

# Fair and Strategy-proof Tournament Rules Design

## Definition. (Tournament)

A tournament  $T = (A, \succ_T)$  is a pair where

- $A$  is a finite set, whose elements are called *agents/alternatives/participants/players*
- $\succ_T$  is a complete asymmetric binary relation over  $A$ ,  
that describes the outcomes of the  $\binom{\#A}{2}$  matches played between each pair of distinct agents.

**Notation.** For agents  $i \neq j \in A$ , we write  $i \succ_T j$  if  $i$  dominates  $j$  in  $T$ .

Let  $\mathcal{T}(A)$  denote the set of tournaments over  $A$ .

If the set of agents is clear we write just  $\mathcal{T}$ .

Since  $A$  is finite we will identify it with  $[n] := \{1, \dots, n\}$ , where  $n = \#A$ .

Let  $\mathcal{T}_n := \mathcal{T}([n])$  denote the set of tournaments on  $n$  agents.

**Remark.** A tournament  $T \in \mathcal{T}(A)$  can also be represented as an oriented graph  $G = (V, \vec{E})$ , where the set of nodes/vertices  $V$  coincide with the set of agents  $A$ , and the edges are given by the tournament relation: the oriented edge from  $i$  to  $j$  is in  $\vec{E}$  if and only if  $i \succ_T j$ . In other words,  $G$  is an orientation of the complete graph on  $A$ .

Given a tournament it is natural to consider some criteria, called *tournament rules* or *choice rules*, that select some kind of winner. The result could be:

- a single alternative (*deterministic rule*)
- a collection of alternatives (*set selection rule*)
- a probability distribution over the set of alternatives (*probabilistic/randomized rule*)

We are interested in designing rules that are both *fair* (selects a meritable alternative) and *resistant to manipulations* (alternatives do not benefit from match-fixing).

---

Each chapter, except the last, is based on one of the references (in order of appearance).

For [3] and [4] I used the ArXiv version.

## Contents

1. Deterministic Rules .....	2
2. Probabilistic Rules .....	6
3. Approximate Non-Manipulability .....	11
4. Partially Transferable Utility .....	15
5. Beyond Condorcet .....	20
Bibliography .....	24

# 1. Deterministic Rules

We start by considering deterministic tournament rules.

**Definition 1.1.** ((Deterministic) Tournament Rule)

A (deterministic) tournament rule over  $A$  is a map  $r : \mathcal{T}(A) \rightarrow A$ , that maps each tournament to a winner.

## Fairness

A fairly intuitive fairness condition is that an alternative which dominates all the others should be the winner.

**Definition 1.2.** (Condorcet Winner)

Alternative  $x$  is a *Condorcet winner* in tournament  $T \in \mathcal{T}(A)$  if  $x \succ_T y$  for all  $y \in A \setminus \{x\}$ .

**Remark.** For a given tournament the Condorcet winner may not exist, but if it does then it is unique (for completeness + asymmetry).

**Definition 1.3.** (Condorcet Consistency)

A (deterministic) tournament rule  $r$  is *Condorcet consistent* if  $r$  always selects the Condorcet winner for every tournament, if one exists.

## Non-manipulability

We consider the following kind of manipulations:

- an individual player can lose on purpose a match that it could have won, or in other words, can reverse outgoing edges (but not incoming ones!);
- two or more players can collude to reverse some of the outcomes of the matches between players in the coalition.

Our goal is to prevent such manipulations.

In either case a manipulation involves switching to a second tournament which is similar to the original, but with some edges reversed.

**Definition 1.4.** (Adjacent Tournaments)

Two tournaments  $T, T' \in \mathcal{T}(A)$  are *S-adjacent* for  $S \subseteq A$  if they differ only on edges in  $S$ , that is for all  $z, w \notin S$  we have  $z \succ_T w \Leftrightarrow z \succ_{T'} w$ .

A successful manipulation of the first kind is that in which a losing player can become the winner of the tournament after reversing an outgoing edge.

**Definition 1.5.** (Monotonicity)

A (deterministic) tournament rule  $r : \mathcal{T}(A) \rightarrow A$  is *monotone* (or *monotonic*) if for all  $T \in \mathcal{T}(A)$ , for all  $x \in A$  such that  $r(T) \neq x$ , and for all  $y \in A$  such that  $x \succ_T y$ , if  $T$  and  $T'$  are  $\{x, y\}$ -adjacent then  $r(T') \neq x$ .

**Example.** Consider the rule which selects the alternative with maximum outdegree (also known as *Copeland rule*). Since this may not be unique, we need a tie-breaking mechanism (for example based on lexicographic order or just pick one at random). If an alternative loses on purpose it only decreases its outdegree, hence this rule is monotone.

On the contrary, the rule which selects the alternative with minimum outdegree is not monotone. Indeed, in this case losing matches is rewarded in order to win the tournament.

A successful manipulation of the second kind for two players is achieved when one of them becomes the winner of the tournament after reversing the outcome of their match.

**Definition 1.6.** (Pairwise Non-Manipulability)

A (deterministic) tournament rule  $r : \mathcal{T}(A) \rightarrow A$  is *pairwise non-manipulable* (PNM) if for all tournaments  $T \in \mathcal{T}(A)$  and all alternatives  $x, y \in A$ , if  $T$  and  $T'$  are  $\{x, y\}$ -adjacent then  $r(T) \in \{x, y\} \Leftrightarrow r(T') \in \{x, y\}$ .

**Remark.** PNM is incomparable with monotonicity.

Indeed, monotonicity clearly does not imply PNM. For example, consider Copeland rule (which is monotone) on a 3-cycle  $x \succ_T y \succ_T z \succ_T x$ : since they have all same outdegree, for tie-breaking  $x$  would be the winner; but  $\{y, z\}$  can collude to reverse their outcome and let  $z$  win.

In the other direction, let  $A = \{a, b\}$  such that  $a \succ_T b$  and  $b \succ_{T'} a$  in the only two possible tournaments  $T$  and  $T'$ . Set  $r(T) = b$  and  $r(T') = a$ ; then  $r$  is vacuously PNM, but is not monotone, since  $a$  can gain by moving from  $T$  to  $T'$ , and  $b$  can gain by moving from  $T'$  to  $T$ .

Analogously, for more than two alternatives we declare the winner based only on the edge between  $a$  and  $b$  in the same way. PNM still holds because no other pair can influence the result by changing their outcome, but this is still not monotone.

## Impossibility result

PNM and Condorcet-consistency are mutually exclusive.

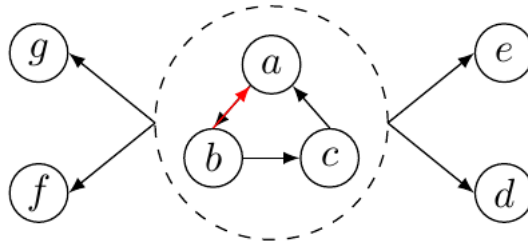
**Theorem 1.7.**

Let  $A$  such that  $\#A \geq 3$ .

Any PNM (deterministic) tournament rule  $r : \mathcal{T}(A) \rightarrow A$  is not Condorcet consistent.

*Proof.* Let  $A = \{a, b, c, \dots\}$ , and assume for contradiction that  $r$  is both PNM and Condorcet consistent; we define a tournament  $T \in \mathcal{T}(A)$  as follows. We let  $C = \{a, b, c\}$  be a subset of  $A$ , where  $C \succ_T x$  for all  $x \in A \setminus C$ . Moreover, inside  $C$  we have that  $a \succ_T b$ ,  $b \succ_T c$  and  $c \succ_T a$ .

Now, there must be at least two alternatives (there might be three) from  $C$  that are different from  $r(T)$ ; without loss of generality  $r(T) \notin \{a, b\}$ . Let  $T'$  be an adjacent tournament such that  $b \succ_{T'} a$ . Crucially,  $b$  is a Condorcet winner in  $T'$ . Since  $r$  is Condorcet consistent, it must hold that  $r(T') = b$ , in contradiction to PNM.  $\square$



## Non-imposition

It is natural to ask whether we can obtain PNM by relaxing the notion of Condorcet consistency.

### Definition 1.8. (Non-Imposition)

A tournament rule  $r : \mathcal{T}(A) \rightarrow A$  is *non-imposing* if it is onto  $A$ ,  
i.e. for every  $x \in A$  there exists  $T \in \mathcal{T}(A)$  such that  $r(T) = x$ .

**Remark.** Clearly Condorcet consistency implies non-imposition.

Indeed, given any alternative  $x \in A$ , consider a tournament  $T$  where  $x$  is a Condorcet winner; then if  $r$  is Condorcet consistent,  $r(T) = x$ .

We ask whether there is a tournament rule that is both PNM and non-imposing. Notice that if there is one alternative or two alternatives, there clearly exists such a rule.

However, it turns out that there is no such rule when there are exactly three alternatives.

### Theorem 1.9.

Let  $A$  such that  $\#A = 3$ .

Any PNM tournament rule  $r : \mathcal{T}(A) \rightarrow A$  is not non-imposing.

