Integrating User Preferences into Gradual Bipolar Argumentation for Personalised Decision Support

Elisa Battaglia[1], Pietro Baroni[1(✉)], Antonio Rago[2], and Francesca Toni[2]

[1] DII, University of Brescia, Brescia, Italy
e.battaglia001@studenti.unibs.it, pietro.baroni@unibs.it
[2] Department of Computing, Imperial College London, London, UK
{a.rago,ft}@imperial.ac.uk

Abstract. Gradual bipolar argumentation has been shown to be an effective means for supporting decisions across a number of domains. Individual user preferences can be integrated into the domain knowledge represented by such argumentation frameworks and should be taken into account in order to provide personalised decision support. This however requires the definition of a suitable method to handle user-provided preferences in gradual bipolar argumentation, which has not been considered in previous literature. Towards filling this gap, we develop a conceptual analysis on the role of preferences in argumentation and investigate some basic principles concerning the effects they should have on the evaluation of strength in gradual argumentation semantics. We illustrate an application of our approach in the context of a review aggregation system, which has been enhanced with the ability to produce personalised outcomes based on user preferences.

Keywords: Gradual argumentation · Preferences · Decision support

1 Introduction

Bipolar Argumentation Frameworks (BAFs) [9] endowed with a gradual argumentation semantics have been shown to provide a suitable formal basis for the development of applications for decision support in a variety of contexts, such as the evaluation of design alternatives [6], multiparty cooperative work [4], forecasting [13] and review aggregation [10]. In a nutshell, BAFs provide an argumentative representation of the network of reasons underlying the uncertain assessment of a given issue, which are related by attack and support relations. A gradual argumentation semantics provides a numerical assessment of the *strength* of the arguments belonging to a BAF. Strength values may then be used as the basis for informed decisions. When decision support concerns some personal choice (e.g. the selection of a product to purchase or a movie to watch) the issue of providing personalised outcomes, taking into account different user

S. Destercke et al. (Eds.): SUM 2024, LNAI 15350, pp. 14–28, 2025.
https://doi.org/10.1007/978-3-031-76235-2_2

preferences, emerges. In the formal context sketched above, this requires user preferences in gradual argumentation semantics for BAFs to be considered, a problem which has not been addressed in previous literature.

This work contributes to fill this gap by investigating some general principles concerning the effects that user preferences should have on the evaluation of argument strength in gradual argumentation semantics and illustrating their application in an enhanced version of the ADA review aggregation system [10].

The paper is organised as follows. After recalling some background notions in Sect. 2, we carry out in Sect. 3 a conceptual analysis on the role of preferences in formal argumentation, pointing out the different natures and uses that can be found in the literature. We then investigate in Sect. 4 general principles concerning the role of user preferences in the evaluation of argument strength and illustrate an application of the proposed approach in Sect. 5, while Sect. 6 concludes.

2 Background

Our work lies in the family of abstract argumentation formalisms, which are focused on the evaluation of the acceptability of arguments based on the relations among them. Dung's argumentation framework [11], considering only an attack relation between arguments, is the simplest model in this area.

Definition 1 [11]**.** *An argumentation framework (AF) is a pair $\mathcal{X}, \mathcal{A}$ where $\mathcal{X}$ is a finite set of arguments and $\mathcal{A} \subseteq \mathcal{X} \times \mathcal{X}$ is a binary (attack) relation.*

In the context of AFs, an *extension-based semantics* is a criterion specifying which sets of arguments, called *extensions*, are collectively acceptable.

We will use the more expressive BAFs, which also encompass a relation of support. Further, we will consider *gradual* argumentation semantics, where argument evaluation is expressed by a *strength* value on a given scale and arguments are equipped with an initial *base score*. These notions are formalised by the following definition [5,9].

Definition 2 [5,9]**.** *A Quantitative Bipolar Argumentation Framework (QBAF) is a quadruple $\langle \mathcal{X}, \mathcal{A}, \mathcal{S}, \tau \rangle$ where $\mathcal{X}$ is a finite set of arguments, $\mathcal{A} \subseteq \mathcal{X} \times \mathcal{X}$ is a binary (attack) relation, $\mathcal{S} \subseteq \mathcal{X} \times \mathcal{X}$ is a binary (support) relation and $\tau : \mathcal{X} \to \mathbb{I}$ is a total function, where $\mathbb{I}$ is a set equipped with a preorder $\leq$. For any $\alpha \in \mathcal{X}$, we call $\tau(\alpha)$ the base score of α. A* gradual semantics *σ is a criterion that, given a QBAF $\mathcal{Q} = \langle \mathcal{X}, \mathcal{A}, \mathcal{S}, \tau \rangle$, returns a strength function $\sigma_{\mathcal{Q}} : \mathcal{X} \to \mathbb{I}$ representing the strength evaluation of the arguments in $\mathcal{Q}$ according to the semantics σ.*

With a minor abuse of notation, for $S \subseteq \mathcal{X}$ we will denote as $\sigma(S)$ the multiset of the strengths of the elements of S, i.e., given $S = \{x_1, \dots x_n\}$, $\sigma(S) = \{\sigma(x_1), \dots \sigma(x_n)\}$. Given $x \in \mathcal{X}$ the set of attackers of x is denoted as $\mathcal{A}(x) \triangleq \{y \mid (y, x) \in \mathcal{A}\}$ and the set of supporters of x is denoted as $\mathcal{S}(x) \triangleq \{y \mid (y, x) \in \mathcal{S}\}$, the set of influencers of x is denoted as $\mathcal{I}(x) \triangleq \mathcal{A}(x) \cup \mathcal{S}(x)$.

We focus on applications of QBAFs for decision support: in this context some of the arguments have a distinguished role, since they represent the possible *answers* (or options) of the decision process, while the reasons in favour or against each option are represented by other arguments (called pro and con arguments), which in turn can be supported or attacked by other reasons corresponding to other pro and con arguments, and so on. QBAFs featuring this structure provide a formal counterpart to the IBIS method for decision making [8,12] as illustrated in [6]. As it emerges from the description sketched above, QBAFs for decision support can be represented as sets of trees, with the root of each tree corresponding to an answer argument, other vertices corresponding to pro and con arguments and the edges corresponding to the attack and support relations. Considering more general topologies, e.g. encompassing cycles of attacks, is left to future work. For brevity, we will consider the treatment of preferences in QBAFs consisting of a single tree, the extension to the case of a set of trees being straightforward. Definition 3 [18] captures the structure of the QBAFs we focus on, namely QBAFs for decision support about an answer r.

