Soc Choice Welf (2018) 51:737–752
https://doi.org/10.1007/s00355-018-1136-2

ORIGINAL PAPER

A paradox of expert rights in abstract argumentation

Nan Li[1]

Received: 26 July 2013 / Accepted: 3 June 2018 / Published online: 26 June 2018

Abstract This paper provides a "liberal paradox" that applies to the framework of abstract argumentation and complements the liberal paradox in preference aggregation. In abstract argumentation, arguments are viewed as abstract entities whose validities are determined according to a binary attack relation. When forming a collective attack relation, parts of it may be reserved to members of the society who hold expert knowledge. I prove that when only a binary evaluation of each argument is permitted, even under a minimal condition of rationality, the assignment of expert rights to two or more agents may be inconsistent with the condition of strong unanimity. Since argumentation aggregation is a particular case of judgement aggregation, this result might be a corollary of Dietrich and List (Soc Choice Welf 31(1):59–78, 2008), if the agenda I consider turns out to be connected in their sense, an issue that this paper has not been able to settle.

1 Introduction

The intent of this study is two-fold. The first objective is to show that the aggregation questions that have been studied thus far, primarily in the context of social choice theory and judgement aggregation, can also be discussed fruitfully in the framework of abstract argumentation.

Argumentation is an indispensable method of shaping the decisions of individuals and societies. It is a dynamic process, a colourful and sometimes unpredictable dialogue, and a ubiquitous phenomenon in many contexts, such as economic behaviours,

✉ Nan Li
nanli@swufe.edu.cn

[1] School of Public Finance and Taxation, Southwestern University of Finance and Economics, Edifice Gezhi, 555 Liutai Avenue, Wenjiang, Chengdu 611130, China

deliberative democracy and judicial practices. As a result, in economics there are many methods of analysing argumentation in general. For example, argumentation and controversy are modelled as part of game play in game theory, and deliberation and cheap talk are important topics in the theory of committees (Austen-Smith 1990; Glazer and Rubinstein 2001; Crawford and Sobel 1982).

According to Walton (2009), when arguing, one must identify the premises and conclusions of an argument, determine the implicit premises or conclusions in the argument, evaluate whether an argument is weak or strong based on general criteria that can be applied to it, and then construct new arguments that can be used to prove a specific conclusion. Considering the variety of tasks involved in and facets of argumentation, some authors of computer science and artificial intelligence have recently begun to model the structure of argumentation in frameworks that facilitate its analysis using the tools of logic. They have also embarked on the project of bridging theories of social choice, games and argumentation.[1]

As part of this endeavour, Dung (1995) introduces *abstract argumentation*, a landmark framework in which arguments are regarded as abstract entities with a binary attack relation among them, thereby resulting in a so-called *argumentation framework*. This relation, which reflects the strength of one argument against another, need not satisfy the standard conditions imposed on preferences. Thus, abstract argumentation theory has proven to be a rather new field distinct from those that economists were previously familiar with.

In this theory, the primitive concept is the attack relation, and the emphasis is placed not on the intrinsic content of the arguments but rather on the relationship among the arguments and the consequences of such relationships with regard to their overall acceptability. To simplify, if an argument is attacked by other arguments that are all unjustified, then the former should be regarded as justified; conversely, if an argument is justified, then any argument that is attacked by it should be regarded as unjustified. Consequently, this theory can help to address problems that are related to the manner in which agents debate, with themselves or with others, and to the manner in which they evaluate and form opinions. Alternatively stated, the theory of abstract argumentation not only facilitates the discussion of the internal validity of certain forms of reasoning but also assists in the formation of collective arguments as a result of aggregating the argumentation framework or arguments of different agents.[2] Hence, the different manner in which argumentation is modelled is worth incorporating into the toolkit of economists, not as an alternative but rather as a complement, and accordingly the types of issues that can be discussed will extend our horizons.

The second objective of this study is to show that the framework can be used to yield a new version of the liberal paradox. In abstract argumentation, in the first phase, an agent typically evaluates attack relations among arguments, then forms her individual argumentation framework in different ways, and hence may decide whether an argument is justified. For a fixed set of arguments, suppose different agents of a society sustain individually consistent yet different attack relations. This paper concentrates

[1] For an overall summary of the state-of-the-art achievements in this field, see Rahwan and Simari (2009).

[2] For several representative works, see Rahwan and Larson (2008), Rahwan et al. (2009), and Rahwan and Tohmé (2010). For abstract argumentation and judgement aggregation, see Caminada and Pigozzi (2011).

on the issues that may arise if this society must reach a consistent attack relation as a joint decision. However, it is not trivial to determine the correct mechanism of aggregating attack relations over a common set of arguments or whether certain desirable properties of possible aggregation mechanisms are compatible. My result reveals an essential difficulty, which may be expressed as follows. Suppose that despite certain differences regarding which set of arguments should prevail, there is basic agreement on two conditions. The first one is that of strong unanimity: broadly speaking, the attack or defence relation among arguments must be collectively accepted if all agents hold the same opinion. Second, there may exist a pair of arguments and a single individual, called the *expert* on that pair, such that her opinion regarding the attack relations between those two arguments must prevail. When only a binary evaluation (i.e., justified or not) of each argument is permitted, we will show that the condition of strong unanimity is incompatible with the assignment of expert rights to different people regarding different issues, even in the minimal case in which only two experts are called to rule on only two pairs of arguments. Thus, a paradox of expert rights holds in abstract argumentation. Taking into account of the importance of expert opinion in argumentation, and in social choice theory and artificial intelligence in general, we may view this result in its proper perspective.

The problem discussed and the result obtained herein bear a resemblance to Sen's (1970) celebrated result regarding the impossibility of a Paretian liberal in preference aggregation,[3] which was later generalized by Dietrich and List (2008) to judgement aggregation, where their main finding is that the liberal paradox holds if and only if the agenda of judgement aggregation is connected. The present work contributes to the classical debate concerning this tension by introducing abstract argumentation into our study. Although argumentation aggregation will be shown to be merely a special case of judgement aggregation, and thus, in principle, the former can be translated into the latter, it will become apparent that the technical requirements and the technique we use for proving the tension in argumentation aggregation are different. However, since we have not unraveled whether argumentation agenda is connected in the sense of Dietrich and List (2008), it is still an open question that whether my result is a consequence of their work. Even so, in a domain of interest in its own right, at least the outcome of the present research complements the preceding conclusion of Sen's (1970). Incidentally it also adds weight to a conjecture offered by Gaertner et al. (1992) that "[i]t is our *belief* that this problem[4] persists under virtually every plausible concept of individual rights that we can think of."

