Optimal Tree Contest Design and Winner-Take-All*

Qian Jiao[†] Zhonghong Kuang[‡] Yiran Liu[§] Yang Yu[¶]

December 30, 2023

Abstract

This paper investigates the effort-maximizing design of multi-stage contests with tree-like architectures through both the architecture and the prize. We first show that given the architecture, the whole budget should be assigned to a single match. This match might not be the final in general, but it must be the final if the architecture is symmetric. If the contest organizer can design the architecture, winner-take-all is optimal, disentangling two design elements. For contest architecture, we use dynamic programming and induction to estimate the optimized total effort level. Our new approach can be applied to extend Gradstein and Konrad (1999)'s result about how the optimal architecture hinges on the noisy level of matches.

JEL Classification: C72, D72

Keywords: Tree contest, Prize allocation, Winner-take-all, Optimal contest architecture

*An earlier version of this paper was circulated under the title "Optimal Contest Design with Tree Architecture". We are especially grateful to Simon Board, Tilman Börgers, Qiang Fu and Jingfeng Lu for detailed comments, which helped us improve the manuscript. For helpful comments and insights, we also thank Mikhail Drugov, Igor Letina, Shuo Liu, Dmitry Ryvkin, Bo Shen, Xiang Sun, Zijia Wang, Fan Wu, Zenan Wu, Mengxi Zhang, seminar participants at the Global Seminar on Contests & Conflict, and conference participants at the 2021 Conference on Contests: Theory and Evidence, 2022 SAET (Society for the Advancement of Economic Theory) Conference, 2022 Conference on Mechanism and Institution Design. This research is funded by the National Natural Science Foundation of China (No. 72103190, 72373171, 72192801). All errors are our own.

†Lingnan College, Sun Yat-sen University, Guangzhou, China, 510275 (jiaoq3@mail.sysu.edu.cn).

‡School of Economics, Renmin University of China, Beijing, China, 100872 (kuang@ruc.edu.cn).

§Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China, 100084 (liu-yr21@mails.tsinghua.edu.cn).

¶Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China, 100084 (yangyu1@mails.tsinghua.edu.cn).

1 Introduction

Contests often include multiple stages. At each stage, parallel single-winner matches occur, and only winners are promoted to the next stage. Contestants put in costly efforts in each stage to reach the final and win prizes. Such contentions, labeled as tree contests (T-contests for short) by Gradstein and Konrad (1999), are used in competitions ranging from sports events to promotions within organizations. For example, an employee vies with intra-group colleagues to be promoted to a group manager, then competes with other group managers for the next promotion opportunity, and so on. In those T-contests, the primary objective is to maximize the total effort of all agents at all stages.

A central question in T-contests is the design of *contest architecture* and *prize structure*. Our analysis of contest architecture is motivated by three features of real-world T-contests. First, every embedded match has a single winner. Second, early efforts are sunk when contestants proceed: Excellent performance in earlier stages cannot compensate for poor performance in later stages. Third, different matches in the same stage can have different numbers of contestants.

In addition to the contest architecture, the effort provision in T-contests is also affected by the prize structure (see Rosen (1986)). Unlike one-stage simultaneous contests with the whole budget going toward the top prize, T-contests are more likely to award an intermediate prize.[1] For example, employees who are eliminated in a promotion competition retain their original positions, and their increased salaries are intermediate prizes. Such an exploration naturally raises several questions. First, given the contest architecture and a fixed prize budget, is a winner-take-all contest necessarily better than one that awards intermediate prizes? Second, how does the (a)symmetry of the contest architecture influence the optimal prize structure? Third, what is the optimal tree architecture?

Nature of the Problem

In this paper, the design elements are contest architecture and prize structure. Both elements, even if considered separately, feature important aspects of real-world contest design and thus attract the attention of researchers. However, in either dimension, previous results are obtained from restricted settings.

[1] Multiple prizes also exist in simultaneous contests. See Xiao (2016) and Fu, Wu and Zhu (2022).

For contest architecture, Gradstein and Konrad (1999) make the first attempt, assuming that the winner-take-all rule is applied exogenously. They consider two alternatives: a simultaneous contest and a contest with *symmetric binary tree architecture*, in which 2^k contestants are sequentially eliminated from the race through k stages of pairwise matches. However, this restricts the number of participants to a power of 2. By contrast, imagine a contest with three players. There are two possible structures: a simultaneous contest, and a two-stage contest in which the winner between two players competes with the third player for victory. Direct comparison between two architectures is simple algebra, while Gradstein and Konrad (1999)'s framework does not apply.

For prize structure, Rosen (1986) concentrates on the symmetric binary tree architecture and asserts that the winner-take-all rule is optimal. Knyazev (2017) generalizes the number of players to have the form of m^N, and the players are eliminated sequentially using m-player matches in N stages. However, this formalization imposes unnecessary requirements on the contest architecture, and even on the number of contestants.

There are additional concerns when two design elements are jointly optimized. Using backward induction, we should first characterize the optimal prize structure in all possible contest architectures. Afterward, we search for the optimal contest architecture. Unfortunately, even if the number of contestants is a power of 2, we cannot conclude that adopting the winner-take-all rule in a symmetric binary tree architecture is optimal. A different contest architecture with a different prize structure might be better.

Moreover, all previous methods cannot handle asymmetric contest architecture where two matches in the same stage have different numbers of participants. Analyzing symmetric architecture is simple since all contestants adopt identical strategies within the same stage. If the architecture is asymmetric, then the subgame perfect equilibrium should be solved match-by-match rather than stage-by-stage.

How does asymmetry influence the outcome of the contest design? Consider the example of promotions within organizations. If the organization is highly asymmetric, is it still optimal to pay a large first-place prize? In technical-oriented unicorns, there are far more employees in technical-related departments than in other departments. Should the CTO be paid a higher salary than the CEO? The previous results fail to address this question.

Our Approach and Results

In our model, an arbitrary number of homogeneous contestants are successively eliminated through a series of stages. In each stage, contestants are divided into predetermined groups and compete within each group, and a unique winner of each group advances to the next stage by the Tullock winner-selection mechanism. The contest organizer designs the tree architecture and allocates the prize money to maximize the total effort. To capture the possibility of asymmetry, we define the T-contest recursively. In particular, a T-contest includes a final match and several sub-contests that are also T-contests and determine who will participate in the final match.

Given the contest architecture, we first show that the contest organizer should devote the entire budget to a single match, which we call the pivotal match. The pivotal match principle springs from two facts about linearity: First, the feasible region of the prize structure is a simplex; Second, the total effort is a linear function of the prize structure. When the contest architecture is symmetric, namely, all matches in the same stage have an identical number of participants, the winner-take-all rule is optimal. We consider the most general form of symmetric architecture where matches in different stages can differ in the number of participants, which, to our knowledge, has never been studied in the literature. Nevertheless, the pivotal match is not necessarily the final match in general when the architecture is asymmetric. This finding demonstrates that the restriction on symmetry architecture is a loss of generality.

We can use the pivotal match principle to disentangle the design of contest architecture and prize structure. To this end, we show that any contest architecture whose pivotal match is not the final is sub-optimal. If the pivotal match is not the final in a given tree architecture, contestants who cannot reach the pivotal match will shirk by always exerting zero effort. Then, if the contest designer can design the architecture, it is better to insert those shirking contestants into other stage-1 matches that may enter the pivotal match. Therefore, the contest designer only needs to adjust the tree architecture, and give the entire prize money to the final champion to maximize the total effort of all contestants.

As for analyzing the optimal tree architecture, it turns out to be a dynamic programming problem since we adopt a recursive definition of tree architecture. A contest architecture is optimal if its final match is appropriately designed and all its sub-contests are optimal. The whole architecture cannot be optimal if any sub-contest is organized poorly. Moreover, even if all sub-contests are optimal, the efficiency of the contest still rides on the final match design. Therefore, the design of the contest architecture ultimately comes down to recursively designing the finals of each sub-contest.

We then provide upper bounds for the expected total effort using induction methods, depending on the underlying winner-selection mechanism. When the contest is relatively discriminatory (i.e., the discriminatory power is at least one in the Tullock framework), the simultaneous contest achieves the upper bound and is thus optimal. With a higher discriminatory power, the equilibrium effort is relatively higher, resulting in a lower contestant's payoff. Hence, the expected payoff of participating in the final match in a multi-stage contest is so small that contestants have little incentive to make an effort in previous stages. Therefore, it is inefficient to organize a contest with multiple stages, and the one-shot simultaneous contest is optimal.

When discriminatory power is less than one, the upper bound we provide is quite tight since the bound can be reached infinitely many times as the number of contestants (n) grows. Only if n is a power of 2, the upper bound is achieved by a symmetric binary tree architecture, which agrees with the findings in Gradstein and Konrad (1999). If not, the optimal architecture is generally asymmetric, and the contest designer faces a trade-off between exploiting contestants and balancing contestants. Although increasing the number of stages could elicit more effort, its side-effect is the imbalance caused by prolonged T-contests, which has not been well recognized since the specific shape of contest architecture is ignored in the literature.

When the contest technology is sufficiently noisy (i.e., discriminatory power is less than 2/3), the architecture solely containing bilateral matches, in which each contestant experiences at most one bye, is optimal. On the one hand, T-contests with bilateral embedded matches have the most stages. On the other hand, the contest organizer should keep the number of nontrivial matches of all contestants as equal as possible for balance concerns. As a result, exploitation and balance align perfectly. However, when the contest technology is moderately noisy, exploitation and balance are conflicting objectives that must be leveled by the contest organizer.

Related Literature

There has been much research on contest design in recent decades.[2] Our paper primarily belongs to the literature on multi-stage sequential elimination contest design. The literature on contests with tree architecture dates back to the seminal paper of Rosen (1986). With winner-take-all, Gradstein and Konrad (1999) endogenize the stage number of T-contests and demonstrate that the stage number depends on the particular Tullock contest technology. We provide a general framework that considers both design elements, which are possibly entangled. Even if we focus on a single design element, our results extend Rosen (1986) and Gradstein and Konrad (1999), respectively.

Given the prize structure, a handful of papers have examined structure design in multi-stage contests. In a two-stage grouping contest setting, Amegashie (1999) derives the optimal number of finalists. Moldovanu and Sela (2006) explore whether a contest should involve a preliminary stage that selects finalists. Cohen, Maor and Sela (2018) study the optimal head start to favor the finalist who is top-ranked in the first stage.

This paper is also linked to the literature on optimal prize allocation by the contest organizer. Several papers have examined optimal prize structures in one-shot contests.[3] In multi-stage contests, including two-stage contest (Krishna and Morgan, 1998) and nested pooling contests (Fu and Lu, 2012), the winner-take-all rule is optimal. Knyazev (2017) attempts to determine the optimal prize structure for a multi-stage contest with m^N players that maximizes the designer's profit, and the players are eliminated sequentially through N stages. Feng et al. (2024) study the prize design in team contests, show that the winner should take all, and examine how to select the winner.

The rest of the paper is organized as follows. Section 2 sets up the model and characterizes the subgame perfect Nash equilibrium. In Section 3, we present the pivotal match principle and rationalize the winner-take-all rule. Section 4 characterizes the optimal contest architecture by exploring its dynamic programming nature. **??** discusses extensions of the model. Appendix A collects some technical proofs.

[2]See Konrad (2009), Fu and Wu (2019) for detailed surveys.

[3]See Krishna and Morgan (1998), Moldovanu and Sela (2001), Cohen, Kaplan and Sela (2008), Drugov and Ryvkin (2020), Letina, Liu and Netzer (2020) for examples.

2 Model

2.1 Setup

There are n homogeneous risk-neutral contestants involved in a T-contest that consists of several stages. In each stage, contestants are embedded into different matches.

Winner-Selection Mechanism in a Match

In a match with $L \geq 2$ contestants, the probability of winning is given by the generalized Tullock contest success function:

$$p_i(e_i, \boldsymbol{e}_{-i}) = \begin{cases} \frac{e_i^\gamma}{\sum_{j=1}^L e_j^\gamma}, & \sum_{j=1}^L e_j > 0, \\ \frac{1}{L}, & \sum_{j=1}^L e_j = 0, \end{cases} \tag{1}$$

where e_i is the effort level of contestant i, and $\boldsymbol{e}_{-i}$ collects the effort of other contestants. γ is the discriminatory power parameter. To ensure pure strategy equilibrium in each match and avoid trivial analysis, we impose the restriction $\gamma \in (0, \frac{n}{n-1})$.[4] Let $c(e_i)$ denote the cost of effort. We assume the cost of effort is linear $c(e_i) = e_i$.[5]

Contest with Tree Architecture

In a T-contest with architecture T, contestants are denoted by the nodes in the lowest level of the tree, and the set of contestants is denoted by $\mathcal{N}(T) = \{1, \cdots, n\}$. The set of matches in T is denoted by $\mathcal{X}(T)$ and represented by nodes that are not in the lowest level. Matches in the same stage occur simultaneously, and the winner of a stage-k match advances to a predetermined match in stage $k+1$. The final match is denoted by x_R. In the T-contest shown in Figure 1, black circles denote contestants, the double circle denotes the final, and hollow circles denote other matches.

To facilitate the analysis, we introduce the following notations. Let $n(x)$ denote the number of participants in match x and $p(x)$ denote the match to which the winner of match x advances. We call $p(x)$ the parent match of x.[6] If $n(x) > 1$, match x is nontrivial,

[4]When $\gamma \geq \frac{n}{n-1}$, the simultaneous contest suffices to fully dissipate the rent (Baye, Kovenock and de Vries, 1999).

[5]Section B.3 considers nonlinear costs.

[6]Note that $p(x_R)$ is not well defined, while $p(\cdot)$ is well defined for a contestant i. In a slight abuse of notation, we say that $p(i)$ represents the first match contestant i takes part in.