Definition 3 [18]**.** *Let $\mathcal{Q}$ be a QBAF $\langle \mathcal{X}, \mathcal{A}, \mathcal{S}, \tau \rangle$. For any arguments $a, b \in \mathcal{X}$, let a* path *from a to b be defined as $(c_0, c_1), ..., (c_{n-1}, c_n)$ for some $n > 0$ (referred to as the length of the path) where $c_0 = a, c_n = b$ and, for any $1 \leq i \leq n, (c_{i-1}, c_i) \in \mathcal{A} \cup \mathcal{S}$. Then, for $r \in \mathcal{X}$, $\mathcal{Q}$ is a* QBAF for r *iff i)$\nexists a \in \mathcal{X} \setminus \{r\}$ such that $(r, a) \in \mathcal{A} \cup \mathcal{S}$ ii)$\forall a \in \mathcal{X} \setminus \{r\}$, there is a unique path from a to r; and iii)$\nexists a \in \mathcal{X}$ with a path from a to a. Given a* QBAF for r $\mathcal{Q}$, *the* pro *arguments and* con *arguments in $\mathcal{Q}$ are defined respectively as: $pro(\mathcal{Q}) = \{a \in \mathcal{X} | \exists p \in paths(a, r)$ where $|p \cap \mathcal{A}|$ is even$\}$; $con(\mathcal{Q}) = \{a \in \mathcal{X} | \exists p \in paths(a, r)$ where $|p \cap \mathcal{A}|$ is odd$\}$.*

In the following, if not otherwise specified, we will assume that all QBAFs are for a root argument r, denoted as $root(\mathcal{Q})$. We will investigate the use of preferences in these QBAFs. In general a preference is a preorder over a set.

Definition 4. *Given a set S, a preference $\preceq_p$ over S is a reflexive and transitive relation on S. As usual, given $a, b \in S$, $a \preceq_p b$ and $a \not\succeq_p b$ will be denoted as $a \prec_p b$, while $a \preceq_p b$ and $b \preceq_p a$ will be denoted as $a \simeq_p b$.*

A preference over a set S can induce a preference over the powerset of S based on some criterion (examples are the `Elitist` and `Democratic` criteria in [16]). When assuming the existence of such a preference-inducing criterion C, we will denote as $\preceq_p^C$ the preference relation between subsets of S induced by $\preceq_p$ according to C. In argumentation frameworks preferences are over arguments, with gradual semantics defined analogously as for QBAFs.

Definition 5. *A preference-based argumentation framework (PAF) [3,14] is a 3-tuple $\langle \mathcal{X}, \mathcal{A}_p, \preceq_p \rangle$ where $\langle \mathcal{X}, \mathcal{A}_p \rangle$ is an AF and $\preceq_p$ is a preference over $\mathcal{X}$. A preference-based QBAF (PQBAF) is a pair $\langle \mathcal{Q}, \preceq_p \rangle$ where $\mathcal{Q} = \langle \mathcal{X}, \mathcal{A}, \mathcal{S}, \tau \rangle$ is a QBAF and $\preceq_p$ is a preference over $\mathcal{X}$. A gradual semantics σ is a criterion that, given a PQBAF $\mathcal{P}$, returns a strength function $\sigma_{\mathcal{P}} : \mathcal{X} \rightarrow \mathbb{I}$ representing the strength evaluation of the arguments in $\mathcal{P}$ according to the semantics σ.*

3 Preferences in Formal Argumentation

Several approaches to the treatment of preferences have been considered in formal argumentation literature. In the context of abstract argumentation a major line of investigation has concerned the treatment of so called *critical attacks* in PAFs. An attack is said to be *critical* [3] if the attacked element is strictly preferred to the attacker. The question then arises on whether and how the preference for the attacked element influences the attack relation. Several approaches use preferences to reduce a PAF $\mathcal{P}$ to an AF, that we will denote as $r(\mathcal{P})$, with the same set of arguments but a different attack relation. The acceptability of arguments is then evaluated by applying a semantics to $r(\mathcal{P})$.

Notation 1. *Given a PAF $\mathcal{P} = \langle \mathcal{X}, \mathcal{A}_p, \succ_p \rangle$, the reduced argumentation framework corresponding to $\mathcal{P}$ is denoted as $r(\mathcal{P}) = \langle \mathcal{X}, \mathcal{A} \rangle$.*

Four main reduction methods have been proposed in the literature (see the relevant references for details):

- [2] $\forall a, b, \in \mathcal{X}$, $(a, b) \in \mathcal{A}$ iff $(a, b) \in \mathcal{A}_p, b \not\succ_p a$.
- [3] $\forall a, b, \in \mathcal{X}$, $(a, b) \in \mathcal{A}$ iff $((a, b) \in \mathcal{A}_p, b \not\succ_p a)$ or $((b, a) \in \mathcal{A}_p, (a, b) \notin \mathcal{A}_p, a \succ_p b)$.
- [14] $\forall a, b, \in \mathcal{X}$, $(a, b) \in \mathcal{A}$ iff $((a, b) \in \mathcal{A}_p, b \not\succ_p a)$ or $((a, b) \in \mathcal{A}_p, (b, a) \notin \mathcal{A}_p)$.
- [14] $\forall a, b, \in \mathcal{X}$, $(a, b) \in \mathcal{A}$ iff $((a, b) \in \mathcal{A}_p, b \not\succ_p a)$ or $((b, a) \in \mathcal{A}_p, (a, b) \notin \mathcal{A}_p, a \succ_p b)$ or $((a, b) \in \mathcal{A}_p, (b, a) \notin \mathcal{A}_p)$.

The first reduction suppresses the critical attack; this technique has been criticized in [3] because it can lead to extensions which are not conflict-free with respect to the original PAF. For this reason, the second reduction, aims to "repair" the AF and avoids that drawback by reversing the direction of the critical attack. In [14], Kaci et al. argued that the second reduction implies a strong constraint since a preferred argument can never be successfully attacked, hence they proposed the third reduction, which deletes a critical attack only if the opposite attack belongs to $\mathcal{A}_p$ too. As a further alternative, the fourth reduction comprises both the second and the third reduction. Each choice corresponds to a different intuition and is subject to potential criticisms.