[3] Sen's paper, especially his formulation of the notion of rights, has faced various challenges since its publication. For some representative work, see Nozick (1974), Bernholz (1974), Gärdenfors (1981), Sugden (1981) and Gaertner et al. (1992), among others. For recent developments, see Deb et al. (1997), van Hees (1999, 2004), and Dowding and van Hees (2003). It is not my intent to clarify the notion of rights in the current paper. Even so, as we will see, because the right involved is that of determining the entirety of the social point of view regarding the attack relation instead of only part of it, Sen's approach may be considered to be more appropriate in this field than it is in preference aggregation.

[4] That is, the conflict between individual rights and Pareto optimality, a concept that is arguably highly relevant to the property "strong unanimity" defined in Sect. 4 (emphasis and footnote added).

2 Abstract argumentation: preliminaries

Dung (1995) presents one of the most influential computational models of argumentation. In his formulation of *argumentation framework*, the internal structure of arguments is ignored, and thus, arguments are regarded as abstract entities, with a binary attack relation among them. However, because my model addresses a given set of arguments if without explicit statement, we adopt the primitive concept below to highlight the essence. Also, compared with the conventional definition, we impose irreflexivity to exclude some inconsistency at the outset.

Definition 1 For a finite set of arguments $\mathcal{A}$, an ***attack relation*** is an irreflexive binary relation $\rightharpoonup$ on $\mathcal{A}$.[5] For any arguments $\alpha, \beta \in \mathcal{A}, \alpha \rightharpoonup \beta$ (or $\beta \leftharpoondown \alpha$) means that α ***attacks*** β, and we call α an ***attacker*** of β.

Remark 1 Based on the definition above, $\alpha \rightharpoonup \beta$ just means that α attacks β, while we do not know whether β attacks α. Accordingly we have to command the latter information to get the global picture as to the attack relations between this pair of arguments. Throughout the current work, unless specified otherwise, we follow this interpretation.

In fact, what we said here is that attack relation may not be symmetric. To avoid any possible false intuition, we should also point out that in some applications, attack relation is neither complete, nor transitive, and nor antisymmetric. Formally,

(1) *symmetric*: $\forall \alpha, \beta \in \mathcal{A}$: $(\alpha \rightharpoonup \beta) \rightarrow (\beta \rightharpoonup \alpha)$;[6]
(2) *complete*: $\forall \alpha, \beta \in \mathcal{A}$, where $\alpha \neq \beta$: $(\alpha \rightharpoonup \beta) \vee (\beta \rightharpoonup \alpha)$;
(3) *transitive*: $\forall \alpha, \beta, \gamma \in \mathcal{A}$: $(\alpha \rightharpoonup \beta) \wedge (\beta \rightharpoonup \gamma) \rightarrow (\alpha \rightharpoonup \gamma)$; and
(4) *antisymmetric*: $\forall \alpha, \beta \in \mathcal{A}$: $(\alpha \rightharpoonup \beta) \wedge (\beta \rightharpoonup \alpha) \rightarrow (\alpha = \beta)$.[7]

For any $\mathcal{B} \subseteq \mathcal{A}$ and $\alpha \in \mathcal{A}$, put $\mathcal{B}^{\rightharpoonup} = \{\gamma \in \mathcal{A} \mid \beta \rightharpoonup \gamma \text{ for some } \beta \in \mathcal{B}\}$ and $\alpha^{\leftharpoondown} = \{\gamma \in \mathcal{A} \mid \gamma \rightharpoonup \alpha\}$.

[5] The Liar Paradox is a famous example that concerns the problem of self-attack that derives from the classical statement "This sentence is false." For convenience, we call this argument κ. If κ is true, then by that argument, κ is false. If κ is false, however, because the argument states precisely that (namely, that it is false), then κ is true. We have thus shown that κ is true if and only if it is false. Although the Liar Paradox invokes the most fundamental challenge of self-reference in language and logic, it is irrelevant to my concerns here. Thus, the constraint that there may be no self-attacking arguments is desirable in my context. However, this postulate is not the general case in abstract argumentation theory, Dung (1995) included.

[6] Thus, not being symmetric means that α attacks β while β does not attack α, which may be the case that α is much powerful than β in argumentation. The two arguments γ and α in Example 1 below vividly illustrate this point.

[7] Thus, not being antisymmetric means that $\exists\, \alpha, \beta$: $(\alpha \rightharpoonup \beta) \wedge (\beta \rightharpoonup \alpha) \wedge (\alpha \neq \beta)$. This requirement is reasonable because it is common for two arguments to attack each other in real argumentations, which is especially true for debates that concern moral values. For example, let $\alpha =$ "Newspapers should not publish the presidential candidates' health information if they do not accede to the request because it concerns their private lives." and $\beta =$ "Newspapers should publish the presidential candidates' health information even if they do not accede to the request because it has public significance." Thus, we can see that $\alpha \rightharpoonup \beta$ and $\beta \rightharpoonup \alpha$ simultaneously. In fact, the foundation of this argumentation is the conflict between private life and public interest.

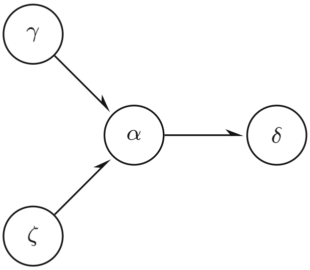

Fig. 1 A murder case. δ: the suspect is innocent; α: there is evidence that he was present at the crime scene *one hour before* the crime; γ: he was witnessed in a nearby town *at the time* of the crime; ζ: the police obtained evidence that *at the same time* he was on the telephone in that town

Definition 2 Let $\mathcal{B} \subseteq \mathcal{A}$ and $\alpha \in \mathcal{A}$. Given an attack relation $\rightharpoonup$, we say that $\mathcal{B}$ ***defends*** argument α if $\alpha^{\leftharpoonup} \neq \emptyset$ and $\alpha^{\leftharpoonup} \subseteq \mathcal{B}^{\rightharpoonup}$, and we then call $\mathcal{B}$ a ***defender*** of α.