*Proof.* Let  $A = \{a, b, c\}$  and assume for contradiction that  $r$  is PNM and non-imposing. We represent tournaments over  $A$  by a string of three bits, where the leftmost, middle and rightmost bits represent the direction of the edges  $(a, b)$ ,  $(a, c)$  and  $(b, c)$ , respectively. Fix an arbitrary tournament as 000.

Assume without loss of generality that  $r(000) = a$ . By PNM we have:

- $r(001) = a$ , otherwise the coalition  $\{b, c\}$  can gain by switching from 000 to 001;
- $r(100) \in \{a, b\}$ , otherwise the coalition  $\{a, b\}$  can gain by switching from 100 to 000;
- $r(010) \in \{a, c\}$ , otherwise the coalition  $\{a, c\}$  can gain by switching from 010 to 000.

We differentiate two cases:

*Case 1:*  $r(111) \in \{b, c\}$ .

By symmetry, we can assume WLOG that  $r(111) = b$ .

By PNM,  $r(101) = b$ ,  $r(110) \in \{b, c\}$ ,  $r(011) \in \{a, b\}$ .

By non-imposition we must have a tournament where  $c$  wins under  $r$ . By the above, we only have two options, 010 and 110.

By PNM, since the two tournaments are  $\{a, b\}$ -adjacent, if the winner in one of these two tournaments is  $c$ , the winner in the other is  $c$  as well, hence it holds that  $r(010) = c$ ,

$r(110) = c$ .

From  $r(110) = c$  and PNM it follows that  $r(100) \in \{a, c\}$ . We already know that

$r(100) \in \{a, b\}$ , hence  $r(100) = a$ .

Now by PNM  $r(101) = a$ , in contradiction to our previous conclusion that  $r(101) = b$ .

*Case 2:*  $r(111) = a$ .

By PNM we conclude that  $r(110) = a$ ,  $r(101) \in \{a, c\}$  and  $r(011) \in \{a, b\}$ .

By non-imposition we must have a tournament where  $b$  wins. We have only two options, 100 and 011.

If  $r(100) = b$ , by PNM  $r(110) = b$ , in contradiction to  $r(110) = a$ .

Symmetrically, if  $r(011) = b$ , we must have  $r(001) = b$ , in contradiction to  $r(001) = a$ .  $\square$

Surprisingly, the above impossibility result is false if  $\#A > 3$ . We establish that for any number of alternatives except three there is a tournament rule that satisfies both properties. Moreover, the tournament rule we construct is also monotone.

**Theorem 1.10.**

Let  $A$  such that  $\#A \neq 3$ .

There exists a tournament rule  $r : \mathcal{T}(A) \rightarrow A$  that is monotone, PNM and non-imposing.

*Proof.* **Omitted.**

*Idea:* The construction is inductive; we design a function that switches between the outcomes in  $\{a, b, c\}$  based only on the direction of the edges  $(a, b)$  and  $(a, c)$ . In order to achieve non-imposition, one of the four configurations of the two edges gives us outcomes in  $A \setminus \{a, b, c\}$ ; in order to decide between alternatives in this set, we use the function for  $\#A - 3$  alternatives whose existence is guaranteed by the induction assumption.

This technique does not work for three alternatives as the set  $A \setminus \{a, b, c\}$  is empty.

The main obstacle is using the same technique to obtain a monotone, PNM, non-imposing rule for six alternatives (despite the impossibility result for three). □

## 2. Probabilistic Rules

Instead of requiring a rule to select a winner we can ask for a normalized score for each alternative, which can be interpreted as the probability of winning the tournament or as a way to divide a prize.

Let  $\Delta^n = \{(x_1, \dots, x_n) \in \mathbb{R}_+^n : x_1 + \dots + x_n = 1\}$  and  $\Delta(X) = \Delta^{\#X}$  for  $X$  finite set.

**Definition 2.1.** ((Probabilistic/Randomized) Tournament Rule)

A (probabilistic/randomized) tournament rule over  $A$  is a map  $r : \mathcal{T}(A) \rightarrow \Delta(A)$ , that maps each tournament to a probability distribution over the alternatives.

A tournament rule on  $n$  agents is a map  $r^{(n)} : \mathcal{T}_n \rightarrow \Delta^n$ .

A tournament rule  $r$  is a family of tournament rules on  $n$  agents for every  $n \in \mathbb{N}$ :  $\{r^{(n)}\}_{n=1}^\infty$ .

Since we consider only tournaments over finite sets, we may use these definitions interchangeably under the name “tournament rule” (it will be clear from context how to disambiguate).

**Notation.** For  $T \in \mathcal{T}_n$  we write  $r(T) := r^{(n)}(T)$ , and for  $i \in [n]$  we write  $r_i(T) := (r(T))_i$  to denote the probability that  $i$  wins  $T$  under  $r$ .

In general we say that a property holds for a tournament rule  $r$  if it holds for every element  $r^{(n)}$  of the family.

**Remark.** A. Altman and R. Kleinberg [2] admit the possibility to select no alternative by modelling rules as maps  $\mathcal{T}(A) \rightarrow \Delta(A \cup \varepsilon)$ ; but since we will be studying rules for any number of agents anyway, this does not change the theory.

Notice that this definition of probabilistic rule generalizes the previous case:

**Definition 2.2.** (Deterministic Tournament Rule)

A (probabilistic) tournament rule over  $A$  is *deterministic* if for every tournament  $T \in \mathcal{T}(A)$  and for every alternative  $a \in A$  we have  $r_a(T) \in \{0, 1\}$ .

From now on we consider only probabilistic rules, so we update our previous definitions and introduce some new ones.

### Fairness

A simple requirement from a probabilistic tournament rule is to be independent of the names of the alternatives. This requirement is reasonable since ties could always be broken using randomization.

**Definition 2.3.** (Anonymity)

A tournament rule  $r$  over  $A$  is called *anonymous* if for every tournament  $T \in \mathcal{T}(A)$  and for every permutation  $\pi$  of  $A$  we have  $\pi(r(T)) \equiv r(\pi(T))$ .

We are not really interested in non anonymous rules.

**Definition 2.4.** (Condorcet Consistency)

A tournament rule  $r$  over  $A$  is *Condorcet consistent* (CC) if for all tournaments  $T \in \mathcal{T}(A)$  where  $c$  is a Condorcet winner in  $T$  we have  $r_c(T) = 1$ .

Equivalently,  $r_z(T) = 0$  for all  $z \in A, z \neq c$ .

The intuition behind Condorcet consistency can be extended to any tournament using the concept of a top cycle.

**Definition 2.5.** (Top Cycle)

A subset of alternatives  $C \subseteq A$  is the *top cycle* in tournament  $T \in \mathcal{T}(A)$  if it is the minimal subset of alternatives such that  $x \succ_T y$  for all  $x \in C, y \in A \setminus C$ . We denote the top cycle of  $T$  as  $TC(T)$ .

**Remark.** The top cycle is well-defined (in particular is not empty) since the set of all alternatives  $A$  satisfies the condition and the minimal set is unique.

Now we can define a condition that requires the winner to be part of the top cycle.

**Definition 2.6.** (Top Cycle Consistency)

A tournament rule  $r$  over  $A$  is *top cycle consistent* (TCC) if for every tournament  $T \in \mathcal{T}(A)$  we have that  $r_x(T) > 0$  implies  $x \in TC(T)$ .

Equivalently, if  $z \notin TC(T)$  then  $r_z(T) = 0$ .

**Example.** The *Top Cycle Rule* (TCR) simply assigns probability  $\frac{1}{\#TC(T)}$  for all elements in  $TC(T)$  and zero otherwise.

This rule is clearly top cycle consistent.

**Remark.** If there is a Condorcet winner, then the top cycle is just a singleton (namely the Condorcet winner itself).

Thus top cycle consistency implies Condorcet consistency.

**Non-manipulability****Definition 2.7.** (Monotonicity)

A tournament rule  $r$  over  $A$  is *monotone* if for all tournaments  $T \neq T' \in \mathcal{T}(A)$  that are  $\{x, y\}$ -adjacent and  $x \succ_T y$  we have  $r_x(T) \geq r_x(T')$ .

When utility is *not transferable*, a manipulation is valid only if neither of the manipulators are worse off and at least one of the manipulators is better off. We call this Pareto manipulation.

**Definition 2.8.** (Pareto Non-Manipulability)

A tournament rule  $r$  over  $A$  is *Pareto non-manipulable for coalitions of size up to  $k$*  ( $k$ -PNM) if for every  $S \subseteq A$  such that  $\#S \leq k$  and for every tournaments  $T, T' \in \mathcal{T}(A)$  that are  $S$ -adjacent we have that  $\exists x \in S : r_x(T) > r_x(T')$  or  $\forall x \in S : r_x(T) \geq r_x(T')$ .

We say that  $r$  is PNM<sup>1</sup> if it is  $k$ -PNM for any  $k \in \mathbb{N}$ .

If we assume *transferable* utility, a much stronger manipulation resistance is required, as the manipulators only care about their combined probability of winning the tournament: that is, a manipulation is valid only if the combined probability of the coalition increases as a result.

**Definition 2.9.** (Strong Non-Manipulability)

A tournament rule  $r$  over  $A$  is *strongly non-manipulable for coalitions of size up to  $k$*  ( $k$ -SNM) if for every  $S \subseteq A$  such that  $\#S \leq k$  and for every tournaments  $T, T' \in \mathcal{T}(A)$  that are  $S$ -adjacent we have  $\sum_{x \in S} r_x(T) \geq \sum_{x \in S} r_x(T')$ .