In the context of the ASPIC+ formalism [16], a rule-based approach to argument construction is proposed, leading to the identification of different forms of attack between arguments, which are classified as preference-dependent or preference-independent. Only preference-dependent attacks are affected by preferences: they are ignored when the attacked argument is strictly preferred to the attacker. Since in ASPIC+ preference-dependent attacks are always symmetrical, this bears some similarity with the third reduction mentioned above.

While the above approaches concern extension-based semantics, the use of preferences in the context of *gradual* argumentation has received lesser attention, as remarked in [15]. The authors propose a property called *Preference Precedence* (PP) stating that the preference relation should have a direct impact on the strength evaluation. In a nutshell, if an argument x_1 is preferred to x_2, then the strength of x_1 should be not lesser than the strength of x_2.

It can be remarked that this principle imposes a rather strong requirement on the evaluation of arguments, independently of the relations holding between them. Moreover, the role of preferences is quite different, as they are not used to modify the attack relation but rather are meant to affect the final evaluation.

The potential twofold role of preferences is also evidenced in [3], where, in addition to handling critical attacks in PAFs, preferences are used to induce an ordering on the extensions prescribed by a given semantics.

With respect to the goals of the present paper, two main limitations emerge from the above surveyed approaches: i) none of them concerns BAFs, i.e. they do not consider the support relation, which is needed in our context; ii) a conceptual analysis about the motivations underlying the different proposals is lacking.

As to the latter point, we remark in particular that different uses of preferences may be required by different application contexts. In this respect, we propose here a simple taxonomy on the uses of preferences in argumentation based on two classification dimensions: i) the origin of preferences, which can be *endogenous* or *exogenous* with respect to the argument construction process; ii) the purpose of the formalization, which can be *normative* or *descriptive*.

Concerning the first point, we call *endogenous* preferences those which are induced on arguments from preferences concerning their constitutive elements, e.g. premises and rules as in [16], while *exogenous* preferences are ascribed to arguments based on elements which are not involved in their construction, like for instance the values they promote, as in [7]. Concerning the second point, a *normative* approach aims to define a standard behavior on the basis of certain rationality principles, while a *descriptive* approach aims to represent how people actually behave, possibly in an unprincipled manner. We suggest that some links can be drawn between these notions and the uses of preferences in the literature.

For instance the PP principle in [15], where, in a sense, preferences determine the evaluation outcomes overriding any relations between arguments, can be justified in a descriptive approach with exogenous preferences. For instance, if some people have a preference for an information source they trust, they may accept all arguments from that source, no matter what their content is. This behavior would be in contrast with a normative approach, where arguments' contents and relations should play a role also in presence of preferences.

Concerning the treatment of critical attacks, suppressing all of them independently of any other condition [2] appears in line with a descriptive approach, where, as above, preferences have a sort of absolute priority over other factors, with the possible production of outcomes which are not conflict-free. On the other hand, the treatment proposed in [16] concerns endogenous preferences with a normative approach, where they have the role of converting mutual attacks into unidirectional ones when appropriate.

While an extended discussion of these aspects is beyond the scope of this paper, the observations above indicate that a proper characterization of the application context is necessary to lay the foundations of the approach we aim to propose. In particular, the preferences we are interested in are exogenous, since they can be provided by users as an additional element with respect to a QBAF representing domain knowledge in a given decision support context. Moreover, we

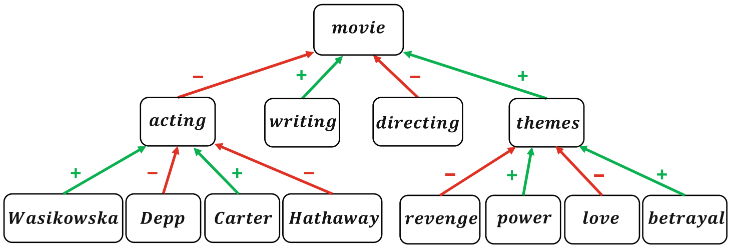

Fig. 1. A simple BAF in the movie domain. Arguments are represented by vertices, attacks by red edges labelled '–' and supports by green edges labelled '+'. Arguments correspond to features and subfeatures of a given movie. An attack (support) relation indicates that a feature has a negative (positive) effect on the assessment of its parent. For more details on the domain modelling, see Sect. 5. (Color figure online)

aim for a normative approach where preferences are used to provide personalised recommendations in a principled manner, as discussed in the next section.

4 Adding Preferences to QBAFs for Decision Support

In this section we illustrate the basic ideas of our approach to encompass user preferences in gradual bipolar argumentation for decision support. To support our presentation, we will use as a running example the simple framework presented in Fig. 1, taken from the movie recommendation domain.

The first question we consider concerns the pairs of arguments on which preferences are given. In this respect, some differences with the approaches reviewed in the previous section have to be underlined. In particular, while, in principle, endogenous preferences can refer to any pair of arguments (since they involve constitutive elements common to all arguments), user-defined exogenous preferences can only be given on arguments whose comparison is meaningful to the user. In the family of frameworks we are considering, it is then natural to consider preferences between sibling nodes (i.e. between influencers of the same node) since they contribute together to the evaluation of the influenced node. For instance, in the example of Fig. 1, it is reasonable to imagine that a user may give more importance to the *themes* of the movie than to the quality of *acting*, or may prefer one actor to another, while it does not seem meaningful to express preferences between an influencer and an influenced node (e.g. between *love* and *themes*) and more generally preferences across different levels of the tree. This represents a significant difference with respect to approaches whose main focus is the treatment of critical attacks.

Towards defining general principles for the treatment of preferences, a further question then concerns identifying the cases where their effect on argument evaluation can be univocally determined. In this respect, we distinguish the cases of preferences concerning arguments of different polarity (e.g. an attacker vs. a supporter) with respect to arguments with the same polarity (e.g. a supporter vs. a supporter). In the first case, the expected effect of preferences can be clearly identified. For instance, if an attacker is preferred to a supporter of a node x, it can be expected that the strength of x should be lower with respect to the case where this preference is the opposite or does not hold. In the example of Fig. 1, if a user gives more importance to *acting* (a negative feature of the considered movie) than to *themes* (a positive feature of the movie) it is reasonable to consider the movie less appropriate for this user with respect to a user who has the opposite preference (or no preference at all). In the second case, the effect of preferences is undetermined: if a user prefers an attacker a_1 to another attacker a_2 and then you consider another user who prefers a_2 to a_1, you have two structurally indistinguishable situations for which a different behavior cannot be prescribed. The same holds for a preference involving two supporters.