Intuitively, a set of arguments $\mathcal{B}$ defends a given argument α if each attacker of α is attacked by some argument in $\mathcal{B}$.

To visualize an attack relation, we can represent it as a directed graph, i.e., a digraph,[8] in which the vertices are arguments and the directed arcs denote attack relations between arguments. As an example, an attack relation and its digraph are presented below.

Example 1 (A MURDER CASE) A murder case is under investigation. Argument δ states that the suspect is innocent. Argument α claims that there is evidence that he was present at the crime scene *one hour before* the crime. However, argument γ declares that he was witnessed in a nearby town *at the time* of the crime. Further, argument ζ asserts that the police obtained evidence that *at the same time* he was on the telephone in that town. The attack relation $\{\alpha \rightharpoonup \delta, \gamma \rightharpoonup \alpha, \zeta \rightharpoonup \alpha\}$ corresponds to the digraph depicted in Fig. 1.

Thus, when we consider an attack relation, although not being the key concern of the current paper, the determination of which arguments involved are justified and which are not is a natural problem. This concept can be expressed through argument labelling, an approach that has been extended by Caminada (2006) among others, on the basis of previous work. In the present work, we adopt only his labels `in` and `out`, and not the label `undec` (undecided). That is, only binary evaluation of each argument is allowed. The formal definitions follow.

Definition 3 A ***labelling*** is a function $l\colon \mathcal{A} \to \{\texttt{in}, \texttt{out}\}$.

Definition 4 Given an attack relation $\rightharpoonup$, a labelling $l\colon \mathcal{A} \to \{\texttt{in}, \texttt{out}\}$ is ***stable*** if

(1) $\forall \alpha \in \mathcal{A}$, $l(\alpha) = \texttt{in}$ if $l(\beta) = \texttt{out}$ for all βs (if any) that attack α, and
(2) $\forall \alpha \in \mathcal{A}$, $l(\alpha) = \texttt{out}$ if there is a β that attacks α and $l(\beta) = \texttt{in}$.

Using this language, the label `in` means that the argument is accepted or justified, and the label `out` the argument is rejected or unjustified.[9]

[8] For a comprehensive treatise on digraph theory, see Bang-Jensen and Gutin (2010).

[9] In the artificial intelligence literature, in addition to the notions of `in` and `out`, like Caminada, many authors also adopt the label `undec` to denote the labelling of an argument whose status, i.e., justified or

This definition works well for simple cases in which we can isolate the arguments that emerge victorious. For example, in the attack relation $\{\alpha \rightharpoonup \beta, \beta \rightharpoonup \gamma\}$, α is `in` because it is not attacked by any argument. Consequently, β is `out`, and γ is `in`. It is clear that this is the only possible stable labelling. However, many attack relations can accommodate multiple stable labellings. As an example, suppose that there are four arguments such that the attack relation among them is $\{\alpha \rightharpoonup \beta, \beta \rightharpoonup \gamma, \gamma \rightharpoonup \delta, \delta \rightharpoonup \alpha\}$. In this case, there exist two stable labellings l_1 and l_2, where $l_1(\alpha) = l_1(\gamma) = \texttt{in}$ and $l_1(\beta) = l_1(\delta) = \texttt{out}$, while $l_2(\alpha) = l_2(\gamma) = \texttt{out}$ and $l_2(\beta) = l_2(\delta) = \texttt{in}$.

However, this is not a universal rule; there also exist attack relations that cannot accommodate any stable labelling.

Example 2 Suppose that $\mathcal{A} = \{\alpha, \beta, \gamma\}$, and the attack relation among them is $\{\alpha \rightharpoonup \beta, \beta \rightharpoonup \gamma, \gamma \rightharpoonup \alpha\}$. Then, there is no stable labelling. In fact, e.g., if we deem argument α to be `in`, then according to Definition 4, argument β is `out`, and argument γ is `in`. Consequently, α should be `out`, posing a contradiction. The same problem arises when we initially deem argument α to be `out`.

Definition 5 We call an attack relation ***stable*** if it can accommodate at least one stable labelling; otherwise, it is ***unstable***.

In this paper, we consider only stable attack relations, assuming that such relation represents the minimal conditions of rationality.

3 The model: argumentation aggregation

In the previous section, we presented a preliminary introduction to the elements of abstract argumentation. This serves the primary objective of the current study, namely, to analyse how to aggregate different individual attack relations over a common set of arguments to achieve a social attack relation and to discuss the inconsistencies among certain desirable properties of that aggregation.[10] In this section, we introduce the model of argumentation aggregation.

We consider a group of agents $N = \{1, 2, \ldots, n\} (n \geq 2)$ and a finite set of arguments $\mathcal{A} = \{\alpha_1, \alpha_2, \ldots, \alpha_m\} (m \geq 2)$. Each agent $i \in N$ has her own attack relation $\rightharpoonup_i$. To show the overall attack relations just between a given pair of arguments $\alpha, \beta \in \mathcal{A}$, each agent can express her opinion by choosing one and only one of the

unjustified, cannot be decided. In some scenarios of the real life, e.g., judicial practices, however, such an undecided argument is not acceptable. In the present work, just as we classify arguments only as justified or unjustified, we do not adopt the label `undec`. In Sect. 4, it will be clear that this constraint is crucial to my paradox. For further discussion of argument labelling, see also Caminada and Pigozzi (2011).

