Social Choice and Welfare
https://doi.org/10.1007/s00355-025-01600-2

ORIGINAL PAPER

Eliciting the deserving winner in the presence of enemies

Pablo Amorós[1]

Received: 28 July 2024 / Accepted: 21 March 2025

Abstract
We analyze the problem of a jury that has to select the deserving winner from a group of candidates when (i) the identity of the deserving winner is known to all jurors but not verifiable, (ii) each juror identifies with a different candidate whom they want to favor, and (iii) some jurors may have enemies among the candidates whom they want to harm. We introduce a necessary condition relating to the jurors' enemies for implementing the deserving winner, called minimal impartiality. The mechanisms proposed in the literature to implement the deserving winner via backward induction fail when jurors have enemies, even though minimal impartiality is satisfied. We propose a simple sequential mechanism that successfully implements the deserving winner via backward induction, whether the jurors have enemies or not, as long as minimal impartiality is satisfied.

1 Introduction

This work analyzes the problem in which a jury has to choose a winner from a group of candidates when (i) the identity of the deserving winner is known to all jurors but not verifiable, (ii) each juror identifies with a different candidate they want to favor, and (iii) some jurors may have an enemy among the candidates they intend to harm. For example, this identification between jurors and candidates occurs in some international competitions, such as those in certain sports, where each country is represented by one candidate and one juror. In such instances, it is reasonable to think that each juror will strive to favor the candidate from their own nationality. At the same time, geopolitical tensions may compel a juror to seek harm to a candidate from a particular foreign

Financial assistance received under project PID2023-147391NB-I00, funded by MCIU/AEI/10.13039/501100011033 / FEDER UE, and funding for open access charge from Universidad de Málaga / CBUA is gratefully acknowledged.

✉ Pablo Amorós
pag@uma.es

[1] Departamento de Teoría e Historia Económica, Universidad de Málaga, Calle El Ejido 6, 29013 Málaga, Spain

country.[1] In order to address this problem, the task is to design a mechanism (or voting system) that incentivizes jurors to choose the deserving winner, whoever they may be. The socially optimal rule is implementable when such a mechanism exists.

Some works, such as Amorós (2011) or Adachi (2014), have proposed mechanisms that implement the socially optimal rule via backward induction when jurors want to favor their friends, but they do not account for situations where some jurors may also have enemies. The existence of enemies alters the strategic behavior of jurors because, in addition to the existing incentives to deceive to ensure that their favorite candidate is chosen, new incentives to deceive emerge to prevent their enemy from being selected when they are the deserving winner. The combination of these incentives could potentially render the implementation of the socially optimal rule impossible. To prevent this, and under the assumption that the planner knows the identity of the candidate each juror identifies with and that of their enemy (if they have one), a specific condition relating to the jurors' enemies must be satisfied. This condition, termed "minimal impartiality," requires that for every pair of candidates, a juror for whom neither of them is an enemy exists. We demonstrate that minimal impartiality is necessary for implementing the socially optimal rule via backward induction.

Unfortunately, the mechanisms of Amorós (2011) and Adachi (2014) mentioned above prove ineffective when jurors may have enemies, despite the fulfillment of minimal impartiality. In this paper, we propose a novel, simple, and natural mechanism that successfully implements the socially optimal rule via backward induction, whether all jurors are impartial regarding all candidates other than their own or some have enemies among the other candidates, as long as minimal impartiality is fulfilled.

In our mechanism, jurors are assigned numbers from 1 to n (where n is the total number of jurors) to announce a winner sequentially. The announcement by juror 1 is implemented if either (1) he has no enemy and does not propose his friend as the winner, or (2) he has an enemy whom he proposes as the winner. If juror 1 has an enemy, e_1, and proposes a candidate x who is neither his friend nor his enemy as the winner, then we move to another stage in which a juror who is impartial between x and e_1 chooses the winner between those two candidates (minimal impartiality guarantees the existence of such an impartial juror). If juror 1 proposes his friend as the winner, the process moves on to juror 2, and this sequence continues. When it is the turn of juror n because jurors 1, 2 , ..., and $n-1$ have announced their friends, the candidate announced by juror n is chosen as the winner, even if that candidate is his friend.

For the proposed mechanism to work, the jurors' order of intervention must satisfy certain conditions. Specifically, agents numbered $n-3$, $n-2$, and $n-1$ must satisfy the following: $n-1$ is not the enemy of $n-2$ or $n-3$, $n-2$ is not the enemy of $n-1$, and the enemies of $n-1$ and $n-2$ (if they exist) are distinct. Fortunately, minimal impartiality guarantees the existence of three jurors who meet these conditions.

Related literature

The first paper analyzing a model similar to that of this work is Amorós et al. (2002). They study the problem of selecting a deserving ranking of candidates (instead of a single winner) that is known to all jurors. Under the assumption that each juror

[1] Other examples are the election of the European Commission's president by the EU countries' heads of government or the Papal election process by the cardinals.

identifies with one candidate to whom they want to favor and has no enemies, they demonstrate that while implementation of the deserving ranking in dominant strategies is not possible, its implementation in Nash equilibrium is possible. Amorós (2023) analyzes the same model as the previous work and proposes two natural sequential mechanisms that implement the deserving ranking via backward induction. Amorós (2009) explores a more general model in which jurors may have enemies and provides a necessary and sufficient condition for the Nash implementability of the deserving ranking. However, the mechanism proposed to demonstrate the sufficient part of this condition is a variation of Maskin's (1999) canonical mechanism, which has been criticized in the literature for being unnatural (see, e.g., Jackson 1992). Yadav (2016) analyzes the problem of implementing the deserving winners in Nash equilibrium when some jurors are "partially honest" in the sense that they have a strict preference for revealing the true state when truth-telling does not lead to a worse outcome for them.

Holzman and Moulin (2013) also analyze the problem of determining a winner when the voters are the candidates. They study a model where each agent has a private opinion about which of his peers is best and yet is indifferent about which of his peers wins, and the objective is to design direct mechanisms that use these opinions as input. The authors prove that no such mechanism (i) has the property that honesty is always a dominant strategy and (ii) respects consensus about who should win and who should lose. It is interesting to compare the current paper's positive result to the negative result of Holzman and Moulin (2013). On the one hand, this paper makes things easier in a few ways: (i) agents do not have arbitrary private opinions, but rather the deserving winner is observed by all agents (although not verifiable), (ii) it involves fewer possible type profiles because, despite some known conflicts of interest, agents otherwise have a preference for the deserving winner, and (iii) it weakens the solution concept from dominant strategy equilibrium to backward induction equilibrium. On the other hand, this paper makes things harder in a few ways: (i) it requires full implementation, and (ii) it relaxes what is common knowledge about preferences. In this paper, the common knowledge is described by a jury configuration specifying all conflicts of interest, which still leaves room for a lot of unknown comparisons (and thus incentives to manipulate), whereas in the simplest interpretation of Holzman and Moulin, the full preference profile is common knowledge.

Niemeyer and Preusser (2023) study a model similar to that of Holzman and Moulin in which an object is allocated among a set of agents, the optimal allocation depends on the agents' information about their peers, and each agent wants the object for himself. However, this model does not consider the possible existence of "enemies". Mackenzie (2015) analyzes a stochastic version of the Holzman and Moulin's model. Tamura (2016) establishes a characterization result in the context of impartial nomination rules that satisfy anonymity, symmetry, and monotonicity. Mackenzie (2020) studies how the pope is elected in the Roman Catholic Church. This problem falls within the cases analyzed by our model since the cardinals are both the jurors and the candidates. Finally, Olckers and Walsh (2024) conduct a survey on peer mechanisms in which the competitors for a prize also determine who wins.

The rest of the paper is organized as follows. Section 2 describes the model and notation. Section 3 states the necessary condition for implementation. Section 4 intro-

duces our mechanism and shows that it implements the deserving winner. Section 5 provides concluding remarks. The Appendix shows that the mechanisms suggested by Amorós (2011) and Adachi (2014) fail when agents have enemies and analyzes two possible extensions.

2 Preliminaries

Let N be a set of $n \geq 4$ candidates participating in a competition. Each candidate has a distinct representative on a jury tasked with selecting one winner from among them. We use agent i to denote candidate i and their corresponding jury representative. There exists a *deserving winner*, denoted as $\omega^d \in N$, who is known to all agents but cannot be verified. The socially optimal outcome is the victory of this deserving winner. However, agents exhibit biases in the following ways:

(1) In line with Amorós (2011), each agent prioritizes their own victory, regardless of the deserving winner.
(2) Unlike Amorós (2011), each agent may have an *enemy* who is always their least preferred candidate, regardless of the deserving winner.
(3) Each agent remains impartial toward all other agents except themselves and their enemy (if they have one). Specifically, if an agent cannot win, and the deserving winner is not their enemy, they prefer the deserving winner to emerge as the victor.

Let us formalize these concepts. Each agent $i \in N$ is characterized by an element $e_i \in (N \backslash \{i\} \cup \{\emptyset\})$, known by the planner. If $e_i \neq \emptyset$, we refer to e_i as the enemy of i. If $e_i = \emptyset$, we say that i has no enemy. Let $\Re$ be the class of preference relations defined over N. Each agent i has a *preference function* $R_i : N \longrightarrow \Re$ that maps each deserving winner ω^d to a preference relation $R_i(\omega^d) \in \Re$. We denote the strict component of $R_i(\omega^d)$ as $P_i(\omega^d)$.

Definition 1 A preference function $R_i : N \to \Re$ is *admissible* for agent $i \in N$ at $e_i \in (N \backslash \{i\} \cup \{\emptyset\})$ if:

(1) For every $\omega^d \in N$ and $j \in N \backslash \{i\}$, we have $i \; P_i(\omega^d) \; j$.
(2) If $e_i \neq \emptyset$, then, for every $\omega^d \in N$ and $j \in N \backslash \{e_i\}$ we have $j \; P_i(\omega^d) \; e_i$.
(3) For every $j, k \in N \backslash \{i, e_i\}$, if $\omega^d = j$ then $j \; P_i(\omega^d) \; k$.

Example 1 (*Admissible preference functions*). Suppose that $N = \{a, b, c, d\}$. Suppose that $e_a = d$ and $e_b = \emptyset$. Table 1 illustrates the constraints the admissible preference functions of agents a and b must satisfy. Agents ranked higher in the table are preferred over those ranked lower. A question mark between two agents means that there are no specific restrictions regarding one being preferred over the other.

Let $\mathcal{R}_i(e_i)$ denote the class of all preference functions that are admissible for agent i at e_i. A *jury configuration* is a profile $e = (e_i)_{i \in N}$ and it is known to the planner. A profile $R \equiv (R_i)_{i \in N}$ of preference functions is admissible at e if $R_i \in \mathcal{R}(e_i)$ for every $i \in N$. Let $\mathcal{R}(e)$ denote the set of admissible profiles of preference functions at e. The planner knows that the agents' preference functions are in $\mathcal{R}(e)$, although he does not know the actual functions.

Table 1 Admissible preference functions for agents a and b in Example 1

	$R_a : N \longrightarrow \Re$				$R_b : N \longrightarrow \Re$			
$\omega^d =$	a	b	c	d	a	b	c	d
	a	a	a	a	b	b	b	b
Pref.	$b?c$	b	c	$b?c$	a	$a?c$	$?d$	c	d
	d	c	b	d	$c?d$		$a?d$	$a?c$
		d	d					

Given a jury configuration e, a *state* is a profile $(\omega^d, R) \in N \times \mathcal{R}(e) \equiv S(e)$. The *socially optimal rule* is the function $\varphi : S(e) \to N$ such that, for each $(\omega^d, R) \in S(e)$, $\varphi(\omega^d, R) = \omega^d$ (*i.e.* , for each admissible state, φ selects the deserving winner).

A normal form mechanism is a pair (M, g), where $M = \times_{i \in N} M_i$, M_i is a message space for agent i, $g : M \to N$ is an outcome function mapping the agents' messages to a final winner, and agents send messages simultaneously. A *sequential mechanism* is a dynamic mechanism in which agents make choices sequentially. Given a jury configuration e, a finite sequential mechanism with perfect information *implements* the socially optimal rule via backward induction if, for each admissible state $(\omega^d, R) \in S(e)$, the only outcome reached via backward induction is ω^d.[2] The socially optimal rule is implementable if a mechanism exists that implements it.

3 A condition for implementation

Our first result states a necessary condition on the jury configuration for implementing the socially optimal rule via backward induction. This condition requires that, for every pair of agents, there is always a third agent who does not have either of them as an enemy.

Definition 2 A jury configuration e is *minimally impartial* if, for every $i, j \in N$, there is some $k_{i,j} \in N \backslash \{i, j\}$ such that $e_{k_{i,j}} \notin \{i, j\}$.[3]

Example 2 (*Minimal impartiality*). Let $N = \{a, b, c, d\}$. Then, there are six possible pairs of agents: ab, ac, ad, bc, bd, and cd. The jury configuration $e = (e_a, e_b, e_c, e_d) = (b, c, a, \emptyset)$ is minimally impartial. To see this, note that: (1) $d \in N \backslash \{a, b\}$ and $e_d \notin \{a, b\}$, (2) $d \in N \backslash \{a, c\}$ and $e_d \notin \{a, c\}$, (3) $b \in N \backslash \{a, d\}$ and $e_b \notin \{a, d\}$, (4) $d \in N \backslash \{b, c\}$ and $e_d \notin \{b, c\}$, (5) $c \in N \backslash \{b, d\}$ and $e_c \notin \{b, d\}$, and (6) $a \in N \backslash \{c, d\}$ and $e_a \notin \{c, d\}$. However, the jury configuration $\hat{e} = (\hat{e}_a, \hat{e}_b, \hat{e}_c, \hat{e}_d) = (d, c, \emptyset, \emptyset)$, despite having fewer agents with enemies than e, is not minimally impartial since, for every $i \in N \backslash \{c, d\}$, we have $\hat{e}_i = c$ or $\hat{e}_i = d$.