While the examples above concern a single preference between a pair of arguments, they can be extended to the case where multiple preferences are given, from which a preference relation on sets of arguments is derived. On this basis, we introduce a property of *local coherence* specifying the effects of preferences between the set of attackers and the set of supporters of a given node. The property refers to the comparison between two PQBAFs which differ only in the preference relation concerning the influencers of a given argument x.

Definition 6. *Given a QBAF, $\mathcal{Q} = \langle \mathcal{X}, \mathcal{A}, \mathcal{S}, \tau \rangle$, let $\mathcal{P} = \langle \mathcal{Q}, \succeq_p \rangle$ and $\mathcal{P}' = \langle \mathcal{Q}, \succeq_{p'} \rangle$ be two PQBAFs. Given $x \in \mathcal{X}$ we say that $\mathcal{P}'$ is a x-*local modification *of $\mathcal{P}$ if $\succeq_p \cap((\mathcal{X} \times \mathcal{X}) \setminus (\mathcal{I}(x) \times \mathcal{I}(x))) = \succeq_{p'} \cap((\mathcal{X} \times \mathcal{X}) \setminus (\mathcal{I}(x) \times \mathcal{I}(x)))$.*

Definition 7. *Given a QBAF, $\mathcal{Q} = \langle \mathcal{X}, \mathcal{A}, \mathcal{S}, \tau \rangle$, let $\mathcal{P} = \langle \mathcal{Q}, \succeq_p \rangle$ and $\mathcal{P}' = \langle \mathcal{Q}, \succeq_{p'} \rangle$ be two PQBAFs, such that $\mathcal{P}'$ is a x-local modification of $\mathcal{P}$ for some $x \in \mathcal{X}$. Let C be a preference inducing criterion and σ a gradual semantics for PQBAFs. The property of* local coherence with preferences *is satisfied iff the following conditions hold:*

- *if $\mathcal{A}(x) \preceq_p^C \mathcal{S}(x)$ and $\mathcal{A}(x) \succeq_{p'}^C \mathcal{S}(x)$ then $\sigma_{\mathcal{P}}(x) \geq \sigma_{\mathcal{P}'}(x)$;*
- *if $\mathcal{A}(x) \succeq_p^C \mathcal{S}(x)$ and $\mathcal{A}(x) \preceq_{p'}^C \mathcal{S}(x)$ then $\sigma_{\mathcal{P}}(x) \leq \sigma_{\mathcal{P}'}(x)$.*

The property of strict local coherence with preferences *is satisfied iff the following conditions hold:*

- *if $\mathcal{A}(x) \preceq_p^C \mathcal{S}(x)$ and $\mathcal{A}(x) \succ_{p'}^C \mathcal{S}(x)$ then $\sigma_{\mathcal{P}}(x) > \sigma_{\mathcal{P}'}(x)$;*
- *if $\mathcal{A}(x) \succeq_p^C \mathcal{S}(x)$ and $\mathcal{A}(x) \prec_{p'}^C \mathcal{S}(x)$ then $\sigma_{\mathcal{P}}(x) < \sigma_{\mathcal{P}'}(x)$.*

Let us illustrate Definitions 6 and 7, with reference to the example of Fig. 1. Consider a relation $\succeq_p = \emptyset$ expressing the absence of any preference, and a relation $\succeq_{p'}$ expressing the view of a user who strictly prefers Depp to Wasikowska and Hathaway to Carter. Then, $\succeq_{p'}$ gives rise to a x-local modification with x=*acting*. Since the user prefers two attackers to two supporters, we may assume that, given any criterion C, $\mathcal{A}(acting) \succ_{p'}^C \mathcal{S}(acting)$. As a consequence, (strict) local coherence requires that $\sigma_{\mathcal{P}'}(acting)$ is (strictly) lesser that $\sigma_{\mathcal{P}}(acting)$.

Under mild requirements on the considered semantics, it is possible to show that local coherence ensures that the effects of preferences are coherent with the roles of pro and con arguments along the structure of the tree. In particular, we show in Proposition 1 that a preference for pros over cons among the influencers of an argument x has the effect of increasing the strength of pros and decreasing the strength of cons in the path from x to the root of the PQBAF, and vice versa in the case of a preference for cons over pros.

Towards this result, we require first of all that the strength of an argument is determined by its base score, the strengths of its attackers and supporters, and the preferences between them. This is the extension to the case of presence of preferences of a property which is common to most gradual argumentation semantics for BAFs in the literature (see e.g. [9]).

Definition 8. *A gradual semantics σ for PQBAFs is* based on local evaluation *if for every PQBAF $\mathcal{P} = \langle \mathcal{Q}, \preceq_p \rangle$, for every $x \in \mathcal{X}$,*
$\sigma_\mathcal{P}(x) = f(\tau(x), \sigma_\mathcal{P}(\mathcal{A}(x)), \sigma_\mathcal{P}(\mathcal{S}(x)), \preceq_p \cap(\mathcal{I}(x) \times \mathcal{I}(x))$ for some function f.

Moreover, we assume a semantics based on a *monotonic* strength function σ. The relevant definitions from [5] are adapted below, in a simplified form which requires a notion of similarity of arguments with respect to their influencers.

Definition 9. *Given two PQBAFs $\mathcal{P} = \langle \mathcal{Q}, \preceq_p \rangle$, $\mathcal{P}' = \langle \mathcal{Q}', \preceq_{p'} \rangle$ two arguments $x \in \mathcal{X}$, $x' \in \mathcal{X}'$ are $\mathcal{I}$-similar iff there is a bijective function $h : \mathcal{I}(x) \rightarrow \mathcal{I}(x')$ such that: i) $\forall y \in \mathcal{I}(x)$, $y \in \mathcal{A}(x)$ iff $h(y) \in \mathcal{A}'(x')$ and $y \in \mathcal{S}(x)$ iff $h(y) \in \mathcal{S}'(x')$; ii) $\forall x_1, x_2 \in \mathcal{I}(x)$, $x_1 \preceq_p x_2$ iff $h(x_1) \preceq_{p'} h(x_2)$.*

In words, two arguments are $\mathcal{I}$-similar if they have the "same" (modulo a bijection) attackers and supporters with the "same" preference relation among them.