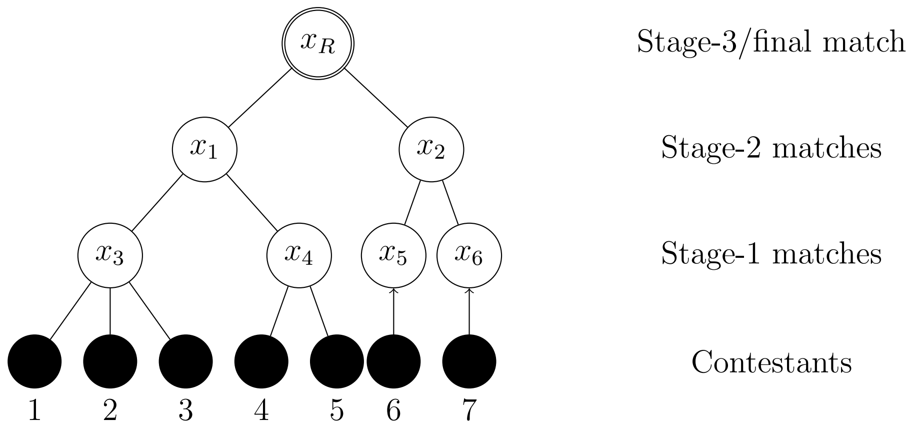

Figure 1: T-Contest

and the winner is selected by Equation 1 and advances to $p(x)$. If $n(x) = 1$, x is *trivial*, and its unique participant advances to $p(x)$ through a *bye*.[7] If match x' occurs after x, and the winner of x has to win x' to win the championship, x' is called a *future match* of x. For each match x, the set of future matches is defined as $\mathcal{F}(x)$, and $p(x) \in \mathcal{F}(x)$. A *sub-contest* includes an initial node and all its successor nodes in lower levels.[8]

Prize Structure

In a T-contest with architecture T, the prize structure $v(\cdot)$ specifies the prize $v(x)$ allocated to the winner of match $x \in \mathcal{X}(T)$, satisfying the budget constraint $\sum_{x \in \mathcal{X}(T)} v(x) \leq 1$. Negative prizes are prohibited to align with examples of sports competitions and organizational promotions. Let $\hat{v}(x)$ denote the *effective prize* for match x, containing dual benefits of winning x: One is the direct benefit of obtaining prize $v(x)$, and the other is the indirect benefit of future prizes by advancing to the next stage. Clearly, $\hat{v}(x) \geq v(x)$.

2.2 Equilibrium

The solution concept is subgame perfect Nash equilibrium. For a match x with effective prize $\hat{v}(x)$, the present payoff of participant i is $p_i(e_i, \boldsymbol{e}_{-i})\hat{v}(x) - e_i$. Since all $n(x)$ participants are homogeneous and have a common effective prize, the symmetric equilibrium effort in match x can be derived as $e^*(x) = \frac{(n(x)-1)\gamma}{n(x)^2}\hat{v}(x)$. Lemma 1 formally

[7] We allow trivial matches because contestants are labeled by nodes at the lowest level. It is equivalent to ruling out trivial matches if different contestants can start from different layers.

[8] In Figure 1, Contestants 1 to 5 compete in two stage-1 matches, and two winners advance to stage 2. Matches x_5, x_6 are trivial, and Contestants 6 and 7 enter the stage-2 match x_2. The set of future matches of Contestant 1 is $\{x_R, x_1, x_3\}$. For match x_5, $p(x_5) = x_2$ and $\mathcal{F}(x_5) = \{x_R, x_2\}$.

presents the equilibrium by pinning down the analytical formula of effective prize $\hat{v}(x)$.

Lemma 1. *There exists a unique subgame perfect Nash equilibrium. In match $x \in \mathcal{X}(T)$, the equilibrium effort of each contestant is $e^*(x) = \frac{(n(x)-1)\gamma}{n(x)^2}\hat{v}(x)$, where the effective prize $\hat{v}(x) = \sum_{x' \in \mathcal{X}(T)} r(x', x)v(x')$ is a **linear** function on $\{v(x)\}_{x \in \mathcal{X}(T)}$. Here*

$$r(x', x) = \mathbf{1}(x' \in \mathcal{F}(x)) \prod_{z \in \mathcal{F}(x) \setminus \mathcal{F}(x')} \left[\frac{n(z) - (n(z)-1)\gamma}{n(z)^2} \right] + \mathbf{1}(x' = x) \tag{2}$$

Proof. See the Appendix. □

Term $r(x', x)$ in Lemma 1 represents the marginal effect of assigning a prize to match x on the effective prize of match x', thus reflecting the transmission mechanism of prize allocation in stimulating effort. Clearly, the prize awarded for a given match boosts effort only in matches that belong to the corresponding sub-contests.

Based on the equilibrium result, we can derive the total effort in the whole T-contest given the contest architecture T and its associated reward scheme v, denoted by $\mathbf{TE}(T, v)$, $\mathbf{TE}(T, v) = \sum_{x \in \mathcal{X}(T)} \mathbf{TE}(x|T, v) = \sum_{x \in \mathcal{X}(T)} \frac{(n(x)-1)\gamma}{n(x)}\hat{v}(x)$, where $\mathbf{TE}(x|T, v)$ denotes the induced effort of match x and is determined by $n(x)e^*(x)$. The contest organizer aims to maximize $\mathbf{TE}(T, v)$ through designing T and v. According to Lemma 1, increasing the final prize value $v(x_R)$ will strictly increase the effective prize $\hat{v}(x)$ for any match x since $x_R \in \mathcal{F}(x)$. Therefore, the budget constraint should be binding, i.e., $\sum_{x \in \mathcal{X}(T)} v(x) = 1$.

3 Optimal Prize Allocation

In this section, we explore the optimal reward scheme for a given contest architecture.

3.1 Pivotal Match Principle

We first investigate how the prize structure affects the total effort of the T-contest.

Given the T and v, the induced total effort is $\boldsymbol{TE}(T, v) = \sum_{x \in \mathcal{X}(T)} H(T, x)v(x)$, which is a **linear** function on $\{v(x)\}_{x \in \mathcal{X}(T)}$ and we can show that $H(T, x) = \sum_{z \in \mathcal{X}(T)} \frac{(n(z)-1)\gamma}{n(z)} r(x, z)$. Here, $H(T, x)$ represents the marginal effect of assigning a prize to match x on the total effort, which aggregates the marginal effects on match effort across all matches. Hence, $H(T, x)$ also represents the total effort level when the entire budget goes to match x.

Based on this observation, we can write out a recursive formula for $H(T, x)$,

$$H(T,x) = \frac{(n(x)-1)\gamma}{n(x)} + \frac{1}{n(x)}\left[1 - \frac{(n(x)-1)\gamma}{n(x)}\right] \sum_{x':p(x')=x} H(T,x'), \tag{3}$$

where the former and latter terms represent the effort level in x and all sub-contests rooted at x, respectively. In each sub-contest, the effective prize of winning is $\frac{1}{n(x)}\left[1 - \frac{(n(x)-1)\gamma}{n(x)}\right]$. To initiate the recursion, we define $H(T, i) = 0$ for contestant $i \in \mathcal{N}(T)$.

Since the domain of $\{v(x)\}_{x\in\mathcal{X}(T)}$ is a simplex, the optimal prize structure should allot the entire prize to a single match. It is called a *pivotal match*, denoted by x_P. The following pivotal match principle establishes the optimal prize allocation. Note that the pivotal match always exists but may not be unique in some corner cases.

Proposition 1 (Pivotal Match Principle)**.** *Given the contest architecture T, the optimal prize structure is $v^*(x_P) = 1$, where x_P maximizes $H(T, x)$.*

Proposition 1 is new in the literature since we allow arbitrary contest architecture.

To illustrate how to identify the pivotal match, consider a 5-player example shown in Figure 2 and calculate $H(T, x)$ of each match x accordingly. The results are summarized in Table 1. When $\gamma \in (0, \frac{5}{4})$, we have $H(T, x_2) > H(T, x_R) > H(T, x_4) = H(T, x_5) > H(T, x_1) = H(T, x_3)$. Therefore, the pivotal match in this tree architecture is $x_P = x_2$.

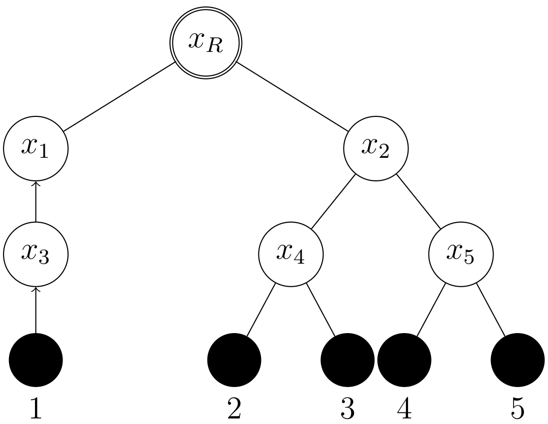

Figure 2: 5-player example

Match	$H(T, \cdot)$
x_R	$\frac{16\gamma-6\gamma^2+\gamma^3}{16}$
x_1	0
x_2	$\frac{4\gamma-\gamma^2}{4}$
x_3	0
x_4	$\frac{\gamma}{2}$
x_5	$\frac{\gamma}{2}$

Table 1: $H(T, x)$ of Matches

Conventional wisdom suggests that a prize at a higher rank in the hierarchy could make a contestant exert more effort because he has to continually work to climb to the top. However, the winner-take-all rule is not necessarily optimal (see Figure 2). In

Proposition 2, we show that as the contest becomes noisier, the winner-take-all is more likely to be optimal.

Proposition 2. *For any contest architecture T, there exists a threshold $\bar{\gamma}(T) \in [0, \frac{n}{n-1}]$ such that the winner-take-all rule is optimal if and only if $\gamma \leq \bar{\gamma}(T)$.*

Proof. Given the contest architecture T and a match $x \neq x_R$, there exists a threshold $\bar{\gamma}(T, x) \in [0, \frac{n}{n-1}]$ such that $H(T, x_R) \geq H(T, x)$ if and only if $\gamma \leq \bar{\gamma}(T, x)$. Then, $\bar{\gamma}(T) = \min_x \bar{\gamma}(T, x)$. See the Appendix for details. □

Proof of Proposition 2. □

Note that it is possible that the winner-take-all rule is optimal for all or none of γ. For the former case ($\bar{\gamma}(T) = \frac{n}{n-1}$), think over the following architecture for example. The final match occurs between two contestants; one is selected by a k-stage symmetric binary tree contest, while the other is selected by a $k+1$-stage one. For the latter case ($\bar{\gamma}(T) = 0$), consider another example. The final match occurs between one pre-determined contestant and another contestant that is selected by a k-stage symmetric binary tree contest. Here, the pivotal match is the sub-final regardless of γ if $k \geq 2$.

This result seems counterintuitive. Recall the recursive formula defined by Equation 3. If we regard $\{H(T, x')\}_{x':p(x')=x}$ as constants, then as γ increases, $H(T, x)$ is more likely to be greater than $\max H(T, x')$ and the winner-take-all rule is more likely to be optimal. However, this intuition ignores the recursive nature of T-contests. As γ increases, $H(T, x)$ increases for all stage-1 matches, and then for all stage-2 matches, and so on.

Recall that the effective prize of each match x comprises two components, the direct benefit of winning a prize and the indirect benefit of being promoted. By Equation 3, the direct benefit dominates as γ increases, while the indirect benefit dominates as γ decreases. Therefore, when γ is relatively low, the winner-take-all rule should be optimal since the winner-take-all rule can exploit indirect benefits to the greatest extent.

3.2 Symmetric Architecture and Winner-Take-All

Although the winner-take-all rule might be sub-optimal in general, when the contest architecture is completely symmetric, we can show that the winner-take-all rule must be optimal. First, we define the concept of symmetric architecture.

Definition 1. *A tree architecture T is* ***symmetric*** *if for any two matches x and x' in the same stage, $n(x) = n(x')$.*

In a symmetric K-stage T-contest, matches in stage k should have the same number of participants (or sub-contests), denoted by N_k. All $\prod_{k=1}^{K} N_k$ homogeneous contestants are treated fairly since their initial seats are equivalent. For the contest architecture mentioned in Rosen (1986) and Gradstein and Konrad (1999), as well as the Wimbledon or FIFA World Cup knockout stages, we have $N_k = 2$ for all k. However, we allow N_k to be contingent on k in general. That is, N_k may differ across stages.

Theorem 1. *It is optimal for the contest organizer to choose the winner-take-all prize allocation rule when the contest architecture is symmetric.*

Theorem 1 is strong since it holds regardless of the number of stages K, the number of participants in each stage $\{N_k\}_{k=1}^{K}$, or the discriminatory power γ.

We then sketch the proof. In a symmetric tree architecture, each sub-contest rooted at a stage-k match has an identical structure. Hence, stage-k matches should have the same $H(T, x)$, denoted by H_k. Then, Equation 3 can be reduced to

$$H_{k+1} = \frac{(N_{k+1} - 1)\gamma}{N_{k+1}} + \frac{N_{k+1} - (N_{k+1} - 1)\gamma}{N_{k+1}} H_k. \tag{4}$$

The former term $\frac{(N_{k+1}-1)\gamma}{N_{k+1}}$ represents the total effort in the stage-$(k+1)$ match with the entire prize, and the latter term represents the total effort in all N_{k+1} sub-contests. In each sub-contest, the effective prize of winning is $\frac{N_{k+1}-(N_{k+1}-1)\gamma}{N_{k+1}^2}$.

Lemma 2 helps rationalize the winner-take-all rule in symmetric contest architectures.

Lemma 2. *$H_{k+1} > H_k$.*

Proof. See the Appendix. □

As a result, $H(T, x_R) = H_K > H_{K-1} > \cdots > H_1$. Since x_R maximizes $H(T, x)$, by Proposition 1, the contest organizer should allocate all prize money to the final. This finishes the proof of Theorem 1.

As first illustrated by Fu and Lu (2012) in a nested pooling contest, the transfer from an earlier stage of the competition to a future stage produces two effects. Similarly in a T-contest, considering a prize transfer from a match x in stage k to its parent match $p(x)$

in stage $k+1$. Contestants would make more effort in $p(x)$, which is a positive effect. While at the same time, this reduces the payoff for the participants in x and all of the matches before it, resulting in less effort being placed into those matches, which is referred to as a negative effect. When the contest architecture is symmetric, the positive effect equals $\frac{(N_{k+1}-1)\gamma}{N_{k+1}}$ times the transferred prize, and the negative effect equals $\frac{(N_{k+1}-1)\gamma}{N_{k+1}}H_k$ times the transferred prize (see Equation 5). Since the positive effect always dominates the negative effect ($H_k < 1$), the winner-take-all rule is optimal.

$$H_{k+1} - H_k = \underbrace{\frac{(N_{k+1}-1)\gamma}{N_{k+1}}}_{\text{positive effect}} - \underbrace{\frac{(N_{k+1}-1)\gamma}{N_{k+1}}H_k}_{\text{negative effect}}. \tag{5}$$

However, when the contest architecture is arbitrarily asymmetric, H_k is no longer well defined and Equation 5 fails. The relative sizes of these two effects are not readily determinable, making winner-take-all no longer necessarily optimal (Proposition 1). In the example shown in Figure 2, if the prize is transferred from x_2 to x_R, the total effort exerted in the final match after the transfer, namely the positive effect, equals $\frac{\gamma}{2}$. However, the transfer reduces the effective prize of x_2 from one to $\frac{2-\gamma}{4}$. Hence, the size of the negative effect would be $\left(1 - \frac{2-\gamma}{4}\right) H(T, x_2) = \frac{\gamma}{2}\frac{(2+\gamma)(4-\gamma)}{8}$. Since the negative effect dominates the positive effect, the winner-take-all rule fails.