The following proposition demonstrates that the socially optimal rule cannot be implemented via backward induction if the jury configuration is not minimally impar-

[2] The feasibility of implementing the socially optimal rule is contingent upon both the feasibility of the implementation mechanism and the feasibility of the employed solution concept. Regarding the latter, we align with Herrero and Srivastava (1992) that backward induction is one of the most compelling solution concepts.

[3] Note that minimal impartiality is trivially satisfied if no agent has enemies.

tial. The reason is that, in that case, there exist two states such that, when moving from one to the other, the individual preferences of each agent remain unchanged even though the deserving winner has changed. Consequently, any backward induction equilibrium in the first state will also be one in the second. Since the deserving winners in the two states are different, implementing the socially optimal rule becomes unattainable.

Proposition 1 *Given a jury configuration e, suppose the socially optimal rule is implementable via backward induction. Then, e is minimally impartial.*

Proof *Claim 1. If e is not minimally impartial, then there are $i, j \in N$ and $R \in R(e)$ such that, for every $k \in N$, $R_k(i) = R_k(j)$.*

If e is not minimally impartial, then there are $i, j \in N$ such that, for every $k \in N\backslash\{i, j\}$, we have $e_k = i$ or $e_k = j$. Abusing notation, for each $k \in N\backslash\{i, j\}$, let $R_k^* \in \Re$ be a preference relation such that:

(1) if $e_k = i$, then $k\ P_k^*\ j\ P_k^*\ l\ P_k^*\ i$ for every $l \in N\backslash\{i, j, k\}$, and
(2) if $e_k = j$, then $k\ P_k^*\ i\ P_k^*\ l\ P_k^*\ j$ for every $l \in N\backslash\{i, j, k\}$.

Similarly, let $R_i^* \in \Re$ be a preference relation such that:

(1) if $e_i = \emptyset$, then $i\ P_i^*\ j\ P_i^*\ l$ for every $l \in N\backslash\{i, j\}$,
(2) if $e_i = j$, then $i\ P_i^*\ l\ P_i^*\ j$ for every $l \in N\backslash\{i, j\}$, and
(3) if $e_i \neq \emptyset$ and $e_i \neq j$, then $i\ P_i^*\ j\ P_i^*\ l\ P_i^*\ e_i$ for every $l \in N\backslash\{i, j, e_i\}$.

Finally, let $R_j^* \in \Re$ be a preference relation such that:

(1) if $e_j = \emptyset$, then $j\ P_j^*\ i\ P_j^*\ l$ for every $l \in N\backslash\{i, j\}$,
(2) if $e_j = i$, then $j\ P_j^*\ l\ P_j^*\ i$ for every $l \in N\backslash\{i, j\}$, and
(3) if $e_j \neq \emptyset$ and $e_j \neq i$, then $j\ P_j^*\ i\ P_j^*\ l\ P_j^*\ e_j$ for every $l \in N\backslash\{i, j, e_j\}$.

By definition of admissible preference function, there exists $R \in \mathcal{R}(e)$ such that (i) $R_i(i) = R_i(j) = R_i^*$, (ii) $R_j(i) = R_j(j) = R_j^*$, and (iii) for each $k \in N\backslash\{i, j\}$, $R_k(i) = R_k(j) = R_k^*$ (see Example 3 below).

Claim 2. If e is not minimally impartial, the socially optimal rule is not implementable via backward induction.

If e is not minimally impartial, by Claim 1, there are $i, j \in N$ such that, for every $k \in N$, $R_k(i) = R_k(j)$, *i.e.*, the preference relation of every agent k at state $(i, R) \in S(e)$ is the same as at state $(j, R) \in S(e)$. Hence, the backward induction equilibria in states (i, R) and (j, R) are identical. If a mechanism implements φ via backward induction, a backward induction equilibrium exists at state (i, R) that results in i. Then, the same equilibrium is a backward induction equilibrium at state (j, R), which contradicts that the mechanism implements φ.[4] □

Example 3 (*Preference functions in Claim 1 of Proposition 1*). Let $N = \{a, b, c, d\}$ and $e = (e_a, e_b, e_c, e_d) = (\emptyset, c, a, b)$. Note that, for every $i \in N\backslash\{a, b\}$, we have

[4] Note that the same argument applies to the implementation of φ in any equilibrium concept that depends only on the ordinal preferences of the agents, such as dominant strategies, Nash equilibrium, subgame perfect equilibrium, etc.

Table 2 Example of admissible preference functions in Example 3

$R_a : N \longrightarrow \Re$				$R_b : N \longrightarrow \Re$				$R_l : N \longrightarrow \Re$				$R_d : N \longrightarrow \Re$			
a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d
a	a	a	a	b	b	b	b	c	c	c	c	d	d	d	d
b	b	c	d	a	a	a	d	b	b	d	d	a	a	c	a
c	c	b	b	d	d	d	a	d	d	b	b	c	c	a	c
d	d	d	c	c	c	c	c	a	a	a	a	b	b	b	b

$e_i = a$ or $e_i = b$ (and therefore, e is not minimally impartial). In this case, as argued in Claim 1 of the proof of Proposition 1, there is $R \in \mathcal{R}(e)$ such that, for every $i \in N$, $R_i(a) = R_i(b)$. Table 2 shows an example of such a profile of admissible preference functions. In general, the only way to ensure that the preference relation of an agent i changes when the deserving winner changes from ω^d to $\hat{\omega}^d$, regardless of his admissible preference function, is that $\omega^d, \hat{\omega}^d \notin \{i, e_i\}$.

Note that, given the characteristics of our model, the condition of minimal impartiality is equivalent to requiring that for each pair of candidates, at least one juror who does not wish to favor or harm either of the candidates exists. In this case, that juror will be "impartial" with respect to both candidates, in the sense that if one of them is the deserving winner, the juror prefers that candidate over the other. Conditions similar to this have frequently appeared in the literature. The first paper to establish a related condition was Amorós (2009, Proposition 1), who demonstrated that, to implement the deserving ranking of candidates in Nash equilibria, it must hold that for each pair of candidates, there exists at least one juror who prefers the two candidates to be ordered according to the deserving ranking.

Minimal impartiality is not only a necessary condition for implementing the socially optimal rule via backward induction but also sufficient. In the following section, we propose a simple and natural mechanism that implements the socially optimal rule via backward induction as long as the jury configuration is minimally impartial.

Remark 1 If at least one agent has an enemy, then the condition of minimal impartiality requires that $n \geq 4$ and is thus an assumption held from the outset of our analysis. On the one hand, if $n = 2$, minimal impartiality is never satisfied. In fact, in this case, the preferences of the two agents will not change with the deserving winner, as either of them will always have themselves as the most preferred alternative and the other agent as the least preferred alternative. Clearly, this makes the implementation of the socially optimal rule impossible. On the other hand, if $n = 3$ and an agent i has an enemy j ($e_i = j$), then the pair of agents j and k (where k is different from i and j) is such that there is no third agent who is not an enemy of either of them, so, by Proposition 1, the implementation of the socially optimal rule via backward induction will also be impossible.

Remark 2 There are situations in which we can ensure minimal impartiality is satisfied without needing additional verification. For example, this occurs if the jury configuration is such that there are at least three agents without enemies. To see this, suppose

that $a, b, c \in N$ are such that $e_a = e_b = e_c = \emptyset$. In this case, for any pair of agents $i, j \in N$, there exists some $k \in \{a, b, c\}$ such that $k \notin \{i, j\}$ and $e_k = \emptyset$, which guarantees that minimal impartiality is satisfied. Similarly, there are other situations in which we know minimal impartiality is not satisfied without needing any verification. For example, this occurs if $n - 2$ agents have the same enemy i. In this case, there is an agent $j \neq i$ such that, for all $k \in N\backslash\{i, j\}$, $e_k = i$, ensuring minimal impartiality.

Remark 3 The most trivial way to satisfy minimal impartiality is that no agent has enemies. Amorós (2011) proposed a natural sequential mechanism for this case that implements the socially optimal rule via backward induction. Adachi (2014) proposed another mechanism that works even if agents have more than one friend they want to favor as long as they have no enemies and certain conditions are met.[5] However, it can be shown that both mechanisms fail when agents have enemies despite the fulfillment of minimal impartiality (see Appendix A).

4 A natural mechanisms that works

In this section, we introduce a mechanism that successfully implements the socially optimal rule via backward induction, whether the agents have enemies or not, as long as minimal impartiality is met.

Before defining the mechanism, we demonstrate that minimal impartiality implies the existence of three agents, i, j, and k, whose enemies (if they exist) are not connected in a particular manner: i is not the enemy of either j or k, j is not the enemy of i, and i and j do not have the same enemy.

Lemma 1 *Suppose that the jury configuration e is minimally impartial. Then:*

(1) There exist three different agents $i, j, k \in N$ such that:

(1.1) $i \notin \{e_j, e_k\}$,

(1.2) $j \neq e_i$, and

(1.3) $e_i \neq e_j$.

(2) For every pair of agents $i, j \in N$, there exists an agent $k_{i,j} \in N\backslash\{i, j\}$ such that, for every $R_{k_{i,j}} \in \mathcal{R}_{k_{i,j}}(e_{k_{i,j}})$, we have $i \; P_{k_{i,j}}(i) \; j$ and $j \; P_{k_{i,j}}(j) \; i$.

Proof To prove point 1, take any agent $i \in N$ such that $e_i \neq \emptyset$. Given agents i and e_i, according to minimal impartiality, there is some $k_{i,e_i} \in N\backslash\{i, e_i\}$ such that $e_{k_{i,e_i}} \notin \{i, e_i\}$. Denoting $j \equiv k_{i,e_i}$, we have $j \neq i$, $j \neq e_i$, $e_j \neq i$, and $e_j \neq e_i$. Additionally, given agents i and $j \equiv k_{i,e_i}$ above, by minimal impartiality, there is some $k_{i,j} \in N\backslash\{i, j\}$ such that $e_{k_{i,j}} \notin \{i, j\}$. Denoting $k \equiv k_{i,j}$, we have $j \neq k \neq i$ and $e_k \neq i$. The proof of point 2 follows directly from the definitions of minimal impartiality and admissible preference function. □

Example 4 (*Agents in* Lemma 1). Let $N = \{a, b, c, d\}$. Let $e = (e_a, e_b, e_c, e_d) = (b, c, a, \emptyset)$. As shown in Example 2, e satisfies minimal impartiality. Next, we propose some agents fulfilling points 1 and 2 of Lemma 1. Point 1 of Lemma 1 is satisfied

[5] The mechanism proposed by Adachi (2014) is intended to choose a ranking of candidates (not necessarily complete) when the agents in charge of choosing it may have several friends.

for agents $i \equiv b$, $j \equiv d$, and $k \equiv c$. To see this, note that $b \notin \{e_d, e_c\}$, $d \neq e_b$, and $e_b \neq e_d$.[6] Point 2 of Lemma 1 is satisfied for agents $k_{a,b} \equiv d$, $k_{a,c} \equiv d$, $k_{a,d} \equiv b$, $k_{b,c} \equiv d$, $k_{b,d} \equiv c$, and $k_{c,d} \equiv a$. To see this note that: for every $R_d \in R_d(e_d)$, $a\ P_d(a)\ b$ and $b\ P_d(b)\ a$; for every $R_d \in R_d(e_d)$, $a\ P_d(a)\ c$ and $c\ P_d(c)\ a$; for every $R_b \in R_b(e_b)$, $a\ P_b(a)\ d$ and $d\ P_b(d)\ a$; for every $R_d \in R_d(e_d)$, $b\ P_d(b)\ c$ and $c\ P_d(c)\ b$; for every $R_c \in R_c(e_c)$, $b\ P_c(b)\ d$ and $d\ P_c(d)\ b$; for every $R_a \in R_a(e_a)$, $c\ P_a(b)\ d$ and $d\ P_a(d)\ c$.

MECHANISM 1 Suppose that the jury configuration e is minimally impartial. To define the mechanism, we first need to establish two elements: (1) a numbering of the agents that will determine their order of intervention and that must satisfy certain conditions about their enemies, and (2) for each pair of agents, an agent who is "impartial" with respect to them in the sense that, if one of them is the deserving winner, the agent prefers him to the other.

(1) Let us number the agents, $N = \{1, ..., n\}$, in such a way that:
  (1.1) $n - 1 \notin \{e_{n-2}, e_{n-3}\}$,
  (1.2) $n - 2 \neq e_{n-1}$, and
  (1.3) $e_{n-1} \neq e_{n-2}$.