Definition 10. *Given two PQBAFs $\mathcal{P} = \langle \mathcal{Q}, \preceq_p \rangle$, $\mathcal{P}' = \langle \mathcal{Q}', \preceq_{p'} \rangle$, for $A \subseteq \mathcal{X}$, $B \subseteq \mathcal{X}'$ A is* strength equivalent *to B, denoted $A = B$ iff $\sigma_\mathcal{P}(A) = \sigma_{\mathcal{P}'}(B)$; A is* at least as strong as *B, denoted $A \geq B$ iff there exists an injective mapping f from B to A such that $\forall \alpha \in B, \sigma_\mathcal{P}(f(\alpha)) \geq \sigma_{\mathcal{P}'}(\alpha)$; A is* stronger than *B, denoted $A > B$ iff $A \geq B$ and $B \not\geq A$.*

Intuitively, two sets of arguments are strength equivalent if the multisets of their strength values are the same (i.e. the two sets have the same cardinality and one can establish a bijection linking arguments with the same strength). A set of arguments A is at least as strong as a set B if A has a greater or equal cardinality than B and for each element of B one can identify a distinct element of A which has a greater or equal strength.

Definition 11. *For any argument x in a PQBAF $\mathcal{P}$, the* shaping triple *of x is $(\tau(x), \mathcal{S}(x), \mathcal{A}(x))$, denoted $ST_\mathcal{P}(x)$. Given two PQBAFs $\mathcal{P} = \langle \mathcal{Q}, \preceq_p \rangle$, $\mathcal{P}' = \langle \mathcal{Q}', \preceq_{p'} \rangle$, let $x_1 \in \mathcal{X}$, $x_2 \in \mathcal{X}'$ be $\mathcal{I}$-similar. $ST_{\mathcal{P}'}(x_2)$ is said to be:* as boosting as *$ST(x_1)$, denoted $ST_\mathcal{P}(x_1) \simeq ST_{\mathcal{P}'}(x_2)$ iff $\tau(x_1) = \tau'(x_2), \mathcal{S}(x_1) = \mathcal{S}'(x_2)$ and $\mathcal{A}(x_1) = \mathcal{A}'(x_2)$;* at least as boosting as *$ST_\mathcal{P}(x_1)$, denoted $ST_\mathcal{P}(x_1) \preceq ST_{\mathcal{P}'}(x_2)$, iff $\tau(x_1) \leq \tau'(x_2), \mathcal{S}(x_1) \leq \mathcal{S}'(x_2), \mathcal{A}(x_1) \leq \mathcal{A}'(x_2)$;* strictly more boosting *than $ST_\mathcal{P}(x_1)$, denoted $ST_\mathcal{P}(x_1) \prec ST_{\mathcal{P}'}(x_2)$, iff $ST_\mathcal{P}(x_1) \preceq ST_{\mathcal{P}'}(x_2)$ and*

$ST_{\mathcal{P}'}(x_2) \not\preceq ST_{\mathcal{P}}(x_1)$. A strength function σ is monotonic *iff the following conditions hold: i) if $ST_{\mathcal{P}}(x_1) \simeq ST_{\mathcal{P}'}(x_2)$, then $\sigma_{\mathcal{P}}(x_1) = \sigma_{\mathcal{P}'}(x_2)$; ii) if $ST_{\mathcal{P}}(x_1) \preceq ST_{\mathcal{P}'}(x_2)$, then $\sigma_{\mathcal{P}}(x_1) \leq \sigma_{\mathcal{P}'}(x_2)$. A strength function σ is* strictly monotonic *iff σ is monotonic and if $ST_{\mathcal{P}}(x_1) \prec ST_{\mathcal{P}'}(x_2)$, then $\sigma_{\mathcal{P}}(x_1) < \sigma_{\mathcal{P}'}(x_2)$.*

The shaping triple collects the elements affecting the strength evaluation of an argument: its base score and its supporters and attackers. The boosting relations are based on an element-wise comparison between shaping triples and essentially check whether two shaping triples are equal or one (strictly) dominates the other with respect to the strength values. A strength function is (strictly) monotonic if its outcomes on arguments (strictly) follow the (in)equalities between the relevant shaping triples.

On this basis, Proposition 1 shows that if a local modification corresponds to a preference for pros over cons then it can only induce an increase of the strength of other pros and a decrease of the strength of other cons, and vice versa in the case of a preference for cons over pros. This can be regarded as a globally coherent behavior induced by the local coherence property.

Proposition 1. *Given a QBAF, $\mathcal{Q} = \langle \mathcal{X}, \mathcal{A}, \mathcal{S}, \tau \rangle$, let $\mathcal{P} = \langle \mathcal{Q}, \succeq_p \rangle$ and $\mathcal{P}' = \langle \mathcal{Q}, \succeq_{p'} \rangle$ be two PQBAFs, such that $\mathcal{P}'$ is a x-local modification of $\mathcal{P}$ for some $x \in \mathcal{X}$. Let C be a preference inducing criterion and σ a monotonic semantics based on local evaluation. If the property of local coherence holds, for any $y \in \{x\} \cup \{z \mid z$ is in the path from x to $root(\mathcal{Q})\}$ it holds that:*

(a) if $\mathcal{I}(x) \cap pro(\mathcal{Q}) \preceq_p^C \mathcal{I}(x) \cap con(\mathcal{Q})$ and $\mathcal{I}(x) \cap pro(\mathcal{Q}) \succeq_{p'}^C \mathcal{I}(x) \cap con(\mathcal{Q})$ then:
- *if $y \in pro(\mathcal{Q})$, then $\sigma_{\mathcal{P}}(y) \leq \sigma_{\mathcal{P}'}(y)$;*
- *if $y \in con(\mathcal{Q})$, then $\sigma_{\mathcal{P}}(y) \geq \sigma_{\mathcal{P}'}(y)$*

(b) if $\mathcal{I}(x) \cap con(\mathcal{Q}) \preceq_p^C \mathcal{I}(x) \cap pro(\mathcal{Q})$ and $\mathcal{I}(x) \cap con(\mathcal{Q}) \succeq_{p'}^C \mathcal{I}(x) \cap pro(\mathcal{Q})$ then:
- *if $y \in pro(\mathcal{Q})$, then $\sigma_{\mathcal{P}}(y) \geq \sigma_{\mathcal{P}'}(y)$;*
- *if $y \in con(\mathcal{Q})$, then $\sigma_{\mathcal{P}}(y) \leq \sigma_{\mathcal{P}'}(y)$*