If we replicate the sub-contest associated with x_2 to match x_1, the T-contest possesses a symmetric tree architecture with eight contestants and three rounds of bilateral matches instead. In this hypothetical architecture, the effective prize of x_1 rises from zero to $\frac{2-\gamma}{4}$, the negative effect of transferring the prize from x_2 to x_R would be $\left(1 - \frac{2-\gamma}{2}\right) H(T, x_2) = \frac{\gamma}{2}\frac{4\gamma-\gamma^2}{4} < \frac{\gamma}{2}$. Clearly, the negative effect shrinks since the hypothetical sub-contest rooted at x_1 generates extra effort. As shown by the above comparison, the asymmetry creates unfairness among the contestants, thereby reducing the motivational effect of the final prize on the contestants, and ultimately leading to the failure of the winner-take-all rule.

The general form of symmetric architecture, which rationalizes winner-take-all, has never been explored in the literature. In the most recent study, Knyazev (2017) requires all matches in the T-contest to have the same number of participants. While we allow the number of participants in matches to vary across different stages. Moreover, according to the result in Theorem 1, the optimality of the winner-take-all rule, which is first stated by Rosen (1986) with 2^k contestants in a k-stage pairwise elimination contest, can be generalized to any symmetric architecture.

3.3 Winner-Take-All in the Joint Design Problem

Another direct implication of the pivotal match principle (Proposition 1) facilitates the optimal joint design in Section 4: Any contest architecture T whose pivotal match is not the final is sub-optimal when the contest architecture is flexible.

Theorem 2. *It is optimal for the contest organizer to choose the winner-take-all prize allocation rule when she can jointly design the contest architecture and prize structure.*

Let $\mathcal{T}(x)$ denote the architecture of the sub-contest whose final match is x. Obviously, $\mathcal{T}(x_R) = T$. Suppose there are m contestants in the sub-contest with architecture $\mathcal{T}(x_P)$, then applying the winner-take-all rule in $\mathcal{T}(x_P)$ will induce the same total effort as in T with $v(x_P) = 1$. Based on the contest architecture $\mathcal{T}(x_P)$, fix the number of stages and the number of matches in each stage, we construct T' by arbitrarily putting those $n - m$ shirking contestants into the stage-1 matches of $\mathcal{T}(x_P)$. Then, the effective prize of each match in T' will be the same as in $\mathcal{T}(x_P)$ since it is solved backward. In stage 1, T' will induce strictly higher effort than $\mathcal{T}(x_P)$ as more contestants are involved in the competition, while the effort induced in the remaining stages remains the same.

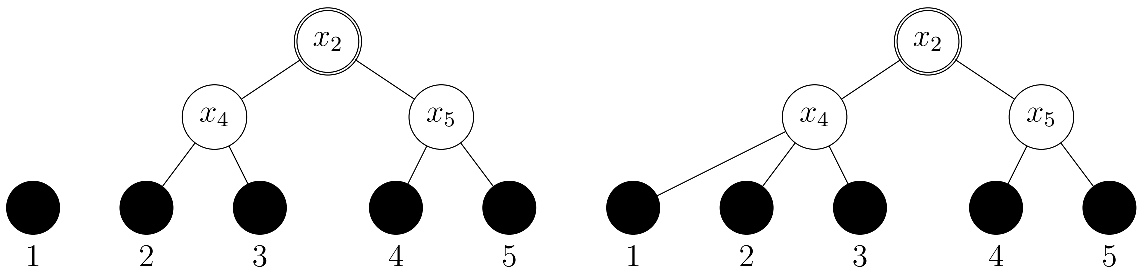

Figure 3: Construction of T' from T

Figure 3 visualizes the above construction process starting from the 5-player example shown in Figure 2 whose architecture is denoted by T. As shown in the left panel, the sub-contest $\mathcal{T}(x_P)$ contains Contestants 2 to 5. Contestant 1 shirks. If Contestant 1 is added into match x_4, as shown in the right panel, the effort level induced in x_4 increases. So does the total effort level, suggesting that the architecture T in Figure 2 is sub-optimal.

Therefore, the total effort maximizing contest organizer should choose the final as the pivotal match if she can freely design the contest architecture. Winner-take-all is always the optimal prize structure, and the joint design problem finally boils down to the design of optimal contest architecture given winner-take-all.

Our results, including Theorem 1 and Theorem 2, provide two rationales (*symmetry* and *flexibility*) for the commonly assumed "winner-take-all" principle in practice.

4 Optimal Contest Architecture

In this section, we restrict our attention to the contest structure with a single prize $v(x_R) = 1$ to determine the optimal contest architecture that maximizes overall effort. Rewriting $\mathbf{TE}(T, v)$ by setting all prizes other than $v(x_R)$ at zero yields

$$\mathbf{TE}(T, v_{\text{WTA}}) = H(T, x_R) = \sum_{z \in \mathcal{X}(T)} \frac{(n(z) - 1)\gamma}{n(z)} r(x_R, z),$$

where v_{WTA} denotes the winner-take-all prize allocation rule. Clearly, the optimal T-contest architecture depends on the discriminatory level of the contest. In the remainder of this section, we investigate noisy contests ($\gamma < 1$) and moderately discriminatory contests ($\gamma \geq 1$) in Section 4.1 and Section 4.2, respectively.

4.1 Noisy Contests: $\gamma < 1$

We first simplify the design problem. In particular, we find that in order to induce the maximum effort, the number of contestants in each nontrivial match must be a prime number. We call this contest architecture a prime-number T-contest.

Lemma 3. *When $\gamma < 1$, the optimal T-contest architecture must be a prime-number T-contest.*

Proof. We prove this by contradiction: When a match includes a composite number (say, $a \times b$) of participants, we can split it into b matches that include a participants each, and let those b winners compete to decide the winner of those $a \times b$ participants. This adjustment induces a higher total effort. The details are relegated to the Appendix. □

Furthermore, given any $\gamma < 1$, the number of participants in any embedded match should not exceed a certain value. Lemma 4 formally presents the result.

Lemma 4. *When $\gamma < 1$, for any prime number p, there exists a threshold $\gamma_p = 1 - \frac{2}{(p+1)(p^2-2)}$ such that (i) when $\gamma < \gamma_p$, the number of participants in any embedded match of all the optimal contest architectures is no more than p; (ii) when $\gamma = \gamma_p$, a T-contest with the number of participants in any embedded match no more than p is optimal.*

Proof. The proof shares a similar logic as Lemma 3. See the Appendix for details. □

Note that a contest organizer should increase the number of stages as much as possible in a noisy contest in order to induce more effort. Lemma 3 illustrates that splitting one match into several matches with an identical number of participants always improves the total effort level. Therefore, the optimal tree structure never includes a match with a composite number of participants. Meanwhile, Lemma 4 demonstrates that splitting is always beneficial to the contest designer if the size of one embedded match is sufficiently large, regardless of whether the number of participants is a composite number.

Among prime-number T-contests, one of the most well-known contest architectures is the binary tree architecture, as formally defined below.

Definition 2. *A contest architecture T is a* ***binary tree*** *if $n(x) \leq 2$ for all $x \in \mathcal{X}(T)$.*

Each embedded match within a binary tree architecture is either a pairwise match ($n(x) = 2$) or a trivial match with only one contestant ($n(x) = 1$). In the latter scenario, the contestant directly advances to the next stage through a *bye*.

4.1.1 Sufficiently Noisy Contests: $\gamma \leq \frac{2}{3}$

By Lemma 4, $\gamma_2 = \frac{2}{3}$, the optimal architecture is just the binary tree.

Corollary 1. *When $\gamma \leq \frac{2}{3}$, the binary tree is the optimal T-contest architecture. Moreover, the optimal T-contest architecture must be a binary tree when $\gamma < \frac{2}{3}$.*

We further investigate the specific architecture of the optimal binary tree, which may not be unique. Based on the definition of $\mathcal{F}(x)$, i.e., the set of future matches, we further define the number of nontrivial future matches for x as $f(x) = \sum_{x' \in \mathcal{F}(x)} \mathbf{1}(n(x') = 2)$, which can describe the difficulty of becoming the grand champion for contestants i.

Conventional wisdom suggests that a balanced contest can better incentivize homogeneous contestants. When the number of contestants (n) is the power of 2, it is possible to design a symmetric binary tree contest without byes to guarantee that $f(i)$ is the same for all participants. However, when n is no longer a power of 2, it is impossible to make all contestants have the same $f(i)$. In this case, for balance concerns, the contest organizer should make the number of nontrivial matches of each contestant as equal as possible.

Definition 3. *A binary tree contest architecture T is **balanced** if*

$$\max_{i \in \mathcal{N}(T)} f(i) - \min_{i \in \mathcal{N}(T)} f(i) = \begin{cases} 0, & \text{if } n \text{ is a power of 2,} \\ 1, & \text{otherwise.} \end{cases} \qquad \text{(Balance Condition)}$$

We can show that the *Balance Condition* can guarantee optimality.

Proposition 3. *When $\gamma \leq \frac{2}{3}$, a binary tree contest is the optimal contest architecture if and only if it is **balanced**.*

Proof. See the Appendix. □

Consider a 6-player example. The following two tree architectures shown in Figure 4 are optimal. The number inside each nontrivial match node represents the value of $f(x)$, i.e., the number of nontrivial future matches. Obviously, both trees are balanced.

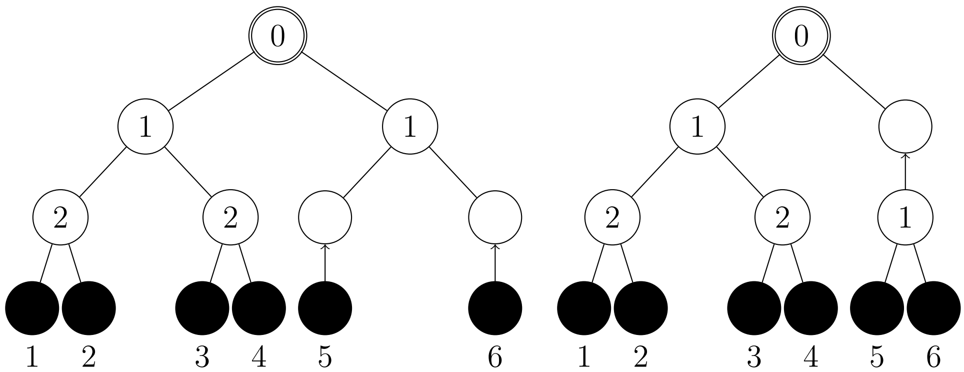

Figure 4: Optimal Architectures in the 6-player Example

Generally, the optimal binary tree architecture is not unique. However, we are able to provide a specific binary tree design to guarantee that it is optimal, in which byes only occur in the first stage (see the left panel of Figure 4). Let k be the largest integer not exceeding $\log_2(n)$. We divide n contestants into two groups: a **Bye** group contains $2^{k+1} - n$ contestants, and a **Competition** group contains $2(n - 2^k)$ contestants. In stage 1, contestants in **Competition** group form pairwise matches and $n - 2^k$ of them advance to stage 2, contestants in **Bye** group advance to stage 2 automatically. From stage 2,

2^k contestants compete through k stages of pairwise matches, and the final winner is selected and obtains the prize purse.

4.1.2 Moderately Noisy Contests: $\gamma \in (\frac{2}{3}, 1)$

When $\gamma \in (\frac{2}{3}, 1)$, we are not able to determine the closed-form optimal contest architecture in general. Nevertheless, we can efficiently solve for the optimal architecture through dynamic programming, which also applies to the cases with $\gamma \in (0, \frac{2}{3}] \cup [1, \frac{n}{n-1})$.

The correctness of dynamic programming stems from subgame optimality. Consider that the number of players in the final match is s, and n contestants are divided into s groups.[9] Assuming the τth group has λ_τ contestants, then the winner of the group final x_τ should be selected by the optimal contest architecture with λ_τ contestants.

The efficiency of dynamic programming is guaranteed by Lemma 4. Since γ_p is monotonically increasing in prime number p and converges to one, for any $\gamma \in (\frac{2}{3}, 1)$, we can always find a minimum prime number $q(\gamma)$ such that $\gamma < \gamma_q$. Hence, the number of participants in each match must be a prime number that is bounded from above.

Given the number of contestants (n), let $\Lambda(n, \gamma) = \langle s, \boldsymbol{\lambda} \rangle$ denote the design of the final match. A s-dimensional non-ascending vector $\boldsymbol{\lambda}$ collects the number of contestants involved in s sub-contests. Namely, $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_s), \lambda_\tau \geq \lambda_{\tau+1}$, and $\sum_{\tau=1}^{s} \lambda_\tau = n$. Therefore, the total effort of the final match is $\frac{(s-1)\gamma}{s}$, and the effective prize of participating in the final competition is $\frac{s-\gamma(s-1)}{s^2}$. Let $\Lambda^*(n, \gamma)$ denote the optimal design of the final match, and $\mathbf{TE}^*(n, \gamma)$ denote the optimized total effort. In the sub-contest with λ_τ contestants, the induced total effort is just $\frac{s-\gamma(s-1)}{s^2}\mathbf{TE}^*(\lambda_\tau, \gamma)$.

Proposition 4. *Given $\mathbf{TE}^*(\hat{n}, \gamma)$ for $\hat{n} = 1, \cdots, n-1$, the optimal design of the final match $\Lambda^*(n, \gamma) = \langle s^*, \boldsymbol{\lambda}^* \rangle$ is determined by*

$$\Lambda^*(n, \gamma) = \arg \max_{\Lambda = \langle s, \boldsymbol{\lambda} \rangle} \left\{ \underbrace{\frac{(s-1)\gamma}{s}}_{\text{Efforts in Final}} + \underbrace{\frac{s-\gamma(s-1)}{s^2} \sum_{\tau=1}^{s} \mathbf{TE}^*(\lambda_\tau, \gamma)}_{\text{Efforts in } s \text{ Sub-contests}} \right\} \tag{6}$$

[9] The number of finalists is denoted by $n(x_R)$ in our previous analysis. However, for the sake of simplicity and clarity, we use s instead throughout designing contest architecture.

where λ_τ is the τth entry in $\boldsymbol{\lambda}$. The total effort induced by the optimal architecture is

$$\mathbf{TE}^*(n,\gamma) = \max_{\langle s,\boldsymbol{\lambda}\rangle} \left\{ \frac{(s-1)\gamma}{s} + \frac{s-\gamma(s-1)}{s^2} \sum_{\tau=1}^{s} \mathbf{TE}^*(\lambda_\tau,\gamma) \right\}. \tag{7}$$

Having derived the optimal design $\Lambda^*(\hat{n},\gamma)$ for each $\hat{n}=1,\cdots,n$, we can recover the optimal contest architecture recursively by unfolding the architecture layer by layer.