Because e is minimally impartial, point 1 of Lemma 1 ensures that such numbering of agents is possible. In particular, given agents i, j, and k in point 1 of Lemma 1, we can take $n - 1 = i$, $n - 2 = j$, and $n - 3 = k$.

(2) For each pair of agents $i, j \in N$, let $k_{i,j} \in N \backslash \{i, j\}$ be an agent such that, for every $R_{k_{i,j}} \in \mathcal{R}_{k_{i,j}}(e_{k_{i,j}})$, $i\ P_{k_{i,j}}(i)\ j$ and $j\ P_{k_{i,j}}(j)\ i$. Agent k_{ij} is "impartial" between agents i and j in that if one of them is the deserving winner, k_{ij} prefers that agent to the other. Since e is minimally impartial, the existence of these agents is guaranteed by point 2 of Lemma 1.

Given a numbering and a notation of the agents as defined in points 1 and 2 above, Mechanism 1 is defined as follows. Agents take turns announcing who they think should win. The announcement of agent 1 is implemented if either (1) he has no enemy and does not propose himself as the winner, or (2) he has an enemy he proposes as the winner. If agent 1 has an enemy and proposes an agent x who is neither himself nor his enemy as the winner, then we move to another stage in which k_{x,e_1} (the "impartial" agent between x and e_1) chooses the winner between those two candidates. If agent 1 proposes himself as the winner, then it is the turn of agent 2, and the process repeats. If it is the turn of agent n because jurors 1, 2, ..., and $n - 1$ have announced themselves, then the agent announced by agent n is chosen as the winner, even if that agent is himself. Formally:

Stage 1: Agent 1 *announces* $m_1 \in N$.

- If $e_1 = \emptyset$:
  - If $m_1 = 1$, then go to Stage 2.
  - If $m_1 \neq 1$, then m_1 is chosen as winner. STOP.
- If $e_1 \neq \emptyset$:

[6] This is not the only choice of agents i, j, and k that satisfy point 1 of Lemma 1. For example, $i = a$, $j = d$, and $k = b$ also meet the requirements.

- If $m_1 = 1$, then go to Stage 2.
- If $m_1 = e_1$, then m_1 is chosen as winner. STOP.
- If $m_1 \notin \{1, e_1\}$, then go to Stage $1.m_1$.

Stage 1.m_1: Agent k_{m_1,e_1} announces $m_{k_{m_1,e_1}} \in \{m_1, e_1\}$.

- Then $m_{k_{m_1,e_1}}$ is chosen as winner. STOP.

Stage 2: Agent 2 announces $m_2 \in N$.

- If $e_2 = \emptyset$:
  - If $m_2 = 2$, then go to Stage 3.
  - If $m_2 \neq 2$, then m_2 is chosen as winner. STOP.
- If $e_2 \neq \emptyset$:
  - If $m_2 = 2$, then go to Stage 3.
  - If $m_2 = e_2$, then m_2 is chosen as winner. STOP.
  - If $m_2 \notin \{2, e_2\}$, then go to Stage $2.m_2$.

Stage 2.m_2: Agent k_{m_2,e_2} announces $m_{k_{m_2,e_2}} \in \{m_2, e_2\}$.

- Then $m_{k_{m_2,e_2}}$ is chosen as winner. STOP.

$\vdots$

Stage $n-1$: Agent $n-1$ announces $m_{n-1} \in N$.

- If $e_{n-1} = \emptyset$:
  - If $m_{n-1} = n-1$, then go to Stage n.
  - If $m_{n-1} \neq n-1$, then m_{n-1} is chosen as winner. STOP.
- If $e_{n-1} \neq \emptyset$:
  - If $m_{n-1} = n-1$, then go to Stage n.
  - If $m_{n-1} = e_{n-1}$, then m_{n-1} is chosen as winner. STOP.
  - If $m_{n-1} \notin \{n-1, e_{n-1}\}$, then go to Stage $n-1.m_1$.

Stage $(n-1).m_1$: Agent $k_{m_{n-1},e_{n-1}}$ announces $m_{k_{m_{n-1},e_{n-1}}} \in \{m_{n-1}, e_{n-1}\}$.

- Then $m_{k_{m_{n-1},e_{n-1}}}$ is chosen as winner. STOP.

Stage n: Juror n announces $m_n \in N$.

- Then m_n is chosen as winner. STOP.

Example 5 (*Mechanism* 1). Let $N = \{a, b, c, d\}$. Let $e = (e_a, e_b, e_c, e_d) = (b, c, a, \emptyset)$. As shown in Example 2, the jury configuration e is minimally impartial. To build Mechanism 1, we proceed as follows:

(1) We must number the agents to determine their order of intervention. In Example 4, we have shown that point 1 of Lemma 1 is satisfied for agents $i \equiv b$, $j \equiv d$, and $k \equiv c$. Then, according to the definition of Mechanism 1, agent $n-1$ is b, agent $n-2$ is d, and agent $n-3$ is c. Hence, since $n = 4$, we have that agent 1 is c, agent 2 is d, agent 3 is b, and agent 4 is a.[7]

[7] Given the numbering $N = \{1, 2, 3, 4\}$, the jury configuration in Example 5 is $e = (e_1, e_2, e_3, e_4) = (4, \emptyset, 1, 3)$.

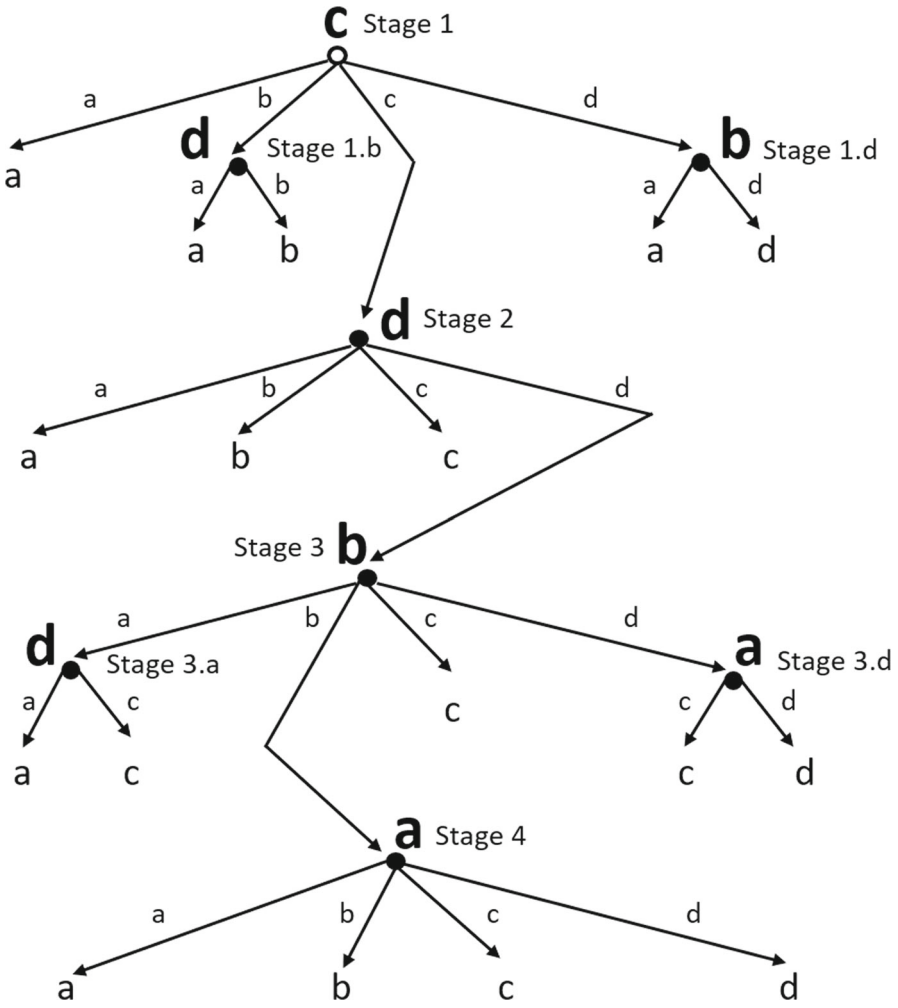

Fig. 1 Mechanism 1 in Example 5

(2) For each pair of agents, i and j, we have to find an agent $k_{i,j} \in N \backslash \{i, j\}$ who is "impartial" between them. In Example 4, we have shown that point 2 of Lemma 1 is satisfied for agents $k_{a,b} \equiv d$, $k_{a,c} \equiv d$, $k_{a,d} \equiv b$, $k_{b,c} \equiv d$, $k_{b,d} \equiv c$, and $k_{c,d} \equiv a$. Then, we take agent d as $k_{a,b}$, $k_{a,c}$, and $k_{b,c}$, agent b is $k_{a,d}$, agent c is $k_{b,d}$, and agent a is $k_{c,d}$.

Then, in this example, Mechanism 1 is as represented in Figure 1. Given the jury configuration e, Table 3 shows the conditions that must satisfy the admissible preference functions of each agent i, $\mathcal{R}_i(e)$.

Next, we explain why the mechanism represented in Figure 1 works. Given any state, $(\omega^d, R) \in S(e)$, consider a profile of strategies that is a backward induction equilibrium of the previous mechanism at (ω^d, R).

Table 3 Admissible preference functions in Example 5

$R_a : N \longrightarrow \Re$				$R_b : N \longrightarrow \Re$				$R_c : N \longrightarrow \Re$				$R_d : N \longrightarrow \Re$			
a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d
a	a	a	a	b	b	b	b	c	c	c	c	d	d	d	d
$c?d$	$c?d$	c	d	a	$a?d$	$a?d$	d	$b?d$	b	$b?d$	d	a	b	c	$a?b?c$
b	b	d	c	d	c	c	a	a	d	a	b	$b?c$	$a?c$	$a?b$	
		b	b	c			c		a		a				

(i) In the intermediate stages, where an agent $k_{i,j}$ has to choose between two other agents i and j (Stages 1.b, 1.d, 3.a, and 3.b), the following will happen: (1) if one of those two agents is the deserving winner, he will be chosen (since $k_{i,j}$ is "impartial" between the two), and (2) if neither of the two agents is the deserving winner, then either of them could be chosen, depending on $k_{i,j}$'s specific preferences in that case.

(ii) If the mechanism reaches Stage 4, agent a will choose himself.

(iii) If the mechanism reaches Stage 3, given point (ii) above, there is nothing agent b can do to be chosen. If $\omega^d = a$, b will ensure that his second-best alternative, a, is chosen (he can do so either by announcing a or b). If $\omega^d = b$, the second-best alternative for b can be either a or d, and it is unclear whether he can guarantee this agent is chosen (it will also depend on a's preferences between c and d). However, b can prevent his enemy c from being chosen (he can do that by announcing b, in which case a will ultimately be chosen). Therefore, in this case. either a or d will ultimately chosen. If $\omega^d = c$, by points (i) and (ii), the only way b can prevent his enemy c from being chosen is by announcing himself, and therefore a will ultimately be chosen. Finally, if $\omega^d = d$, b's second-best alternative is d, and then he will ensure d is chosen by announcing d.

(iv) Based on the above, if the mechanism reaches Stage 2, the only case where d can take action to ensure he is chosen is when $\omega^d = d$ (in this case, d will announce himself and will ultimately be chosen in Stage 3.d). If $\omega^d = a$, d will ensure that his second-best alternative, a, is chosen (he can do so by announcing a). Similarly, if $\omega^d = b$, d will ensure that his second-best alternative, b, is chosen, and if $\omega^d = c$, d will ensure that his second-best alternative, c , is chosen (he can do so by announcing b and c, respectively).

(v) Based on the above points, the only case where c can act in Stage 1 to ensure he is chosen is when $\omega^d = c$ (in this case, c will announce himself and will ultimately be chosen in Stage 2). If $\omega^d = a$, by points (i) and (iv), c cannot prevent a from being chosen, regardless of his announcement. If $\omega^d = b$, c's second-best alternative is b, and he can ensure b is chosen (either in Stage 1.b by announcing b or in Stage 2 by announcing c. Finally, if $\omega^d = d$, c's second-best alternative is d, and he can ensure d is chosen (either in Stage 1.d by announcing d or in Stage 3.d by announcing c).

Thus, any backward induction equilibrium of the mechanism always results in ω^d.

The following theorem demonstrates that the observation in Example 5 holds true in general. Whenever the necessary condition of minimal impartiality is satisfied, Mechanism 1 implements the socially optimal rule via backward induction.

Theorem 1 *Suppose that the jury configuration e is minimally impartial. Then, Mechanism 1 implements φ via backward induction.*

Proof Let $N = \{1, ..., n\}$ be the numbering of the agents as defined in Mechanism 1. Furthermore, for each pair $i, j \in N$, let $k_{i,j} \in N\backslash\{i, j\}$ be as defined in Mechanism 1. Given any state, $(\omega^d, R) \in S(e)$, consider a profile of strategies that is a backward induction equilibrium of Mechanism 1 at (ω^d, R). For each $k \in N$, let x_k denote the agent who would be ultimately chosen by Mechanism 1 in case Stage k is reached, given the previous profile of equilibrium strategies. Similarly, for each $k \in N\backslash\{n\}$ such that $e_k \neq \emptyset$ and each $m_k \in N\backslash\{k, e_k\}$, let $x_{k.m_k}$ denote the agent who would be ultimately chosen by Mechanism 1 in case Stage $k.m_k$ is reached, given the previous profile of equilibrium strategies. The proof proceeds by stating and proving a series of eight claims.