[10] Distinct from my focus in the present work, Bodanza and Auday (2009) analyse the problem of aggregating individual attack relations over a common set of arguments to "obtain a unique socially justified set of arguments." They articulate the difference between the aggregation methods involved. That is, their work "can be done in two different ways: a social attack relation is built up from the individual ones, and then is used to produce a set of justified arguments, or this set is directly obtained from the sets of individually justified arguments." Their primary concern is "whether these two procedures can coincide or under what conditions this could happen." My task here starts from the first step of the first approach, although with a completely different destination.

four alternatives: (1) both arguments are perfectly compatible; (2) α attacks β, but not vice versa; (3) β attacks α, but not vice versa; or (4) they attack each other (expressing that they are in conflict but are the same powerful in argumentation). Formally, it is (1) $\alpha \not\rightleftharpoons \beta$; (2) $\alpha \rightharpoonup\!\!\not\leftharpoondown \beta$; (3) $\alpha \not\rightharpoonup\!\!\leftharpoondown \beta$; or (4) $\alpha \rightleftharpoons \beta$, respectively. Based on their individual opinions, this group of agents must form a collective opinion, as defined in the following two definitions.

Definition 6 The ***argumentation agenda*** X for $\mathcal{A}$ is the set of attack propositions on which judgements are made, i.e., $X = \{p_{\alpha\beta}, \neg p_{\alpha\beta} \mid \alpha, \beta \in \mathcal{A}, \alpha \neq \beta\}$, where proposition $p_{\alpha\beta}$ stands for $\alpha \rightharpoonup \beta$, and $\neg p_{\alpha\beta}$ stands for $\alpha \not\rightharpoonup \beta$, i.e., α does not attack β.[11]

Remark 2 According to this definition, accepting $\alpha \not\rightleftharpoons \beta$ corresponds to justifying $\neg p_{\alpha\beta} \wedge \neg p_{\beta\alpha}$, while accepting $\alpha \rightharpoonup\!\!\not\leftharpoondown \beta$ corresponds to justifying $p_{\alpha\beta} \wedge \neg p_{\beta\alpha}$, and accepting $\alpha \rightleftharpoons \beta$ corresponds to justifying $p_{\alpha\beta} \wedge p_{\beta\alpha}$.[12] Thus, the argumentation agenda can be recognized as a standard judgement agenda as defined by Dietrich and List (2008).

We introduce the argumentation aggregation mechanism in the following, and leave the discussion on the relationship between argumentation aggregation and judgement aggregation in greater detail to Sect. 5.

Definition 7 Let AR denote the set of all attack relations over $\mathcal{A}$, where $\mathcal{A}$ is a set of arguments. A ***social argumentation function*** is a mapping $f : AR^n \rightarrow AR$. We call $f(\rightharpoonup_1, \ldots, \rightharpoonup_i, \ldots, \rightharpoonup_n)$ the ***social attack relation***, where $i \in N$ and $\rightharpoonup_i$ is the attack relation of agent i.[13]

Example 3 (Social Attack Relation with Majority Rule) Suppose that there are three agents considering a set of three arguments, α, β, and γ. Their individual attack relations are

$$\begin{aligned} &\rightharpoonup_1 : \{\alpha \rightharpoonup \beta, \beta \rightharpoonup \gamma\}; \\ &\rightharpoonup_2 : \{\alpha \leftharpoondown \beta, \beta \rightharpoonup \gamma\}; \text{ and} \\ &\rightharpoonup_3 : \{\alpha \rightharpoonup \beta, \beta \leftharpoondown \gamma\}, \end{aligned}$$

[11] This formulation was suggested by the Editor.

[12] This was reminded by Juan Carlos García-Bermejo Ochoa.

[13] Bodanza and Auday (2009) define this concept in a similar manner. The difference are as follows: (1) they do not define the social argumentation function explicitly. Instead, they call it an aggregation of individual argumentation frameworks "according to some specified mechanism M"; (2) they impose no requirement of irreflexivity on either individual or social level such as that implied by the concept of attack relation here. On the other hand, Dunne et al. (2012) call this procedure "argument aggregation" and this mechanism "argument aggregation function". In our daily usage, argument means both a reason given in proof or rebuttal, and the act or process of arguing. Since we use "argument" in the first sense from the outset of the current work, we adopt "argumentation" to avoid ambiguity. In fact, just as footnote 10 above shows, there are two ways to obtain socially justified set of arguments. We feel that "argument aggregation" risks the implication of the second way of Bodanza and Auday (2009). However, that is not what we want to explore.

respectively. If this society adopts the majority rule as their social argumentation function, then the social attack relation is $\{\alpha \rightharpoonup \beta, \beta \rightharpoonup \gamma\}$, which coincides with $\rightharpoonup_1$ here.

4 Impossibility theorem

Now, suppose that we wish to find a social argumentation function f with the following intuitive properties:

Universal Domain (Condition D): The domain of f is the set of all profiles such that each agent's attack relation is stable.

Attack-Unanimity (Condition AU): For any $\alpha, \beta \in \mathcal{A}$, α attacks β for the society if α attacks β for each agent.

Defence-Unanimity (Condition DU): For any $\mathcal{B} \subseteq \mathcal{A}$ and $\beta \in \mathcal{A}$, $\mathcal{B}$ defends β for the society if $\mathcal{B}$ defends β for each agent.

Minimal Rights (Condition R): There are at least *two* agents such that for each of them, there is at least one pair of arguments between which she is decisive over the attack relation. That is, for all admissible profiles $(\rightharpoonup_1, \ldots, \rightharpoonup_n)$, writing $\rightharpoonup$ for $f(\rightharpoonup_1, \ldots, \rightharpoonup_n)$, we have $\alpha \rightharpoonup \beta \Leftrightarrow \alpha \rightharpoonup_i \beta$, and $\beta \rightharpoonup \alpha \Leftrightarrow \beta \rightharpoonup_i \alpha$, where i is the expert of the attack relation between argument α and β.

Remark 3 If we were to take the same approach as Dietrich and List (2008), there are two possible definitions of "unanimity principle" in my framework: we should have defined it either as Condition AU or DU. In some places below, when referring to these two conditions in combination, i.e., $AU \wedge DU$, we call it *strong unanimity*, or Condition SU in short, in the sense that this merged property is stronger than each of the two independent conditions above.

The following theorem reveals an inherent tension between liberal values and the Pareto principle in the most general scenario of abstract argumentation.

Theorem 1 *Given any set of two or more arguments and any number of two or more individuals, there exists no social argumentation function that satisfies Conditions D, SU, and R while generates stable social attack relation.*

Proof For some permissible profiles, I will show that any social argumentation function generates an unstable social attack relation, or violates some aforementioned conditions in this process.