Consider a 10-player example with $\gamma = 0.9$. We sequentially derive the optimal design of the final match with $\hat{n} = 1,\cdots,10$ contestants, as listed in Table 2. The last row of Table 2 shows that two contestants participate in the final when $n = 10$: One is selected from a sub-contest with 4 contestants, and the other is chosen from a sub-contest with 6 contestants, as shown in Figure 5(a). Furthermore, based on the optimal tree architecture with $n = 4$ and $n = 6$, we can pin down the final of these two sub-contests, which is the sub-final of the original contest, as shown in Figure 5(b). Finally, we know that the optimal contest architecture is simultaneous when $n = 2$ or 3. Hence, we have recovered the optimal contest architecture according to Table 2, as shown in Figure 5(c).

$n=$	s^*	$\boldsymbol{\lambda}^*$	Total Effort	$n=$	s^*	$\boldsymbol{\lambda}^*$	Total Effort
3	3	$(1,1,1)$	0.6000	7	2	$(4,3)$	0.8068
4	2	$(2,2)$	0.6975	8	2	$(4,4)$	0.8336
5	2	$(3,2)$	0.7387	9	2	$(5,4)$	0.8450
6	2	$(3,3)$	0.7800	10	2	$(6,4)$	0.8563

Table 2: A 10-player Example for Dynamic Programming

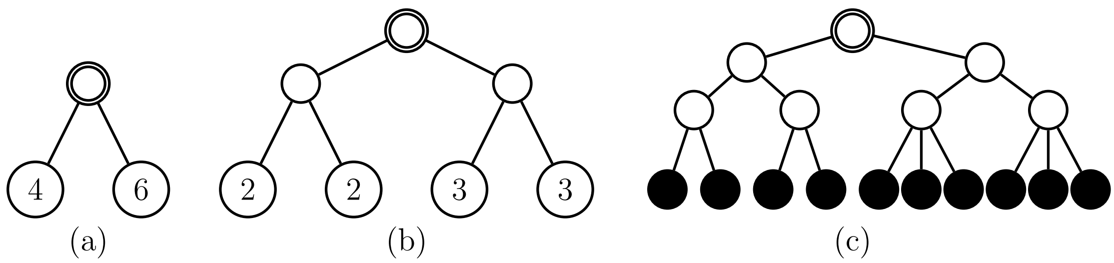

Figure 5: Recovering Optimal Contest Architecture when $\gamma = 0.9$ and $n = 10$

4.1.3 An Upper Bound of $\mathbf{TE}^*(n,\gamma)$

We can use the recursive formula for $\mathbf{TE}^*(n,\gamma)$ (Equation 7), derived from dynamic programming, to provide an upper bound for the total effort level with $\gamma < 1$. This upper bound is quite tight since as n grows, (i) the difference between the bound and $\mathbf{TE}^*(n,\gamma)$ converges to 0, and (ii) the bound can be reached infinitely many times.

Lemma 5. *When $\gamma < 1$, $\mathbf{TE}^*(n,\gamma) \le 1-(1-\frac{\gamma}{2})^{\log_2 n}$. Equality holds if and only if n is a power of 2.*

Proof. We can prove this proposition by induction.

Base Case. When $n = 2$, the total effort induced is $\frac{\gamma}{2}$. The equality holds.

Induction Process. Assume that $\mathbf{TE}^*(\hat{n},\gamma) \le 1-(1-\frac{\gamma}{2})^{\log_2 \hat{n}}$ holds for $\hat{n} = 1, \cdots, n-1$. For simplicity, we define the contestant surplus: $\mathbf{CS}^*(n,\gamma) \triangleq 1 - \mathbf{TE}^*(\hat{n},\gamma)$. Then,

$$\begin{aligned}\mathbf{CS}^*(n,\gamma) &= \min_{\langle s,\boldsymbol{\lambda}\rangle}\left\{\left[1-\frac{(s-1)\gamma}{s}\right]\frac{1}{s}\sum_{\tau=1}^{s}\mathbf{CS}^*(\lambda_\tau,\gamma)\right\}\\ &\ge \min_{\langle s,\boldsymbol{\lambda}\rangle}\left\{\left[1-\frac{(s-1)\gamma}{s}\right]\left[\frac{1}{s}\sum_{\tau=1}^{s}(1-\frac{\gamma}{2})^{\log_2\lambda_\tau}\right]\right\},\end{aligned}$$

where the inequality holds by induction hypothesis.

Step 1. We aim to show that $\frac{1}{s}\sum_{\tau=1}^{s}(1-\frac{\gamma}{2})^{\log_2\lambda_\tau} \ge (1-\frac{\gamma}{2})^{\log_2\frac{n}{s}}$.

Consider $(1-\frac{\gamma}{2})^{\log_2 n}$ as a function of n. It is convex since the second-order derivative $\frac{\ln(1-\frac{\gamma}{2})}{(n\ln 2)^2}(1-\frac{\gamma}{2})^{\log_2 n}\left[\ln(1-\frac{\gamma}{2})-\ln 2\right] > 0$. According to Jensen's inequality, we have

$$\frac{1}{s}\sum_{\tau=1}^{s}(1-\frac{\gamma}{2})^{\log_2\lambda_\tau} \ge (1-\frac{\gamma}{2})^{\log_2\frac{\sum_{\tau=1}^{s}\lambda_\tau}{s}} = (1-\frac{\gamma}{2})^{\log_2\frac{n}{s}}.$$

Here, equality holds if and only if $\lambda_1=\lambda_2=\cdots=\lambda_\tau=\frac{n}{s}$ and $\frac{n}{s}$ is an integer.

Step 2. We aim to show that $1-\frac{(s-1)\gamma}{s} \ge (1-\frac{\gamma}{2})^{\log_2 s}$. Equality holds if and only if $s = 2$. The details are relegated to the Appendix.

With the above two inequalities, we immediately have

$$\left[1-\frac{(s-1)\gamma}{s}\right]\left[\frac{1}{s}\sum_{\tau=1}^{s}(1-\frac{\gamma}{2})^{\log_2\lambda_\tau}\right] \ge (1-\frac{\gamma}{2})^{\log_2 s}(1-\frac{\gamma}{2})^{\log_2\frac{n}{s}} = (1-\frac{\gamma}{2})^{\log_2 n}.$$

Here, equality holds if and only if $s=2$, $\lambda_1=\lambda_2=\frac{n}{2}$, and $\frac{n}{2}$ is an integer.

Therefore, $\mathbf{CS}^*(n,\gamma) \ge (1-\frac{\gamma}{2})^{\log_2 n}$ and thus $\mathbf{TE}^*(n,\gamma) \le 1-(1-\frac{\gamma}{2})^{\log_2 n}$. Equality holds if and only if (i) $s=2$, (ii) $\lambda_1=\lambda_2=\frac{n}{2}$, (iii) $\frac{n}{2}$ is an integer, and (iv) equality also holds for $\frac{n}{2}$. $\square$

When a certain contest architecture can reach the upper bound, it must be optimal.

The upper bound in Lemma 5 is tight when n is a power of 2, where the optimal contest architecture is a *symmetric* (*balanced*) binary tree with $\log_2 n$ stages.

Proposition 5. *When $\gamma < 1$ and n is a power of 2, the* ***symmetric*** *binary tree contest is the* ***unique optimal*** *contest architecture.*

For the 10-player example with $\gamma = 0.9$ in Table 2, we compare the total effort and the upper bound, and summarize them into Table 3. When $n \geq 6$, the total effort is lower than the upper bound by $< 1\%$, indicating that the bound is quite accurate.

$n =$	$\mathbf{TE}^*(n, 0.9)$	Upper Bound	Gap	$n =$	$\mathbf{TE}^*(n, 0.9)$	Upper Bound	Gap
3	0.6000	0.6123	0.0123	7	0.8068	0.8133	0.0065
4	0.6975	0.6975	0	8	0.8336	0.8336	0
5	0.7387	0.7504	0.0117	9	0.8450	0.8497	0.0047
6	0.7800	0.7868	0.0068	10	0.8563	0.8628	0.0065

Table 3: A 10-player Example for the Upper Bound

4.1.4 Comparisons with Gradstein and Konrad (1999)

In Section 4.1, we study the optimal contest architecture under the winner-take-all rule when $\gamma < 1$. This topic is initially explored by Gradstein and Konrad (1999). They suggest that for $\gamma < 1$, the optimal contest architecture is a binary tree. By contrast, Proposition 3 in this paper demonstrates that when $\gamma < \frac{2}{3}$, the optimal contest architecture is indeed a binary tree. If $\gamma \in [\frac{2}{3}, 1)$, it is possible that contest architectures other than a binary tree may be optimal. Notably, we prove that the threshold $\frac{2}{3}$ is tight. On one hand, the threshold is no less than $\frac{2}{3}$ by Corollary 1. On the other hand, we can construct a scenario in which contest architecture other than binary tree is optimal when $\gamma = \frac{2}{3}$.

The discrepancy arises from limitations in Gradstein and Konrad (1999)'s proof methodology. Specifically, their approach involves three key steps: first, for a given number of stages, they insist on an equal number of contestants in each embedded match; second, they express the total effort as a function of the stage number; third, they optimize the number of stages, ultimately concluding that more stages are optimal to induce higher effort, hence suggesting a binary tree architecture as ideal for T-contests.

During their analysis, they relax the assumption that the number of participants in each match must be an integer. If the group size is g in the match x, the total effort induced is still assumed to be $\frac{(g-1)\gamma}{g}\hat{v}(x)$, even if g is no longer an integer. However, this relaxation is a loss of generality. Given the number of stages, it is generally impossible to construct a contest architecture with the same group size across all embedded matches. Consequently, this loss of generality leads to different results.[10]

Note that the optimal solution in a relaxed optimization problem must be optimal in the original problem if it meets the condition being dropped. Under the context of designing contest architecture, when n is a power of 2, the dropped integer constraint is satisfied. Hence, we anticipate that Gradstein and Konrad (1999)'s claim is valid if n is a power of 2, which can be corroborated by Proposition 5. Our proof methodology, rooted in dynamic programming and mathematical induction, boasts two advantages over their proof. First, our concise proof in Lemma 5 derives the upper bound directly from the Bellman equation, bypassing the reliance on specific contest architectures and significantly simplifying the proof. Second, for rigorous concerns, the integer constraints on the group size and the stage number are fully considered in our proof.

4.2 Relatively Discriminatory Contests: $\gamma \geq 1$

For $\gamma \geq 1$, the optimal contest architecture is a simultaneous contest. When $\gamma > 1$, it is the unique optimal one. In doing so, we use the induction method to show that simultaneous contest is optimal for $n = 2, 3, \cdots$, using the dynamic programming iteration formula Equation 7.

Proposition 6. *When $\gamma > 1$, $\mathbf{TE}^*(n,\gamma) = \frac{n-1}{n}\gamma$, and the* ***simultaneous*** *contest is the* ***unique optimal*** *T-contest architecture.*

Proof. We can prove this proposition by induction. See the Appendix for details. □

We now explain why we should gather all contestants in a grand static contest. Since the effective prize in each match is merely the equilibrium payoff of its parent match, a higher equilibrium payoff indicates a higher efficiency of adopting multi-stage contests. In a Tullock contest, the total effort of a match increases with the discriminatory power,

[10] For example, in a three-player contest, a completely symmetric binary tree architecture does not exist. Instead, there are two possible contest architectures: a simultaneous contest and an asymmetric binary tree contest. In the binary tree contest, the winner between two contestants contends with the third contestant for the championship. The total effort in simultaneous contest is $\frac{2\gamma}{3}$, and in the binary tree contest is $\frac{\gamma}{2} + \frac{2-\gamma}{4}\frac{\gamma}{2}$. The binary tree contest is optimal only if $\gamma \leq \frac{2}{3}$.

while the resulting equilibrium payoff decreases. Therefore, the stage number should be fixed at one when the contest is relatively discriminatory.

Proposition 6 is not new in the literature. Gradstein and Konrad (1999) first state that when the discriminatory power $\gamma \geq 1$, the optimal contest is a simultaneous contest. We provide an alternative approach, grounded on dynamic programming and mathematical induction, to reach the same result.

5 Concluding Remark

This paper studies the optimal joint design of contest architecture and prize structure in T-contests. We pin down the optimal contest rule through two steps. First, we present the pivotal match principle to rationalize the winner-take-all rule. Second, we characterize the optimal contest architecture under the winner-take-all rule.

Our benchmark analysis considers a linear effort cost and homogeneous contestants. As well-documented in the literature, it is challenging to relax these two assumptions. Yet, in Appendix B, we discuss these two assumptions. Another possible direction is allowing multiple contestants to advance to the next stage in each embedded match, which can be viewed as a combination of T-contests and pooling elimination contests. Subsequently, a second prize could be assigned to the runner-up of a single match.

Our analysis also raises the question of shortlisting, a significant issue that is overlooked in multi-stage contest designs. In the shortlisting problem, the contest organizer precludes a subset of contestants and applies the winner-take-all rule. Contrary to the conventional wisdom that shortlisting is detrimental to eliciting effort, our study implies that shortlisting is beneficial when the T-contest is unbalanced. Although the theoretical approach of this study contributes to the shortlisting problem (see Appendix B for our analysis), the full characterization of optimal shortlisting is beyond the scope of this study, and a comprehensive investigation should be attempted in future research.

A Proofs

Proof of Lemma 1. We prove Lemma 1 by induction. Let $h(x)$ denote the stage number of match x, then $h(x_R)$ is the total number of stages. We further let $d(x) = h(x_R) - h(x)$, which is the remaining stages left after match x.

Let $\mathcal{D}_k(T)$ denote the set of matches with $d(x) = k$, namely $\mathcal{D}_k(T) = \{x \in \mathcal{X}(T) : d(x) = k\}$. Then, $\mathcal{D}_0(T) = \{x_R\}$ and $\mathcal{D}_1(T) = \{x : p(x) = x_R\}$. For any $x \in \mathcal{D}_k(T)$, given the effective prizes $\hat{v}(x)$, the equilibrium effort of each contestant in match x is given by $e^*(x) = \frac{(n(x)-1)\gamma}{n(x)^2}\hat{v}(x)$. Then, we can further calculate the effective prizes of matches in $\mathcal{D}_{k+1}(T)$. This induction process consists of several stages that start at the base case $k = 0$.

Base Case. After the final, the winner will take the prize and the contest is over. So the effective prize of the final is just $v(x_R)$. Any contestant who advances to the final will exert a level of effort $e^*(x_R) = \frac{(n(x_R)-1)\gamma}{n(x_R)^2}\hat{v}(x_R)$. Therefore, Lemma 1 holds for the final match $\mathcal{D}_0$.