We say that  $r$  is SNM if it is  $k$ -SNM for any  $k \in \mathbb{N}$ .

**Remark.** Note that in the case of deterministic tournament rules the above definitions of non-manipulability are equivalent, and both variants of non-manipulability are extensions of the non-manipulability property for deterministic tournament rules.

<sup>1</sup>Not to be confused with [Definition 1.6](#).

## Results

### Proposition 2.10.

TCR satisfies monotonicity and PNM.

*Proof.* Notice that reversing an edge may never decrease the size of the top cycle without removing the loser from the top cycle, and may never allow the loser into the top cycle unless it was already in the top cycle. So TCR is monotone.

To see TCR satisfies PNM, note that a successful manipulation must either reduce the size of the top cycle while all coalition members remain within, or add a new manipulator to the top cycle while keeping the top cycle of the same size. Both cases require the manipulators to remove a non-manipulator from the top cycle. For that to happen, all alternatives directly or indirectly dominated by that non-manipulator must also be removed from the top cycle, and this is only possible if at least one manipulator is removed from the top cycle, in contradiction to PNM.  $\square$

We exhibited a rule which is monotone, Pareto non-manipulable and top cycle consistent (thus Condorcet consistent).

Unfortunately, we cannot do the same for the notion of strong non-manipulability.

### Proposition 2.11.

There exist no 2-SNM Condorcet consistent tournament rules.

*Proof.* Same as [Theorem 1.7](#), with the appropriate changes of notation.  $\square$

This impossibility result requires us to weaken either the non-manipulability or Condorcet consistency requirement.

## Approximate Condorcet Consistency

We start by weakening the notion of Condorcet consistency: we allow a Condorcet winner to be selected only with high probability.

### Definition 2.12. (Condorcet Consistency Value)

A tournament rule  $r$  on  $n$  alternatives has *Condorcet consistency value*  $\gamma \in [0, 1]$  if for all  $T \in \mathcal{T}_n$  such that  $c$  is a Condorcet winner in  $T$  we have  $r_c(T) \geq \gamma$ .

A tournament rule  $r = \{r^{(n)}\}_{n=1}^{\infty}$  has Condorcet consistency value  $\gamma : \mathbb{N} \rightarrow [0, 1]$  if the rule on  $n$  alternatives  $r^{(n)}$  has Condorcet consistency value  $\gamma(n)$  for every  $n \in \mathbb{N}$ .

The following simple tournament rule provides a lower bound for the Condorcet consistency value of SNM tournament rules.

### Proposition 2.13.

There exists a monotone SNM tournament rule which has Condorcet consistency value of  $\frac{2}{n}$ .

*Proof.* The required rule simply selects two alternatives uniformly at random, and then chooses the winner of the pairwise match between them. Equivalently, every alternative is chosen with probability

$$r_x(T) = \frac{2}{n(n-1)} \cdot \#\{y : x \succ_T y\}$$

If  $x$  is a Condorcet winner, then  $\#\{y : x \succ_T y\} = n - 1$  and  $r_x(T) = \frac{2}{n}$ .



The rule is (strictly) monotone, since losing on purpose reduces the number of winning matches. It is also SNM: since we can only reverse edges in  $S$ , we are just moving the counting of the winning matches from one addend to the other; for example, if we have the tournaments  $T, T'$  such that they are  $\{x, x'\}$ -adjacent,  $x \succ_T x'$  and  $x' \succ_{T'} x$ , then  $\#\{y : x \succ_T y\} = \#\{y' : x' \succ_{T'} y'\}$ .  $\square$

This lower bound is tight in the case where non-manipulability is required for coalitions of size 3.

**Theorem 2.14.**

There exists no tournament rule that is 3-SNM and has a Condorcet consistency value  $\gamma > \frac{2}{n}$ .

*Proof.* Assume for contradiction tournament rule  $r$  satisfying the conditions of the theorem.

Let  $T$  be a tournament of size  $n$  where for all  $i < j \in A$  we have  $i \prec_T j$ . In particular alternative  $n$  is a Condorcet winner in  $T$ , and by consistency we have  $r_n(T) > \frac{2}{n}$ . Consider any three agents  $i-1, i, i+1 \in A$ . Consider the tournament  $T'$  which is  $\{i-1, i+1\}$ -adjacent to  $T$  such that  $(i-1) \succ_{T'} (i+1)$ . This manipulation creates a 3-cycle in the graph:  $i \succ_{T'} (i-1) \succ_{T'} (i+1) \succ_{T'} i$ , where agents  $i-1, i, i+1$  are completely symmetric. Since this is a manipulation by  $i-1$  and  $i+1$ , we have

$$r_{i-1}(T') + r_{i+1}(T') = r_{i-1}(T) + r_{i+1}(T)$$

and by symmetry

$$r_{i-1}(T') = r_{i+1}(T') = \frac{1}{2}[r_{i-1}(T) + r_{i+1}(T)]$$

However, this could also be seen as a manipulation by the trio  $(i-1, i, i+1)$ , hence

$$r_{i-1}(T') + r_i(T') + r_{i+1}(T') = r_{i-1}(T) + r_i(T) + r_{i+1}(T)$$

and by symmetry

$$r_{i-1}(T') = r_i(T') = r_{i+1}(T') = \frac{1}{3}[r_{i-1}(T) + r_i(T) + r_{i+1}(T)]$$

Equating the two expressions for  $r_{i-1}(T')$

$$\begin{aligned} \frac{1}{2}[r_{i-1}(T) + r_{i+1}(T)] &= \frac{1}{3}[r_{i-1}(T) + r_i(T) + r_{i+1}(T)] \\ \cancel{2}[r_{i-1}(T) + r_{i+1}(T)] &= 2[\cancel{r_{i-1}(T)} + r_i(T) + \cancel{r_{i+1}(T)}] \\ 2r_i(T) &= r_{i+1}(T) + r_{i-1}(T) \end{aligned}$$

This recurrence means the sequence  $(r_1(T), \dots, r_n(T))$  is an arithmetic progression:

$$\begin{cases} r_1(T) = a \\ r_k(T) = a + (k-1)d, \quad k = 1, \dots, n \end{cases}$$

$$2[a + (i-1)d] = [a + id] + [a + (i-2)d]$$

Our goal is to find an upper bound on the value of  $r_n(T)$ .

Combined with conditions  $r_k(T) \geq 0 \quad \forall k \in \{1, \dots, n\}$  and  $\sum_{k=1}^n r_k(T) \leq 1$ , we get that

$$\sum_{k=1}^n r_k(T) = a \cdot n + d \cdot \frac{n(n-1)}{2} \leq 1$$

The maximum value of  $a$ , given a fixed  $d$ , is

$$a_{\max} = \frac{1 - d \cdot \frac{n(n-1)}{2}}{n}$$

and the condition  $a_{\max} \geq a = r_1(T) \geq 0$  is equivalent to  $d \leq \frac{2}{n(n-1)}$ . Plugging in

$$\begin{aligned} r_n(T) &\leq a_{\max} + (n-1)d \\ &= \frac{1}{n} - d \cdot \frac{n-1}{2} + (n-1)d \\ &= \frac{1}{n} + d \cdot \frac{n-1}{2} \end{aligned}$$

Now we maximize by plugging the maximum value of  $d$

$$\begin{aligned} r_n(T) &\leq \frac{1}{n} + d_{\max} \cdot \frac{n-1}{2} \\ &= \frac{1}{n} + \frac{2}{n(n-1)} \cdot \frac{(n-1)}{2} \\ &= \frac{2}{n} \end{aligned}$$

in contradiction to our assumption. □

For coalitions of size 2, we know the following upper bound on  $\gamma$ .

**Proposition 2.15.**

There exist no 2-SNM tournament rules that have Condorcet consistency value  $\gamma > \frac{2}{3}$ .

*Proof.* Assume for contradiction a tournament rule  $r$  that satisfies the requirements, and consider a tournament  $T$  of arbitrary size  $n$  with a top cycle of three elements  $\{a, b, c\}$ . Any pair of these elements may manipulate to make one of them a Condorcet winner, gaining at least  $\gamma$ . Hence,

$$\begin{aligned} r_a(T) + r_b(T) &> \frac{2}{3} \\ r_a(T) + r_c(T) &> \frac{2}{3} \\ r_b(T) + r_c(T) &> \frac{2}{3} \end{aligned}$$

and therefore, by summing

$$\begin{aligned} 2[r_a(T) + r_b(T) + r_c(T)] &> 3 \cdot \frac{2}{3} = 2 \\ r_a(T) + r_b(T) + r_c(T) &> 1 \end{aligned}$$

which is a contradiction. □

### 3. Approximate Non-Manipulability

Motivated by match-based applications such as sporting events, where the outcome of pair-wise matches is explicitly learned and used to select a winner, we consider instead the design of tournament rules that are exactly Condorcet-consistent, but only approximately 2-SNM.

The focus of this chapter is to quantify how much manipulable certain tournament formats are (i.e. how much can teams possibly gain by fixing matches), the idea being that it is easier to discourage manipulation in tournaments that are less manipulable.