Claim 1. Let $k \in N\backslash\{n\}$ be such that $e_k \neq \emptyset$ and let $m_k \in N\backslash\{k, e_k\}$.
(1.1) If $\omega^d = e_k$, then $x_{k.m_k} = e_k$.
(1.2) If $\omega^d = m_k$, then $x_{k.m_k} = m_k$.
(1.3) If $\omega^d \notin \{e_k, m_k\}$, then $x_{k.m_k} \in \{e_k, m_k\}$.

If Stage $k.m_k$ is reached, it is because $e_k \neq \emptyset$ and the announcement of agent k at Stage k was such that $m_k \notin \{k, e_k\}$. In this case, agent k_{m_k,e_k} chooses between m_k and e_k, the choice of agent k_{m_k,e_k} will be selected as the winner, and the mechanism will stop. By definition, agent k_{m_k,e_k} is such that $m_k \; P_{k_{m_k,e_k}}(m_k) \; e_k$ and $e_k \; P_{k_{m_k,e_k}}(e_k) \; m_k$. Therefore, if $\omega^d = e_k$, agent k_{m_k,e_k} will announce e_k, and if $\omega^d = m_k$, he will announce m_k. If $\omega^d \notin \{e_k, m_k\}$, it can happen either $m_k \; R_{k_{m_k,e_k}}(\omega^d) \; e_k$ or $e_k \; P_{k_{m_k,e_k}}(\omega^d) \; m_k$, so agent k_{m_k,e_k} can announce either m_k or e_k.

Claim 2. Let $k \in N\backslash\{n\}$.
(2.1) If $\omega^d = e_k$, we have $x_k = x_{k+1}$.
(2.2) If $\omega^d = k$, then:
(2.2.1) if $x_{k+1} \neq k$, we have $x_k \neq k$, and
(2.2.2) if $x_{k+1} = k$, we have $x_k = k$.
(2.3) If $\omega^d \notin \{e_k, k\}$, then:
(2.3.1) if $x_{k+1} \neq k$, we have $x_k = \omega^d$, and
(2.3.2) if $x_{k+1} = k$, we have $x_k = k$.

Suppose that $\omega^d = e_k$. Then, by definition of Mechanism 1 and by Claim 1, at Stage k, agent k has to decide between selecting e_k at that stage or selecting x_{k+1} at some later stage: if $m_k = e_k$, then e_k is chosen at Stage k; if $m_k \notin \{k, e_k\}$, then e_k is chosen at Stake $k.m_k$; if $m_k = k$, then x_{k+1} is ultimately chosen at some later stage. Because e_k is the worst alternative for agent k, he will announce $m_k = k$, and agent x_{k+1} will be ultimately chosen (if $x_{k+1} = e_k$, he is indifferent between all announcements).

Suppose that $\omega^d = k$ and $x_{k+1} \neq k$. In this case, agent k cannot do anything to be selected neither at Stage k nor at any subsequent stage: if his announcement is $m_k = e_k$, then e_k is chosen at Stage k; if his announcement is $m_k \notin \{e_k, k\}$, then the mechanism advances to Stage $k.m_k$, in which either m_k or e_k will be chosen

(depending on the preferences of agent k_{m_k,e_k}); if his announcement is $m_k = k$, then the mechanism advances to Stage $k+1$, and agent $x_{k+1} \neq k$ is ultimately selected. Among the previous agents that k can make to be chosen once Stage k is reached (all different from k), x_k will be the one he prefers the most.

Suppose that $\omega^d = k$ and $x_{k+1} = k$. Then, because k is the best alternative for agent k, he will announce $m_k = k$ and agent $x_{k+1} = k$ will be chosen at some subsequent stage.

Suppose that $\omega^d \notin \{e_k, k\}$ and $x_{k+1} \neq k$. In this case, using the same argument as when $\omega^d = k$ and $x_{k+1} \neq k$, we conclude that agent k cannot do anything to be selected neither at Stage k nor at any subsequent stage. However, by Claim 1, now agent k can ensure that ω^d is chosen at Stage k by announcing $m_k = \omega^d$. Since $\omega^d \notin \{e_k, k\}$, then ω^d is the second most preferred alternative for agent k. Therefore, agent k will announce $m_k = \omega^d$, and ω^d will be selected at Stage k.

Suppose that $\omega^d \notin \{e_k, k\}$ and $x_{k+1} = k$. Because k is the best alternative for agent k, he will announce $m_k = k$, and agent $x_{k+1} = k$ will be ultimately chosen at some subsequent stage.

Claim 3. Suppose that Stage n is reached. *Then,* $x_n = n$.

It follows from the fact that n is the most preferred alternative for agent n.

Claim 4. Suppose that Stage $n-1$ *is reached.*

(4.1) If $\omega^d = e_{n-1}$*, then* $x_{n-1} = n$.

(4.2) If $\omega^d = n-1$*, then* $x_{n-1} \neq n-1$.

(4.3) If $\omega^d \notin \{e_{n-1}, n-1\}$*, then* $x_{n-1} = \omega^d$.

It follows from Claims 2 and 3 and the fact that $n-1 \neq n$.

Claim 5. Suppose that Stage $n-2$ *is reached.*

(5.1) If $\omega^d = n-1$ *and* $x_{n-1} = n-2$*, then* $x_{n-2} = n-2$.

(5.2) Otherwise, $x_{n-2} = \omega^d$.

Suppose that $\omega^d = e_{n-2}$. Since Mechanism 1 is such that $n-1 \neq e_{n-2}$ and $e_{n-1} \neq e_{n-2}$, then $\omega^d \notin \{e_{n-1}, n-1\}$. Therefore, by point 4.3 of Claim 4, $x_{n-1} = \omega^d$. Then, because $\omega^d = e_{n-2}$, by point 2.1 of Claim 2, we have $x_{n-2} = x_{n-1} = \omega^d$.

Suppose that $\omega^d = n-2$. Since Mechanism 1 is such that $n-2 \neq e_{n-1}$ and $n-2 \neq n-1$, then $\omega^d \notin \{e_{n-1}, n-1\}$. Therefore, by point 4.3 of Claim 4, $x_{n-1} = \omega^d$. Then, because $\omega^d = n-2$ and $x_{n-1} = \omega^d = n-2$, by point 2.2.2 of Claim 2, we have $x_{n-2} = n-2 = \omega^d$.

Suppose that $\omega^d \notin \{e_{n-2}, n-2\}$. Then we have three possibilities:

(i) If $\omega^d = e_{n-1}$, by point 4.1 of Claim 4, $x_{n-1} = n$. Then, because $\omega^d \notin \{e_{n-2}, n-2\}$ and $x_{n-1} = n \neq n-2$, by point 2.3.1 of Claim 2, we have $x_{n-2} = \omega^d$.

(ii) If $\omega^d = n-1$, by point 4.2 of Claim 4, $x_{n-1} \neq n-1$. If, in addition, $x_{n-1} \neq n-2$, by point 2.3.1 of Claim 2, we have $x_{n-2} = \omega^d$. If, on the contrary, $x_{n-1} = n-2$, by point 2.3.2 of Claim 2, we have $x_{n-2} = n-2$.
(iii) If $\omega^d \notin \{e_{n-1}, n-1\}$, by point 4.3 of Claim 4, $x_{n-1} = \omega^d$. Then, because $\omega^d \notin \{e_{n-2}, n-2\}$ and $x_{n-1} = \omega^d \neq n-2$, by point 2.3.1 of Claim 2, we have $x_{n-2} = \omega^d$.

Claim 6. Suppose that Stage $n-3$ is reached. Then, $x_{n-3} = \omega^d$.

Suppose that $\omega^d = e_{n-3}$. By definition of Mechanism 1, we have $e_{n-3} \neq n-1$, and then $\omega^d \neq n-1$. Therefore, by point 5.2 of Claim 5, $x_{n-2} = \omega^d$. Hence, by point 2.1 of Claim 2, $x_{n-3} = x_{n-2} = \omega^d$.

Suppose that $\omega^d = n-3$. By definition of Mechanism 1, we have $n-1 \neq n-3$, and then $\omega^d \neq n-1$. Therefore, by point 5.2 of Claim 5, $x_{n-2} = \omega^d$. Hence, by point 2.2.2 of Claim 2, $x_{n-3} = x_{n-2} = \omega^d$.

Suppose that $\omega^d \notin \{e_{n-3}, n-3\}$. Then we have two possibilities:

(i) Suppose that $\omega^d = n-1$. If, in addition, $x_{n-1} = n-2$, by point 5.1 of Claim 5, we have $x_{n-2} = n-2$. Then, since $n-2 \neq n-3$, by point 2.3.1 of Claim 2, $x_{n-3} = \omega^d$. If, on the contrary, $x_{n-1} \neq n-2$, by point 5.2 of Claim 5, we have $x_{n-2} = \omega^d = n-1$. Since $n-1 \neq n-3$, again by point 2.3.1 of Claim 2, $x_{n-3} = \omega^d$.
(ii) Suppose that $\omega^d \neq n-1$. Then, by point 5.2 of Claim 5, $x_{n-2} = \omega^d$. Because $\omega^d \neq n-3$, then $x_{n-2} \neq n-3$. Hence, by point 2.3.1 of Claim 2, $x_{n-3} = \omega^d$.

Claim 7. Suppose that $n \geq 5$. Suppose that Stage $n-4$ is reached. Then, $x_{n-4} = \omega^d$.

Suppose that $\omega^d = e_{n-4}$. Then, by point 2.1 of Claim 2 and by Claim 6, $x_{n-4} = x_{n-3} = \omega^d$.

Suppose that $\omega^d = n-4$. By Claim 6, $x_{n-3} = \omega^d$. Then, since $\omega^d = n-4$ and $x_{n-3} = n-4$, by point 2.2.2 of Claim 2, $x_{n-4} = n-4 = \omega^d$.

Suppose that $\omega^d \notin \{e_{n-4}, n-4\}$. By Claim 6, $x_{n-3} = \omega^d$. Then, since $\omega^d \neq n-4$ and $x_{n-3} \neq n-4$, by point 2.3.1 of Claim 2, $x_{n-4} = \omega^d$.

Claim 8. Suppose that $n \geq 6$. Suppose that Stage $n-t$ is reached for some $t \geq 5$. Then, $x_{n-t} = \omega^d$.

The demonstration of this claim follows from repeatedly applying the same argument used in the proof of Claim 7.

From Claims 6, 7, and 8, we have that, given any state $(\omega^d, R) \in S(e)$, every profile of backward induction equilibrium strategies of Mechanism 1 at (ω^d, R) is such that ω^d is ultimately chosen as the winner. □

The following example illustrates the importance of the numbering of the agents in Mechanism 1 satisfying the requirements listed in points 1 and 2 of its definition.

Example 6 (*The order of the agents in Mechanism 1 matters*). Let $N = \{a, b, c, d\}$. Let $e = (e_a, e_b, e_c, e_d) = (b, c, a, \emptyset)$. As shown in Example 2, the jury configuration e is minimally impartial. Then, by Lemma 1 and Theorem 1, we can find an order of intervention of the agents such that Mechanism 1 implements the socially optimal rule

via backward induction.[8] However, if the order of intervention of the agents does not satisfy those requirements, the mechanism fails. For example, consider the order where $1 \equiv c$, $2 \equiv d$, $3 \equiv a$, and $4 \equiv b$ (see Figure 2).[9] In this case, $n-1 = a = e_c = e_{n-3}$, and then point 1.1 in the definition of Mechanism 1 is not satisfied. As we show next, this failure to meet the required numbering specifications for the agents causes the mechanism not to work.

Let $(\omega^d, R) \in S(e)$ be a state such that $\omega^d = a$, $b\ P_d(a)\ c$, and $d\ P_c(a)\ b$ (there are admissible preference functions for agents d and c that satisfy these conditions). Then, any profile of strategies that is a backward induction equilibrium of the mechanism depicted in Figure 1 at (ω^d, R) is such that:

1. Since b is the most preferred alternative for agent b, $x_4 = b$.
2. Since $b\ P_d(a)\ c$, $x_{3.c} = b$.
3. Since $d\ P_c(a)\ b$, $x_{3.d} = d$.
4. By points 1, 2, and 3, and since b is the worst alternative for agent a, he will announce d at Stage 3, and then $x_3 = d$.
5. By point 4, since d is the most preferred alternative for agent d, he will announce d at Stage 2, and then $x_2 = d$.
6. Since a is preferred to b for agent d, $x_{1.b} = a$.
7. Since a is preferred to d for agent b, $x_{1.d} = a$.
8. By points 5, 6, and 7, since a is the worst alternative for agent c, he will announce c at Stage 1, and then $x_1 = d$.

Therefore, the outcome reached via backward induction is $d \neq \omega^d$, and so the mechanism in Figure 1 does not implement the socially optimal rule via backward induction.

One can find similar examples to demonstrate that if the numbering of the agents does not satisfy any of the other requirements listed in the definition of Mechanism 1 ($n-1 \neq e_{n-2}, n-2 \neq e_{n-1}$, and $e_{n-1} \neq e_{n-2}$), then Mechanism 1 fails to implement the socially optimal rule.