Induction Process. Assume Lemma 1 holds for all matches in $\mathcal{D}_k$. Let $x \in \mathcal{D}_{k+1}$ and hence $p(x) \in \mathcal{D}_k$. By induction assumption, the equilibrium individual effort level in match $p(x)$ is $e^*(p(x)) = \frac{(n(p(x))-1)\gamma}{n(p(x))^2}\hat{v}(p(x))$ for all $n(p(x))$ contestants, and each contestant wins match $p(x)$ with probability $\frac{1}{n(p(x))}$. Hence, the expected payoff of participating match $p(x)$, i.e., $\frac{1}{n(p(x))}\hat{v}(p(x)) - e^*(p(x))$, is the indirect benefit of winning match x. According to the definition of $r(x', x)$, $r(p(x), x) = \frac{n(p(x))-(n(p(x))-1)\gamma}{n(p(x))^2}$ and for any $x' \in \mathcal{F}(p(x))$,

$$r(x', x) = r(x', p(x)) \cdot r(p(x), x) = \frac{n(p(x)) - (n(p(x)) - 1)\gamma}{n(p(x))^2} r(x', p(x)).$$

Then, $\hat{v}(x)$ can be obtained in the following way,

$$\hat{v}(x) = v(x) + \underbrace{\left[\frac{1}{n(p(x))}\hat{v}(p(x)) - e^*(p(x))\right]}_{\text{Indirect benefit of winning match } x} = v(x) + r(p(x), x)\hat{v}(p(x)) = v(x) + \sum_{x' \in \mathcal{F}(x)} r(x', x)v(x').$$

$\square$

Proof of Lemma 2. We first prove that $H_k < 1$ by the induction method. It is apparent that $H_1 < 1$. If $H_k < 1$, then $H_{k+1} < \frac{(N_{k+1}-1)\gamma}{N_{k+1}} + \frac{N_{k+1}-(N_{k+1}-1)\gamma}{N_{k+1}} = 1$. Finally, $H_{k+1} - H_k = \frac{(N_{k+1}-1)\gamma}{N_{k+1}}(1 - H_k) > 0$, which finishes the proof. $\square$

Proof of Lemma 3. According to the pivotal match principle, we only need to focus on the winner-take-all prize allocation rule. Consider the sub-contest with final match x, denoted by $\mathcal{T}(x)$. If applying the winner-take-all rule on $\mathcal{T}(x)$ induces total effort $\alpha \in [0, 1)$, then the total effort induced in sub-contest $\mathcal{T}(x)$ in the original architecture T with the winner-take-all rule should be $\alpha\hat{v}(x)$, where α can be defined as the **rent dissipation rate** of sub-contest $\mathcal{T}(x)$.

We prove Lemma 3 by contradiction. Assume the contest architecture T is optimal and there exists one match $z \in \mathcal{X}(T)$ in the kth stage such that $n(z)$ is not a prime number. Then $n(z)$ can be decomposed into the product of two integers greater than 1, i.e., $n(z) = ab$, where $a, b > 1$. Denote those $n(z)$ matches whose parent match is z in T as $x_1, \cdots, x_{n(z)}$. That is, in stage k, the winner of matches $x_1, \cdots, x_{n(z)}$ advances to match z. Then, we could adjust the contest architecture T by adding an additional stage between stages $k-1$ and k as follows:

1. Divide winners of $x_1, \cdots, x_{n(z)}$ into b matches of a members each. The members in each match compete with each other, and the winners of those b matches form a new match, which is denoted by z'. They compete with each other in stage k to advance further.
2. Winners of matches in stage $k-1$ excluding $x_1, \cdots, x_{n(z)}$ are advanced to stage k directly.
3. In stage $k+1$, all the matches are the same as before, except the winner of match z is replaced by z'.
4. All other stages remain unchanged.

Thus, we get a new contest architecture T'.

In the following, we will show $\mathbf{TE}(T|v_{\text{WTA}}) < \mathbf{TE}(T'|v_{\text{WTA}})$, which is equivalent to proving $\mathbf{TE}(\mathcal{T}(z)|v_{\text{WTA}}) < \mathbf{TE}(\mathcal{T}(z')|v_{\text{WTA}})$.

Clearly,

$$\mathbf{TE}(\mathcal{T}(z)|v_{\text{WTA}}) = \frac{\gamma(ab-1)}{ab} + \frac{ab-\gamma(ab-1)}{a^2b^2}\sum_{i=1}^{n(z)}\alpha_i,$$

where $\alpha_i = \mathbf{TE}(\mathcal{T}(x_i)|v_{\text{WTA}})$ is a constant in $[0, 1)$.

Similarly,

$$\mathbf{TE}(\mathcal{T}(z')|v_{\text{WTA}}) = \frac{\gamma(b-1)}{b} + \frac{b-\gamma(b-1)}{b}\frac{\gamma(a-1)}{a} + \frac{a-\gamma(a-1)}{a^2}\frac{b-\gamma(b-1)}{b^2}\sum_{i=1}^{n(z)}\alpha_i.$$

Then

$$\mathbf{TE}(\mathcal{T}(z')|v_{\text{WTA}}) - \mathbf{TE}(\mathcal{T}(z)|v_{\text{WTA}}) = \frac{\gamma(1-\gamma)(a-1)(b-1)}{a^2b^2}\left(ab - \sum_{i=1}^{n(z)}\alpha_i\right) > 0,$$

which contradicts to the optimality of T.

Therefore, the optimal T-contest must be a prime-number T-contest. □

Proof of Lemma 4. The main logic of this proof is similar to the proof of Lemma 3. Assume the contest architecture T is optimal and there exists one match $z \in \mathcal{N}(T)$ in the kth stage such that $n(z) = q$ is an odd prime number. Denote those q matches whose parent match is z as $x_1, \cdots, x_q$. As in the proof of Lemma 3, define α_i as the rent dissipation rate of match x_i. Without loss of generality, we assume $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_q$. Then we could adjust contest architecture T by adding an additional stage between stages $k-1$ and k as follows:

1. Group winners of matches $x_1, \cdots, x_{\frac{q+1}{2}}$ together in a new match, which is denoted by y_1, and group winners of matches $x_{\frac{q+3}{2}}, \cdots, x_q$ together in another new match, which is denoted by y_2. The winners of matches y_1 and y_2 form a new match, which is denoted by z', and they compete with each other in stage k to advance further.
2. The winners of matches in stage $k-1$ excluding $x_1, \cdots, x_q$ are advanced to stage k directly.
3. In stage $k+1$, all the matches are the same as before, except the winner of match z is replaced by z'.
4. All other stages remain unchanged.

Thus, we obtain a new contest architecture T'.

Let p denote a prime number that is strictly lower than q. Then, we need to show that when $\gamma < \gamma_p = 1 - \frac{2}{(p+1)(p^2-2)}$, the match with q participants could not be optimal. That is to prove $\mathbf{TE}(T|v_{\text{WTA}}) < \mathbf{TE}(T'|v_{\text{WTA}})$, which is equivalent to proving $\mathbf{TE}(\mathcal{T}(z)|v_{\text{WTA}}) < \mathbf{TE}(\mathcal{T}(z')|v_{\text{WTA}})$.

Clearly,

$$
\begin{aligned}
\mathbf{TE}(\mathcal{T}(z)|v_{\text{WTA}}) &= \frac{\gamma(q-1)}{q} + \frac{q-\gamma(q-1)}{q^2}\sum_{i=1}^{q}\alpha_i, \\
\mathbf{TE}(\mathcal{T}(z')|v_{\text{WTA}}) &= \frac{\gamma}{2} + \frac{2-\gamma}{4}\frac{\gamma(q-1)}{q+1} + \frac{2-\gamma}{2}\frac{q+1-\gamma(q-1)}{(q+1)^2}\sum_{i=1}^{\frac{q+1}{2}}\alpha_i \\
&\quad + \frac{2-\gamma}{4}\frac{\gamma(q-3)}{q-1} + \frac{2-\gamma}{2}\frac{q-1-\gamma(q-3)}{(q-1)^2}\sum_{i=\frac{q+3}{2}}^{q}\alpha_i.
\end{aligned}
$$

Then we have

$$\mathbf{TE}(\mathcal{T}(z')|v_{\text{WTA}}) - \mathbf{TE}(\mathcal{T}(z)|v_{\text{WTA}}) = \mathcal{C}_0 + \mathcal{C}_1 \sum_{i=1}^{\frac{q+1}{2}} \alpha_i + \mathcal{C}_2 \sum_{i=\frac{q+3}{2}}^{q} \alpha_i,$$

where $\mathcal{C}_0 = \frac{\gamma(1-\gamma)}{2} - \frac{\gamma(1-\gamma)q^2+\gamma}{(q^2-1)q}$, $\mathcal{C}_1 = -\frac{\frac{\gamma(1-\gamma)}{2}(q-1+\frac{\gamma(1+\gamma)}{2})+(1-\gamma)q+\gamma}{(q+1)^2 q} < 0$ and $\mathcal{C}_2 = \frac{2-\gamma}{2}\frac{q-1-\gamma(q-3)}{(q-1)^2} - \frac{q-\gamma(q-1)}{q^2}$.

Since $q \geq p+1$ and hence $p \leq q-1$, then $\gamma < 1 - \frac{2}{(p+1)(p^2-2)} \leq 1 - \frac{2}{q(q^2-2q-1)}$, which implies that $\mathcal{C}_0 > 0$.

- If $\mathcal{C}_2 \leq 0$, then

$$\mathcal{C}_0 + \mathcal{C}_1 \sum_{i=1}^{\frac{q+1}{2}} \alpha_i + \mathcal{C}_2 \sum_{i=\frac{q+3}{2}}^{q} \alpha_i > \mathcal{C}_0 + \mathcal{C}_1 \frac{q+1}{2} + \mathcal{C}_2 \frac{q-1}{2} = 0.$$

- If $\mathcal{C}_2 > 0$, let some $\alpha \in [\alpha_{\frac{q+1}{2}}, \alpha_{\frac{q+3}{2}}]$, then

$$\mathcal{C}_0 + \mathcal{C}_1 \sum_{i=1}^{\frac{q+1}{2}} \alpha_i + \mathcal{C}_2 \sum_{i=\frac{q+3}{2}}^{q} \alpha_i > \mathcal{C}_0 + \mathcal{C}_1 \frac{q+1}{2}\alpha + \mathcal{C}_2 \frac{q-1}{2}\alpha.$$

  Note that $\mathcal{C}_0 + \mathcal{C}_1 \frac{q+1}{2}\alpha + \mathcal{C}_2 \frac{q-1}{2}\alpha$ is linear with α. It is strictly positive for any $\alpha \in [0,1)$ since $\mathcal{C}_0 + \mathcal{C}_1 \frac{q+1}{2}\alpha + \mathcal{C}_2 \frac{q-1}{2}\alpha = \mathcal{C}_0 > 0$ when $\alpha = 0$ and $\mathcal{C}_0 + \mathcal{C}_1 \frac{q+1}{2}\alpha + \mathcal{C}_2 \frac{q-1}{2}\alpha = 0$ when $\alpha = 1$.

Therefore, we can conclude that $\mathbf{TE}(\mathcal{T}(z')|v_{\text{WTA}}) - \mathbf{TE}(\mathcal{T}(z)|v_{\text{WTA}}) > 0$.

When $\gamma = 1 - \frac{2}{(p+1)(p^2-2)}$, we have $\mathcal{C}_0 \geq 0$ and $\mathbf{TE}(\mathcal{T}(z')|v_{\text{WTA}}) - \mathbf{TE}(\mathcal{T}(z)|v_{\text{WTA}}) \geq 0$. □

Proof of Proposition 3. Given the winner-take-all rule, the effective prize $\hat{v}(x)$ in Lemma 1 can be simplified as $\hat{v}(x) = r(x_R, x) = \left(\frac{2-\gamma}{4}\right)^{f(x)}$, and $\mathbf{TE}(x|T,v) = \frac{(n(x)-1)\gamma}{n(x)}\left(\frac{2-\gamma}{4}\right)^{f(x)}$, where $n(x) = 1$ or 2 and $f(x)$ denotes the number of nontrivial future matches.

Let $\mathcal{X}_m = \{x \in \mathcal{X}(T) : n(x) = 2, f(x) = m\}$ denote the set of nontrivial matches whose number of nontrivial future matches is m. Clearly, $\mathbf{TE}(x|T,v) > 0$ if and only if $n(x) = 2$ and x belongs to some $\mathcal{X}_m$. Then, we can calculate $\mathbf{TE}(T,v)$ by summing all $\mathbf{TE}(x|T,v)$ over different values of $f(x)$:

$$\mathbf{TE}(T,v) = \sum_{x \in \mathcal{X}(T)} \frac{(n(x)-1)\gamma}{n(x)}\left(\frac{2-\gamma}{4}\right)^{f(x)} = \sum_{m=0}^{\infty} \frac{\gamma}{2}\left(\frac{2-\gamma}{4}\right)^m |\mathcal{X}_m|, \tag{8}$$

which can be viewed as a linear function of $\{|\mathcal{X}_m|\}_{m=0}^{\infty}$. We further impose three linear constraints on $\{|\mathcal{X}_m|\}_{m=0}^{\infty}$:

- **Nonnegativity**. $|\mathcal{X}_m| \geq 0$ for all m.
- **Binarity**. $|\mathcal{X}_0| \leq 1$ and $|\mathcal{X}_{m+1}| \leq 2|\mathcal{X}_m|$. The former holds since the final match is unique. The latter holds since the winner of each match in $\mathcal{X}_{m+1}$ must take part in a match in $\mathcal{X}_m$ in the future and the matches in $\mathcal{X}_m$ only have $2|\mathcal{X}_m|$ contestants altogether.
- **Regularity**. $\sum_{m=0}^{\infty} |\mathcal{X}_m| = n - 1$. On the one hand, each nontrivial match must belong to exactly one $\mathcal{X}_m$. On the other hand, the total number of nontrivial matches should be $n-1$ since each nontrivial match eliminates one contestant.