**Definition 3.1.** (2-Strong Non-Manipulability at probability  $\alpha$ )

A tournament rule  $r$  is *2-strongly non-manipulable at probability  $\alpha$*  (2-SNM- $\alpha$ ) if, for all  $i$  and  $j$  and pairs of  $\{i, j\}$ -adjacent tournaments  $T$  and  $T'$ ,

$$r_i(T') + r_j(T') - r_i(T) - r_j(T) \leq \alpha.$$

**Definition 3.2.** ( $k$ -Strong Non-Manipulability at probability  $\alpha$ )

A tournament rule  $r$  is  *$k$ -strongly non-manipulable at probability  $\alpha$*  ( $k$ -SNM- $\alpha$ ) if, for all subsets  $S$  of players of size at most  $k$ , for all pairs of  $S$ -adjacent tournaments  $T$  and  $T'$ ,

$$\sum_{i \in S} r_i(T') - \sum_{i \in S} r_i(T) \leq \alpha.$$

#### Lower bounds

**Theorem 3.3.**

There is no Condorcet-consistent tournament rule on  $n$  players (for  $n \geq 3$ ) that is 2-SNM- $\alpha$  for  $\alpha < \frac{1}{3}$ .

*Proof.* Consider the tournament  $T$  on three players  $A$ ,  $B$ , and  $C$  where  $A$  beats  $B$ ,  $B$  beats  $C$ , and  $C$  beats  $A$ . Note that, while this tournament has no Condorcet winner, changing the result of any of the three games results in a Condorcet winner. For example, if  $A$  bribes  $C$  to lose to  $A$ , then  $A$  becomes the Condorcet winner.

If we have a tournament rule  $r$  that is 2-SNM- $\alpha$ , then combining this with the above fact gives rise to the following three inequalities.

$$\begin{aligned} r_A(T) + r_B(T) &\geq 1 - \alpha \\ r_B(T) + r_C(T) &\geq 1 - \alpha \\ r_C(T) + r_A(T) &\geq 1 - \alpha \end{aligned}$$

Together these imply  $r_A(T) + r_B(T) + r_C(T) \geq \frac{3}{2}(1 - \alpha)$ .

But  $r_A(T) + r_B(T) + r_C(T) = 1$ ; it follows that  $\alpha \geq \frac{1}{3}$ , as desired.

We can extend this counterexample to  $n > 3$  players by introducing  $n - 3$  dummy players who all lose to  $A$ ,  $B$ , and  $C$ ; the argument above continues to hold, since

$$r_A(T) + r_B(T) + r_C(T) \leq 1.$$

□

**Theorem 3.4.**

There is no Condorcet-consistent tournament rule on  $n$  players (for  $n \geq 2k - 1$ ) that is  $k$ -SNM- $\alpha$  for  $\alpha < \frac{k-1}{2k-1}$ .

*Proof.* Consider the following tournament  $T$  on the  $2k - 1$  players labelled 1 through  $2k - 1$ . Each player  $i$  wins their match versus the  $k - 1$  players  $i + 1, i + 2, \dots, i + (k - 1)$ , and loses their match versus the  $k - 1$  players  $i - 1, i - 2, \dots, i - (k - 1)$  (indices taken modulo  $2k - 1$ ). Note that the coalition of players  $S_i = \{i, i - 1, \dots, i - (k - 1)\}$  of size  $k$  can cause  $i$  to become a Condorcet winner if all players in the coalition agree to lose their games with  $i$ .

If we have a tournament rule  $r$  that is  $k$ -SNM- $\alpha$ , then this implies the following  $2k - 1$  inequalities (one for each  $i \in [2k - 1]$ ):

$$\sum_{j \in S_i} r_j(T) \geq 1 - \alpha$$

Summing these  $2k - 1$  inequalities, we obtain

$$k \sum_{j=1}^{2k-1} r_j(T) \geq (2k - 1)(1 - \alpha)$$

Since  $\sum_{j=1}^{2k-1} r_j(T) \leq 1$ , this implies that  $\alpha \geq \frac{k-1}{2k-1}$ , as desired.

Again, it is possible to extend this example to any number of players  $n \geq 2k - 1$  by introducing dummy players who lose to all  $2k - 1$  of the above players. □

## The Random Single-Elimination Bracket Rule

**Definition 3.5.** (Single-Elimination Bracket)

A *single-elimination bracket* (or *bracket*, for short)  $B$  on  $n = 2^h$  players is a complete binary tree of height  $h$  whose leaves are labelled with some permutation of the  $n$  players. The outcome of a bracket  $B$  under a tournament  $T$  is the labelling of internal nodes of  $B$  where each node is labelled by the winner of its two children under  $T$ . The winner of  $B$  under  $T$  is the label of the root of  $B$  under this labelling.

**Definition 3.6.** (Random Single-Elimination Bracket Rule)

The *random single-elimination bracket rule*  $r$  is a tournament rule on  $n = 2^h$  players where  $r_i(T)$  is the probability player  $i$  is the winner of  $B$  under  $T$  when  $B$  is chosen uniformly at random from the set of  $n!$  possible brackets.

If  $n$  is not a power of 2, we define the random single-elimination bracket rule on  $n$  players by introducing  $2^{\lceil \log_2 n \rceil} - n$  dummy players who lose to all of the existing  $n$  players.

It is straightforward to check that the random single-elimination bracket rule is both Condorcet-consistent and monotone. It can be proved that in addition to these properties, it is also 2-SNM-1/3 (which is the best possible, by [Theorem 3.3](#)).

**Theorem 3.7.**

The random single elimination bracket rule is 2-SNM-1/3.

*Proof.* Omitted. □

## Other formats

Finally, there are many other tournament formats that are either used in practice or have been previously studied. In this section we show that many of these formats are more susceptible to manipulation than the random single elimination bracket rule; in particular, all of the following formats are at best 2-SNM-1/2.

By far the most common tournament rule for round robin tournaments is some variant of a ‘scoring’ rule, where the winner is the player who has won the most games (with ties broken in some fashion if multiple players have won the same maximum number of games). In voting theory, this rule is often called Copeland’s rule, or Copeland’s method.

**Definition 3.8.** (Copeland Rule)

A tournament rule  $r$  is a *Copeland rule* if the winner is always selected from the set of players with the maximum number of wins.

We begin by showing that no Copeland rule can be 2-SNM- $\alpha$  for any  $\alpha < 1$  (regardless of how the rule breaks ties).

**Theorem 3.9.**

There is no Copeland rule on  $n$  players that is 2-SNM- $\alpha$  for  $\alpha < 1 - \frac{2}{n-1}$ .

*Proof.* Assume to begin that  $n = 2k + 1$  is odd, and let  $r$  be a Copeland rule on  $n$  players.

Let  $T$  be the tournament where each player  $i$  beats the  $k$  players  $\{i + 1, i + 2, \dots, i + k\}$  but loses to the  $k$  players  $\{i - 1, i - 2, \dots, i - k\}$ , with indices taken modulo  $n$  (similar to the tournament in the proof of [Theorem 3.4](#)).

Since  $\sum_{i=1}^n r_i(T) = 1$ , there must be some  $i$  such that  $r_{i-1}(T) + r_i(T) \leq \frac{2}{n}$ . On the other hand, if player  $i - 1$  throws their match to player  $i$ , then player  $i$  becomes the unique Copeland winner (winning  $k + 1$  games) and  $r_i(T') = 1$ .

It follows that, for such a rule, if  $r$  is 2-SNM- $\alpha$ , then  $\alpha \geq 1 - \frac{2}{n}$ .

If  $n$  is even, then we can embed the above example for  $n - 1$  by assigning one player to be a dummy player that loses to all teams. This immediately implies  $\alpha \geq 1 - \frac{2}{n-1}$  in this case.  $\square$

Other examples of tournament rules that are Condorcet-consistent and monotone are:

the top cycle rule, the iterative Condorcet rule, and the randomized voting caterpillar rule.

The following tournament will provide lower bounds on  $\alpha$  against all of them.

**Definition 3.10.** (Superman-Kryptonite Tournament)

The *superman-kryptonite* tournament on  $n$  players has  $i$  beat  $j$  whenever  $i < j$ , except that player  $n$  beats player 1. That is, player 1 beats everyone except for player  $n$ , who loses to everyone except for player 1.

**Definition 3.11.** (Top Cycle Rule)

The *top cycle* of a tournament  $T$  is the minimal set of players who beat all other players.

The *top cycle rule* is a tournament rule which assigns the winner to be a uniformly random element of this set.

**Theorem 3.12.**

The top cycle rule on  $n$  players is not 2-SNM- $\alpha$  for any  $\alpha < 1 - \frac{2}{n}$ .

*Proof.* Let  $T$  be the superman-kryptonite tournament on  $n$  players.

The top cycle in  $T$  contains all the players, so  $r_1(T) + r_n(T) = \frac{1}{n} + \frac{1}{n} = \frac{2}{n}$ .

However, if player  $n$  throws their match to player 1, player 1 becomes a Condorcet winner and  $r_1(T') = 1$ . It follows that  $\alpha \geq 1 - \frac{2}{n}$ .  $\square$

**Definition 3.13.** (Iterative Condorcet rule)

The *iterative Condorcet rule* is a tournament rule that uniformly removes players at random until there is a Condorcet winner, and then assigns that player to be the winner.

