provided arguments in NA with an advantage 1.

Definition 9 (**Argument acceptability and balanced extension**) *Argument A is acceptable in $\mathcal{BAAF}$ if and only if the strength of the argument A is greater than 0. The set of acceptable arguments is called balanced extension (BE) in the balanced semantics.*

Example 1 *Let us consider the argument graph in the left side of Fig. 8. Since argument A is in NA, it is associated with +1. Argument B has -1 since it is attacked by A in NA, that is, $(A, B) \in DR$. Argument C has +1 since it receives a cognitive support from A, i.e., (A, C) $\in CS$. Argument D has +1 since (B, D)$\in CS$ and -1 since it receives a cognitive attack from A, i.e., (A, D)$\in CA$. We represent these analyses as in the right side of Fig. 8 in which we denote CA and CS simply by CR, and call it the valuation for each argument. The balanced extension is therefore $\{A, C\}$. In this example, it coincides with the grounded extension for the argumentation framework with OA only.*

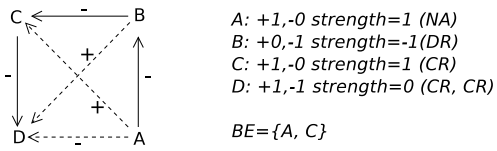

Figure 8. Valuation and acceptability of arguments

4.3. Some pathological or baffling arguments

We have presented the basic part of the balanced semantics based on Heider's balance theory. So far so good. In the argument community, there are so many well-known pathological arguments [7] that deserve attention and should be challenged. At this point, we address the question how to deal with those baffling cases in the balanced abstract argumentation framework, and we confirm its expressiveness at this stage.

4.3.1. Bi-directional attack

Even cycle [7] Triads are basic constituents for the balanced semantics. The even cycle $A \overset{-}{\longleftrightarrow} B$ does not have any explicit form of triads. However, we can calculate its valuation for each node simply by applying the notion of NA as A: +0, -0 $strength$=0 and B: +0, -0 $strength$=0. The balanced extension is thus ϕ, which coincides with the grounded extension.

Zombie argument [7] The balanced extension for Zombie argument coincides with the grounded extension ϕ.

4.3.2. Self-defeating argument

The self-defeating argument is one that attacks itself. We uncoil this as $A \overset{-}{\longleftrightarrow} A$. Then its valuation is A: +0, -0 $strength$=0, resulting in the extension ϕ.

4.4. Conflict resolution in extension

In Dung's argumentation semantics, it was essential or absolute that the extension be conflict-free. In the balanced abstract argumentation framework, however, it is not a primary requirement, but a collateral one to be restored later. For conflict resolution in extension, we introduce the following definition.

Definition 10 (**Conflict resolution**) *Let E be a balanced extension which includes a conflicting pair of arguments A with strength s_1 and B with strength s_2 such that $(A, B) \in OA$ (original attack) or CA (cognitive attack). If $s_1 \geq s_2$, then $E' = \{A\} \cup S$, where $S \subseteq E$ is the set of arguments whose elements do not have the attack relation with A, else $E' = \{B\} \cup S$, where $S \subseteq E$ is the set of arguments whose elements do not have the attack relation with B. If E' is conflict-free, then we let E' be a conflict-resolved extension. Otherwise, we repeat the above process for the other conflicting pair of arguments in E' until conflicts are fully resolved.*

4.5. Presence of imbalanced triads

The balanced abstract argumentation frameworks may include imbalanced triads from the start. In the balanced abstract argumentation frameworks with imbalanced triads, there appear pairs of arguments such that $(A, B) \in OR$, and $(A, B) \in CR$ or $(B, A) \in CR$. For the imbalanced triads, we need a special handling.

Definition 11 (**Undercutting cognitive sign**) *The pair $(A, B) \in CR$ generated in the imbalanced triads is counted as invalid.*

5. Concluding Remark and Future Work

In this paper, we preferred the socio-psychological view to a logical view, and considered a new acceptability of argumentation based on it. We described an initial attempt at a new acceptability notion for argumentation by observing that Dung's starting idea for argument acceptability is one of the four-cornered alternatives based on Heider's balance theory for the socio-psychological relation. The mindset of the Dungean argumentation semantics and ours are very different. The mathematical notions such as ordering, maximality, fixpoint theory, etc. play a crucial role in his theory construction to stipulate argument acceptability. On the other hand, the mindset of our approach based on Heider's balance theory in socio-psychology consists of the comparison of the number of yeas and nays (majority principle in democracy or collective choice theory), postponed conflict-freeness, etc. However, it is interesting to know that both bring us almost the same result for the unipolar argumentation framework. The PIRIKA system [5] with the idiosyncrasy of $\mathcal{BAAF}$ is currently under development. [2]

References

[1] L. Amgoud, C. Cayrol, M.-C. Lagasquie-Schiex, and P. Livet. On bipolarity in argumentation frameworks. *Int. J. Intell. Syst.*, 23(10):1062–1093, 2008.

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

[3] F. Heider. Attitudes and cognitive organizations. *Journal of Psycology*, 21:107–112, 1946.

[4] F. Heider. *The Psychology of Interpersonal Relations*. John Wiley, 1958.

[5] Y. Katsura, H. Sawamura, T. Hagiwara, and J. Riche. Asynchronous argumentation with pervasive personal communication tools. In *Proceedings of the 6th International Conference on Agents and Artificial Intelligence*, pages 105–114. SciTePress, 2014.

[6] S. C. Lee, R. G. Muncaster, and D. A. Zinnes. The friend of my enemy is my enemy: Modeling triadic international relationships. *Synthese*, 100:333–358, 1994.

[7] H. Prakken and G. Vreeswijk. Logical systems for defeasible argumentation. In *D. Gabbay and F. Guenthner, editors, Handbook of Philosophical Logic*, pages 219–318. Kluwer, 2002.

[8] I. Rahwan and G. R. Simari, editors. *Argumentation in Artificial Intelligence*. Springer, 2009.

[2] An acronym for PIlot for the RIght Knowledge and Argument. The open source software and the video clip of PIRIKA on iPad are available at URL `http://pirika.cs.ie.niigata-u.ac.jp/pirika_project/index.html`. PIRIKA is now freely available from the Apple store.