Therefore, the optimization problem is a linear programming problem with decision variables $\{|\mathcal{X}_m|\}_{m=0}^{\infty}$, and the sum of $|\mathcal{X}_m|$ is fixed by regularity. Note that in Equation 8, the coefficient of variable $|\mathcal{X}_m|$ decreases with m. Therefore, let k be the largest integer not exceeding $\log_2(n)$, in order to maximize Equation 8, we need to make $\mathcal{X}_m$ with a smaller $m(\leq k)$ as large as possible so that the binarity constraints are binding. Combined with the regularity constraint, we can derive the unique optimal solution of Equation 8:

$$|\mathcal{X}_m| = 2^m, \forall m \leq k-1, \quad |\mathcal{X}_k| = n - 2^k, \quad |\mathcal{X}_m| = 0, \forall m > k. \tag{9}$$

Let $\mathcal{N}_m$ denote the set of contestants whose number of nontrivial future matches is m: $\mathcal{N}_m = \{i \in \mathcal{N}(T) : f(i) = m\}$. Then, for those participants of matches in $\mathcal{X}_m$, they either advance all through byes in previous matches, which belong to $\mathcal{N}_{m+1}$, or experience another nontrivial match prior to this match, namely, the winners of matches in $\mathcal{X}_{m+1}$. Note that each match in $\mathcal{X}_m$ has two participants; hence, we have the following equality

$$2|\mathcal{X}_m| = |\mathcal{X}_{m+1}| + |\mathcal{N}_{m+1}|.$$

Therefore, we can derive $\{|\mathcal{N}_m|\}_{m=0}^{\infty}$ from $\{|\mathcal{X}_m|\}_{m=0}^{\infty}$.

Only if. By Equation 9, we have $|\mathcal{N}_k| = 2^{k+1} - n$, $|\mathcal{N}_{k+1}| = 2(n - 2^k)$, and $|\mathcal{N}_m| = 0, \forall m \neq k, k+1$, where k is the largest integer not exceeding $\log_2(n)$. The contest is balanced.

If. Assume there exists an integer k such that for all natural numbers $m \neq k, k+1$, $|\mathcal{N}_m| = 0$. Then, based on the relationship $2|\mathcal{X}_m| = |\mathcal{X}_{m+1}| + |\mathcal{N}_{m+1}|$, we can obtain

1. For all $m \geq k+1$, $|\mathcal{X}_m| = 2^{m-(k+1)}|\mathcal{X}_{k+1}|$. To ensure the regularity condition, $|\mathcal{X}_{k+1}|$ must be zero and hence $|\mathcal{X}_m| = 0$ for all $m \geq k+1$.

2. For all $m \leq k-1$, $|\mathcal{X}_m| = 2^m|\mathcal{X}_0| = 2^m$, and $|\mathcal{X}_k| = 2^k - |\mathcal{N}_k|$. By regularity, we can conclude that k is the largest integer not exceeding $\log_2(n)$ and $|\mathcal{N}_k| = 2^{k+1} - n$. Then, we recover the condition in Equation 9.

$\square$

Proof of Lemma 5. It remains to prove the following inequality,

$$1 - \frac{(s-1)\gamma}{s} \geq (1 - \frac{\gamma}{2})^{\log_2 s}. \tag{10}$$

Let $\eta = \log_2 s \in [1, \infty)$ and $t = 1 - \gamma \in (0, 1)$. Then, proving Equation 10 is equivalent to prove $2^\eta t - t + 1 \geq (1+t)^\eta$. By Taylor expansion, we have $(1+t)^\eta = 1 + \sum_{i=1}^{\infty} \binom{\eta}{i} t^i$, where $\binom{\eta}{i}$ is generalized binomial that is defined by $\binom{\eta}{i} = \frac{\eta(\eta-1)\cdots(\eta-i+1)}{i(i-1)\cdots 1}$. In particular, $2^\eta = 1 + \sum_{i=1}^{\infty} \binom{\eta}{i}$ and thus $2^\eta t - t + 1 = 1 + \sum_{i=1}^{\infty} \binom{\eta}{i} t$. Hence, it remains to prove

$$\sum_{i=1}^{\infty} \binom{\eta}{i} t \geq \sum_{i=1}^{\infty} \binom{\eta}{i} t^i. \tag{11}$$

Let $\lfloor \eta \rfloor$ be the largest integer no more than η. When $i \leq \lfloor \eta \rfloor$, $\binom{\eta}{i} \geq 0$ and thus $\binom{\eta}{i} t^i \geq \binom{\eta}{i} t$ because $t \in (0,1)$, and the equality holds if and only if $i = 1$. We then move to those terms with $i > \lfloor \eta \rfloor$. For $l \geq 0$, we consider two consecutive terms, $i = \lfloor \eta \rfloor + 2l + 1$ and $i = \lfloor \eta \rfloor + 2l + 2$. On the right-hand side of Equation 11,

$$\begin{aligned}
&\binom{\eta}{\lfloor \eta \rfloor + 2l + 1} t^{\lfloor \eta \rfloor + 2l + 1} + \binom{\eta}{\lfloor \eta \rfloor + 2l + 2} t^{\lfloor \eta \rfloor + 2l + 2} \\
=&\binom{\eta}{\lfloor \eta \rfloor + 2l + 1} t^{\lfloor \eta \rfloor + 2l + 1} - \frac{\lfloor \eta \rfloor + 2l + 1 - \eta}{\lfloor \eta \rfloor + 2l + 2} \binom{\eta}{\lfloor \eta \rfloor + 2l + 1} t^{\lfloor \eta \rfloor + 2l + 2} \\
=&\binom{\eta}{\lfloor \eta \rfloor + 2l + 1} \left(t^{\lfloor \eta \rfloor + 2l + 1} - \frac{\lfloor \eta \rfloor + 2l + 1 - \eta}{\lfloor \eta \rfloor + 2l + 2} t^{\lfloor \eta \rfloor + 2l + 2} \right)
\end{aligned}$$

Similarly, on the left-hand side of Equation 11,

$$\binom{\eta}{\lfloor \eta \rfloor + 2l + 1} t + \binom{\eta}{\lfloor \eta \rfloor + 2l + 2} t = \binom{\eta}{\lfloor \eta \rfloor + 2l + 1} \left(t - \frac{\lfloor \eta \rfloor + 2l + 1 - \eta}{\lfloor \eta \rfloor + 2l + 2} t \right)$$

Notice that $\binom{\eta}{\lfloor \eta \rfloor + 2l + 1} \geq 0$, and the equality holds if and only if η is an integer. Furthermore,

$$0 < t^{\lfloor \eta \rfloor + 2l + 1} - \frac{\lfloor \eta \rfloor + 2l + 1 - \eta}{\lfloor \eta \rfloor + 2l + 2} t^{\lfloor \eta \rfloor + 2l + 2} < t^{\lfloor \eta \rfloor + 2l + 2} \left(1 - \frac{\lfloor \eta \rfloor + 2l + 1 - \eta}{\lfloor \eta \rfloor + 2l + 2} \right) < t - \frac{\lfloor \eta \rfloor + 2l + 1 - \eta}{\lfloor \eta \rfloor + 2l + 2} t.$$

This ends the proof of Equation 11:

$$\sum_{i=1}^{\infty}\binom{\eta}{i}t^i = \sum_{i=1}^{\lfloor\eta\rfloor}\binom{\eta}{i}t^i + \sum_{l=0}^{\infty}\left\{\binom{\eta}{\lfloor\eta\rfloor+2l+1}t^{\lfloor\eta\rfloor+2l+1} + \binom{\eta}{\lfloor\eta\rfloor+2l+2}t^{\lfloor\eta\rfloor+2l+2}\right\}$$
$$\leq \sum_{i=1}^{\lfloor\eta\rfloor}\binom{\eta}{i}t + \sum_{l=0}^{\infty}\left\{\binom{\eta}{\lfloor\eta\rfloor+2l+1}t + \binom{\eta}{\lfloor\eta\rfloor+2l+2}t\right\} = \sum_{i=1}^{\infty}\binom{\eta}{i}t.$$

Equality holds if and only if η is an integer and $\lfloor\eta\rfloor < 2$, suggesting that $\eta = 1$ and $s = 2$. □

Proof of Proposition 6. Base Case. When $n = 2$, the total effort induced is $\frac{\gamma}{2}$. The simultaneous contest is the only well-defined contest architecture.

Induction Process. Assume that $\mathbf{TE}^*(\hat{n},\gamma) = \frac{\hat{n}-1}{\hat{n}}\gamma$ holds for $\hat{n} = 1, \cdots, n-1$. We need to prove that $\mathbf{TE}^*(n,\gamma) = \frac{n-1}{n}\gamma$. Replacing $\mathbf{TE}^*(\lambda_\tau,\gamma)$ by $\frac{\lambda_\tau-1}{\lambda_\tau}\gamma$ (induction hypothesis) in Equation 7, we have

$$\mathbf{TE}^*(n,\gamma) = \max_{\langle s,\boldsymbol{\lambda}\rangle}\left\{\frac{(s-1)\gamma}{s} + \frac{s-\gamma(s-1)}{s^2}\sum_{\tau=1}^{s}\frac{\lambda_\tau-1}{\lambda_\tau}\gamma\right\}.$$

Since the total effort level of $\frac{n-1}{n}\gamma$ can be reached by the simultaneous contest, it remains to prove that any other design of the final match can not exceed this level,

$$\forall\langle s,\boldsymbol{\lambda}\rangle, \frac{(s-1)\gamma}{s} + \frac{s-\gamma(s-1)}{s^2}\sum_{\tau=1}^{s}\frac{\lambda_\tau-1}{\lambda_\tau}\gamma \leq \frac{n-1}{n}\gamma.$$

Since $\frac{\lambda-1}{\lambda}$ is concave in λ, $\frac{1}{s}\sum_{\tau=1}^{s}\frac{\lambda_\tau-1}{\lambda_\tau} \leq \frac{n/s-1}{n/s} = \frac{n-s}{n}$ by Jensen's inequality. Thus, it remains to show that $\frac{(s-1)\gamma}{s} + \frac{s-\gamma(s-1)}{s}\frac{n-s}{n}\gamma \leq \frac{n-1}{n}\gamma$. The inequality can be rearranged as $\gamma(\frac{1}{s}-\frac{1}{n})(s-1)(1-\gamma) \leq 0$. Since $s \in [1,n]$ and $\gamma \geq 1$, this inequality holds. □

B More Discussions

B.1 Shortlisting

In the short run, the contest organizer may not be able to adjust the tree architecture. Section 3 thus considers a reduced-form design problem that allows prize design only. We now consider another short-term design problem that allows the organizer to shortlist a subset of the contestants to participate under the winner-take-all rule.

Note that if the pivotal match is not the final match, namely $x_P \neq x_R$, those contestants whose future match is not x_P will make zero effort. In this case, the contest organizer can be strictly better off by excluding contestants that are not in $\mathcal{T}(x_P)$, provided that the entire prize goes to the top position.[11] We obtain the following.

Proposition 7. *Given the contest architecture T and the winner-take-all rule, excluding all contestants that are not in $\mathcal{T}(x_P)$ can strictly increase the total effort when the pivotal match x_P in the original architecture is not the final match.*

The conventional wisdom holds that a contest elicits higher bids when it involves more contestants. However, there exist studies pointing out that shortlisting can heat up the competition (Che and Gale, 2003, Fu, Jiao and Lu, 2015). Under the winner-take-all prize allocation rule, Proposition 7 reveals that shortlisting could be a better choice in some T-contests, which coincides with previous studies. Our result thus provides an alternative rationale for excluding contestants in contests with tree architecture.

However, excluding all contestants that are not in $\mathcal{T}(x_P)$ is not necessarily optimal. Consider a contest architecture T_{short} consisting of three sub-contests, as shown in the left panel of Figure 6: the first sub-contest (rooted at x_1) consists of only one contestant; the second sub-contest (rooted at x_2) consists of 2^k contestants through k stages of bilateral matches; and the third sub-contest (rooted at x_3) consists of $2^k - 1$ contestants through k stages of bilateral matches. With a sufficiently large k and discriminatory power $\gamma < \frac{2}{3}$, the pivotal match x_P is x_2. However, removing the first sub-contest (the right panel of Figure 6) is better than retaining $\mathcal{T}(x_2)$ only. Moreover, shortlisting may still be profitable even if the pivotal match is already the final match. Consider the same architecture T_{short} with small k and γ, in this case, the pivotal match is the final match, and removing the first sub-contest will make the contest organizer strictly better off.

Furthermore, following the same construction process when proving Theorem 2, putting those shirking contestants into the stage 1 matches of $\mathcal{T}(x_P)$ will increase the total effort. Hence, the total effort level induced in the short term by shortlisting must not exceed the level induced in the long run under joint design in our baseline setting.

[11] For the organizational example introduced before, shortlisting corresponds to reducing the staff or streamlining the organizational structure.

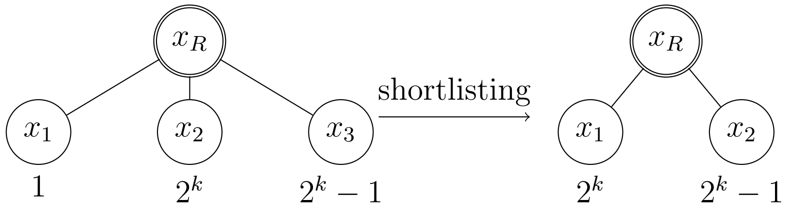

Figure 6: Shortlisting

Corollary 2. *Given the number of contestants and the winner-take-all rule, shortlisting cannot increase the total effort when contest architecture can be freely adjusted.*

B.2 Contestant Heterogeneity

Our paper assumes homogeneous players, which enables a tractable analysis but limits the scope of the study. Considering heterogeneous contestants is technically challenging for the following reasons. First, the equilibrium characterization of a single match with more than two contestants is intractable under Tullock contest technologies. Moreover, since part of the contestants may exert zero effort, it is likely that boundary (corner) solutions widely exist, which makes equilibrium characterization more difficult.[12]

Second, the subgame perfect equilibrium in each sub-contest is no longer well defined. Since each contestant is not sure of the capability of their future opponents, he could only adopt a contingent strategy depending on the realization of future opponents. Clearly, the distribution of participants is jointly determined by the strategies of all contestants that advance to the match. So do the effective prizes. Therefore, the equilibrium strategy of each contestant in match x depends not only on his own position in the contest architecture but also on the probability distribution of his opponents in the current match and all future matches. These strategies are intertwined in such a complex way that they lose the subgame optimality property as in the baseline analysis.[13]

Third, when contestants are heterogeneous, the optimal contest design also contains seeding—the initial assignment of contestants in the tree architecture. However, taking seeding as the unique design instrument is already complicated enough for analytical

[12] Cornes and Hartley (2005) and Ryvkin (2013) investigate the properties of a simultaneous Tullock contest with heterogeneous players. Gurtler and Kräkel (2010) and Parreiras and Rubinchik (2006) consider tournament settings with heterogeneous agents.