**Theorem 3.14.**

The iterative Condorcet rule on  $n$  players is not 2-SNM- $\alpha$  for any  $\alpha < \frac{1}{2} - \frac{1}{n(n-1)}$ .

*Proof.* Let  $T$  be the superman-kryptonite tournament on  $n$  players.

Note that no Condorcet winner will appear until either player 1 is removed, player  $n$  is removed, or all other  $n - 2$  players are removed.

If all the other  $n - 2$  players are removed before players 1 or  $n$  (which occurs with probability  $\frac{2 \cdot (n-2)!}{n!} = \frac{2}{n(n-1)}$ ), then player  $n$  wins.

If this does not happen and player  $n$  is removed before player 1 (which occurs with

probability  $\frac{1}{2} \left(1 - \frac{2}{n(n-1)}\right) = \frac{1}{2} - \frac{1}{n(n-1)}$ , then player 1 becomes the Condorcet winner and wins.

Otherwise, player 1 will be removed before player  $n$ , while some players in 2 through  $n-1$  remain, and one of them will become the Condorcet winner (the remaining player in  $\{2, \dots, n-1\}$  with lowest index).

It follows that  $r_1(T) = \frac{1}{2} - \frac{1}{n(n-1)}$  and  $r_2(T) = \frac{2}{n(n-1)}$ , so  $r_1(T) + r_2(T) = \frac{1}{2} + \frac{1}{n(n-1)}$ .

On the other hand, if player  $n$  throws their match to player 1, then again player 1 becomes a Condorcet winner and  $r_1(T') = 1$ . It follows that  $\alpha \geq \frac{1}{2} - \frac{1}{n(n-1)}$ .  $\square$

**Definition 3.15.** (Randomized Voting Caterpillar Rule)

The *randomized voting caterpillar rule* is a tournament rule which chooses a winner as follows. Choose a random permutation  $\pi$  of  $[n]$ . Start by matching  $\pi(1)$  and  $\pi(2)$ , and choose a winner according to  $T$ . Then for all  $i \geq 3$ , match  $\pi(i)$  with the winner of the most recent match. The player that wins the last match (against  $\pi(n)$ ) is declared the winner.

**Theorem 3.16.**

The randomized voting caterpillar rule on  $n$  players is not 2-SNM- $\alpha$  for any  $\alpha < \frac{1}{2} - \frac{n-2}{n(n-1)}$ .

*Proof.* Let  $T$  be the superman-kryptonite tournament on  $n$  players.

The only way player 1 loses is if either player  $n$  occurs later in  $\pi$  than player 1 (which happens with probability  $\frac{1}{2}$ ) or if  $\pi(n) = 1$  and  $\pi(1) = 2$  and they play in the first round (which happens with probability  $\frac{1}{n(n-1)}$ ).

The only way player  $n$  can win is if  $\pi(n) = n$  (i.e., they only play the very last game), in which case they will play player 1 and win (this happens with probability  $\frac{1}{n}$ ).

It follows that  $r_1(T) = \frac{1}{2} - \frac{1}{n(n-1)}$  and  $r_n(T) = \frac{1}{n}$ , so  $r_1(T) + r_n(T) = \frac{1}{2} + \frac{n-2}{n(n-1)}$ .

On the other hand, if player  $n$  throws their match to player 1, then again player 1 becomes a Condorcet winner and  $r_1(T') = 1$ . It follows that  $\alpha \geq \frac{1}{2} - \frac{n-2}{n(n-1)}$ .  $\square$

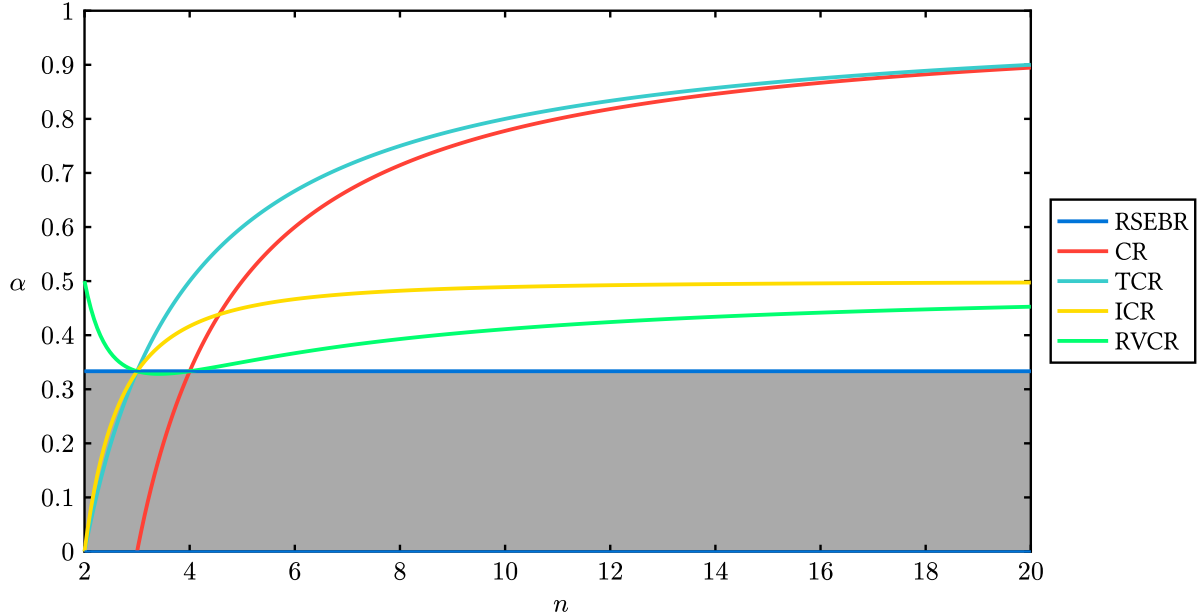


Figure 2: Lower bounds for  $\alpha$  for some 2-SNM- $\alpha$  rules (remember the universal  $1/3$  lower bound)

## 4. Partially Transferable Utility

We recall the previous definitions about manipulations and then move to a more general setting.

### Definition 4.1. ( $S$ -adjacent)

Tournaments  $T, T' \in \mathcal{T}_n$  are  $S$ -adjacent where  $S \subseteq [n]$  if  $i \succ_T j \Leftrightarrow i \succ_{T'} j$  for  $i \neq j \in [n] \setminus S$ . In other words,  $T$  and  $T'$  are  $S$ -adjacent if they coincide on every match except possibly those between agents in  $S$ .

When utilities are non-transferable, two agents are willing to collude only if one of them can strictly improve her probability of winning at no cost to the other. Formally, distinct agents  $i, j \in [n]$  collude from tournament  $T$  to tournament  $T'$  only if both

$$\begin{aligned} \max\{r_i(T') - r_i(T), r_j(T') - r_j(T)\} &> 0 \text{ and} \\ \min\{r_i(T') - r_i(T), r_j(T') - r_j(T)\} &\geq 0. \end{aligned}$$

Thus, to incentivize agents against such manipulations, a tournament rule must satisfy the following notion of non-manipulability.

### Definition 4.2. (2-Pareto Non-Manipulability)

A tournament rule  $r$  is *2-Pareto non-manipulable* (2-PNM) if for all  $i \neq j \in [n]$  and  $\{i, j\}$ -adjacent tournaments  $T \neq T' \in \mathcal{T}_n$ , either

- $\min\{r_i(T') - r_i(T), r_j(T') - r_j(T)\} < 0$  or
- $\max\{r_i(T') - r_i(T), r_j(T') - r_j(T)\} \leq 0$ .

There exists a rule that is monotone, TCC, and 2-PNM (see [Proposition 2.10](#)).

The barrier to pairwise manipulation is much lower when utilities are completely transferable since two agents only care about the probability that at least one of them wins the tournament. In other words,  $i$  and  $j$  collude from  $T$  to  $T'$  only if  $r_i(T') + r_j(T') > r_i(T) + r_j(T)$ .

Under this utility model, an agent may be willing to sacrifice and shift a significant portion of her probability to her partner in crime. Thus, tournament rules must satisfy a stronger notion of non-manipulability in this setting.

### Definition 4.3. (2-Strong Non-Manipulability)

A tournament rule  $r$  is *2-strongly non-manipulable* (2-SNM) if  $r_i(T') + r_j(T') \leq r_i(T) + r_j(T)$  for all  $i \neq j \in [n]$  and  $\{i, j\}$ -adjacent tournaments  $T \neq T' \in \mathcal{T}_n$ .

**Remark.** By symmetry, we need to have equality:  $r_i(T') + r_j(T') = r_i(T) + r_j(T)$ .

We have seen that no Condorcet consistent tournament rule is 2-SNM (see [Proposition 2.11](#)).

However, despite this strong impossibility result, instances of collusion are relatively infrequent in the real world, suggesting that settings in which utilities are completely transferable are uncommon. On the other hand, instances of collusion are not unheard of, suggesting that utility is neither always non-transferable.

