Equilibria in Quantitative Bipolar Argumentation Dialogues

Extended Abstract

Andreas Brännström
Umeå University
Umeå, Sweden
andreas.brannstrom@umu.se

Timotheus Kampik
Umeå University
Umeå, Sweden
timotheus.kampik@umu.se

ABSTRACT

We introduce MQBAFs, multi-agent extensions of quantitative bipolar argumentation frameworks (QBAFs). In MQBAFs, agents have objectives to maximise or minimise the strengths of some arguments, and establish preferences over these objectives. Agents make utterances by adding or removing arguments or attacks to/from a QBAF, or by changing arguments' initial strengths. We then define equilibria for MQBAFs, in which no rational agent would make any additional utterance. The notion of MQBAF opens quantitative bipolar argumentation to principled strategic analysis.

KEYWORDS

Formal Argumentation, Rationality, Equilibrium

ACM Reference Format:
Andreas Brännström and Timotheus Kampik. 2026. Equilibria in Quantitative Bipolar Argumentation Dialogues: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/VBDV4500

1 INTRODUCTION

Formal argumentation [1, 2, 5, 19] studies inference over graphs representing arguments and their relationships. Modeling inference as a dynamic exchange of arguments until a conclusion is reached is natural from a human perspective. Thus, formal argumentation may help bridge the gap between human and machine reasoning [13], for example by allowing humans to *contest* claims made by machines [17]. In recent years, the *quantitative bipolar argumentation* variant has attracted considerable attention. Here, arguments with numerical *base scores* are connected by support and attack relations, and a semantics determines their final strengths based on the graph topology and base scores in a *Quantitative Bipolar Argumentation Framework* (QBAF) [1, 5, 11]. A *QBAF* is a tuple $\langle X, R^-, R^+, \tau \rangle$ where X is a finite set of arguments, $R^- \subseteq X \times X$ is a binary (attack) relation, $R^+ \subseteq X \times X$ is a binary (support) relation, and $\tau : X \longrightarrow [0,1]$ is a total function; for any $x \in X$, we call $\tau(x)$ the base score of x. For $x \in X$, let $R^-(x) = \{y \mid (y,x) \in R^-\}$ and $R^+(x) = \{y \mid (y,x) \in R^+\}$. Given a QBAF $Q = \langle X, R^-, R^+, \tau \rangle$, a (possibly partial) strength function is defined as $\sigma : X \longrightarrow [0,1]$, where $\sigma_Q(x)$ denotes the strength of $x \in X$ w.r.t. Q.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/VBDV4500

Although an assumption of formal argumentation is its multi-agent nature [6, 15, 21], the exchange of arguments between agents is left implicit in QBAFs. This raises the question of how strategic considerations affect settings where agents alter the graph to maximize or minimize the strength of certain *topic* arguments.

In this paper, we introduce *multi-agent QBAFs* (MQBAFs), where agents pursue maximisation or minimisation objectives over arguments. A QBAF reaches an *equilibrium* whenever no agent can advance its objectives by a new move. MQBAFs can, for example, address the need for principled methods to design and verify Dispute Trees (DTs) for argumentation-based explanations [9, 20]. Intuitively, such DTs can be viewed as an MQBAF with a single topic argument and two agents, i.e., a *proponent* who aims to maximize the strength of the topic and an *opponent* aiming to minimize it. With MQBAFs, we can assess the rationality of agents constructing a DT, which can then be considered a soundness property.

2 MULTI-AGENT QBAFS

In the setting of a global QBAF $Q = \langle X, R^-, R^+, \tau \rangle$, let X denote the vocabulary of available arguments, and R^- and R^+ the potential attack and support relations among them. We consider a finite set of agents A, where each agent $Ag \in A$ is associated with (i) a set of objectives specifying arguments whose strength it aims to increase or decrease, and (ii) a preference ordering over these objectives. An *utterance profile* maps each uttered argument to an agent. Given such a profile, the corresponding *induced QBAF* is obtained by restricting X, R^-, R^+, and τ to the uttered arguments, forming the basis of a *Multi-Agent QBAF (MQBAF)*.