Remember that for any pair of arguments, say α and β, an agent can express one of four alternatives: (1) $\alpha \not\rightleftharpoons \beta$, (2) $\alpha \rightharpoonup\!\!\not\leftharpoondown \beta$, (3) $\alpha \not\rightharpoonup\!\!\leftharpoondown \beta$, or (4) $\alpha \rightleftharpoons \beta$. For any society that respects expert rights, such an alternative must form the social opinion of the relation between α and β if this agent is decisive over the attack relation between these two arguments.

Let the two agents referred to in Condition R be 1 and 2, and let the two pairs of arguments be $\{\alpha, \beta\}$ and $\{\gamma, \delta\}$. Because it is sufficient to construct a counterexample to prove the theorem, we assume that there are no other arguments.

Let α, β, γ and δ all be distinct. Suppose that in addition to deeming that α attacks β, agent 1 also deems that β attacks γ, and γ attacks δ. Let everyone else in the

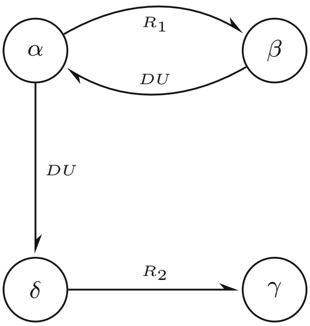

Fig. 2 A paradox of expert rights with four arguments

community, agent 2 included, deems that β attacks α, α attacks δ, and δ attacks γ. That is,

$$\text{agent } 1\colon \{\alpha \rightharpoonup \beta, \beta \rightharpoonup \gamma, \gamma \rightharpoonup \delta\};$$
$$\text{agent } 2, \ldots, \text{n}\colon \{\beta \rightharpoonup \alpha, \alpha \rightharpoonup \delta, \delta \rightharpoonup \gamma\}.$$

All attack relations are stable, and thus Condition D is satisfied. By Condition R, for the society, α attacks β and δ attacks γ based on the expert rights of agents 1 and 2, respectively. For the society, however, $\{\alpha\}$ defends γ and $\{\beta\}$ defends δ according to Condition DU.

Because δ is an attacker of γ, consequently α must attack δ. Following the same logic, it is clear that β must attack α. Without further exploration, e.g., regardless of whether β attacks γ, or whether δ attacks α, we can show the social attack relation up to this stage in Fig. 2.[14]

The result, however, violates the expert right of agent 1, who deems that α attacks β.

Now, let $\{\alpha, \beta\}$ and $\{\gamma, \delta\}$ have one argument in common, say $\alpha = \delta$. Assume now that agent 1 deems that α attacks β and that everyone else in the community, agent 2 included, deems that γ attacks δ $(= \alpha)$. Additionally, let everyone in the community, agents 1 and 2 included, deems that β attacks γ. That is,

$$\text{agent } 1\colon \{\alpha \rightharpoonup \beta, \beta \rightharpoonup \gamma\};$$
$$\text{agent } 2, \ldots, \text{n}\colon \{\beta \rightharpoonup \gamma, \gamma \rightharpoonup \delta \ (= \alpha)\}.$$

All attack relations are stable, and thus, Condition D is satisfied. By Condition R, for the society, α attacks β, and γ attacks δ $(= \alpha)$, whereas by Condition AU, β must attack γ. Consequently, we obtain the same unstable social attack relation as the one in Example 2.

[14] We indicate the force that determines the social attack relation between two arguments by a label next to the corresponding harpoon, where DU denotes Condition DU and R with a subscript the expert right of the corresponding agent.

Finally, if $\{\alpha, \beta\}$ and $\{\gamma, \delta\}$ are the same pair of arguments, then it is impossible to respect the rights of agents 1 and 2 simultaneously when they hold different opinions as to the attack relation between the two arguments. □

Remark 4 Although the case in which the two pairs of arguments have one argument in common bears some resemblance, at first glance, to the famous liberal paradox of Sen (1970), this connection is tangential. We know that when a preference is reflexive and complete, acyclicity of the preference relation is a sufficient and necessary condition for a choice function to be defined over a finite set of alternatives. Hence, acyclicity plays a determining role in the liberal paradox in preference aggregation. Acyclicity of attack relation, however, is only a sufficient but not a necessary condition for a stable labelling to be obtained. That is, in some cases, we can obtain a stable labelling even for a cyclic attack relation. An example of such a case was presented at the end of Sect. 2, namely, attack relation $\{\alpha \rightharpoonup \beta, \beta \rightharpoonup \gamma, \gamma \rightharpoonup \delta, \delta \rightharpoonup \alpha\}$, which is a cycle but accommodates two stable labellings.

The following example provides a good motivation for the present work.

Example 4 (Ghen v. Rich, 8 F. 159 (Mass. 1881) Adapted) Ghen v. Rich is a famous American property law case. In this case, plaintiff Ghen killed a whale at sea, leaving his identifying bomb lance in the whale. The custom and usage in the whaling industry in Cape Cod at that time was that the individual who killed a whale using a specially marked bomb lance owned that whale. If such a whale were found on a beach, the finder would notify the killer and receive a finder's fee.

The whale later washed up on shore 17 miles away and was discovered by Ellis. Ellis knew or should have known of the prevailing custom and usage in the whaling industry regarding the finding of a lost whale killed by another. He, however, sold the whale at an auction to defendant Rich, who then shipped off the blubber. Ghen discovered the fate of the whale and initiated a libel action against Rich to recover the value of the whale.

Bench-Capon (2002) shows that abstract argumentation can indeed represent case law, and Ghen v. Rich is one of his examples. For convenience of illustration, however, only four arguments from the case of Ghen v. Rich are abstracted here. Using Bench-Capon's phrasing, except for the first argument, they are as follows:

α: Failure of the plaintiff to receive compensation is unfair;
β: Defendant has the right to pursue his livelihood;
γ: Effort promising success to secure animal made by pursuer;
δ: Pursuer not in possession.

Accurately, what these arguments mean are: (1) argument α states that because Ghen executed the dangerous aspect of whaling by landing the harpoons, it is unfair for him not to receive any compensation; (2) argument β states that defendant Rich had the right to pursue his livelihood and that this pursuit also enhanced social welfare by providing food for others; (3) argument γ states that Ghen's effort to seize the whale was, in some sense, adequate; and (4) argument δ states that the pursuer Ghen failed to possess the whale physically.