Proof. Proof is by induction on the length of the path from any element of $\mathcal{I}(x)$ to y. *Induction base.* Suppose that $y = x$ and let $w \in \mathcal{I}(x) \cap pro(\mathcal{Q})$. Then, by Definition 3, the number of attacks in the path from w to $root(\mathcal{Q})$ is even, while for $u \in \mathcal{I}(x) \cap con(\mathcal{Q})$ the number of attacks in the path from u to $root(\mathcal{Q})$ is odd. Therefore, if $y = x \in pro(\mathcal{Q})$ then $w \in \mathcal{S}(x)$ and $u \in \mathcal{A}(x)$. As a consequence $\mathcal{I}(x) \cap pro(\mathcal{Q}) = \mathcal{S}(x)$ and $\mathcal{I}(x) \cap con(\mathcal{Q}) = \mathcal{A}(x)$. In the case (a), this entails $\mathcal{S}(x) \preceq_p^C \mathcal{A}(x)$ and $\mathcal{S}(x) \succeq_{p'}^C \mathcal{A}(x)$. It then follows from the local coherence property that $\sigma_{\mathcal{P}}(x) \leq \sigma_{\mathcal{P}'}(x)$. If instead $y = x \in con(\mathcal{Q})$ we get $\mathcal{I}(x) \cap pro(\mathcal{Q}) = \mathcal{A}(x)$ and $\mathcal{I}(x) \cap con(\mathcal{Q}) = \mathcal{S}(x)$ and we analogously derive $\sigma_{\mathcal{P}}(y) \geq \sigma_{\mathcal{P}'}(y)$. The proof for case (b) is similar.

Induction step. We inductively suppose that the statement above is valid for every y in the path from any child w of x to $root(\mathcal{Q})$ such that the path has length n, and we want to show that it holds for length $n + 1$.

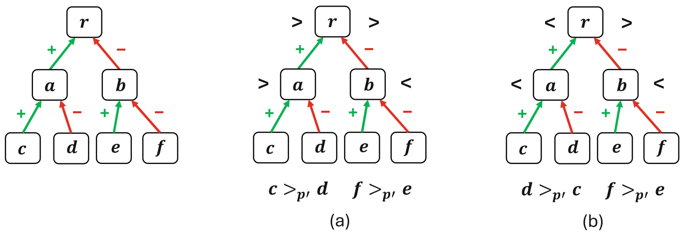

Fig. 2. Two simple examples of the effects of multiple preferences. The non-leaf nodes are labeled by $>$ and $<$ to indicate, respectively, an increase and a decrease in their strengths, assuming the strict version of Proposition 1. The effects on the root strength of the two preferences added in case (a) are concordant, while they are discordant in case (b), hence, the overall effect on the strength of r is not determined in the second case.

Considering the case (a), for every z such that the path from w to z has length n, we know from the induction hypothesis that if $z \in pro(\mathcal{Q})$ then $\sigma_{\mathcal{P}}(z) \leq \sigma_{\mathcal{P}'}(z)$.

Let now y be the argument attacked or supported by z. We get that: (i) if $y \in pro(\mathcal{Q})$, then $z \in \mathcal{S}(y)$, hence $ST_{\mathcal{P}'}(y) \geq ST_{\mathcal{P}}(y)$ (y has a stronger supporter in $\mathcal{P}'$ while all other elements of its shaping triple are the same) and, by the monotonicity property, $\sigma_{\mathcal{P}'}(y) \geq \sigma_{\mathcal{P}}(y)$; (ii) if $y \in con(\mathcal{Q})$, then $z \in \mathcal{A}(y)$, hence $ST_{\mathcal{P}'}(y) \leq ST_{\mathcal{P}}(y)$ and, by the monotonicity property $\sigma_{\mathcal{P}'}(y) \leq \sigma_{\mathcal{P}}(y)$. The treatment of the case (b) is analogous. $\square$

Assuming a stricly monotonic semantics and the property of strict local coherence, a strict version of Proposition 1 where the inequalities between $\sigma_{\mathcal{P}}(y)$ and $\sigma_{\mathcal{P}'}(y)$ are strict can be derived. We omit the obvious statement and the analogous proof due to space limitations.

While the above results concern only preferences over the influencers of a specific argument they can be used as a basis to reason about the addition of an arbitrary set of preferences to a QBAF. In particular a QBAF without preferences can be regarded as a special case of *preference uniform* PQBAF where for every pair x, y of sibling nodes it holds that $x \simeq_p y$. A preference uniform PQBAF can be transformed into a generic PQBAF by a series of local modifications, each concerning the preferences over the children of a single node.

By iterated application of Proposition 1 it is possible to characterize the variations induced by the preferences on the argument strengths, when they are determined. In particular, when the hypotheses of Proposition 1 are satisfied, a x-local modification determines effects on the arguments in the path from x to the root. If there are multiple modifications, they affect concurrently the arguments shared in the relevant paths: these include at least the root and may involve also other arguments (see Fig. 2). For the arguments shared among multiple paths, multiple variations are induced by multiple preferences. If the variations are all concordant the overall effect on the arguments is determined, otherwise it is not.

Figure 2 presents a simple case where the propagation of the effects of preferences on the strength of the root node is determined, and a case where it is not.

5 Integrating Preferences in the ADA System

ADA is a review aggregation system introduced in [10] within the movie domain The ADA pipeline is organised as follows (see [10] for details):

– ADA employs natural language processing to break down the reviews into a *feature-based review aggregation* where each feature corresponds to an argument and is assigned a polarity, based on sentiment analysis;
– ADA generates a tree-structured QBAF to represent the feature-based review aggregation and the identified polarities (Fig. 1 shows a simple example);
– the base score of each feature is derived from the review aggregation, in a nutshell the base score reflects how much the reviews about that feature are coherent: if a feature has consistently positive or negative reviews its base score will be higher, if reviews are more mixed its base score will be lower;
– A gradual semantics is applied to the QBAF to compute argument strengths;
– The extracted QBAF provides the underlying structure for generating dialogical explanations for users, taking into account the computed strengths.