Remark 4 In a recent contribution to full subgame perfect implementation, Mackenzie and Zhou (2022) consider menu mechanisms where at each history, an agent selects from a subset of his consumption space called a menu. Interestingly, Mechanism 1 fits this description. That said, Mackenzie and Zhou require several conditions that are violated here. For example, the rule should be strategy-proof, any strict ranking of the consumption space should be possible, and the set of possible type profiles should have a Cartesian product structure.

Remark 5 Moore and Repullo (1988) characterize the choice rules implementable in subgame perfect equilibria. Similarly, Herrero and Srivastava (1992) characterize the set of choice functions that are implementable via backward induction. Unfortunately,

[8] In particular, the numbering of the agents where $1 \equiv c$, $2 \equiv d$, $3 \equiv b$, and $4 \equiv a$, satisfies requirement (1) of the definition of Mechanism 1. Additionally, the agents $k_{a,b} \equiv d$, $k_{a,c} \equiv d$, $k_{a,d} \equiv b$, $k_{b,c} \equiv d$, $k_{b,d} \equiv c$, and $k_{c,d} \equiv a$ satisfy requirement (2) of that definition.

[9] In this case, the agents $k_{1,2} \equiv 3$, $k_{1,3} \equiv 2$, $k_{1,4} \equiv 2$, $k_{2,3} \equiv 4$, $k_{2,4} \equiv 1$, and $k_{3,4} \equiv 2$ satisfy requirement (2) of the definition of Mechanism 1.

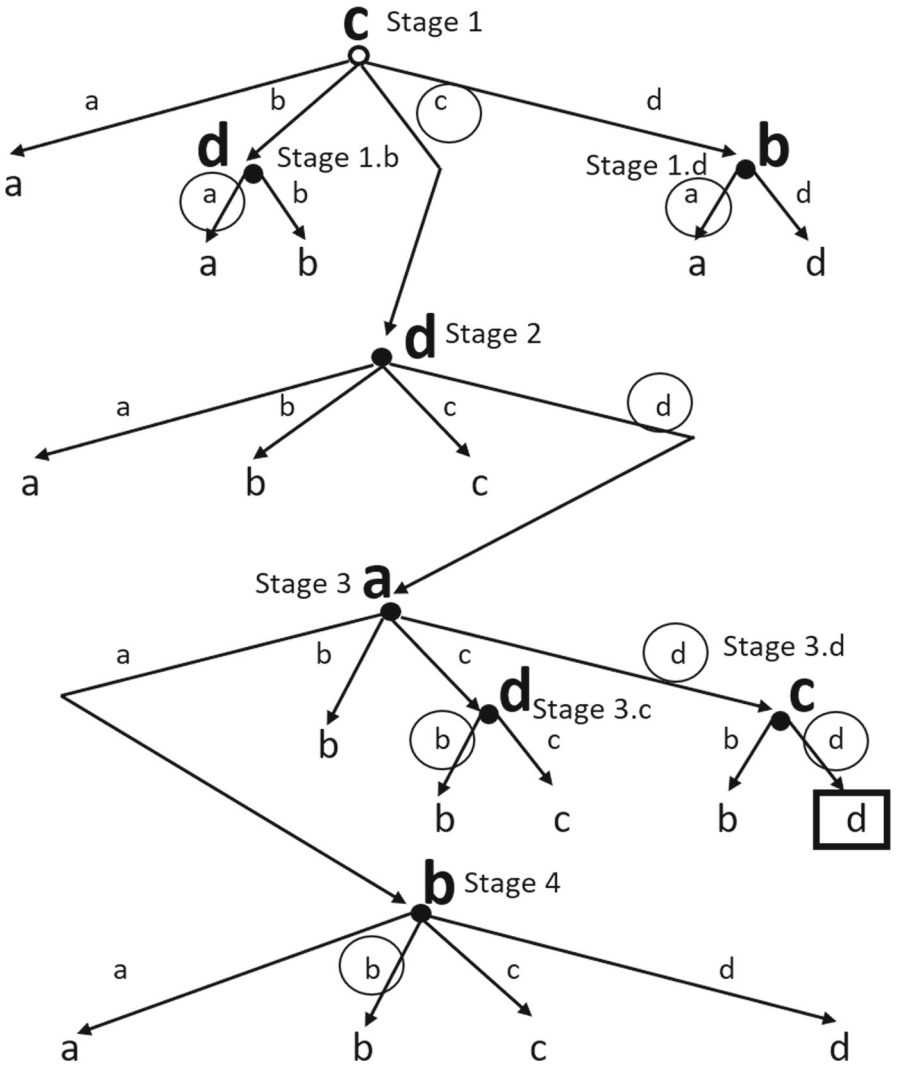

Fig. 2 Mechanism in Example 6

these characterizations are cumbersome, and the sufficient conditions are virtually impossible to check.

5 Concluding comments

We have studied the problem of selecting the deserving winner from a group of candidates when each juror identifies with a different friend they wish to favor, and some jurors may have enemies they intend to harm. We have identified a necessary and

sufficient condition on the jurors' enemies to implement the deserving winner by backward induction. The mechanisms suggested in the literature to implement the deserving winner via backward induction when jurors lack enemies fail when jurors may indeed have enemies, even when the previous condition is satisfied. To address this problem, we have proposed a simple mechanism that successfully implements the deserving winner via backward induction in such cases.

Here are some suggestions for potential lines of extensions.

a. The feasibility of implementing the deserving winner is contingent upon both the feasibility of the implementation mechanism and the feasibility of the solution concept utilized. Regarding the latter, we concur with Herrero and Srivastava (1992) that, besides dominant strategies, backward induction is among the most attractive solution concepts. Using an argument similar to that of Amorós et al. (2002, Theorem 1), it can be demonstrated that if there are only three agents, the deserving winner is not implementable in dominant strategies even if minimal impartiality is satisfied. However, the question of whether this impossibility extends to cases where there are more than three agents is still open.

b. Exploring a more general model than the one examined in this work could involve eliminating the identification between candidates and jurors. This modification would allow each juror to have multiple friends they wish to favor and enemies they aim to harm. In Appendix B, we provide a preliminary approach to this problem. However, extending our Mechanism 1 to this setting remains an open question.

c. Continuing with the previous extension, it would be interesting to study the case where the mechanism designer does not know each juror's friends and enemies. In this case, we should investigate whether it is possible to design a mechanism in which the jurors have incentives to reveal such information.

d. There are situations in which being selected is a bad thing, so (i) the worst alternative for each agent is for himself to be chosen, and (ii) having someone as an enemy means that I want him to always be selected, even when he does not "deserve" it. For example, this could occur when identifying actual criminals from suspects. In Appendix C, we analyze this extension where the roles of the agents and their enemies are reversed.

Funding Funding for open access publishing: Universidad de Málaga/CBUA

Table 4 $R_i(a)$in Example 7

$R_a(a)$	$R_b(a)$	$R_c(a)$	$R_d(a)$
a	b	c	d
c	a	b	a
b	d	d	b
d	c	a	c

Appendix A

Natural mechanisms that do not work

In this Appendix, we analyze two natural mechanisms proposed in the literature that implement the socially optimal rule via backward induction when agents have no enemies: the mechanisms suggested by Amorós (2011) and Adachi (2014). We propose examples showing that these mechanisms may fail when agents have enemies, even if the jury configuration is minimally impartial.

Amorós' (2011) **mechanism**. Agents are numbered to announce a winner sequentially. If agent 1 declares a winner other than himself, his announcement is enacted. Otherwise, the turn passes to agent 2 , and the process is repeated. If agent n's turn arrives, his announcement is selected, even if he announced that he is the winner. This mechanism works if the agents have no enemies but may fail when they do.

Example 7 (*Amorós' mechanism does not work*). Suppose that $N = \{a, b, c, d\}$ and $e = (e_a, e_b, e_c, e_d) = (d, \emptyset, a, c)$. The jury configuration e is minimally impartial. To see this, note that: (1) $d \in N\backslash\{a, b\}$ and $e_d \notin \{a, b\}$, (2) $b \in N\backslash\{a, c\}$ and $e_b \notin \{a, c\}$, (3) $b \in N\backslash\{a, d\}$ and $e_b \notin \{a, d\}$, (4) $a \in N\backslash\{b, c\}$ and $e_a \notin \{b, c\}$, (5) $c \in N\backslash\{b, d\}$ and $e_c \notin \{b, d\}$, and (6) $b \in N\backslash\{c, d\}$ and $e_b \notin \{c, d\}$. Let us number the agents so that $1 = a$, $2 = b$, $3 = c$, and $4 = d$. Let $(\omega^d, R) \in S(e)$ be such that $\omega^d = a$ and, for each $i \in N$, $R_i(a)$ is as represented in Table 4 (note that, given e, these preference relations can be derived from some admissible preference functions).

We demonstrate below that Amorós' (2011) mechanism has a backward induction equilibrium at state (ω^d, R) that results in c, and then, because $\omega^d = a$, the mechanism fails to implement the socially optimal rule. More precisely, the only profile of strategies that is a backward induction equilibrium of the mechanism at (ω^d, R) is like that represented by the solid circles and squares in Figure 3 (the solid circles represent the agents' actions, and the solid square represents the outcome of this equilibrium). Note that:

1. Since d is the most preferred alternative for agent d, if the mechanism reaches the last stage, agent d will be selected.
2. Since agent c prefers b to d and a, by point 1, if the mechanism reaches the third stage, agent b will be ultimately selected.
3. Since b is the most preferred alternative for agent b, by point 2, if the mechanism reaches the second stage, agent b will be ultimately selected.

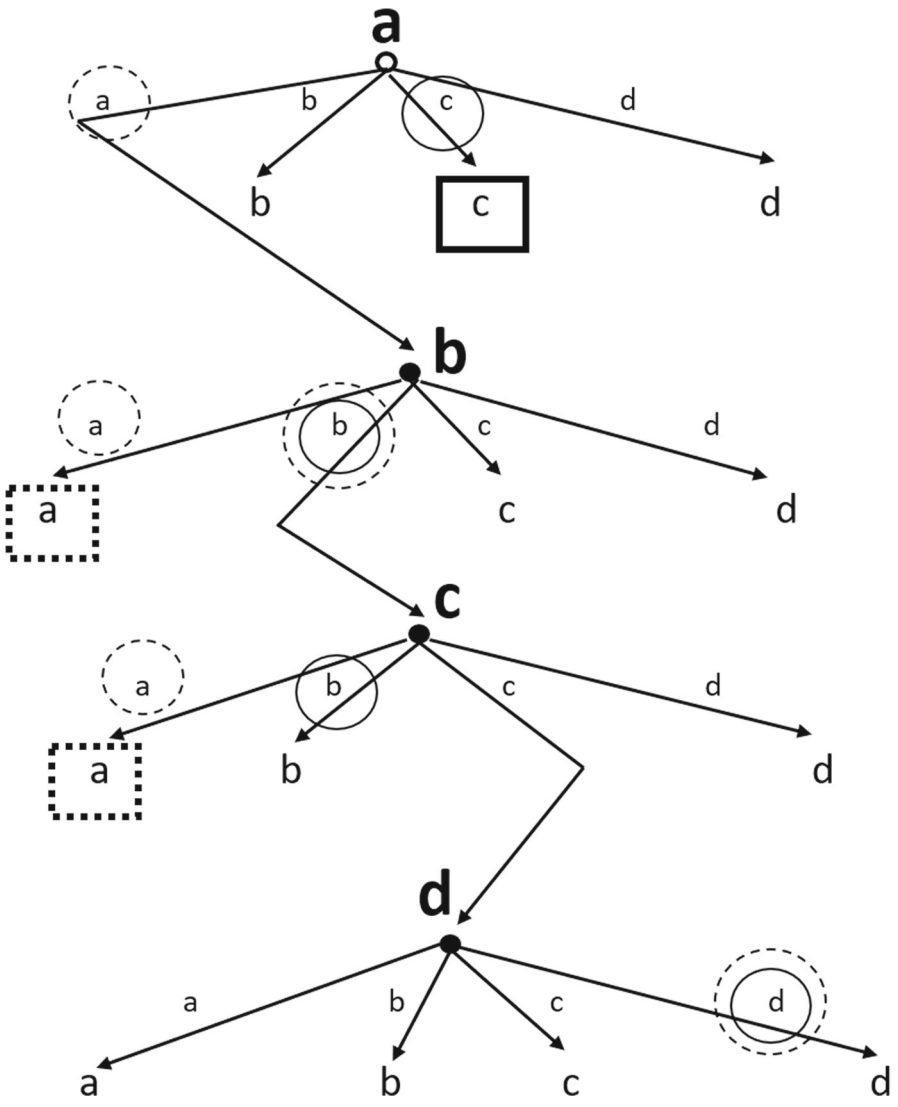

Fig. 3 Amorós' (2011) mechanism in Example 7

4. Since agent a prefers c to b and d, by point 3, agent a will announce c at the first stage, and then c will be selected.