Having borrowed this case as the prototype for our discussion, let us now deviate from it by fabricating the story left. Assume that there are two Judges, Clark and

Daniel, in the court. Their individual attack relations are as follows:

$$\text{Clark}\colon \{\alpha \rightharpoonup \beta, \beta \rightharpoonup \gamma, \gamma \rightharpoonup \delta\};$$
$$\text{Daniel}\colon \{\beta \rightharpoonup \alpha, \alpha \rightharpoonup \delta, \delta \rightharpoonup \gamma\}.$$

Consequently, Clark supports the claim of plaintiff Ghen, and Daniel supports the claim of defendant Rich.

Suppose that the court assigns Clark to determine the collective attack relation between arguments α and β, i.e., a discourse involving fairness, right and welfare, and assigns Daniel to determine the collective attack relation between arguments γ and δ, i.e., a discourse involving possession and ownership. In addition, the court accepts the unanimous attack and defence relation among arguments. Then, as shown by the proof of Theorem 1, the final collective attack relation violates the expert right of Clark.

In the domain of classical preference aggregation, we can show that for a seemingly similar problem, it is straightforward to find a desirable solution.

Example 5 In a society, there are n agents who must decide the social welfare ranking among four alternatives A, B, C and D. Suppose that their preferences are

$$\text{agent } 1\colon A \succ B \succ C \succ D;$$
$$\text{agent } 2, \ldots, \text{n}\colon B \succ A \succ D \succ C.$$

Let the preference of agent 1 concerning alternatives A and B and the preference of agent 2 concerning alternatives C and D be regarded as decisive. The social preference should also satisfy the classical constraints of unrestricted domain and the Pareto principle. It is trivial to show that the only solution is

$$A \succ B \succ D \succ C.$$

Remark 5 By contrast, in the case of my proof above, a seemingly similar attack relation $\{\alpha \rightharpoonup \beta, \beta \rightharpoonup \delta, \delta \rightharpoonup \gamma\}$ violates Condition DU on argument pairs $\{\alpha, \gamma\}$ and $\{\beta, \delta\}$. In Remark 4, we demonstrated a fundamental difference between the postulate of a preference and that of an attack relation. Here, from another perspective, the example above highlights this difference again. In preference aggregation, we require that individual preferences be transitive and that the resulting collective preference be acyclic. In abstract argumentation, however, once there exists an attack relation $\{\alpha \rightharpoonup \beta, \beta \rightharpoonup \gamma\}$, whether isolated or as a subset of a larger attack relation, argument α belongs to the defender of argument γ if such a defender exists. In short, we do not need attack relation to be transitive.

Remark 6 Although we have demonstrated the paradox only in the context of expert rights, we regard this as a convenient interpretation in argumentation aggregation. Formally, expert rights share the same structure as liberal rights in the literature on preference aggregation, where certain individuals are assigned rights to determine the social statuses of certain alternatives.

5 Discussion and related work

The liberal paradox not only haunts preference aggregation but also appears in judgement aggregation, an active emerging multidisciplinary field. In Dietrich and List (2008), the authors identify a problem that generalizes Sen's liberal paradox. Under plausible conditions, they prove that the assignment of rights to two or more agents or subgroups is also inconsistent with the unanimity principle.

Simply speaking, they consider a group of agents $N = \{1, 2, \ldots, n\}(n \geq 2)$ and an agenda, i.e., a non-empty subset X of logic $\mathbf{L}$ expressed as $X = \{p, \neg p\colon\ p \in X_+\}$ for a set $X_+ \subseteq \mathbf{L}$ of unnegated propositions on which binary judgements (i.e., yes or no) are made. They call propositions $p, q \in X$ *conditionally dependent* if there exist some $p^* \in \{p, \neg p\}$ and $q^* \in \{q, \neg q\}$ such that $\{p^*, q^*\} \cup Y$ is inconsistent for some $Y \subseteq X$ that is consistent with each of p^* and q^*. The agenda X is considered to be *connected* if any two propositions $p, q \in X$ are conditionally dependent. Their main finding is that if and only if the agenda is connected, then there exists no aggregation function F that generates consistent collective judgement sets and simultaneously satisfies the conditions of universal domain, minimal rights and the unanimity principle.[15]

Moreover, after a trivial transformation from the question of whether alternative a is strictly better than alternative b to the question of whether the proposition "alternative a is strictly better than alternative b" is true, they prove that the preference agenda is connected. Consequently, Sen's liberal paradox naturally becomes a corollary of their finding.

5.1 Argumentation aggregation: a special case of judgement aggregation

Because judgement aggregation and abstract argumentation share certain common interests and both depend on the toolkit of logic in different senses, especially because of the seeming relationship between a *connected* judgement agenda and a digraph, it may be conjectured that the result of Dietrich and List (2008) already implies the findings of the current paper.

To put things into perspective, we must first elucidate the relationship between argumentation aggregation and judgement aggregation. Just as Definition 6 and Remark 2 indicate, if considering this issue from the perspective of the definition of argumentation agenda, we see that argumentation aggregation is merely a special case of judgement aggregation.

Another approach is to associate each of the four possibilities of attack relation noted above for any pair of arguments α and β with a certain issue and to explicitly enumerate the space of all stable attack relations. Thus, my framework can be embedded into binary aggregation. In this manner, by Proposition 2.1 of Dokow and Holzman

[15] To be specific, they define these three properties as follows: Universal Domain: the domain of the aggregation function F is the set of all possible profiles of consistent and complete individual judgement sets; Minimal Rights: there exist (at least) two agents who are each decisive regarding (at least) one proposition-negation pair $\{p, \neg p\} \subseteq X$; Unanimity Principle: for any profile $(A_1, \ldots, A_n)$ in the domain of F and any proposition $p \in X$, if $p \in A_i$ for all agents i, then $p \in F(A_1, \ldots, A_n)$, where A_i is the judgement set of agent i.