We consider a third utility model in which utilities are partially transferable: distinct agents  $i$  and  $j$  collude from tournament  $T$  to tournament  $T'$  only if

$$r_i(T') + r_j(T') > r_i(T) + r_j(T) + \lambda \max\{r_i(T) - r_i(T'), r_j(T) - r_j(T')\}.$$

In this model, two agents always collude if both of them improve their chances of winning and never collude if both of their chances decrease. The interesting case is when one agent improves her chances at the expense of the other. One interpretation of this necessary condition is that agents would rather win the tournament themselves but are still willing to collude if the gain in probability

is significantly larger than each agent's loss.

Here,  $\lambda$  is a parameter that measures how transferable utility is. Note that when  $\lambda$  is low, utilities are more transferable. We will later see how  $\lambda$  can be interpreted as agents' level of selfishness.

We now define a notion of non-manipulability for this model.

**Definition 4.4.** (2-Non-Manipulability for  $\lambda$ )

A tournament rule  $r$  is *2-non-manipulable* for  $\lambda \geq 0$  (2-NM $_{\lambda}$ ) if

$$r_i(T') + r_j(T') \leq r_i(T) + r_j(T) + \lambda \max\{r_i(T) - r_i(T'), r_j(T) - r_j(T')\}$$

for all  $i \neq j \in [n]$  and  $\{i, j\}$ -adjacent tournaments  $T \neq T' \in \mathcal{T}_n$ .

We say  $r$  is 2-NM $_{\infty}$  if

$$r_i(T') + r_j(T') \leq r_i(T) + r_j(T) + \lim_{\lambda \rightarrow \infty} \lambda \max\{r_i(T) - r_i(T'), r_j(T) - r_j(T')\}$$

(where the limit is taken in the extended reals).

Observe that when  $\lambda = 0$ , our notion of non-manipulability coincides with strong non-manipulability. Moreover, we show that our notion coincides with Pareto non-manipulability when  $\lambda = +\infty$ .

We remark that we do not interpret 2-NM $_{\lambda}$  as an approximation to 2-SNM. Unlike approximation algorithms, a tournament designer who finds herself faced with e.g., agents who value their opponents chances of winning as much as their own (completely transferable utility) may not find it in her best interest to use a 2-NM $_{\lambda}$  tournament rule for some  $\lambda > 1$ . Rather,  $\lambda$  is meant to model the behavior of the agents.

**Proposition 4.5.**

A tournament rule is 2-PNM if and only if it is 2-NM $_{\infty}$ .

*Proof.* Suppose tournament rule  $r$  is not 2-PNM. Then, there exist distinct agents  $i, j \in [n]$  and a pair of  $\{i, j\}$ -adjacent tournaments  $T \neq T' \in \mathcal{T}_n$  such that WLOG  $r_i(T') - r_i(T) > 0$  and  $r_j(T') - r_j(T) \geq 0$ . Thus,

$$r_i(T') + r_j(T') - r_i(T) - r_j(T) > 0 \geq \lim_{\lambda \rightarrow \infty} \lambda \max\{r_i(T) - r_i(T'), r_j(T) - r_j(T')\}$$

so  $r$  is not 2-NM $_{\infty}$ .

Conversely, take  $r$  to be 2-PNM, so for all distinct agents  $i, j \in [n]$  and  $\{i, j\}$ -adjacent  $T \neq T' \in \mathcal{T}_n$ , either

(1) WLOG  $r_i(T') - r_i(T) < 0$ , in which case

$$r_i(T') + r_j(T') - r_i(T) - r_j(T) < +\infty = \lim_{\lambda \rightarrow \infty} \lambda \max\{r_i(T) - r_i(T'), r_j(T) - r_j(T')\}$$

(2) or  $\max\{r_i(T') - r_i(T), r_j(T') - r_j(T)\} \leq 0$ , in which case

$$r_i(T') + r_j(T') - r_i(T) - r_j(T) \leq 0 \leq \lim_{\lambda \rightarrow \infty} \lambda \max\{r_i(T) - r_i(T'), r_j(T) - r_j(T')\}$$

Thus,  $r$  is 2-NM $_{\infty}$ . □

We are interested in approximately non-manipulable tournament rules; that is, rules under which no two agents can collude to gain in joint probability more than  $\alpha$  more than each agent's loss (weighted by  $\lambda$ ).

We will see later that there is a range of  $\lambda$  for which fair and non-manipulable tournament rules do not exist. For  $\lambda$  in this range, it may be better to design approximately non-manipulable tournament rules tailored to  $\lambda$  than use a 2-NM $_{\lambda'}$  tournament rule for some  $\lambda' > \lambda$ .



**Definition 4.6.** (2-Non-Manipulability up to  $\alpha$  for  $\lambda$ )

A tournament rule  $r$  is *2-non-manipulable up to  $\alpha$*  for  $\lambda \geq 0$  ( $2\text{-NM}_\lambda\text{-}\alpha$ ) if

$$r_i(T') + r_j(T') \leq r_i(T) + r_j(T) + \lambda \max\{r_i(T) - r_i(T'), r_j(T) - r_j(T')\} + \alpha$$

for all  $i \neq j \in [n]$  and  $\{i, j\}$ -adjacent tournaments  $T \neq T' \in \mathcal{T}_n$ .

In addition to being robust against pairwise manipulations, a tournament rule should be robust to the intentional throwing of matches.

**Definition 4.7.** (Monotonicity)

A tournament rule is *monotone* if  $r_i(T) \geq r_i(T')$  for all  $i \neq j \in [n]$  and  $\{i, j\}$ -adjacent tournaments  $T \neq T' \in \mathcal{T}_n$  such that  $i \succ_T j$ .

Intuitively, monotonicity says that no agent should be able to improve her chances of winning by deliberately losing one of her matches. Thus, agents have an incentive to win each of their matches under monotone rules. Violations of this property should be seen as quite severe.

**Proposition 4.8.**

Let  $r$  be a  $2\text{-NM}_\lambda$  tournament rule for some  $\lambda > 0$ , then the following two statements are equivalent.

1.  $r$  is monotone
2. For all  $i \neq j \in [n]$  and  $\{i, j\}$ -adjacent tournaments  $T \neq T' \in \mathcal{T}_n$  such that  $i \prec_T j$

$$r_i(T') - r_i(T) \leq (\lambda + 1)(r_j(T) - r_j(T'))$$

*Proof.* Let  $T \neq T' \in \mathcal{T}_n$  be  $\{i, j\}$ -adjacent tournaments such that  $i \prec_T j$  (and thus  $i \succ_{T'} j$ ), and suppose  $r$  is monotone.

Then, by monotonicity,  $r_i(T') \geq r_i(T)$  and  $r_j(T) \geq r_j(T')$ ; in other words  $r_i(T) - r_i(T') \leq 0 \leq r_j(T) - r_j(T')$ , so

$$\begin{aligned} r_i(T') - r_i(T) &\leq r_j(T) - r_j(T') + \lambda \max\{r_i(T) - r_i(T'), r_j(T) - r_j(T')\} \\ &= (\lambda + 1)(r_j(T) - r_j(T')) \end{aligned}$$

where the inequality follows from the fact that  $r$  is  $2\text{-NM}_\lambda$ .

Now, suppose the second statement holds. Applying the second statement twice yields

$$\begin{aligned} r_i(T') - r_i(T) &\leq (\lambda + 1)(r_j(T) - r_j(T')) \\ r_j(T) - r_j(T') &\leq (\lambda + 1)(r_i(T') - r_i(T)) \end{aligned}$$

Multiplying the second inequality by  $\lambda + 1$ , adding the result to the first inequality, and simplifying yields

$$\lambda(\lambda + 2)(r_i(T') - r_i(T)) \geq 0$$

Since  $r$  is  $2\text{-NM}_\lambda$  for  $\lambda > 0$ , this inequality implies  $r_i(T') \geq r_i(T)$ . □

This proposition offers a natural interpretation of the parameter  $\lambda$  and the  $2\text{-NM}_\lambda$  property for monotone tournament rules:  $\lambda$  is how much each agent weighs her own probability of winning over others' probabilities of winning and, if a tournament rule is  $2\text{-NM}_\lambda$ , then switching the outcome of a match does not increase the probability of winning for the new winner by more than a  $\lambda + 1$  factor over the loss of the new loser. Note that the proposition does not hold for  $\lambda = 0$ . Indeed, monotonicity and  $2\text{-SNM}$  are independent properties: neither implies the other.

## Lower Bounds

We have seen that no Condorcet consistent tournament rule is  $2\text{-NM}_0\text{-}\alpha$  for  $\alpha < 1/3$  (see [Theorem 3.3](#)). The same lower bound construction yields:

### Theorem 4.9.

No Condorcet consistent tournament rule is  $2\text{-NM}_\lambda\text{-}\alpha$  for  $\lambda < 1 - 3\alpha$ .

*Proof.* Suppose tournament rule  $r$  is CC and  $2\text{-NM}_\lambda\text{-}\alpha$ , and consider any tournament  $T$  on  $[n]$  in which 1 dominates 2, 2 dominates 3, and 3 in turn dominates 1; all the others lose to 1, 2, 3. Note that any two agents among  $\{1, 2, 3\}$  can collude so that one of them becomes undefeated.