We next define the objective function that specifies, for each agent, the arguments it seeks to increase or decrease.

Definition 2.1 (Objectives). Let $Q = \langle X, R^-, R^+, \tau \rangle$ be a QBAF, A be a finite set of agents, and σ a strength function. An *objective function* is a mapping $obj : A \longrightarrow 2^{X \times \{+,-\}}$ where $\forall Ag \in A : obj(Ag) \subseteq X \times \{+,-\}$ is non-empty.

Given an *objective function* obj and an agent $Ag \in A$:

- $(x,+) \in obj(Ag)$ means Ag aims to maximise $\sigma(x)$;
- $(x,-) \in obj(Ag)$ means Ag aims to minimise $\sigma(x)$.

Agents may hold multiple objectives. A preference function captures the priority ordering among these objectives.

Definition 2.2 (Preferences). Let $Q = \langle X, R^-, R^+, \tau \rangle$ be a QBAF, A be a finite set of agents, σ be a strength function, and obj be an objective function. A *preference function* is a function $pref : A \longrightarrow 2^{(X \times \{+,-\}) \times (X \times \{+,-\})}$ where $\forall Ag \in A : pref(Ag) \subseteq obj(Ag) \times obj(Ag)$, and $pref(Ag)$ is assumed to be a total preorder on $obj(Ag)$, such that for all $(x,d), (y,d'), (z,d'') \in obj(Ag)$:

(1) $(x,d) \preceq_{Ag} (x,d)$ (reflexivity),

(2) if $(x,d) \preceq_{Ag} (y,d')$ and $(y,d') \preceq_{Ag} (z,d'')$ then $(x,d) \preceq_{Ag} (z,d'')$ (transitivity),
(3) $(x,d) \preceq_{Ag} (y,d')$ or $(y,d') \preceq_{Ag} (x,d)$ (totality).

We can now define MQBAF as a structure of agents, arguments, objectives, and preferences as a shared framework.

Definition 2.3 (Multi-Agent QBAF). Let $Q = \langle X, R^-, R^+, \tau\rangle$ be a QBAF, A be a finite set of agents, obj be an objective function, and $pref$ be a preference function. A *Multi-Agent QBAF (MQBAF)* is a tuple

$$M = (\langle X, R^-, R^+, \tau\rangle, A, obj, pref)$$

where $\forall Ag \in A : obj(Ag) \subseteq X \times \{+,-\}$ and $\forall Ag \in A : pref(Ag) \subseteq obj(Ag) \times obj(Ag)$. We constrain obj by requiring that $\forall Ag \in A$, $\forall x \in X$ if $(x,+) \in obj(Ag)$ then $(x,-) \notin obj(Ag)$ and vice versa.

This structure provides a setting for analysing agents' decisions over argument strengths. By considering utterances of agents, we analyse rationality and equilibrium.

An utterance profile (static and complete) specifies which agent puts forward which arguments.

Definition 2.4 (Utterance Profile). Let $M = (\langle X, R^-, R^+, \tau\rangle, A, obj, pref)$ be an MQBAF. An *utterance profile for* M is a partial function $utt^M : X \rightharpoonup A$, where $utt^M(x) = a$ means that agent $Ag \in A$ utters argument $x \in X$ w.r.t. M. When M is clear from the context, we simply write utt.

To analyse only the arguments explicitly introduced by agents, in contrast to the global QBAF containing all available arguments, we construct the so-called induced QBAF.

Definition 2.5 (Induced QBAF). Let $M = (\langle X, R^-, R^+, \tau\rangle, A, obj, pref)$ be an MQBAF and let $utt : X \to A$ be an utterance profile. The *QBAF induced by* utt is the tuple $G_{utt} = \langle X', R^{-'}, R^{+'}, \tau'\rangle$, where $X' = \{x \in X \mid \exists Ag \in A : utt(x) = Ag\}$, $R^{-'} = R^- \cap (X' \times X')$, $R^{+'} = R^+ \cap (X' \times X')$, and $\tau' = \tau \cap (X' \times [0,1])$.