(2009), this framework can, in principle, ultimately be transformed into judgement aggregation, although this task may not be easy in practice for certain specific cases.[16]

5.2 The relationship between the ranges and the axioms

As a second step, we compare the technical requirements of the two results.

First, since a stable attack relation corresponds exactly to a complete and consistent judgement set, my condition of universal domain is equivalent to the counterpart of Dietrich and List (2008). And since they only impose consistency on the collective judgement set, my social rationality requirement is stronger than theirs. That is, my range is smaller.

Second, as explained in Remark 3, strong unanimity considered here is stronger than the unanimity principle of Dietrich and List (2008).

And third, stated in my framework, their minimal rights condition requires that there are at least two agents such that for each of them, say i, there is at least one pair of arguments α and β such that $\alpha \rightharpoonup \beta \Leftrightarrow \alpha \rightharpoonup_i \beta$ for all admissible profiles $(\rightharpoonup_1, \ldots, \rightharpoonup_n)$. This is in stark contrast to my minimal rights condition as defined in Sect. 4. Or put it another way, observing from the perspective of the definition of argumentation agenda, the expert rights of Dietrich and List (2008) mean that the expert is decisive on proposition $p_{\alpha\beta}$, while my expert rights mean that the expert is decisive on propositions $p_{\alpha\beta}$ and $p_{\beta\alpha}$. In this sense, the condition of minimal rights considered in this work is also stronger than theirs.

5.3 The relationship between existing liberal paradoxes: an open question

The main finding of Dietrich and List (2008) is that if and only if the agenda is connected, then there exists no aggregation function F that generates consistent collective judgement sets and simultaneously satisfies the conditions of universal domain, minimal rights and the unanimity principle. That is, the function F exists whenever the agenda is not connected. Therefore, besides the discussion above, to show the relationship between my result and the one of Dietrich and List (2008), we should answer whether argumentation agenda falls under their impossibility condition of the agenda, viz., connectedness. In fact, this also depends on the generality of the agenda.

On the other hand, just as Sect. 5.2 shows, for range, rights and unanimity, my conditions are stronger and thus Theorem 1 is a corollary of their result if argumentation agenda is "connected" in their sense. Even so, my direct proof has its own significance. While if argumentation agenda is not "connected" in their sense, however, taking into account stronger social rationality requirement and axioms, the result of my current work does not conflict with theirs either. Unfortunately until now we haven't figured out whether my agenda defined in Definition 6, as translated into judgement aggregation formalism of Dietrich and List (2008), is "connected" in their sense. As the result, we have not yet succeeded in showing the relationship between these two results.

[16] This perspective and the related literature were suggested by a referee.

Finally, regarding the difference between my result and that of Sen (1970), my analysis is somewhat scattered throughout the current paper. Here, a short summary is in order. First, in the aggregation mechanism, from the perspective of the input, preference aggregation requires transitivity of preference, whereas the present work does not; moreover, from the perspective of the output, preference aggregation requires acyclicity, whereas this condition is also not required in this work. Second, because the structure of the binary relations involved and the stability requirements that are imposed in my case are nonidentical in nature, the proof of my result, especially the case where four arguments are different, implies a clear distinction from Sen's (1970) counterpart. Therefore, this work is complementary to the one of Sen (1970).

5.4 Related work

In addition to the studies reviewed above, there are several others in the literature related to the interface between social choice theory and abstract argumentation. Among them, Rahwan and Tohmé (2010) is closely related to the present work. However, there is a salient fundamental divergence between my framework and theirs: whereas each agent may have different attack relation in my model, in that of Rahwan and Tohmé (2010), the agents all have the same attack relation and the only difference lies in their labellings. Based on this treatment, the two authors reveal an impossibility in the same vein as that of Arrow. In addition, they focus on argument-wise plurality voting and fully characterize the space of individual judgements that guarantees collective rationality via that aggregation mechanism.

Caminada and Pigozzi (2011) apply abstract argumentation to judgement aggregation by introducing operators that do not violate any of the agents' views. Nevertheless, what they aggregate is the labellings of the arguments instead of the attack relations between pairs of arguments. In fact, in their scenario, all attack relations are the same for all agents. They demonstrate how to map a judgement aggregation problem into an argumentation framework for certain cases. Even so, they acknowledge that "whether such mapping exists for all kinds of judgement aggregation problems is still an open question."

Similar to my model, Dunne et al. (2012) explore the scenario in which each agent has a different attack relation and the mechanism for achieving the social position is voting. Although these authors present some impossibility results and introduce the concept of unanimous attack, among others, which is reminiscent of my condition of attack-unanimity, their main focus is driven by a discussion of computational complexity.

Coste-Marquis et al. (2007) focus on the merging of Dung's argumentation frameworks. Unlike my fixed set of arguments, even the set of arguments for each agent may be distinct in their system. They develop a procedure to accommodate this divergence by expanding these sets of argumentation frameworks to frameworks with the same set of arguments. Then they merge all these expanded frameworks based on their concept of minimized distance, rather than through voting. Whereas this is a deliberative procedure with some desirable properties, our setup and intentions diverge considerably.

Finally, while it is less general than the frameworks of judgement aggregation, graph aggregation is an important basic problem. In this context, although with no direct focus on argumentation, Endriss and Grandi (2014) formulate desirable axioms and define certain aggregators. More importantly, they refine the ultrafilter method and relate it to the axioms and collective rational requirements to prove Arrovian impossibilities.

6 Conclusion

In the context of expert rights, this paper shows that three desirable properties, viz., universal domain, strong unanimity and minimal rights, are inconsistent in abstract argumentation. That is, we find that the liberal paradox, which captures a tension between liberal values and the Pareto principle, also exists in a new and significantly different form in argumentation aggregation. Although this aggregation is a special domain of judgement aggregation, it is distinct from preference aggregation. And compared to the existing liberal paradox in preference aggregation, this paper provides a complementary result. However, to get a clear perspective of the significance of this work, in the future we still need to study whether my agenda defined in the form of judgement aggregation is connected in the sense of Dietrich and List (2008), and thus whether my upshot is a corollary of theirs.

Gaertner et al. (1992) conjecture that the inherent conflict between liberal values and the Pareto principle should be pervasive throughout many different contexts. In light of the detailed analysis above, my finding bears witness to the speculation proposed more than two decades ago.