Consider the  $\{1, 2\}$ -adjacent tournament  $T'$  where  $1 \prec_{T'} 2$ . Since 2 is the Condorcet winner, by CC we have  $r_2(T') = 1$  and  $r_1(T') = 0$ . By  $2\text{-NM}_\lambda\text{-}\alpha$  we have

$$r_1(T') + r_2(T') - r_1(T) - r_2(T) \leq \lambda \max\{r_1(T) - r_1(T'), r_2(T) - r_2(T')\} + \alpha$$

But  $r_1(T) - r_1(T') = r_1(T) \geq 0$  and  $r_2(T) - r_2(T') = r_2(T) - 1 \leq 0$ , so the condition becomes

$$1 - r_1(T) - r_2(T) \leq \lambda r_1(T) + \alpha$$

Analogously for the other manipulations,

$$1 - r_2(T) \leq (\lambda + 1)r_1(T) + \alpha$$

$$1 - r_3(T) \leq (\lambda + 1)r_2(T) + \alpha$$

$$1 - r_1(T) \leq (\lambda + 1)r_3(T) + \alpha$$

Adding these three inequalities together and isolating  $\lambda$  yields

$$3 - [r_1(T) + r_2(T) + r_3(T)] \leq (\lambda + 1)[r_1(T) + r_2(T) + r_3(T)] + 3\alpha$$

$$3 - 3\alpha \leq (\lambda + 2)[r_1(T) + r_2(T) + r_3(T)]$$

$$\frac{3(1 - \alpha)}{r_1(T) + r_2(T) + r_3(T)} - 2 \leq \lambda$$

Since  $r_1(T) + r_2(T) + r_3(T) \leq 1$ , this inequality implies that  $\lambda \geq 1 - 3\alpha$ . □

### Corollary 4.10.

No Condorcet consistent tournament rule is  $2\text{-NM}_\lambda$  for  $\lambda < 1$ .

**Conjecture.** There exists a tournament rule that is monotone, Condorcet consistent, and  $2\text{-NM}_1$ .

## Tournament rules

We now consider several tournament rules and examine their trade-offs between  $\alpha$  and  $\lambda$ .

Interestingly, the superman-kryptonite tournament (and its variants) is responsible for all our lower bounds, suggesting that it is especially problematic.

### Definition 4.11. (Superman-kryptonite tournament)

The *superman kryptonite tournament* on  $[n]$  has  $i \succ_T j$  whenever  $i < j$ , except  $n \succ_T 1$ . In particular, superman 1 dominates all agents but kryptonite  $n$ , and  $n$  is dominated by all agents except 1.

**Theorem 4.12.**

Assume the following rules to satisfy  $2\text{-NM}_\lambda\text{-}\alpha$  for some  $\lambda \geq 0$ .

- RSEB, then  $\alpha \geq \Omega(1/n)$ .
- RKotH or PR, then  $\alpha \geq \Omega(1)$ .

That is, RSEB, RKotH and PR are always pairwise manipulable regardless of  $\lambda$ .

*Proof.* **To do.** □

**Theorem 4.13.**

Assume the following rules to satisfy  $2\text{-NM}_\lambda\text{-}\alpha$  for some  $\lambda \geq 0$ .

- ICR, then  $\lambda \geq (1 - O(\alpha))\Omega(n^2)$ .
- RDM, then  $\lambda \geq (1 - O(n\alpha))\Omega(n)$ .
- RVC, TCR or PRSL, then  $\lambda \geq (1 - O(\alpha))\Omega(n)$ .

*Proof.* **To do.** □

Table 1: Performance summary

Rule	Monotone?	Fairness	$2\text{-NM}_\lambda\text{-}\alpha$	$2\text{-NM}_\lambda\text{-}f(n)$
ICR [1]	Yes [1]	DSTC	$\lambda \geq (\frac{1}{2} - \alpha) \Omega(n^2)$	$f(n) \geq 1/2$
RVC [1]	Yes [1]	DSTC	$\lambda \geq (\frac{1}{2} - \alpha) \Omega(n)$	$f(n) \geq 1/2$
TCR [1]	Yes [1]	DSTC	$\lambda \geq (1 - \alpha)\Omega(n)$	$f(n) \geq 1$
RSEB [9]	Yes [9]	TCC	$\alpha \geq 1/n$	$f(n) \geq \varepsilon(\lambda) > 0$
RKotH [10]	Yes [10]	CovC [10], DSTC	$\alpha \geq 1/10$	$f(n) \geq 1/10$
RDM [6]	Yes	DSTC	$\lambda \geq (1 - \frac{n\alpha}{2}) \Omega(n)$	$f(n) \geq \frac{\lambda-1}{\lambda(2\lambda+1)}$
PR [3]	?	DSTC	$\alpha \geq 1/13$	$f(n) \geq 1/13$
PRSL [3]	?	DSTC	$\lambda \geq (1 - \alpha)\Omega(n)$	$f(n) \geq 1$

2

<sup>2</sup>This table is from an alternate version of [4] available at:  
<https://comsoc-community.org/archive/comsoc-2023/comsoc/papers/paper59.pdf>

## 5. Beyond Condorcet

⚠ **Warning.** This is based on original research.

At this point it may be clear that Condorcet consistency is a bit problematic. In some ways it is too strong, in that it automatically selects the Condorcet winner even if he may have won all matches by small margins; in others it is too weak, like in the superman-kryptonite example where we may have a clear winner but not being a Condorcet winner prevents it to be chosen.

We need a framework which takes in account not only the matches won, but also by how much.

**Definition 5.1.** (Weighted Tournament)

A *weighted tournament* is a tournament with pairs of weights for each match: the match between alternatives  $i$  and  $j$  has weights  $\{w_{ij}, w_{ji}\}$  such that  $w_{ij}, w_{ji} \in [0, 1]$  and

$w_{ij} + w_{ji} = 1$  for every  $i \neq j \in [n]$ .

It is represented as a complete directed weighted graph, with two edges for each pair of nodes which satisfies the above condition.

Here the weights could represent the number of points scored in the match or the number of repeated matches won, followed by a normalization.

The definitions of tournament rules and manipulations still apply, but we want to replace Condorcet consistency. Instead of establishing a new fairness criterion (axiomatic way), we present a rule that in some sense is intuitively fair. The idea is that the probability of winning should be proportional to the performance in the tournament (which in our model depends on the weights of the edges).

We use as a proxy of performance an aggregation of the weights, which we can interpret as the number of total points/wins in the tournament.

**Definition 5.2.** (Score)

The *score*  $s_i$  of alternative  $i$  is the outdegree of the node in the tournament graph,

that is  $s_i = \sum_{j=1}^n w_{ij}$ .

Now we need to turn these in a probability distribution over the alternatives. We can do this thanks to the softmax function.

**Definition 5.3.** (Proportional Score Rule)

The *proportional score rule* assigns to alternative  $i$  probability of winning  $r_i = e^{s_i} / \sum_{j=1}^n e^{s_j}$ .

The properties of the softmax guarantee that this is well posed, in particular  $r_i \in [0, 1]$  and  $\sum_{i=1}^n r_i = 1$ .

### Fairness

Consider the tournament on  $n$  agents where player 1 is a Condorcet winner with weight  $0.5 < a \leq 1$  and all the other players have weight 0.5 among them. Then

$$\begin{aligned} s_1 &= a \cdot (n - 1) \\ s_i &= \frac{1}{2} \cdot (n - 2) + (1 - a) \quad \forall i \neq 1 \end{aligned}$$

For small  $a$  (close to 0.5) the score of player 1 is close to that of the other players, but increasing  $a$  the gap grows; this is consistent with the idea that being a Condorcet winner is not enough to win the tournament, but we need big margins to have high probability of winning.

Moreover increasing the number of players, further increase the gap; indeed, winning more matches against many players by a big margin should give higher probability of winning the tournament.

In the limit for  $n \rightarrow \infty$  we get that  $s_1 = 1$ , that is we recover Condorcet consistency.

Now consider the superman-kryptonite tournament: player 1 wins with weight  $a$  against all but player  $n$ , to which he loses with weight  $c$ ; the other players beat each other with weight 0.5 and player  $n$  with weight  $d$ . Then

$$\begin{aligned} s_1 &= a \cdot (n-2) + (1-c) \\ s_n &= c + (1-d) \cdot (n-2) \\ s_i &= \frac{1}{2} \cdot (n-3) + (1-a) + d, \quad \forall i \neq 1, n \end{aligned}$$

As before we can observe a behaviour consistent with our intuition: player 1 (superman) can get a high score (and so high probability of winning the tournament) despite the existence of player  $n$  (kryptonite), if the parameter  $a$  is big enough compared to the others, and especially if there are many players.

### Non-manipulability

Now we want to estimate how much manipulable is this tournament rule.

In this framework a manipulation is a redistribution of the weights among the members of the coalition such that we still have a valid weighted tournament.

#### Theorem 5.4.

The proportional score rule is 2-SNM-tanh( $\frac{1}{4}$ ).

*Proof.* Consider the coalition  $\{1, 2\}$  and the manipulation that changes the pair  $(w_{12}, w_{21})$  from  $(\omega, 1-\omega)$  to  $(\omega', 1-\omega')$ , with  $\omega \neq \omega' \in [0, 1]$ .