We can now define preference relations over induced QBAFs. For each agent $Ag \in A$, the relation $\preceq_{(x,d)}$ for $(x,d) \in obj(Ag)$ by setting that Ag *weakly prefers* G_1 over G_2, denoted $G_1 \preceq_{(x,d)} G_2$, iff ($d = +$ and $\sigma_{G_1}(x) \geq \sigma_{G_2}(x)$) or ($d = -$ and $\sigma_{G_1}(x) \leq \sigma_{G_2}(x)$), and define that Ag *weakly prefers* G_1 over G_2 w.r.t. its objective set, denoted $G_1 \preceq_{obj(Ag)} G_2$, iff for every $(x,d) \in obj(Ag)$ either $G_1 \preceq_{(x,d)} G_2$ holds or there exists $(y,d') \in obj(Ag)$ with $(y,d') \prec_{Ag} (x,d)$ such that $G_1 \preceq_{(y,d')} G_2$.

Rationality for an agent's utterance requires that, given the utterances of all other agents, the agent cannot improve its evaluated outcome by unilaterally changing its utterance.

Definition 2.6 (Utterance Rationality). Let $M = (\langle X, R^-, R^+, \tau\rangle, A, obj, pref)$ be an MQBAF, utt be an utterance profile, and G_{utt} its induced QBAF. utt is *rational for agent* $Ag \in A$ iff, for every *unilateral deviation* utt^* such that $\forall x \in X \setminus \{y \in X \mid utt(y) = Ag\} : utt^*(x) = utt(x)$, for every G_{utt^*} that is the induced QBAF of utt^* it, holds that agent Ag weakly prefers G_{utt} to G_{utt^*}.

Utterance equilibria characterise utterance profiles where no agent has an incentive to deviate.

Definition 2.7 (Utterance Equilibrium). Let $M = (\langle X, R^-, R^+, \tau\rangle, A, obj, pref)$ be an MQBAF and utt an utterance profile. utt is an *utterance equilibrium* iff utt is rational for every agent $Ag \in A$.

An utterance equilibrium, evaluated via induced QBAFs and agents' objectives, ensures every agent's current utterance is at least as good as any unilateral deviation, illustrated next.

Example 2.8. Let $M = \langle X, R^-, R^+, \tau\rangle, A, obj, pref\rangle$ be an MQBAF with $X = \{a,b,c\}$, $R^- = \{(c,b)\}$, $R^+ = \{(a,b), (c,a)\}$, and $\tau = \{(a,0.4), (b,0.2), (c,0.3)\}$. Let $A = \{Ag1, Ag2\}$, $obj(Ag1) = \{(a,+), (b,+)\}$ and $pref(Ag1) = \{((a,+),(b,+))\}$; let $obj(Ag2) = \{(a,+)\}$ and $pref(Ag2) = \emptyset$. Let σ follow DFQuAD semantics [19].

Fix the utterance profile $utt = \{(c,Ag1), (a,Ag2), (b,Ag2)\}$. For $Ag1$'s possible deviation, consider $utt^* = utt \setminus \{(c,Ag1)\}$. The corresponding induced QBAFs are:

- $G_{utt} = \langle\{a,b,c\},\ \{(c,b)\},\ \{(a,b),(c,a)\},\ \{(a,0.4),(b,0.2),(c,0.3)\}\rangle$,
- $G_{utt^*} = \langle\{a,b\},\ \emptyset,\ \{(a,b)\},\ \{(a,0.4),(b,0.2)\}\rangle$.

In G_{utt} we obtain $\sigma_{G_{utt}}(c) = 0.3$, $\sigma_{G_{utt}}(a) = 0.58$, and $\sigma_{G_{utt}}(b) = 0.424$. In G_{utt^*} we have $\sigma_{G_{utt^*}}(a) = 0.4$ and $\sigma_{G_{utt^*}}(b) = 0.52$. Thus the changes from G_{utt^*} to G_{utt} are $\sigma_{G_{utt}}(a) - \sigma_{G_{utt^*}}(a) = 0.18$, and $\sigma_{G_{utt}}(b) - \sigma_{G_{utt^*}}(b) = -0.096$.