As a preliminary application of the concepts we previously introduced, we investigated a method to integrate user preferences in ADA in compliance with the properties introduced in Sect. 4. We assume that, for a given user u, preferences are given as a set $\mathcal{P}_u$ of pairs of features, where each pair (x, y) is such that x and y are sibling nodes in the QBAF, and indicates that x is strictly preferred over y. As a simple example, that we will use also later, we consider a user who, with reference to the framework shown in Fig. 1, prefers the actor $Johnny\ Depp$ over the actress $Mia\ Wasikowska$, and the movie's *writing* over its *directing*.

Our main goal was to carry out an initial analysis of the issues to be faced when putting our general notions into practice, before developing further investigations. To this purpose, the idea was to adopt a simple parametric approach to combine preferences with existing gradual argumentation semantics and then carry out a preliminary assessment of the impact of the use of preferences on strength evaluation. The approach we adopted consists in decreasing the base score of the arguments which are less preferred, by multiplying them by a given *discount factor* $\delta < 1$. Formally, given a set of preferences $\mathcal{P}_u$, for each $(x, y) \in \mathcal{P}_u$, we let $\tau_p(y) = \delta \cdot \tau(y)$. Then a gradual semantics is applied to the framework with these modified base scores. Simple as it is, this method is coherent with the properties introduced in Sect. 4 whenever the preferences over the influencers of a given argument are uniform in terms of polarity (i.e. all indicate the superiority of an attacker over a supporter or vice versa) and the adopted semantics is monotonic. In this context, in order to carry out a preliminary evaluation of the effects on the argument strengths, we considered two main choices affecting the final outcome: the factor δ used to decrease the base scores of the less preferred arguments and the chosen semantics.

As to the first choice, we considered the following values for δ: 0.25, 0.50, 0.75. As to the second choice we considered four gradual semantics, namely, QuAD [6], DF-QuAD [19], REB [1] and Quadratic Energy Model (QEM) [17]. All these semantics produce strength values belonging to the $[0, 1]$ real interval.

We draw some general considerations on the roles of the elements involved in our approach, before presenting some examples of quantitative assessments.

Role of the Base Scores. The idea of using preferences to adjust the value of the base scores has the advantage of simplicity and of enabling the use of existing gradual semantics as interchangeable alternatives, while ensuring the satisfaction of the properties introduced in Sect. 4. However, it makes the effect of preferences dependent on the way the base score of arguments is determined. In the case of ADA, the base score of an argument corresponding to a feature is derived from a normalised count of the positive and negative judgments about that feature found in the set of reviews under consideration (see [10] for details). The normalisation method used in the original version of ADA, gave rise typically to rather small base scores (around 0.05 with the dataset in the movie domain). While these small base scores were in line with the original purposes of ADA, they turned out to be somehow problematic, since the quantitative effects of preferences were in fact dampened since they were conveyed through the adjustment of small values. To avoid this problem we experimented with a different normalisation method which led to base scores' with values on average ten times bigger, leading in turn to a greater quantitative impact of the preferences.

Role of the Gradual Semantics. As mentioned above, we experimented our approach with four gradual semantics, which, while sharing the property of being monotonic, are rather different by design (a comparison is beyond the scope of the present paper). This entails that a given change in base scores, as determined by preferences in our approach, may have significantly different impacts on strength evaluation outcomes, depending on the semantics adopted. In a sense, we can say that a gradual semantics can be more or less sensitive to preferences in the context of our approach. While the study of a formal notion of sensitivity to preferences is an interesting issue for future work, we will draw some preliminary considerations concerning the semantics we experimented with.

Definition of the Discount Factor. The value of the discount factor δ modulates the entity of the modification of base scores due to preferences, the bigger this modification, the bigger the expected impact on argument strengths, though it depends also on the sensitivity of semantics, as mentioned previously. Indeed, one can regard as an open question, whose answer is context dependent, how heavily preferences should affect strength evaluation with respect to the rest of the framework. In the specific case of ADA, the question concerns balancing the individual inclinations of a user and the indications emerging from the reviews produced by the community. It can be imagined that a proper balance is in turn user dependent, with some users being more radical in their preferences and others who are more open to take into account the opinions of the crowd. While a

Table 1. Strength values and percent variations with $Depp \succ_p Wasikowska$ and $writing \succ_p directing$, $\delta = 0.75$.

	old base score normalisation				new base score normalisation			
	QuAD	DF-QuAD	REB	QEM	QuAD	DF-QuAD	REB	QEM
Depp	0.0111	0.0111	0.0111	0.0111	0.2	0.2	0.2	0.2
Wasikowska	0.0167	0.0167	0.0167	0.0167	0.375	0.375	0.375	0.375
directing	0.0250	0.0250	0.0250	0.0250	0.45	0.45	0.45	0.45
writing	0.0444	0.0444	0.0444	0.0444	0.5	0.5	0.5	0.5
acting	0.0275	0.0407	0.0223	0.0223	0.0894	0.3288	0.0915	0.0910
movie	0.5429	0.5279	0.5219	0.5171	0.6194	0.5336	0.5465	0.5700
Δ% acting	–16,17	–6.13	–0.53	–0.41	–41.00	–10.33	–9.89	–14.81
Δ% movie	1.19	0.52	0.27	0.04	14.22	6.01	5.14	6.48

Table 2. Focus on *acting* and *movie* for $\delta = 0.25$ and $\delta = 0.5$.

		old base score normalisation				new base score normalisation			
		QuAD	DF-QuAD	REB	QEM	QuAD	DF-QuAD	REB	QEM
$\delta = 0.25$	acting	0.0021	0.0354	0.0221	0.0222	0.0742	0.2530	0.0743	0.0859
	movie	0.5533	0.5334	0.5247	0.5178	0.7560	0.6030	0.6023	0.6605
	Δ% acting	–32.71	–18.38	–1.58	–0.54	–51.00	–30.99	–26.81	–19.57
	Δ% movie	3.14	1.57	0.81	0.17	39.41	19.80	15.89	23.38
$\delta = 0.5$	acting	0.0222	0.0380	0.0222	0.0222	0.0818	0.2909	0.0825	0.0897
	movie	0.5493	0.5306	0.5222	0.5174	0.6872	0.5668	0.5741	0.6131
	Δ% acting	–32.34	–12.25	–1.06	–0.54	–46.00	–20.66	–18.80	–16.01
	Δ% movie	2.39	1.04	0.54	0.10	26.72	12.61	10.46	14.52

proper modelling of these different attitudes represents another interesting subject of future investigation, the use of different discount factors can be regarded as a first crude method to give them a counterpart in our approach.