Note that if no agent had enemies, Amorós' (2011) mechanism would not fail. Specifically, if the jury configuration were $\hat{e} = (\hat{e}_a, \hat{e}_b, \hat{e}_c, \hat{e}_d) = (\emptyset, \emptyset, \emptyset, \emptyset)$, then every $\hat{R} \in \mathcal{R}(\hat{e})$ would be such that, for each $i \in N$, $\hat{R}_i(a)$ is as represented in Table 5. In particular, the preference relation $R_c(a)$ in Table 4 is no longer possible. In this case, the backward induction equilibria of the mechanism in Figure 3 at any state $(\omega^d, \hat{R}) \in S(\hat{e})$ such that $\omega^d = a$ are represented by the dashed circles and

Table 5 $\hat{R}_i(a)$ in Example 7

$\hat{R}_a(a)$	$\hat{R}_b(a)$	$\hat{R}_c(a)$	$\hat{R}_d(a)$
a	b	c	d
$c?b?d$	a	a	a
	$c?d$	$b?d$	$b?c$

Table 6 $R_i(c)$ in Example 8

$R_a(c)$	$R_b(c)$	$R_c(c)$	$R_d(c)$
a	b	c	d
c	c	b	b
b	a	d	a
d	d	a	c

squares (dashed circles represent the agents' actions, and dashed squares represent the equilibria outcomes). Note that any of these equilibria results in $a = \omega^d$.

Adachi's (2014) **mechanism**. Agents are numbered from 1 to n. In the first stage, agents who are not friends of agent 1 sequentially vote whether they think 1 should be chosen as the winner. If any of them vote yes, then agent 1 is selected. If everyone votes no, then we move on to the second stage, where the process is repeated with agent 2 . If we reach stage n, agent n is elected without the need for any vote.[10] Although this mechanism works if the agents have no enemies, it may fail when they do.

Example 8 (*Adachi's mechanism does not work*). Consider the same case analyzed in Example 7, where $N = \{a, b, c, d\}$ and $e = (e_a, e_b, e_c, e_d) = (d, \emptyset, a, c)$. Let us number the agents so that $1 = a$, $2 = b$, $3 = c$, and $4 = d$. Let $(\omega^d, R) \in S(e)$ be such that $\omega^d = c$ and, for each $i \in N$, $R_i(c)$ is as represented in Table 6 (given e, these preference relations can be derived from some admissible preference functions).

Next, we show that Adachi's (2014) mechanism has a backward induction equilibrium at state (ω^d, R) that results in b, and then, because $\omega^d = c$, the mechanism fails to implement the socially optimal rule. Specifically, every profile of strategies that is a backward induction equilibrium of the mechanism at (ω^d, R) is as represented in Figure 4, where the solid circles represent the agents' actions, and the solid squares represent the possible equilibrium outcomes.

These equilibria are as follows.

1. Suppose that the mechanism reaches Stage 3. Then, agent c will be ultimately selected. To see this, note that:

1.1. If agent d's turn comes, he announces "no", since he prefers d to c.

1.2. By point 1.1, if agent b's turn comes, he announces "yes", since he prefers c to d.

[10] The mechanism proposed originally by Adachi (2014) is intended to choose a ranking of candidates (not necessarily complete) when the agents in charge of choosing it may have several friends. Here, we present its version for the case in which only one candidate has to be chosen, and the agents are also the candidates.

Table 7 $\hat{R}_i(c)$ in Example 8

$\hat{R}_a(c)$	$\hat{R}_b(c)$	$\hat{R}_c(c)$	$\hat{R}_d(c)$
a	b	c	d
c	c	$a?b?d$	c
$b?d$	$a?d$		$b?a$

1.3. By point 1.2, agent a is indifferent between announcing "yes" or "no" since, in both cases, c is selected.

2. Suppose that the mechanism reaches Stage 2. Then, agent b will be ultimately selected. To see this, note that:

2.1. By point 1, if agent d's turn comes, he announces "yes", since he prefers b to c.

2.2. By point 2.1, if agent c's turn comes, he is indifferent between announcing "yes" or "not" since, in both cases, b is selected.

2.3. By point 2.2, agent a is indifferent between announcing "yes" or "no" since, in both cases, b is selected.

3. At Stage 1, all agents will announce "no", and then, by point 2, agent b will be ultimately selected. To see this, note that:

3.1. By point 2, if agent d's turn comes, he announces "no", since he prefers b to a.

3.2. By point 3.1, if agent c's turn comes, he announces "no", since he prefers b to a.

3.3. By point 3.2, agent b announces "no", since he prefers b to a.

Like Amorós' (2011) mechanism, Adachi's (2014) mechanism would not fail if no agent had enemies. If the jury configuration were $\hat{e} = (\hat{e}_a, \hat{e}_b, \hat{e}_c, \hat{e}_d) = (\emptyset, \emptyset, \emptyset, \emptyset)$, every profile of admissible preference functions $\hat{R} \in \mathcal{R}(\hat{e})$ would be such that, for each $i \in N$, $\hat{R}_i(c)$ is as represented in Table 7. Note that the preference relation $R_d(c)$ in Table 6 is no longer possible. In this case,, for every $(\omega^d, \hat{R}) \in S(\hat{e})$ such that $\omega^d = c$, the backward induction equilibria of the mechanism in Figure 4 at state $(\omega^d, \hat{R})$ are represented by the dashed circles and squares (dashed circles represent the agents' actions, and dashed squares represent the outcome of these equilibria). All these equilibria result in $a = \omega^d$.

Appendix B

Multiple friends and enemies

In this Appendix, we propose a way to extend the model analyzed in this work to the situation where each juror may have multiple friends he wishes to favor and enemies he aims to harm. While generalizing the necessary condition for implementation established in Proposition 1 to this setting is straightforward, extending Mechanism 1 (and therefore Theorem 1) to the case of multiple friends and enemies is not obvious and would require extensive new analysis, which we leave for future research.

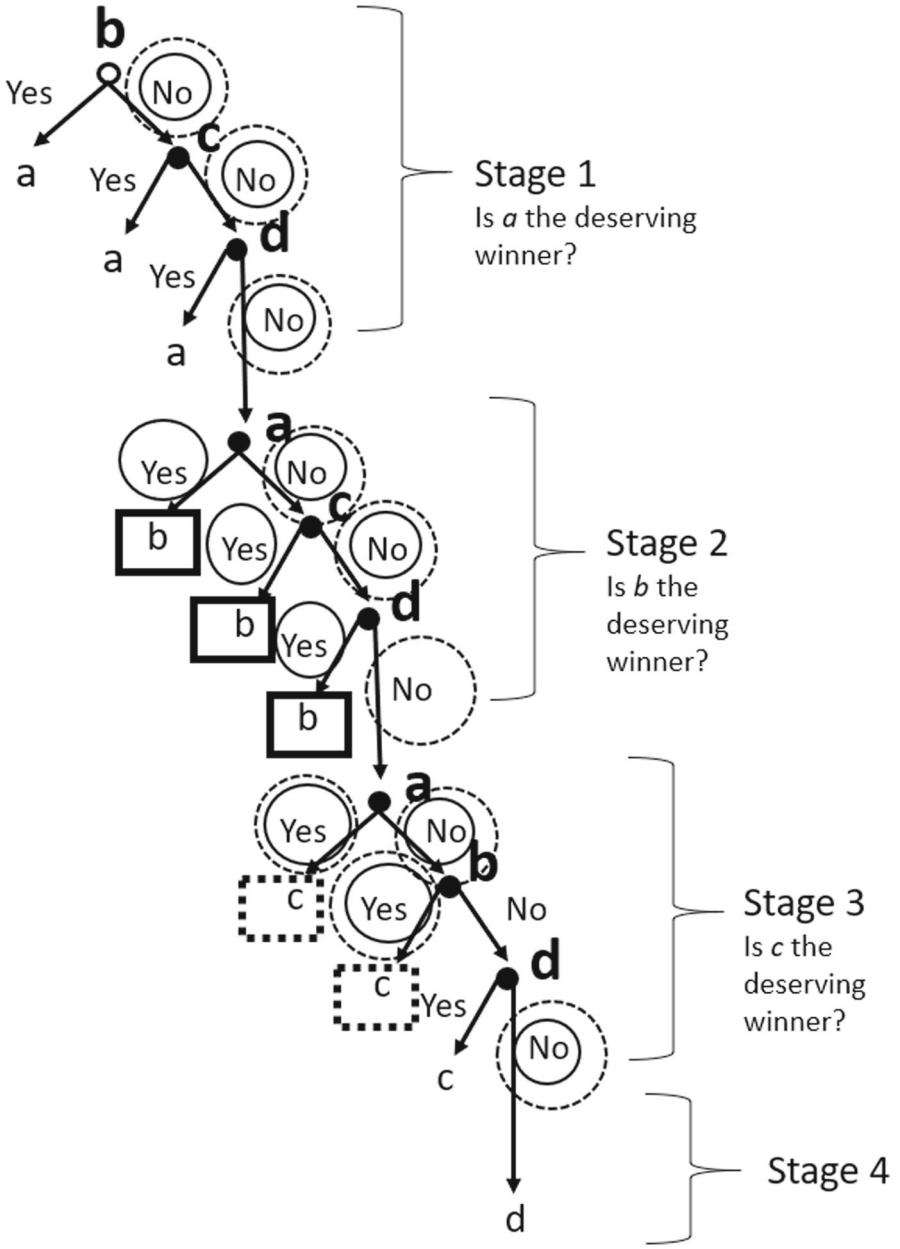

Fig. 4 Adachi's (2014) mechanism in Example 8

If each juror can have multiple friends and enemies, it is no longer possible to establish a one-to-one identification between candidates and jurors. Therefore, we will now distinguish between a group J of $n \geq 2$ jurors and a group C of $m \geq 2$ candidates. Let $\Re$ the class of preference relations over C. A preference function for juror $i \in J$ is a mapping $R_i : C \to \Re$ which associates with each possible deserving winner $\omega^d \in C$ a preference relation $R_i(\omega^d)$. Each juror i is characterized by a partition of C into three subsets: *friends* ($F_i \subset C$), *enemies* ($E_i \subset C$), and *impartial* candidates ($I_i \subseteq C$). Some of these subsets may be empty. Specifically, if I_i is non-empty, it must include more than one candidate. The planner knows that: (1) i prefers any candidate in F_i over any candidate not in that group, regardless of the deserving winner; (2) i prefers any candidate not included in E_i over those in it, regardless of the deserving winner; (3) whenever one candidate in I_i is the deserving winner, i prefers that candidate over any other in that group.

Definition 3 A preference function $R_i : C \to \Re$ is *admissible* for juror i at (F_i, E_i, I_i) if:

(1) For every $x \in F_i$, $\omega^d \in C$, and $y \in C \backslash F_i$ we have $x \; P_i(\omega^d) \; y$.
(2) For every $x \in E_i$, $\omega^d \in C$, and $y \in C \backslash E_i$ we have $y \; P_i(\omega^d) \; x$.
(3) For every $x, y \in I_i$, if $x = \omega^d$ then $x \; P_i(\omega^d) \; y$.

Let $\mathcal{R}(F_i, E_i, I_i)$ be the class of all preference functions that are admissible for i at (F_i, E_i, I_i). A jury configuration is a profile $(F, E, I) \equiv (F_i, E_i, I_i)_{i \in J}$. A state is a profile (R, ω^d), where $R = (R_i)_{i \in J}$. The set of admissible states when the jury configuration is (F, E, I) is $S(F, E, I) \equiv \times_{i \in J} \mathcal{R}(F_i, E_i, I_i) \times C$. The socially optimal rule is a function $\psi : S(F, E, I) \to C$ such that, for each $(R, \omega^d) \in S(F, E, I)$, $\psi(R, \omega^d) = \omega^d$. Given a jury configuration (F, E, I), a finite sequential mechanism with perfect information implements the socially optimal rule via backward induction if, for each admissible state $(R, \omega^d) \in S(F, E, I)$, the only outcome reached via backward induction is ω^d.

We now propose an extension of the necessary condition for implementation in this setting, where a juror may have multiple friends and enemies. The condition is a generalization of minimal impartiality and requires that, for each pair of candidates, at least one juror includes them among their impartial candidates.

Definition 4 A jury configuration (F, E, I) is *minimally neutral* if, for every $x, y \in C$ there exists some $i \in J$ such that $x, y \in I_i$.

Proposition 2 *Given a jury configuration* (F, E, I)*, suppose the socially optimal rule is implementable via backward induction. Then,* (F, E, I) *is minimally neutral.*

Proof Suppose by contradiction that there exist $x, y \in C$ such that, for every $i \in J$, either $x \notin I_i$ or $y \notin I_i$. Then, by definition of admissible preference function, there exists $R \in \times_{i \in J} \mathcal{R}(F_i, E_i, I_i)$ such that, for every $i \in J$, $R_i(x) = R_i(y)$ (*i.e.*, the preference relation of each juror i when the deserving winner is x is the same as when the deserving winner is y). Then, the backward induction equilibria at state (R, x) are the same as those at state (R, y). If a mechanism implements ψ via backward induction, a backward induction equilibrium exists at state (R, x) that results in x.