By $((a,+),(b,+))$, we have $0.18 > -0.096$, so $Ag1$ strictly prefers G_{utt} to G_{utt^*}. Hence $Ag1$ has no profitable unilateral deviation. For $Ag2$'s possible deviations, we keep $utt(c) = Ag1$ fixed and consider: $utt_1^{**} := utt \setminus \{(b,Ag2)\}$, $utt_2^{**} := utt \setminus \{(a,Ag2)\}$, $utt_3^{**} := utt \setminus \{(a,Ag2),(b,Ag2)\}$.

The corresponding induced QBAFs are:

- $G_1 = \langle\{a,c\},\ \emptyset,\ \{(c,a)\},\ \{(a,0.4),(c,0.3)\}\rangle$,
- $G_2 = \langle\{b,c\},\ \{(c,b)\},\ \emptyset,\ \{(b,0.2),(c,0.3)\}\rangle$,
- $G_3 = \langle\{c\},\ \emptyset,\ \emptyset,\ \{(c,0.3)\}\rangle$.

In G_{utt} we have $\sigma_{G_{utt}}(a) = 0.58$. In G_1, $\sigma_{G_1}(a) = 0.58$. In G_2, $a \notin X$ so $\sigma_{G_2}(a) = 0$. In G_3, $a \notin X$ so $\sigma_{G_3}(a) = 0$. Since $obj(Ag2) = \{(a,+)\}$, $Ag2$ weakly prefers G_{utt} to all alternatives (tie for $\{a,c\}$, strictly better than $\{b,c\}$ or $\{c\}$). Thus, since no agent has a profitable unilateral deviation, the utterance profile $utt = \{(c,Ag1), (a,Ag2), (b,Ag2)\}$ is an utterance equilibrium.

3 DISCUSSION AND RELATED WORK

We have used quantitative bipolar argumentation [1, 5, 19] as a substrate for modelling multi-agent interaction with objectives and preferences, capturing agents' maximisation and minimisation goals, and equilibrium notions for strategic stability. In contrast to qualitative treatments of preferences and strategy [2, 3, 10, 16], MQBAFs provide the first systematic account of equilibria under quantitative argumentation with quantitative objectives. Strategic dialogue formalism has been extensively explored [4, 6, 7, 12, 14]. More recent work has proposed argumentation with beliefs [21], later extended to QBAFs with belief [8], conceptually close to our approach of modelling objectives.

We restricted attention to strategies with myopic deviations, leaving aside more refined notions of rationality. We also assumed observable moves and fixed objectives. Extensions could consider game-theoretic refinements such as mixed or subgame-perfect equilibria [22]. Moreover, future work can extend MQBAFs to formal dialogue games [6, 7, 18], where dialogue protocols may define admissible changes to a QBAF. This would enable analysing how the existence of equilibrium depends on protocol constraints, allowing us to highlight interesting intricacies in MQBAF dialogues.

ACKNOWLEDGMENTS

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP).

REFERENCES

[1] Leila Amgoud and Jonathan Ben-Naim. 2018. Evaluation of arguments in weighted bipolar graphs. *International Journal of Approximate Reasoning* 99 (2018), 39–55. https://doi.org/10.1016/j.ijar.2018.05.004

[2] Leila Amgoud and Claudette Cayrol. 2013. On the acceptability of arguments in preference-based argumentation. *arXiv preprint arXiv:1301.7358* (2013).

[3] Leila Amgoud and Srdjan Vesic. 2014. Rich preference-based argumentation frameworks. *International Journal of Approximate Reasoning* 55, 2 (2014), 585–606.

[4] Pietro Baroni, Giulia Comini, Antonio Rago, and Francesca Toni. 2017. Abstract games of argumentation strategy and game-theoretical argument strength. In *International Conference on Principles and Practice of Multi-Agent Systems*. Springer, 403–419.