[13] In a two-stage lottery rent-seeking group contest, Stein and Rapoport (2004) study $k(\geq 2)$ groups with asymmetric valuations of the rent (the members within each group are homogeneous) and compare the contest structures of Between-Group and Semi-Finals models.

results. The existing literature is limited to the discussion of the two-stage competition. Groh et al. (2012) study optimal seeding with four heterogeneous participants in an all-pay auction. Stracke (2013) studies the case of four players of two types, and compares the equilibrium results between a static one-stage contest and a dynamic two-stage contest. The experimental work done by Hörtnagl et al. (2013) further explores how heterogeneity in contestants' investment costs affects competition intensity in a two-stage contest.

Nevertheless, the expected total effort in equilibrium is approximately unchanged after averaging the marginal costs of weakly heterogeneous players. Ryvkin (2009) argues that weak heterogeneity is a more reasonable assumption than arbitrary heterogeneity under an evolutionary perspective.[14] Moreover, Fang, Noe and Strack (2020) claim that preselected contestants, such as employees who all passed employment screening tests, can be treated as homogeneous players. We simply follow their assumption of homogeneity.

B.3 Nonlinear Costs

The previous literature points out that the optimal contest design often depends on the shape of contestants' effort cost curves in dynamic contests (e.g., Moldovanu and Sela (2001)). In this section, we relax the assumption of linear effort cost functions and investigate the impact of the curvature of cost functions on the optimal design scheme.

Consider a power form cost function $c(e) = e^{\beta}$ with $\beta > 0$. Therefore $\beta > 1$ ($\beta = 1$ or $\beta < 1$) implies increasing (constant or decreasing) marginal cost of effort. Similar to Lemma 1, the subgame perfect Nash equilibrium of the T-contest is unique, and the equilibrium effort of each contestant in match $x \in \mathcal{X}(T)$ is $e^*(x) = \left(\frac{\gamma}{\beta}\frac{n(x)-1}{n(x)^2}\hat{v}(x)\right)^{1/\beta}$, where $\hat{v}(x) = v(x) + \sum_{x' \in \mathcal{F}(x)} r(x', x)v(x')$ denotes the effective prize of match x, and $r(x', x) = \prod_{z \in \mathcal{F}(x) \setminus \mathcal{F}(x')} \left[\frac{n(z) - \frac{\gamma}{\beta}(n(z)-1)}{n(z)^2}\right]$ denotes the coefficient of $v(x')$ for $x' \in \mathcal{F}(x)$.

Furthermore, we can extend $r(x', x)$ in a similar way as Equation 2: define $r(x', x) = 1$ if $x' = x$, and $r(x', x) = 0$ if $x' \notin \{x\} \cup \mathcal{F}(x)$. Therefore, given the contest architecture T, the total effort induced by prize structure v is

$$\mathbf{TE}(T, v) = \sum_{x \in \mathcal{X}(T)} n(x) \left(\sum_{x' \in \mathcal{X}(T)} \frac{\gamma}{\beta} \frac{n(x) - 1}{n^2(x)} r(x', x) v(x') \right)^{1/\beta}. \tag{12}$$

[14] Ryvkin (2009) studies a binary elimination contest with weakly heterogeneous players. It shows that the expected total effort in equilibrium under linear approximation remains unchanged after averaging their marginal costs. Similarly, weak heterogeneity will not alter the main results of our paper.

B.3.1 Pivotal Match or Prize Splitting?

Given the contest architecture T, the total effort is a quasi-convex function of $\{v(x)\}_{x\in\mathcal{X}(T)}$ when $\beta \leq 1$.[15] Since the domain of $\{v(x)\}_{x\in\mathcal{X}(T)}$ is a simplex, the maximum point must be located at a vertex, which validates the pivotal match principle.

Proposition 8. *When $\beta \leq 1$, (i) Given the contest architecture T, the optimal prize structure is $v^*(x_P) = 1$, where x_P maximizes $\sum_{z\in\mathcal{X}(T)} n(z)\left(\frac{\gamma}{\beta}\frac{n(z)-1}{n(z)^2}r(x,z)\right)^{1/\beta}$. (ii) It is optimal for the contest organizer to choose the winner-take-all prize allocation rule when she can jointly design the contest architecture and prize structure.*

Proof. (i) We first show that $\mathbf{TE}(T,v)^\beta$ is a convex function with respect to v, i.e., for any $\theta \in (0,1)$, $\mathbf{TE}(T,\theta v_1 + (1-\theta)v_2)^\beta \leq \theta\mathbf{TE}(T,v_1)^\beta + (1-\theta)\mathbf{TE}(T,v_2)^\beta$.

Let $H(T)_x$ denote the column vector $H(T)_x = \left(n(x)^\beta\frac{\gamma}{\beta}\frac{n(x)-1}{n(x)^2}r(x',x)\right)_{x'\in\mathcal{X}(T)}$ and $H(T)$ denote the matrix $H(T) = (H(T)_x)_{x\in\mathcal{X}(T)}$. Let V denote the column vector of $v(\cdot)$, i.e., $V = (v(x))_{x\in\mathcal{X}(T)}$. Hence, $\mathbf{TE}(T,v)^\beta = \|H(T)V\|_{\frac{1}{\beta}}$ is a $\frac{1}{\beta}$-norm.[16]

According to triangle inequality in L_p norm, we have

$$\|H(T)\left[\theta V_1 + (1-\theta)V_2\right]\|_{\frac{1}{\beta}} \leq \|H(T)\theta V_1\|_{\frac{1}{\beta}} + \|H(T)(1-\theta)V_2\|_{\frac{1}{\beta}} = \theta\|H(T)V_1\|_{\frac{1}{\beta}} + (1-\theta)\|H(T)V_2\|_{\frac{1}{\beta}},$$

which implies $\mathbf{TE}(T,v)^\beta$ is convex in v.

Under the budget constraint, the feasible region of prize structure v is a simplex with the dimension of $|\mathcal{X}(T)|$. Therefore, $\mathbf{TE}(T,v)^\beta$ is maximized when v is a vertex of the simplex, namely

$$v^*(x) = \begin{cases} 1, & x = x_P, \\ 0, & \text{otherwise}. \end{cases}$$

(ii) Apply a method similar to that used to prove Theorem 2, when $x_P \neq x_R$, we can construct another contest architecture T' that induces strictly higher effort than T: Suppose there are m contestants in the sub-contest with architecture $\mathcal{T}(x_P)$, then applying the winner-take-all rule in $\mathcal{T}(x_P)$ will induce same total effort as in T with $v(x_P) = 1$. Based on the contest architecture $\mathcal{T}(x_P)$, fixing the number of stages and the number of matches in each stage, we construct T' by randomly selecting one contestant

[15] In the proof of Proposition 8, we will show that $\mathbf{TE}(T,v)^\beta$ is a convex function with respect to v, then it must be a quasi-convex function with respect to v, so does $\mathbf{TE}(T,v)$.

[16] Since $\frac{1}{\beta} > 1$ only if $\beta < 1$, this norm is no longer well defined when $\beta > 1$.

(say contestant i) in the first stage of $T(x_P)$ and grouping him with those $n-m$ shirking contestants into a new match of stage 0. Then, the effective prize of each match in T' except the new match of stage 0 will be the same as in $\mathcal{T}(x_P)$ since it is solved backward. And T' will induce strictly higher effort than $\mathcal{T}(x_P)$ since it includes an extra match in stage 0. $\square$

However, when $\beta > 1$, the total effort function is no longer quasi-convex, the same argument fails, and a pivotal match may not necessarily exist. Nevertheless, in the following, we present a necessary and sufficient condition that assigning all prizes to a single match is optimal, which also determines whether prize splitting is beneficial to the contest organizer.

Proposition 9. *Given the contest architecture T, (i) When $\beta \geq 1$, x is a pivotal match if and only if*

$$\sum_{z \in \mathcal{X}(T)} \frac{n(z)-1}{n(z)} (r(x',z) - r(x,z)) \left(\frac{n(z)-1}{n^2(z)} r(x,z) \right)^{\frac{1}{\beta}-1} \leq 0, \forall x'. \tag{13}$$

If no such x exists, the prize should be split into multiple matches. (ii) When $\beta > 1$, the optimal prize structure is either winner-take-all or prize splitting.

Proof. (i) Without loss of generality, assume the nodes of tree architecture T, $x_1, x_2, \cdots, x_n$, are topologically sorted. That is, $x_j \notin \mathcal{F}(x_i), \forall j > i$. Clearly, the first node x_1 must be the root x_R. Define $b_{ij} = \frac{\gamma}{\beta} \frac{n(x_i)-1}{n^2(x_i)} r(x_j, x_i)$, and let the matrix B collect all entries,

$$B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix}.$$

Then, we can define a column vector $\boldsymbol{w} = (w_1, \cdots, w_n)$ such that $\boldsymbol{w} = B\boldsymbol{v}$, where $\boldsymbol{v} = (v(x_1), \cdots, v(x_n))$ denotes the column vector representing the prize allocation rule $v(\cdot)$. Hence, the total effort of all players in the contest can be expressed as

$$\mathbf{TE}(T, v) = \sum_i n(x_i) \left(\sum_j b_{ij} v(x_j) \right)^{1/\beta} = \sum_i n(x_i) w_i^{1/\beta} \tag{14}$$

We then prove Proposition 9(i) in two steps: Firstly, we prove that the Hessian matrix, denoted by $\nabla^2\mathbf{TE}$, is negative definite, if we treat $\mathbf{TE}$ as a function of $\{v_i\} = \{v(x_i)\}$; Secondly, we prove that Equation 13 is a necessary and sufficient condition for a match to be pivotal by calculating the Karush-Kuhn-Tucker (KKT) conditions.

Step 1. Recall the definition of $r(x', x)$ in Section B.3, we can determine the sign of b_{ij}: $b_{ij} = 0$ for $i > j$; $b_{ij} \geq 0$ for $i < j$; and $b_{ij} > 0$ for $i = j$.

Since B is an upper triangular matrix with positive real-number diagonals, B is invertible. Note that $\frac{\partial \mathbf{TE}}{\partial w_i \partial w_j} = 0$ for $i \neq j$. According to the compound function derivation rule, the Hessian matrix can be written as $\nabla^2\mathbf{TE} = B^T\mathbf{diag}(\zeta_1, \cdots, \zeta_n)B$, where the entries of the diagonal matrix $\mathbf{diag}(\zeta_1, \cdots, \zeta_n)$ are $\zeta_i = \sum_i \frac{1}{\beta}(\frac{1}{\beta} - 1)n(x_i)w_j^{\frac{1}{\beta}-2} < 0$. Thus, $\nabla^2\mathbf{TE}$ is negative definite.

Step 2. Treating $\mathbf{TE}$ as a function of $\{v(x_i)\}$, the Lagrangian of optimization problem $\max \mathbf{TE}(T, v)$ s.t. $v_i \geq 0, \sum_i v_i \leq 1$ is

$$\min L(v, \mu, \lambda) = -\mathbf{TE}(T, v) + \mu(\sum v_i - 1) - \lambda_i v_i,$$

$$\begin{cases} v_i \geq 0, \sum_i v_i \leq 1, \\ \mu, \lambda_i \geq 0, \\ \lambda_i v_i = 0, \mu(\sum v_i - 1) = 0. \end{cases} \tag{15}$$

Since $\nabla^2\mathbf{TE}$ is negative definite, the prize structure $v^*(x_k) = 1$ maximizes $\mathbf{TE}(T, v)$ if and only if v^* satisfies the KKT condition

$$\begin{cases} \nabla_v L(v, \mu, \lambda) = 0, \\ v_i \geq 0, \sum_i v_i \leq 1, \\ \mu, \lambda_i \geq 0, \\ \lambda_i v_i = 0, \mu(\sum v_i - 1) = 0. \end{cases} \tag{16}$$

Substitute $v^*(x_k) = 1$ into Equation 16, the KKT conditions can be simplified to

$$\begin{cases} -\frac{\partial}{\partial v_i}\mathbf{TE}(T, v)|_{v=v^*} + \mu - \lambda_i = 0 \\ \mu, \lambda_i \geq 0, \\ \lambda_k = 0. \end{cases} \tag{17}$$

Then

$$\frac{\partial}{\partial v_i}\mathbf{TE}(T, v)|_{v=v^*} - \frac{\partial}{\partial v_k}\mathbf{TE}(T, v)|_{v=v^*} = (\mu - \lambda_i) - (\mu - \lambda_k) = -\lambda_i \leq 0, \forall i. \tag{18}$$

Recall the expression of $\mathbf{TE}(T, v)$ in Equation 12. Then

$$\frac{\partial}{\partial v_i}\mathbf{TE}(T, v)|_{v=v^*} = \sum_{j=1}^{n} \frac{1}{\beta}(\frac{\gamma}{\beta})^{1/\beta} \frac{n(x_j) - 1}{n(x_j)} r(x_i, x_j) \left(\frac{n(x_j) - 1}{n^2(x_j)} r(x_k, x_j)\right)^{\frac{1}{\beta}-1},$$
$$\frac{\partial}{\partial v_k}\mathbf{TE}(T, v)|_{v=v^*} = \sum_{j=1}^{n} \frac{1}{\beta}(\frac{\gamma}{\beta})^{1/\beta} \frac{n(x_j) - 1}{n(x_j)} r(x_k, x_j) \left(\frac{n(x_j) - 1}{n^2(x_j)} r(x_k, x_j)\right)^{\frac{1}{\beta}-1}.$$

Therefore, Equation 18 is equivalent to

$$\begin{aligned} & \frac{\partial}{\partial v_i}\mathbf{TE}(T, v)|_{v=v^*} - \frac{\partial}{\partial v_k}\mathbf{TE}(T, v)|_{v=v^*} \\ = \; & \frac{1}{\beta}(\frac{\gamma}{\beta})^{1/\beta} \sum_{j=1}^{n} \frac{n(x_j) - 1}{n(x_j)} (r(x_i, x_j) - r(x_k, x_j)) \left(\frac{n(x_j) - 1}{n^2(x_j)} r(x_k, x_j)\right)^{\frac{1}{\beta}-1} \leq 0, \forall i. \end{aligned}$$

Setting $x_k = x, x_i = x', x_j = z$ and removing the constant $\frac{1}{\beta}(\frac{\gamma}{\beta})^{1/\beta}$, we obtain Equation 13.

(ii) Assume a pivotal match x is not the final x_R. Then there must exist a node z that is not the descendant of x, which satisfies $r(x_R, z) > 0$ and $r(x, z) = 0$. Thus,

$$\frac{n(z) - 1}{n(z)} (r(x_R, z) - r(x, z)) \left(\frac{n(z) - 1}{n^2(z)} r(x, z)\right)^{\frac{1}{\beta}-1} = +\infty.$$

This contradicts to Equation 13 when we set $x' = x_R$. Therefore, only x_R could be the pivotal match. □

Equation 13 shows that the contest organizer is more likely to split the prize when the cost structure becomes more convex. For a match x meeting Equation 13, its associated $\{r(x, z)\}_{z \in \mathcal{X}(T)}$ needs to be averagely larger than $\{r(x', z)\}_{z \in \mathcal{X}(T)}$ for any other matches.