We want an upper bound (as tight as possible) on  $r'_1 + r'_2 - r_1 - r_2$ .

Let  $\Delta := \omega' - \omega$ ; WLOG we can assume  $\Delta > 0$  (otherwise swap the players). Notice that

$$s'_1 = s_1 + \Delta, \quad s'_2 = s_2 - \Delta$$

and all the others remain unchanged:  $s'_i = s_i, \forall i \neq 1, 2$ .

Let  $A := e^{s_1}, \quad B := e^{s_2}, \quad C := \sum_{i \geq 3} e^{s_i}$  so

$$R' := r'_1 + r'_2 = \frac{e^\Delta A + e^{-\Delta} B}{e^\Delta A + e^{-\Delta} B + C}, \quad R := r_1 + r_2 = \frac{A + B}{A + B + C}$$

We perform a change of coordinates:

$$u := \frac{A}{C} \geq 0, \quad v := \frac{B}{C} \geq 0, \quad t := e^\Delta > 1$$

so that we can write

$$R' - R = \frac{tu + t^{-1}v}{tu + t^{-1}v + 1} - \frac{u + v}{u + v + 1} \leq \tanh\left(\frac{\ln t}{4}\right) = \tanh\left(\frac{\Delta}{4}\right)$$

where we used [Lemma 5.5](#). Since  $\Delta \leq 1$ , we get the result.  $\square$

#### Lemma 5.5.

Let  $t > 1$  and for  $u, v \geq 0$

$$g(u, v) = \frac{tu + t^{-1}v}{tu + t^{-1}v + 1} - \frac{u + v}{u + v + 1}$$

Then  $\sup_{u, v \geq 0} g(u, v) = \tanh\left(\frac{\ln t}{4}\right)$ .

*Proof.* Let  $D' := tu + t^{-1}v + 1$  and  $D := u + v + 1$  so that

$$g = \frac{tu + t^{-1}v}{D'} - \frac{u + v}{D}$$

We compute the derivatives.

$$\begin{aligned}\frac{\partial g}{\partial u} &= \frac{t \cdot D' - (tu + t^{-1}v) \cdot t}{D'^2} - \frac{1 \cdot D - (u + v) \cdot 1}{D^2} = \frac{t}{D'^2} - \frac{1}{D^2} \\ \frac{\partial g}{\partial v} &= \frac{t^{-1} \cdot D' - (tu + t^{-1}v) \cdot t^{-1}}{D'^2} - \frac{1 \cdot D - (u + v) \cdot 1}{D^2} = \frac{t^{-1}}{D'^2} - \frac{1}{D^2}\end{aligned}$$

A stationary point must satisfy

$$\begin{aligned}\frac{t}{D'^2} &= \frac{1}{D^2} \quad \text{and} \quad \frac{t^{-1}}{D'^2} = \frac{1}{D^2} \\ tD^2 &= D'^2 \quad \text{and} \quad t^{-1}D^2 = D'^2\end{aligned}$$

and subtracting gives

$$(t - t^{-1})D^2 = 0$$

For  $t > 1$  we have  $t - t^{-1} \neq 0$ , and  $D^2 > 0$ , so this is impossible.  
So stationary points must be on the boundary.

We can write

$$\begin{aligned}g(u, v) &= \left(1 - \frac{1}{tu + t^{-1}v + 1}\right) + \left(1 - \frac{1}{u + v + 1}\right) \\ &= \frac{1}{u + v + 1} - \frac{1}{tu + t^{-1}v + 1}\end{aligned}$$

which goes to 0 when  $(u, v) \rightarrow \infty$  along any path.

On  $v = 0$  we have

$$g(u, 0) = \frac{tu}{1 + tu} - \frac{u}{1 + u} = \frac{(t - 1)u}{(1 + tu)(1 + u)} =: h(u)$$

Compute the derivative

$$\begin{aligned}h'(u) &= \frac{(t - 1) \cdot (1 + tu)(1 + u) - (t - 1)u \cdot [t \cdot (1 + u) + (1 + tu) \cdot 1]}{(1 + tu)^2(1 + u)^2} \\ &= \frac{(t - 1)[1 + tu + tu^2 - tu - tu^2 - tu^2]}{(1 + tu)^2(1 + u)^2} \\ &= \frac{(t - 1)(1 - tu^2)}{(1 + tu)^2(1 + u)^2}\end{aligned}$$

Since  $t > 1$ , the solution of  $h'(u) = 0$  is  $u = 1/\sqrt{t}$  and

$$h\left(\frac{1}{\sqrt{t}}\right) = \frac{(t - 1)\frac{1}{\sqrt{t}}}{\left(1 + t\frac{1}{\sqrt{t}}\right)\left(1 + \frac{1}{\sqrt{t}}\right)} = \frac{(t - 1)}{(1 + \sqrt{t})(\sqrt{t} + 1)} = \frac{t - 1}{(1 + \sqrt{t})^2}$$

With analogous computations we can show that on  $u = 0$  there is another stationary point  $v = \sqrt{t}$  which attains the same value but with opposite sign.

So this quantity is the maximum and we can rewrite it as

$$\frac{t-1}{(1+\sqrt{t})^2} = \frac{\sqrt{t}-1}{\sqrt{t}+1} = \frac{e^{\frac{\ln t}{2}} - 1}{e^{\frac{\ln t}{2}} + 1} = \tanh\left(\frac{\ln t}{4}\right)$$

□

Since  $\tanh(\frac{1}{4}) \approx 0.245 < 0.333... = 1/3$ , we improved the lower bound of [Theorem 3.3](#).

Surprisingly, the same bound applies for coalitions of arbitrary size, unlike the Condorcet case (see [Theorem 3.4](#))

**Theorem 5.6.**

The proportional score rule is  $k$ -SNM- $\tanh(\frac{1}{4})$  for every  $k \geq 2$ .

*Proof.* Let  $S = \{1, \dots, k\}$  and

$$A = \sum_{i \in S} e^{s_i}, \quad B = \sum_{j \notin S} e^{s_j}, \quad R = \frac{A}{A+B}$$

Inside the coalition define the probability weights

$$p_i = \frac{e^{s_i}}{A}, \quad i \in S$$

so that  $\sum_{i \in S} p_i = 1$ .

After the manipulation we have  $s'_i = s_i + \Delta_i$  for some  $\Delta_i \in [0, 1]$  such that  $\sum_{i \in S} \Delta_i = 1$  so

$$A' = \sum_{i \in S} e^{s_i + \Delta_i} = A \sum_{i \in S} p_i e^{\Delta_i} = As, \quad s := \sum_{i \in S} p_i e^{\Delta_i}$$

while  $B' = B$  is unchanged. Therefore

$$R' = \frac{A'}{A' + B} = \frac{As}{As + B} = \frac{\alpha s}{\alpha s + 1}, \quad \alpha := \frac{A}{B}$$

and the difference is a function of the scalar  $s$  only

$$R' - R = \frac{\alpha s}{\alpha s + 1} - \frac{\alpha}{\alpha + 1} = \frac{(s-1)\alpha}{(1+s\alpha)(1+\alpha)}$$

Notice this is the function  $h(u)$  of [Lemma 5.5](#) with  $u = \alpha$  and  $t = s$ .

In particular, if we define  $\delta := \|\Delta\|_\infty = \max_{i \in S} \Delta_i$  then

$$\begin{aligned} e^{-\delta} &\leq e^{\Delta_i} \leq e^\delta, \quad \forall i \in S \\ e^{-\delta} &\leq s = \sum_{i \in S} p_i e^{\Delta_i} \leq e^\delta \end{aligned}$$

Hence  $|\ln s| \leq \delta$  and we get the bound

$$R' - R \leq \tanh\left(\frac{\ln s}{4}\right) \leq \tanh\left(\frac{\delta}{4}\right)$$

Since  $\delta \leq 1$ , we get the result.

□

## Bibliography

- [1] A. Altman, A. D. Procaccia, and M. Tennenholtz, “Nonmanipulable Selections from a Tournament,” in *Computational Foundations of Social Choice*, F. Brandt, V. Conitzer, L. A. Hemaspaandra, J.-F. Laslier, and W. S. Zwicker, Eds., in Dagstuhl Seminar Proceedings (DagSemProc), vol. 10101. Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2010, pp. 1–6. doi: 10.4230/DagSemProc.10101.5.
- [2] A. Altman and R. Kleinberg, “Nonmanipulable Randomized Tournament Selections,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 24, no. 1, pp. 686–690, July 2010, doi: 10.1609/aaai.v24i1.7613.
- [3] J. Schneider, A. Schwartzman, and S. M. Weinberg, “Condorcet-Consistent and Approximately Strategyproof Tournament Rules,” in *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, C. H. Papadimitriou, Ed., in Leibniz International Proceedings in Informatics (LIPIcs), vol. 67. Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2017, pp. 35:1–35:20. doi: 10.4230/LIPIcs.ITCS.2017.35.
- [4] D. Pennock, A. Schwartzman, and E. Xue, “Toward Fair and Strategyproof Tournament Rules for Tournaments with Partially Transferable Utilities,” in *Algorithmic Decision Theory*, R. Freeman and N. Mattei, Eds., Cham: Springer Nature Switzerland, 2025, pp. 174–188.