Table 8 Admissible preference functions for agents a and b in Example 9

$R_a : N \longrightarrow \Re$				$R_b : N \longrightarrow \Re$			
a	b	c	d	a	b	c	d
d	d	d	d	a	$a\ ?\ c\ ?\ d$	c	d
$b\ ?\ c$	b	c	$b\ ?\ c$	$c\ ?\ d$	b	$a\ ?\ d$	$a\ ?\ c$
a	c	b	a	b		b	b
	a	a					

Then, the same equilibrium is a backward induction equilibrium at state (R, y), which contradicts that the mechanism implements ψ.[11] □

The intuition behind the proof of this result is the same as that of Proposition 1. If the jury configuration is not minimally neutral, there exist two states such that, when moving from one to the other, the individual preferences of all jurors remain unchanged even though the deserving winner has changed. Consequently, any backward induction equilibrium in the first state will also be one in the second. Since the deserving winners in the two states are different, implementing the socially optimal rule becomes unattainable.

Unfortunately, extending Mechanism 1 (and therefore Theorem 1) to this setting with multiple friends and enemies is not obvious and would require extensive new analysis.

Appendix C

Reversal of roles between agents and enemies

In this Appendix, we consider the situation in which being selected is a "bad" thing so the worst alternative for every agent is for himself to be chosen while the enemy being chosen is his best alternative. In this case, the class of admissible preference function for an agent is defined as follows.

Definition 5 A preference function $R_i : N \rightarrow \Re$ is *admissible* for agent $i \in N$ at $e_i \in (N \backslash \{i\} \cup \{\emptyset\})$ if:

(1) For every $\omega^d \in N$ and $j \in N \backslash \{i\}$, we have $j\ P_i(\omega^d)\ i$.
(2) If $e_i \neq \emptyset$, then, for every $\omega^d \in N$ and $j \in N \backslash \{e_i\}$ we have $e_i\ P_i(\omega^d)\ j$.
(3) For every $j, k \in N \backslash \{i, e_i\}$, if $\omega^d = j$ then $j\ P_i(\omega^d)\ k$.

Example 9 Let us consider the same situation analyzed in Example 1, where $N = \{a, b, c, d\}$, $e_a = d$ and $e_b = \emptyset$. Now that being chosen is a bad thing, the admissible preference functions of agents a and b are those represented in Table 8.

In this context, the same minimal impartiality condition from Definition 2 remains necessary for implementing the socially optimal rule via backward induction. The

[11] Note that the same argument applies to the implementation of ψ in any equilibrium concept that depends only on the ordinal preferences of the agents, such as dominant strategies, Nash equilibrium, subgame perfect equilibrium, etc.

proof of this result is very similar to that of Proposition 1, and we include it for completeness.

Proposition 3 *Given a jury configuration e, suppose the socially optimal rule is implementable via backward induction. Then, e is minimally impartial.*

Proof *Claim 1. If e is not minimally impartial, then there are* $i, j \in N$ *and* $R \in R(e)$ *such that, for every* $k \in N$, $R_k(i) = R_k(j)$.

If e is not minimally impartial, then there are $i, j \in N$ such that, for every $k \in N\backslash\{i, j\}$, we have $e_k = i$ or $e_k = j$. Abusing notation, for each $k \in N\backslash\{i, j\}$, let $R_k^* \in \Re$ be a preference relation such that:

(1) if $e_k = i$, then $i \; P_k^* \; j \; P_k^* \; l \; P_k^* \; k$ for every $l \in N\backslash\{i, j, k\}$, and
(2) if $e_k = j$, then $j \; P_k^* \; i \; P_k^* \; l \; P_k^* \; k$ for every $l \in N\backslash\{i, j, k\}$.

Similarly, let $R_i^* \in \Re$ be a preference relation such that:

(1) if $e_i = \emptyset$ or $e_i = j$, then $j \; P_i^* \; l \; P_i^* \; i$ for every $l \in N\backslash\{i, j\}$, and
(2) if $e_i \neq \emptyset$ and $e_i \neq j$, then $e_i \; P_i^* \; j \; P_i^* \; l \; P_i^* \; i$ for every $l \in N\backslash\{i, j, e_i\}$.

Finally, let $R_j^* \in \Re$ be a preference relation such that:

(1) if $e_j = \emptyset$ or $e_j = i$, then $i \; P_j^* \; l \; P_j^* \; j$ for every $l \in N\backslash\{i, j\}$, and
(2) if $e_j \neq \emptyset$ and $e_j \neq i$, then $e_j \; P_j^* \; i \; P_j^* \; l \; P_j^* \; j$ for every $l \in N\backslash\{i, j, e_j\}$.

By definition of admissible preference function, there exists $R \in \mathcal{R}(e)$ such that (i) $R_i(i) = R_i(j) = R_i^*$, (ii) $R_j(i) = R_j(j) = R_j^*$, and (iii) for each $k \in N\backslash\{i, j\}$, $R_k(i) = R_k(j) = R_k^*$.

Claim 2. If e is not minimally impartial, the socially optimal rule is not implementable via backward induction.

The proof is identical to that of Claim 2 in Proposition 1. □

Next, assuming that minimal impartiality holds, we analyze the modification of Mechanism 1 to make it work when being selected is a bad thing. To do so, we distinguish between two scenarios: one where there is at least one agent without enemies and another where all agents have enemies. This distinction is because, in the first scenario, a very simple mechanism with only two stages works. Unfortunately, if all agents have enemies, the modified mechanism is more complex and closely resembles the original Mechanism 1.

MECHANISM 2 *($e_i = \emptyset$ for some $i \in N$)*

Suppose that e is such that it satisfies minimal impartiality and at least one agent i is such that $e_i = \emptyset$. By minimal impartiality, we know that for every agent $j \in N\backslash\{i\}$, there exists an agent $k_{i,j} \in N\backslash\{i, j\}$ such that, for every $R_{k_{i,j}} \in \mathcal{R}_{k_{i,j}}(e_{k_{i,j}})$, we have $i \; P_{k_{i,j}}(i) \; j$ and $j \; P_{k_{i,j}}(j) \; i$. Mechanism 2 works as follows. In the first stage, agent i announces who he believes should win. If he announces himself, the mechanism selects him. If he announces another agent $j \neq i$, a second stage follows, where $k_{i,j}$ chooses the winner between i and j. Formally:

Stage 1: Agent i announces $m_i \in N$.

- If $m_i = i$, then m_i is chosen as winner. STOP.
- If $m_i \neq i$, then go to Stage 2.m_i.

Stage 2.m_i: Agent k_{i,m_i} announces $m_{k_{i,m_i}} \in \{i, m_i\}$.

- Then $m_{k_{i,m_i}}$ is chosen as winner. STOP.

Theorem 2 *Suppose that the jury configuration e is minimally impartial and that there is some $i \in N$ with $e_i = \emptyset$. Then, Mechanism 2 implements φ via backward induction.*

Proof For each agent $j \in N\backslash\{i\}$, let $k_{i,j} \in N\backslash\{i, j\}$ be as defined in Mechanism 2. Given any state, $(\omega^d, R) \in S(e)$, consider a profile of strategies m_i^*, $(m^*_{k_{i,m_i}})_{m_i \in N\backslash\{i\}}$ that is a backward induction equilibrium of Mechanism 2 at (ω^d, R).

Claim 1. Let $m_i \in N\backslash\{i\}$.

(1.1) If $\omega^d \in \{i, m_i\}$, then $m^*_{k_{i,m_i}} = \omega^d$.

(1.2) If $\omega^d \notin \{i, m_i\}$, then $m^*_{k_{i,m_i}} \in \{i, m_i\}$.

Suppose Stage 2.m_i is reached. By definition, agent k_{i,m_i} is such that $i \; P_{k_{i,m_i}}(i) \; m_i$ and $m_i \; P_{k_{i,m_i}}(m_i) \; i$. Therefore, if $\omega^d = i$, agent k_{i,m_i} will announce i, and if $\omega^d = m_i$, he will announce m_i. If $\omega^d \notin \{i, m_i\}$, it can happen either $i \; P_{k_{i,m_i}}(\omega^d) \; m_i$ or $m_i \; P_{k_{i,m_i}}(\omega^d) \; i$, so agent k_{i,m_i} can announce either i or m_i .

Claim 2. If $\omega^d \neq i$, then $m_i^* = \omega^d$ and ω^d will be ultimately chosen.

This follows from Claim (1.1) and the fact that, since $e_i = \emptyset$ and $\omega^d \neq i$, then ω^d is the best alternative for agent i.

Claim 3. If $\omega^d = i$, then $m_i^* \in N$ and i will be ultimately chosen.

From the definition of Mechanism 2 and Claim (1.1), it follows that, regardless of the announcement m_i made by i in Stage 1, ω^d will be ultimately chosen. □

MECHANISM 3 *($e_i \neq \emptyset$ for every $i \in N$)*

Suppose that e is such that it satisfies minimal impartiality and $e_i \neq \emptyset$ for every $i \in N$.[12] By minimal impartiality, we know that for every agent $j \in N\backslash\{i\}$, there exists an agent $k_{i,j} \in N\backslash\{i, j\}$ such that, for every $R_{k_{i,j}} \in \mathcal{R}_{k_{i,j}}(e_{k_{i,j}})$, we have $i \; P_{k_{i,j}}(i) \; j$ and $j \; P_{k_{i,j}}(j) \; i$.

For the mechanism that we are going to propose below to work, the following property must be fulfilled:

Definition 6 A jury configuration e satisfies *enemy diversity* if there exist four different agents $i, j, k, l \in N$ such that $e_i \notin \{e_j, e_k, e_l, j\}$ and $e_j \notin \{e_i, e_k, e_l, i\}$.

Let us number the agents, $N = \{1, ..., n\}$, in such a way that:

(1) $e_{n-1} \notin \{e_n, e_{n-2}, e_{n-3}, n-2, n-3\}$, and
(2) $e_{n-2} \notin \{e_n, e_{n-1}, e_{n-3}, n-1\}$.

In particular, given agents i, j, k, and l in Definition 6, we can take $n - 1 = i$, $n - 2 = j$, and (i) if $e_i \neq l$, we can take $n = k$ and $n - 3 = l$, and (ii) if $e_i = l$ (and therefore $e_i \neq k$), we can take $n = l$ and $n - 3 = k$.

[12] For this to be possible, it must be the case that $n \geq 5$.

Mechanism 3 is defined as follows. Agents take turns announcing who they think should win. If agent 1 proposes himself, his announcement is implemented. If agent 1 proposes an agent x who is neither himself nor his enemy, then we move to another stage in which $k_{x,1}$ chooses between those two candidates. If agent 1 proposes his enemy as the winner, the turn passes to agent 2 and the process repeats. If agents 1, 2, ..., and $n-1$ all propose themselves and it is agent n's turn, then the agent announced by agent n is chosen, even if that agent is his enemy. Formally:
Stage 1: Agent 1 *announces* $m_1 \in N$.

- If $m_1 = 1$, then m_1 is chosen as winner. STOP.
- If $m_1 = e_1$, then go to Stage 2.
- If $m_1 \notin \{1, e_1\}$, then go to Stage 1.m_1.

Stage 1.m_1*: Agent* $k_{m_1,1}$ *announces* $m_{k_{m_1,1}} \in \{m_1, 1\}$.

- Then $m_{k_{m_1,1}}$ is chosen as winner. STOP.

Stage 2: Agent 2 *announces* $m_2 \in N$.

- If $m_2 = 2$, then m_2 is chosen as winner. STOP.
- If $m_2 = e_2$, then go to Stage 3.
- If $m_2 \notin \{2, e_2\}$, then go to Stage 2.m_2.

Stage 2.m_2*: Agent* $k_{m_2,2}$ *announces* $m_{k_{m_2,2}} \in \{m_2, 2\}$.

- Then $m_{k_{m_2,2}}$ is chosen as winner. STOP.

$\vdots$

Stage $n-1$*: Agent* $n-1$ *announces* $m_{n-1} \in N$.

- If $m_{n-1} = n-1$, then m_{n-1} is chosen as winner. STOP.
- If $m_{n-1} = e_{n-1}$, then go to Stage n.
- If $m_{n-1} \notin \{n-1, e_{n-1}\}$, then go to Stage $n-1.m_1$.

Stage $(n-1).m_1$*: Agent* $k_{m_{n-1},n-1}$ *announces* $m_{k_{m_{n-1},n-1}} \in \{m_{n-1}, n-1\}$.

- Then $m_{k_{m_{n-1},n-1}}$ is chosen as winner. STOP.

Stage n: Juror n announces $m_n \in N$.

- Then m_n is chosen as winner. STOP.

The proof that Mechanism 3 implements the socially optimal rule via backward induction is very similar to that of Theorem 1, and we include it for completeness.

Theorem 3 *Suppose that the jury configuration e satisfies enemy diversity and minimal impartiality and is such that* $e_i \neq \emptyset$ *for every agent i. Then, Mechanism 3 implements* φ *via backward induction.*[13]

[13] Although we do not have a formal proof of it, we conjecture that if e satisfies minimal impartiality and is such that $e_i \neq \emptyset$ for every $i \in N$, then it satisfies enemy diversity. Therefore, enemy diversity would not impose any additional constraints on the jury configuration

Proof Let $N = \{1, ..., n\}$ be the a numbering of the agents as defined in Mechanism 3. For each pair $i, j \in N$, let $k_{i,j} \in N\backslash\{i, j\}$ be as defined in Mechanism 3. Given any state, $(\omega^d, R) \in S(e)$, consider a profile of strategies that is a backward induction equilibrium of Mechanism 3 at (ω^d, R). For each $k \in N$, let x_k denote the agent who would be ultimately chosen by Mechanism 3 in case Stage k is reached, given the previous profile of equilibrium strategies. Similarly, for each $k \in N\backslash\{n\}$ and each $m_k \in N\backslash\{k, e_k\}$, let $x_{k.m_k}$ denote the agent who would be ultimately chosen by Mechanism 3 in case Stage $k.m_k$ is reached, given the previous profile of equilibrium strategies.