[5] Pietro Baroni, Antonio Rago, and Francesca Toni. 2019. From Fine-Grained Properties to Broad Principles for Gradual Argumentation: A Principled Spectrum. *International Journal of Approximate Reasoning* 105 (feb 2019), 252–286. https://doi.org/10.1016/j.ijar.2018.11.019

[6] Elizabeth Black and Anthony Hunter. 2009. An inquiry dialogue system. *Autonomous Agents and Multi-Agent Systems* 19 (2009), 173–209.

[7] Andreas Brännström, Virginia Dignum, and Juan Carlos Nieves. 2025. Goal-hiding information-seeking dialogues: A formal framework. *International Journal of Approximate Reasoning* 177 (2025), 109325.

[8] Anders Brännström, Chiaki Sakama, and Juan Carlos Nieves. 2025. Formal Verification of Manipulation Dialogues. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025)*. IFAAMAS.

[9] Kristijonas Čyras, Timotheus Kampik, and Qingtao Weng. 2022. Dispute trees as explanations in quantitative (bipolar) argumentation. In *ArgXAI 2022, 1st International Workshop on Argumentation for eXplainable AI, Cardiff, Wales, September 12, 2022*, Vol. 3209. CEUR-WS.

[10] Kristijonas Čyras and Francesca Toni. 2016. ABA+: assumption-based argumentation with preferences. *arXiv preprint arXiv:1610.03024* (2016).

[11] Louise Dupuis De Tarlé, Elise Bonzon, and Nicolas Maudet. 2022. Multiagent dynamics of gradual argumentation semantics. In *21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022)*.

[12] Joseph Devereux and Chris Reed. 2009. Strategic argumentation in rigorous persuasion dialogue. In *International Workshop on Argumentation in Multi-Agent Systems*. Springer, 94–113.

[13] Emmanuelle Dietz, Antonis C. Kakas, and Loizos Michael. 2022. Argumentation: A calculus for Human-Centric AI. *Frontiers Artif. Intell.* 5 (2022). https://doi.org/10.3389/FRAI.2022.955579

[14] Guido Governatori. 2021. Strategic argumentation. *Journal of Applied Logics* (2021).

[15] Emmanuel Hadoux, Anthony Hunter, and Sylwia Polberg. 2023. Strategic argumentation dialogues for persuasion: Framework and experiments based on modelling the beliefs and concerns of the persuadee. *Argument & Computation* 14, 2 (2023), 109–161.

[16] Souhila Kaci and Leendert van der Torre. 2008. Preference-based argumentation: Arguments supporting multiple values. *International Journal of Approximate Reasoning* 48, 3 (2008), 730–751.

[17] Francesco Leofante, Hamed Ayoobi, Adam Dejl, Gabriel Freedman, Deniz Gorur, Junqi Jiang, Guilherme Paulino-Passos, Antonio Rago, Anna Rapberger, Fabrizio Russo, Xiang Yin, Dekai Zhang, and Francesca Toni. 2024. Contestable AI Needs Computational Argumentation. https://doi.org/10.24963/KR.2024/83

[18] Peter McBurney and Simon Parsons. 2009. Dialogue games for agent argumentation. In *Argumentation in artificial intelligence*. Springer, 261–280.

[19] Nico Potyka. 2019. Extending Modular Semantics for Bipolar Weighted Argumentation. In *18th International Conference on Autonomous Agents and MultiAgent Systems*, Noa Agmon, Edith Elkind, Matthew E. Taylor, and Manuela Veloso (Eds.). IFAAMAS, Montreal, 1722–1730.

[20] Antonio Rago, Hengzhi Li, and Francesca Toni. 2023. Interactive Explanations by Conflict Resolution via Argumentative Exchanges. In *Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning*. 582–592. https://doi.org/10.24963/kr.2023/57

[21] Chiaki Sakama. 2024. Argumentation and Belief. In *Proceedings of the 10th International Conference on Computational Models of Argument (COMMA 2024)*.

[22] Bruno Yun, Srdjan Vesic, and Nir Oren. 2022. Representing pure Nash equilibria in argumentation. *Argument & Computation* 13, 2 (2022), 195–208.