The term $r(x',z) - r(x,z)$ is negative for some z with a large $r(x,z)$ and is positive otherwise. Notice that $\left(\frac{n(z)-1}{n^2(z)} r(x,z)\right)^{\frac{1}{\beta}-1}$ is relatively lower for a larger $r(x,z)$, and it goes down further as β grows. When β becomes larger, the weight of $r(x',z) - r(x,z)$ is lower (higher) if it is negative (positive). Therefore, under the more demanding condition, splitting the prize is more likely to occur.

In particular, when $\beta = 1$, Equation 13 can be simplified to

$$\sum_{z \in \mathcal{X}(T)} \frac{n(z)-1}{n(z)} \big(r(x',z) - r(x,z)\big) \leq 0, \forall x'. \tag{19}$$

Recall that $H(T,x)$ is defined as $\sum_{z \in \mathcal{X}(T)} \frac{(n(z)-1)\gamma}{n(z)} r(x,z)$. Hence, the left-hand side of Equation 19 can be further expressed as

$$\sum_{z \in \mathcal{X}(T)} \frac{n(z)-1}{n(z)} r(x',z) - \sum_{z \in \mathcal{X}(T)} \frac{n(z)-1}{n(z)} r(x,z) = \frac{H(T,x') - H(T,x)}{\gamma}.$$

Then, the match $x_P = \arg\max_{x \in \mathcal{X}(T)} H(T,x)$ satisfies the Equation 19. Proposition 9(i) is reduced to Proposition 1.

Above analysis suggests that the winner-take-all rule can be rationalized under concave but not convex costs. In addition, due to the lack of quasi-convexity under convex costs, numerical methods cannot solve the general-form optimal prize rule, since the output may be locally optimal but not globally optimal. In the following, we first validate the robustness of the dynamic programming method given the winner-take-all rule, and then work on a joint design problem with four contestants to obtain corresponding intuitions.

B.3.2 Optimal Contest Architecture under Winner-Take-All Rules

Following the notations in the baseline analysis, let $\Lambda(n,\gamma) = \langle s, \boldsymbol{\lambda} \rangle$ denote the design of the final match in the T-contest. Then, the total effort of the final match is $\left(\frac{\gamma}{\beta}\frac{s-1}{s^2}\right)^{1/\beta}$, and the effective prize of participating in the final competition is $\frac{s-\frac{\gamma}{\beta}(s-1)}{s^2}$. Let $\Lambda^*(n,\gamma)$ and $\mathbf{TE}^*(n,\gamma)$ denote the optimal design of the final match and the maximum total effort, respectively. Given the maximum total effort of the sub-contest with λ_τ contestants $\mathbf{TE}^*(\lambda_\tau,\gamma)$, the total effort induced in the sub-contest with λ_τ contestants in the original contest architecture is just $\left(\frac{s-\frac{\gamma}{\beta}(s-1)}{s^2}\right)^{1/\beta} \mathbf{TE}^*(\lambda_\tau,\gamma)$. The Bellman Equation 20 shown in Proposition 10 characterizes the optimal T-contest architecture with n contestants under nonlinear costs, so that we can store the optimal design of the final match $\Lambda^*(n,\gamma)$ in the dynamic programming process.

Proposition 10. *Given* $\mathbf{TE}^*(\hat{n},\gamma)$ *for* $\hat{n}=1,\cdots,n-1$*, the optimal design of the final match* $\Lambda^*(n,\gamma)=\langle s^*,\boldsymbol{\lambda}^*\rangle$ *is determined by*

$$\Lambda^*(n,\gamma)=\arg\max_{\Lambda=\langle s,\boldsymbol{\lambda}\rangle:s\in\{2,\cdots,n\}}\left\{\underbrace{\left(\frac{\gamma}{\beta}\frac{s-1}{s^2}\right)^{1/\beta}}_{\text{Efforts in Final}}+\underbrace{\left(\frac{s-\frac{\gamma}{\beta}(s-1)}{s^2}\right)^{1/\beta}\sum_{\tau=1}^{s}\mathbf{TE}^*(\lambda_\tau,\gamma)}_{\text{Efforts in } s \text{ Sub-contests}}\right\} \tag{20}$$

where λ_τ *is the* τ*th entry in* $\boldsymbol{\lambda}$*.*

The total effort induced by the optimal contest architecture is

$$\mathbf{TE}^*(n,\gamma)=\left(\frac{\gamma}{\beta}\frac{s^*-1}{(s^*)^2}\right)^{1/\beta}+\left(\frac{s^*-\frac{\gamma}{\beta}(s^*-1)}{(s^*)^2}\right)^{1/\beta}\sum_{\tau=1}^{s^*}\mathbf{TE}^*(\lambda^*_\tau,\gamma). \tag{21}$$

B.3.3 Optimal Joint Design: an Example

Consider the joint design problem with four players in a T-contest.

Given a binary tree architecture shown in Figure 7, we assume $\frac{\gamma}{\beta}\in(0,2)$ to ensure pure strategy equilibrium in each match and avoid trivial analysis. Consider the prize structure v_s such that $v_s(x_R)=s\in[0,1]$ and $v_s(x_1)=v_s(x_2)=\frac{1-s}{2}$. Then, v_0 means halving all prizes to both semifinals, and v_1 means winner-take-all.[17] According to Equation 12, the total effort induced by prize structure v_s is $\mathbf{TE}(T,v_s)=2\left(\frac{\gamma s}{4\beta}\right)^{\frac{1}{\beta}}+4\left[\frac{\gamma}{4\beta}\left(\frac{1-s}{2}+(\frac{2\beta-\gamma}{4\beta})s\right)\right]^{\frac{1}{\beta}}$.

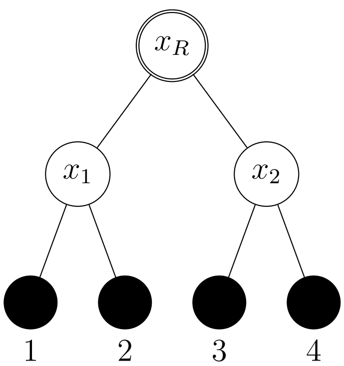

Figure 7: 4-player T-Contest

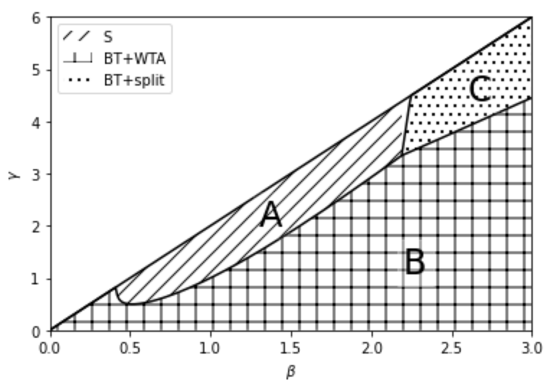

Figure 8: Joint Design

[17]Clearly, any prize allocation rule with $v_s(x_1)\neq v_s(x_2)$ is dominated. Hence, assigning all budgets to one semifinal is strictly dominated by v_0. Only the final x_R could be a pivotal match.

Clearly, solving the optimal prize structure based on a transcendental equation is difficult, and incorporating the design of the contest architecture is even more challenging. In the following, we use numerical software to find the optimal contest architecture and reward scheme, thereby gaining some insight for general joint design. Among possible tree architectures, only the simultaneous contest and the strong balanced binary tree contest could be optimal. The simultaneous contest is necessarily associated with the winner-take-all rule. While for the strong balanced binary tree, either winner-take-all or prize splitting could be optimal. Therefore, we need to compare three potential strategies.

Figure 8 summarizes the optimal strategy for different values of β and γ with $\gamma < 2\beta$. Given β, when γ is relatively small such that (β, γ) falls in area **B**, applying the winner-take-all rule in a complete binary tree architecture (`BT+WTA` for short) is optimal. While with a relatively large γ, if β is small such that (β, γ) falls in area **A**, the simultaneous contest (`S` for short) is optimal. Otherwise, (β, γ) falls in area **C**, splitting the prize in a complete binary tree architecture (`BT+Split` for short) is optimal.

Figure 8 shows that `BT+WTA` is optimal with a relatively small γ. On the one hand, given β, `S` is better than `BT+WTA` for a large γ, since the simultaneous contest tends to dissipate more rent.[18] On the other hand, `BT+Split` is better than `BT+WTA` for a large γ when $\beta > 1$. According to Proposition 9, the winner-take-all rule is better than prize splitting if and only if $(\frac{\gamma}{2\beta})(\frac{1}{2} - \frac{\gamma}{4\beta})^{\frac{1}{\beta}-1} < 1$. Since $(\frac{\gamma}{2\beta})(\frac{1}{2} - \frac{\gamma}{4\beta})^{\frac{1}{\beta}-1}$ increases with γ when $\beta > 1$, `BT+WTA` is better when γ is small.[19]

Under the winner-take-all rule, contestants have more incentive to exert effort when competitions become more discriminatory ($\gamma \to 2\beta^-$). A concave cost with decreasing marginal cost further strengthens this competition effect. However, with a convex cost of effort, the increasing marginal cost will discourage contestants from making an intense effort. Therefore, splitting the prize, which reduces the high marginal cost associated with winner-take-all, is optimal.

Further note that when the cost function is sufficiently concave (i.e., β is sufficiently small), `BT+WTA` is always the optimal design. This is because concave cost implies economies of scale, and monotonicity under linear costs (namely, the total effort is monotonic with the number of contestants) becomes less effective for effort elicitation.

[18] For example, when $\beta = 1$, the simultaneous contest is better if $\gamma > 1$ (Proposition 6).

[19] For any $\beta > 1$ and $\gamma \to 2\beta^-$, the winner-take-all rule is no longer optimal. When $\gamma \to 2\beta^-$, $\frac{1}{2} - \frac{\gamma}{4\beta} \to 0^+$. Since $\frac{1}{\beta} - 1$ is negative, $(\frac{1}{2} - \frac{\gamma}{4\beta})^{\frac{1}{\beta}-1}$ goes to positive infinity and $(\frac{\gamma}{2\beta})(\frac{1}{2} - \frac{\gamma}{4\beta})^{\frac{1}{\beta}-1} > 1$.

Therefore, aggregating all contestants in a simultaneous contest is never optimal.

References

Amegashie, J. A. 1999. "The design of rent-seeking competitions: Committees, preliminary and final contests." *Public Choice*, 99: 63–76.

Baye, M. R., D. Kovenock, and C. G. de Vries. 1999. "The incidence of overdissipation in rent-seeking contests." *Public Choice*, 99(3-4): 439–454.

Che, Y., and I. Gale. 2003. "Optimal design of research contests." *American Economic Review*, 93(3): 646–671.

Cohen, C., T. Kaplan, and A. Sela. 2008. "Optimal rewards in contests." *RAND Journal of Economics*, 39: 434–451.

Cohen, N., G. Maor, and A. Sela. 2018. "Two-stage elimination contests with optimal head starts." *Review of Economic Design*, 22(3-4): 177–192.

Cornes, R., and R. Hartley. 2005. "Asymmetric contests with general technologies." *Economic Theory*, 26: 923–946.

Drugov, M., and D. Ryvkin. 2020. "Tournament rewards and heavy tails." *Journal of Economic Theory*, 190: 105116.

Fang, D., T. Noe, and P. Strack. 2020. "Turing up the heat: the discouraging effect of competition in contests." *Journal of Political Economy*, 128(5): 1940–1975.

Feng, X., Q. Jiao, Z. Kuang, and J. Lu. 2024. "Optimal prize design in team contests with pairwise battles." *Journal of Economic Theory*, 215: 105765.

Fu, Q., and J. Lu. 2012. "The optimal multi-stage contest." *Economic Theory*, 51: 351–382.

Fu, Q., and Z. Wu. 2019. "Contests: Theory and topics." *Oxford Research Encyclopedia of Economics and Finance.*

Fu, Q., Q. Jiao, and J. Lu. 2015. "Contests with endogenous entry." *International Journal of Game Theory*, 44: 387–424.

Fu, Q., Z. Wu, and Y. Zhu. 2022. "On equilibrium existence in generalized multi-prize nested lottery contests." *Journal of Economic Theory*, 200: 105377.

Gradstein, M., and K. A. Konrad. 1999. "Orchestrating rent seeking contests." *Economic Journal*, 109: 536–545.

Groh, C., B. Moldovanu, A. Sela, and U. Sunde. 2012. "Optimal seedings in elimination tournaments." *Economic Theory*, 49: 59–80.

Gurtler, O., and M. Kräkel. 2010. "Optimal tournament contracts for heterogeneous workers." *Journal of Economic Behavior & Organization*, 75: 180–191.

Hörtnagl, T., R. Kerschbamer, R. Stracke, and U. Sunde. 2013. "Heterogeneity in rent-seeking contests with multiple stages: theory and experimental evidence." *working paper.*

Knyazev, D. 2017. "Optimal prize structures in elimination contests." *Journal of Economic Behavior & Organization*, 139: 32–48.

Konrad, K. A. 2009. *Strategy and dynamics in contests.* Oxford University Press.

Krishna, V., and J. Morgan. 1998. "The winner-take-all principle in small tournaments." *Advances in Applied Microeconomics*, 7: 61–74.

Letina, I., S. Liu, and N. Netzer. 2020. "Delegating performance evaluation." *Theoretical Economics*, 15: 477–509.

Moldovanu, B., and A. Sela. 2001. "The optimal allocation of prizes in contests." *American Economic Review*, 91(3): 542–558.

Moldovanu, B., and A. Sela. 2006. "Contest architecture." *Journal of Economic Theory*, 126: 70–96.

Parreiras, S., and A. Rubinchik. 2006. "Contests with three or more heterogeneous agents." *Games and Economic Behavior*, 68: 703–715.

Rosen, S. 1986. "Prizes and incentives in elimination tournaments." *American Economic Review*, 76(4): 701–715.

Ryvkin, D. 2009. "Tournaments of weakly heterogeneous players." *Journal of Public Economic Theory*, 11(5): 819–855.

Ryvkin, D. 2013. "Heterogeneity of players and aggregate effort in contests." *Journal of Economics & Management Strategy*, 22(4): 728–743.

Stein, W. E., and A. Rapoport. 2004. "Asymmetric two-stage group rent-seeking: comparison of two contest structures." *Public Choice*, 118: 151–167.

Stracke, R. 2013. "Contest design and heterogeneity." *Economic Letters*, 121: 4–7.

Xiao, J. 2016. "Asymmetric all-pay contests with heterogeneous prizes." *Journal of Economic Theory*, 163: 178–221.