As a specific but informative illustration of the issue discussed above, Tables 1 and 2 show the results of applying our approach to the example of Fig. 1 with different choices of base score normalisation, discount factor and semantics. Table 1 shows strength values of the arguments affected by preferences and evidences the variations of *acting* and *movie* with respect to the case of no preferences, for δ = 0.75. Table 2 focuses on *acting* and *movie* for other values of δ.

We discuss the effects on the strength of the *movie* representing the root of the framework, and of the feature *acting*, which are both affected by the preferences between the underlying elements. The main comments are as follows.

With the old normalisation method the dampening effect of small base scores values is evident on the variations of the *movie* strength which, in all cases are below 4%. A more significant effect is visible on the *acting* strength in the case of QuAD and DF-QuaD semantics, while the other semantics are less sensitive.

With the new normalisation method, as expected, the impact of preferences on the strength of both *movie* and *acting* is more significant and its amplitude is modulated by the choice of the discount factor, thus achieving the goal of taking preferences into account more effectively. In all cases QuAD is definitely more sensitive to preferences than the other semantics. Depending on the discount factor, the variation of the *movie* strength with QuAD ranges from 14.22% to 39.41%, while REB is the least sensitive semantics in all cases, with a range from 5.14% to 15.89%. As a side remark, we note that DF-QuAD is more sensitive than QEM with small base scores, while the converse holds with the new normalisation method.

6 Conclusions

With the aim of enabling personalised recommendations, we explored a normative approach for the use of exogenous preferences in QBAFs for decision support. In particular, we introduced a property of local coherence concerning the expected effects of preferences on argument strength and proved that, under the assumption of a monotonic semantics, it ensures that these effects are in line with the roles of pro and con arguments along the structure of the framework. Based on this approach, we extended a review aggregation system with the ability to deal with user preferences and carried out a preliminary experiment, showing how the quantitative effects of preferences are significantly affected by alternative design choices. Among future research directions we mention the study of further methods to deal with preferences in gradual argumentation semantics and a more extensive experimentation in the domain of review aggregation. We also plan to investigate the relationships of our argumentative approach with methods adopted in other fields, like Multi-Criteria Decision Analysis and Bayesian decision theory.

Acknowledgements. Baroni was partially supported by MUR project PRIN 2022 EPICA "Enhancing Public Interest Communication with Argumentation" (CUP D53D23008860006) funded by the European Union - Next Generation EU. Rago and Toni were partially funded by J.P. Morgan and by the Royal Academy of Engineering under the Research Chairs and Senior Research Fellowships scheme. Rago and Toni were partially funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 101020934). Any views or opinions expressed herein are solely those of the authors listed.

References

1. Amgoud, L., Ben-Naim, J.: Evaluation of arguments in weighted bipolar graphs. Int. J. Approx. Reason. **99**, 39–55 (2018)
2. Amgoud, L., Cayrol, C., Lagasquie-Schiex, M.C.: On the bipolarity in argumentation frameworks. In: Proceedings of the 10th International Workshop on Non-Monotonic Reasoning (NMR), pp. 1–9 (2004)

3. Amgoud, L., Vesic, S.: Rich preference-based argumentation frameworks. Int. J. Approx. Reason. **55**(2), 585–606 (2014)
4. Aurisicchio, M., Baroni, P., Pellegrini, D., Toni, F.: Comparing and integrating argumentation-based with matrix-based decision support in *Arg&Dec*. In: Black, E., Modgil, S., Oren, N. (eds.) TAFA 2015. LNCS (LNAI), vol. 9524, pp. 1–20. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-28460-6_1
5. Baroni, P., Rago, A., Toni, F.: From fine-grained properties to broad principles for gradual argumentation: a principled spectrum. Int. J. Approx. Reason. **105**, 252–286 (2019)
6. Baroni, P., Romano, M., Toni, F., Aurisicchio, M., Bertanza, G.: Automatic evaluation of design alternatives with quantitative argumentation. Argument Comput. **6**(1), 24–49 (2015)
7. Bench-Capon, T.: Persuasion in practical argument using value-based argumentation frameworks. J. Log. Comput. **13**(3), 429–448 (2003)
8. Buckingham Shum, S., Hammond, N.: Argumentation-based design rationale: what use at what cost? Int. J. Hum.-Comput. Stud. **40**(4), 603–652 (1994)
9. Cayrol, C., Lagasquie-Schiex, M.C.: Gradual valuation for bipolar argumentation frameworks. In: Godo, L. (ed.) ECSQARU 2005. LNCS (LNAI), vol. 3571, pp. 366–377. Springer, Heidelberg (2005). https://doi.org/10.1007/11518655_32
10. Cocarascu, O., Rago, A., Toni, F.: Extracting dialogical explanations for review aggregations with argumentative dialogical agents. In: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, (AAMAS), pp. 1261–1269 (2019)
11. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artif. Intell. **77**(2), 321–358 (1995)
12. Fischer, G., Lemke, A.C., McCall, R., Morch, A.I.: Making argumentation serve design. Hum.-Comput. Interact. **6**(3), 393–419 (1991)
13. Irwin, B., Rago, A., Toni, F.: Forecasting argumentation frameworks. In: Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning (KR), pp. 533–543 (2022)
14. Kaci, S., van der Torre, L.W.N., Villata, S.: Preference in abstract argumentation. In: Proceedings of the 7th International Conference on Computational Models of Argument (COMMA), pp. 405–412 (2018)
15. Mailly, J., Rossit, J.: Argument, I choose you! preferences and ranking semantics in abstract argumentation. In: Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning, (KR), pp. 647–651 (2020)
16. Modgil, S., Prakken, H.: A general account of argumentation with preferences. Artif. Intell. **195**, 361–397 (2013)
17. Potyka, N.: Continuous dynamical systems for weighted bipolar argumentation. In: Proceedings of the 16th International Conference on Principles of Knowledge Representation and Reasoning (KR), pp. 148–157 (2018)
18. Rago, A., Li, H., Toni, F.: Interactive explanations by conflict resolution via argumentative exchanges. In: Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning, (KR), pp. 582–592 (2023)
19. Rago, A., Toni, F., Aurisicchio, M., Baroni, P.: Discontinuity-free decision support with quantitative argumentation debates. In: Proceedings of the 15th International Conference on Principles of Knowledge Representation and Reasoning (KR), pp. 63–73 (2016)