Claim 1. Let $k \in N\backslash\{n\}$ and $m_k \in N\backslash\{k, e_k\}$.
(1.1) If $\omega^d = k$, then $x_{k.m_k} = k$.
(1.2) If $\omega^d = m_k$, then $x_{k.m_k} = m_k$.
(1.3) If $\omega^d \notin \{k, m_k\}$, then $x_{k.m_k} \in \{k, m_k\}$.

In this case, Stage $k.m_k$ is reached, agent $k_{m_k,k}$ chooses between m_k and k, the choice of agent $k_{m_k,k}$ will be selected as the winner, and the mechanism will stop. By definition, agent $k_{m_k,k}$ is such that $m_k \; P_{k_{m_k,k}}(m_k) \; k$ and $k \; P_{k_{m_k,k}}(k) \; m_k$. Therefore, if $\omega^d = k$, agent $k_{m_k,k}$ will announce k, and if $\omega^d = m_k$, he will announce m_k. If $\omega^d \notin \{k, m_k\}$, it can happen either $m_k \; R_{k_{m_k,k}}(\omega^d) \; k$ or $k \; P_{k_{m_k,k}}(\omega^d) \; m_k$, so agent $k_{m_k,k}$ can announce either m_k or k.

Claim 2. Let $k \in N\backslash\{n\}$.
(2.1) If $\omega^d = k$, we have $x_k = x_{k+1}$.
(2.2) If $\omega^d = e_k$, then:
(2.2.1) if $x_{k+1} \neq e_k$, we have $x_k \neq e_k$, and
(2.2.2) if $x_{k+1} = e_k$, we have $x_k = e_k$.
(2.3) If $\omega^d \notin \{k, e_k\}$, then:
(2.3.1) if $x_{k+1} \neq e_k$, we have $x_k = \omega^d$, and
(2.3.2) if $x_{k+1} = e_k$, we have $x_k = e_k$.

Suppose that $\omega^d = k$. Then, by definition of Mechanism 3 and by Claim 1, at Stage k, agent k has to decide between selecting k at that stage or selecting x_{k+1} at some later stage: if $m_k = k$, then k is chosen at Stage k; if $m_k \notin \{k, e_k\}$, then k is chosen at Stake $k.m_k$; if $m_k = e_k$, then x_{k+1} is ultimately chosen at some later stage. Because k is the worst alternative for agent k, he will announce $m_k = e_k$, and agent x_{k+1} will be ultimately chosen (if $x_{k+1} = k$, he is indifferent between all announcements).

Suppose that $\omega^d = e_k$ and $x_{k+1} \neq e_k$. In this case, agent k cannot do anything to ensure that e_k gets selected neither at Stage k nor at any subsequent stage: if his announcement is $m_k = k$, then k is chosen at Stage k; if his announcement is $m_k \notin \{k, e_k\}$, then the mechanism advances to Stage $k.m_k$, in which either m_k or k will be chosen (depending on the preferences of agent $k_{m_k,k}$); if his announcement is $m_k = e_k$, then the mechanism advances to Stage $k + 1$, and agent $x_{k+1} \neq e_k$ is ultimately selected. Among the previous agents that k can make to be chosen once Stage k is reached (all different from e_k), x_k will be the one he prefers the most.

Suppose that $\omega^d = e_k$ and $x_{k+1} = e_k$. Then, because e_k is the best alternative for agent k, he will announce $m_k = e_k$ and agent $x_{k+1} = e_k$ will be chosen at some subsequent stage.

Suppose that $\omega^d \notin \{k, e_k\}$ and $x_{k+1} \neq e_k$. In this case, using the same argument as when $\omega^d = e_k$ and $x_{k+1} \neq e_k$, we conclude that agent k cannot do anything to ensure

that e_k is selected neither at Stage k nor at any subsequent stage. However, by Claim 1, now agent k can ensure that ω^d is chosen at Stage k by announcing $m_k = \omega^d$. Since $\omega^d \notin \{k, e_k\}$, then ω^d is the second most preferred alternative for agent k. Therefore, agent k will announce $m_k = \omega^d$, and ω^d will be selected at Stage k.

Suppose that $\omega^d \notin \{k, e_k\}$ and $x_{k+1} = e_k$. Because e_k is the best alternative for agent k, he will announce $m_k = e_k$, and agent $x_{k+1} = e_k$ will be ultimately chosen at some subsequent stage.

Claim 3. Suppose that Stage n is reached. *Then,* $x_n = e_n$.

It follows from the fact that e_n is the most preferred alternative for agent n.

Claim 4. Suppose that Stage $n-1$ *is reached.*

(4.1) If $\omega^d = n-1$, *then* $x_{n-1} = e_n$.

(4.2) If $\omega^d = e_{n-1}$, *then* $x_{n-1} \neq e_{n-1}$.

(4.3) If $\omega^d \notin \{n-1, e_{n-1}\}$, *then* $x_{n-1} = \omega^d$.

It follows from Claims 2 and 3 and the fact that Mechanism 3 is such that $e_{n-1} \neq e_n$.

Claim 5. Suppose that Stage $n-2$ *is reached.*

(5.1) If $\omega^d = e_{n-1}$ *and* $x_{n-1} = e_{n-2}$, *then* $x_{n-2} = e_{n-2}$.

(5.2) Otherwise, $x_{n-2} = \omega^d$.

Suppose that $\omega^d = e_{n-2}$. Since Mechanism 3 is such that $n-1 \neq e_{n-2}$ and $e_{n-1} \neq e_{n-2}$, then $\omega^d \notin \{n-1, e_{n-1}\}$. Therefore, by point 4.3 of Claim 4, $x_{n-1} = \omega^d$. Then, because $\omega^d = e_{n-2}$, by point 2.2.2 of Claim 2, we have $x_{n-2} = e_{n-2} = \omega^d$.

Suppose that $\omega^d = n-2$. Since Mechanism 3 is such that $n-2 \neq e_{n-1}$ and $n-2 \neq n-1$, then $\omega^d \notin \{n-1, e_{n-1}\}$. Therefore, by point 4.3 of Claim 4, $x_{n-1} = \omega^d$. Then, because $\omega^d = n-2$, by point 2.1 of Claim 2, we have $x_{n-2} = x_{n-1} = \omega^d$.

Suppose that $\omega^d \notin \{n-2, e_{n-2}\}$. Then we have three possibilities:

(i) If $\omega^d = e_{n-1}$, by point 4.2 of Claim 4, $x_{n-1} \neq e_{n-1}$. If, in addition, $x_{n-1} \neq e_{n-2}$, by point 2.3.1 of Claim 2, we have $x_{n-2} = \omega^d$. If, on the contrary, $x_{n-1} = e_{n-2}$, by point 2.3.2 of Claim 2, we have $x_{n-2} = e_{n-2}$.

(ii) If $\omega^d = n-1$, by point 4.1 of Claim 4, $x_{n-1} = e_n$. Then, because $\omega^d \notin \{n-2, e_{n-2}\}$ and $x_{n-1} = e_n$, and since Mechanism 3 is such that $e_n \neq e_{n-2}$, by point 2.3.1 of Claim 2, we have $x_{n-2} = \omega^d$.

(iii) If $\omega^d \notin \{n-1, e_{n-1}\}$, by point 4.3 of Claim 4, $x_{n-1} = \omega^d$. Then, because $\omega^d \notin \{n-2, e_{n-2}\}$ and $x_{n-1} = \omega^d \neq e_{n-2}$, by point 2.3.1 of Claim 2, we have $x_{n-2} = \omega^d$.

Claim 6. Suppose that Stage $n-3$ *is reached. Then,* $x_{n-3} = \omega^d$.

Suppose that $\omega^d = e_{n-3}$. By definition of Mechanism 3, we have $e_{n-3} \neq e_{n-1}$, and then $\omega^d \neq e_{n-1}$. Therefore, by point 5.2 of Claim 5, $x_{n-2} = \omega^d$. Hence, by point 2.2 2 of Claim 2, $x_{n-3} = e_{n-3} = \omega^d$.

Suppose that $\omega^d = n-3$. By definition of Mechanism 3, we have $n-3 \neq e_{n-1}$, and then $\omega^d \neq e_{n-1}$. Therefore, by point 5.2 of Claim 5, $x_{n-2} = \omega^d$. Hence, by point 2.1 of Claim 2, $x_{n-3} = x_{n-2} = \omega^d$.

Suppose that $\omega^d \notin \{n-3, e_{n-3}\}$. Then we have two possibilities:

(i) Suppose that $\omega^d = e_{n-1}$. If, in addition, $x_{n-1} = e_{n-2}$, by point 5.1 of Claim 5, we have $x_{n-2} = e_{n-2}$. By definition of Mechanism 3, we have $e_{n-2} \neq e_{n-3}$. Then, by point 2.3.1 of Claim 2, $x_{n-3} = \omega^d$. If, on the contrary, $x_{n-1} \neq e_{n-2}$, by point 5.2 of Claim 5, we have $x_{n-2} = \omega^d$. By definition of Mechanism 3, we have $e_{n-1} \neq e_{n-3}$. Then, $x_{n-2} = \omega^d = e_{n-1} \neq e_{n-3}$, and by point 2.3.1 of Claim 2, $x_{n-3} = \omega^d$.

(ii) Suppose that $\omega^d \neq e_{n-1}$. Then, by point 5.2 of Claim 5, $x_{n-2} = \omega^d$. Because $\omega^d \neq e_{n-3}$, then $x_{n-2} \neq e_{n-3}$. Hence, by point 2.3.1 of Claim 2, $x_{n-3} = \omega^d$.

Claim 7. Suppose that Stage $n-4$ is reached. Then, $x_{n-4} = \omega^d$.

Suppose that $\omega^d = n-4$. Then, by point 2.1 of Claim 2 and by Claim 6, $x_{n-4} = x_{n-3} = \omega^d$.

Suppose that $\omega^d = e_{n-4}$. By Claim 6, $x_{n-3} = \omega^d$. Then, since $\omega^d = e_{n-4}$ and $x_{n-3} = e_{n-4}$, by point 2.2.2 of Claim 2, $x_{n-4} = e_{n-4} = \omega^d$.

Suppose that $\omega^d \notin \{n-4, e_{n-4}\}$. By Claim 6, $x_{n-3} = \omega^d$. Then, since $\omega^d \neq e_{n-4}$, by point 2.3.1 of Claim 2, $x_{n-4} = \omega^d$.

Claim 8. Suppose that $n \geq 6$. Suppose that Stage $n-t$ is reached for some $t \geq 5$. Then, $x_{n-t} = \omega^d$.

The demonstration of this claim follows from repeatedly applying the same argument used in the proof of Claim 7.

From Claims 6, 7, and 8, we have that, given any state $(\omega^d, R) \in S(e)$, every profile of backward induction equilibrium strategies of Mechanism 1 at (ω^d, R) is such that ω^d is ultimately chosen as the winner. □

Declaration of generative AI and AI-assisted technologies in the writing process. During the preparation of this work the author used ChatGPT in order to improve the language and readability of the article. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References

Adachi T (2014) A natural mechanism for eliciting rankings when jurors have favorites. Games Econ Behav 87:508–518

Amorós P, Corchón LC, Moreno B (2002) The scholarship assignment problem. Games Econ Behav 38:1–18

Amorós P (2009) Eliciting socially optimal rankings from unfair jurors. J Econ Theory 144:1211–1226

Amorós P (2011) A natural mechanism to choose the deserving winner when the jury is made up of all contestants. Econ Lett 110:241–244

Amorós P (2023) Implementing optimal scholarship assignments via backward induction. Math Soc Sci 125:1–10

Herrero MJ, Srivastava S (1992) Implementation via backward induction. J Econ Theory 56:70–88

Holzman R, Moulin H (2013) Impartial nominations for a prize. Econometrica 81:173–196

Jackson MO (1992) Implementation in undominated strategies: a look at bounded mechanisms. Rev Econ Stud 59:757–775

Mackenzie A (2015) Symmetry and impartial lotteries. Games Econ Behav 94:15–28

Mackenzie A (2020) An axiomatic analysis of the papal conclave. Econ Theory 69:713–743

Mackenzie A, Zhou Y (2022) Menu mechanisms. J Econ Theory 204:105511

Maskin E (1999) Nash equilibrium and welfare optimality. Rev Econ Stud 66:23–38

Moore J, Repullo R (1988) Subgame perfect implementation. Econometrica 56:1191–1220

Niemeyer A, Preusser J (2023) Simple allocation with correlated types. In: Discussion Paper Series—Collaborative Research Center Transregio 224 Discussion Paper No. 486

Olckers M, Walsh, T (2024) Manipulation and peer mechanisms: a survey. Artificial Intelligence 104196

Tamura S (2016) Characterizing minimal impartial rules for awarding prizes. Games Econ Behav 95:41–46

Yadav S (2016) Selecting winners with partially honest jurors. Math Soc Sci 83:35–43