Acknowledgements My greatest debt is to Salvador Barberà. Although in a form different from that presented in this paper, it was he who introduced me to the problem of argumentation. I have benefited immensely from the many helpful discussions I have had with him. I am also grateful for discussions with Juan Carlos García-Bermejo Ochoa, Franz Dietrich and Xavier Vilà; comments from Carmen Beviá, Richard Booth, Martin Caminada, Gregory Gutin, Paola Manzini, Marco Mariotti, Jordi Massó, Prasanta Pattanaik, Ariel Rubinstein, Leon van der Torre, Jörgen Weibull, Yongsheng Xu, and Jesús Zamora Bonilla; and advice concerning references from Paul E. Dunne and Iyad Rahwan. I would like to thank the three anonymous referees of the IJCAI 2011 Workshop on Social Choice and Artificial Intelligence for their insightful comments on a previous version of this paper with the title "A Liberal Impossibility of Abstract Argumentation" as well as the participants of presentations given in Madrid, London, Stockholm, Luxembourg, Barcelona and New Delhi for their helpful suggestions. In the final stage of preparation, I am very grateful to two anonymous referees for their thoughtful suggestions. Especially I owe a great deal to an editor for his extraordinarily detailed critique and unusual professional dedication. The study is based on one part of my dissertation. I gratefully acknowledge support from the Ministerio de Ciencia e Innovación de España and FEDER through the project ECO2008-04756. All errors are my own.

References

Austen-Smith D (1990) Information transmission in debate. Am J Polit Sci 34(1):124–152

Bang-Jensen J, Gutin G (2010) Digraphs: theory, algorithms and applications, 2nd edn. Springer, London

Bench-Capon TJM (2002) Representation of case law as an argumentation framework. In: Bench-Capon TJM, Daskalopulu A, Winkels RGF (eds) Legal knowledge and information systems: JURIX 2002: the fifteenth annual conference. IOS Press, Amsterdam, pp 103–112

Bernholz P (1974) Is a Paretian liberal really impossible? Public Choice 20(1):99–107

Bodanza GA, Auday MR (2009) Social argument justification: some mechanisms and conditions for their coincidence. In: Sossai C, Chemello G (eds) Lecture Notes in Computer Science: symbolic and quantitative approaches to reasoning with uncertainty, vol 5590/2009. Springer, Berlin, pp 95–106

Caminada M (2006) On the issue of reinstatement in argumentation. In: Fisher M, van der Hoek W, Konev B, Lisitsa A (eds) Lecture Notes in Computer Science: logics in artificial intelligence, 10th European conference, JELIA 2006, Liverpool, UK, September 13–15, 2006 proceedings, vol 4160/2006. Springer, Berlin, pp 111–123

Caminada M, Pigozzi G (2011) On judgement aggregation in abstract argumentation. J Auton Agents Multi-Agent Syst 22(1):64–102

Coste-Marquis S, Devred C, Konieczny S, Lagasquie-Schiex M-C, Marquis P (2007) On the merging of Dung's argumentation systems. Artif Intell 171(10–15):730–753

Crawford VP, Sobel J (1982) Strategic information transmission. Econometrica 50(6):1431–1451

Deb R, Pattanaik PK, Razzolini L (1997) Game forms, rights, and the efficiency of social outcomes. J Econ Theory 72(1):74–95

Dietrich F, List C (2008) A liberal paradox for judgement aggregation. Soc Choice Welf 31(1):59–78

Dokow E, Holzman R (2009) Aggregation of binary evaluations for truth-functional agendas. Soc Choice Welf 32(2):221–241

Dowding K, van Hees M (2003) The construction of rights. Am Polit Sci Rev 97(2):281–293

Dung PM (1995) On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and N-person games. Artif Intell 77(2):321–358

Dunne PE, Marquis P, Wooldridge M (2012) Argument aggregation: basic axioms and complexity results. Computational Models of Argument (COMMA 2012), pp 129–140

Endriss U, Grandi U (2014) Collective rationality in graph aggregation. In: Schaub T, Friedrich G, O'Sullivan B (eds) ECAI 2014: 21st European conference on artificial intelligence. IOS Press, Amsterdam, pp 291–296

Gaertner W, Pattanaik P, Suzumura K (1992) Individual rights revisited. Economica 59(234):161–77

Gärdenfors P (1981) Rights, games and social choice. Noûs 15(3):341–56

Glazer J, Rubinstein A (2001) Debates and decisions: on a rationale of argumentation rules. Games Econ Behav 36(2):158–173

Nozick R (1974) Anarchy, State, and Utopia. Basil Blackwell, Oxford

Rahwan I, Larson K (2008) Pareto optimality in abstract argumentation. In: Cohn A (ed) Proceedings of the 23rd AAAI conference on artificial intelligence, vol I. AAAI Press, Chicago, pp 150–155

Rahwan I, Simari GR (eds) (2009) Argumentation in artificial intelligence. Springer, Dordrecht

Rahwan I, Tohmé F (2010) Collective argument evaluation as judgement aggregation. In: van der Hoek W, Kaminka GA, Lespérance Y, M Luck, Sen S (eds) Proceedings of 9th international conference on autonomous agents and multiagent systems (AAMAS 2010), vol I. International Foundation for Autonomous Agents and Multiagent Systems, Toronto, pp 417–424

Rahwan I, Larson K, Tohmé F (2009) A characterisation of strategy-proofness for grounded argumentation semantics. In: Kitano H (ed) IJCAI'09 proceedings of the 21st international joint conference on artificial intelligence. Morgan Kaufmann, San Francisco, pp 251–256

Sen AK (1970) The impossibility of a Paretian liberal. J Polit Econ 78(1):152–157

Sugden R (1981) The political economy of public choice. John Wiley, New York

van Hees M (1999) Liberalism, efficiency, and stability: some possibility results. J Econ Theory 88(2):294–309

van Hees M (2004) Freedom of choice and diversity of options: some difficulties. Soc Choice Welf 22(1):253–266

Walton D (2009) Argumentation theory: a very short introduction. In: Rahwan I, Simari GR (eds) Argumentation in artificial intelligence. Springer, Dordrecht, pp 1–22