# Optimal Seedings in Elimination Tournaments

Christian Groh, Benny Moldovanu, Aner Sela and Uwe Sunde*

March, 2008

**Abstract**

We study an elimination tournament with heterogenous contestants whose ability is common-knowledge. Each pair-wise match is modelled as an all-pay auction. Equilibrium efforts are in mixed strategies, yielding complex dynamics: endogenous win probabilities in each match depend on other matches' outcome through the identity of the expected opponent in the next round. The designer seeds competitors according to their ranks. For tournaments with four players we find optimal seedings for three different criteria: 1) maximization of total tournament effort; 2) maximization of the probability of a final among the two top ranked teams; 3) maximization of the win probability for the top player. We also find the seedings ensuring that higher ranked players have a higher winning probability. We compare our predictions with data from NCAA basketball tournaments.

*JEL Classification Numbers:* D72, D82, D44

*Keywords:* Elimination tournaments, Seedings, All-Pay Auctions

---

* We wish to thank Alex Ilek, Mike Baye, Dan Kovenock and two anonymous referees for very helpful comments. Groh and Moldovanu: Department of Economics, University of Bonn, Lennestr. 37, 53113 Bonn, Germany. Email: groh@uni-bonn.de, mold@uni-bonn.de. Sela: Department of Economics, Ben-Gurion University of the Negev Beer–Sheva 84105, Israel. Email: anersela@bgu.ac.il; Sunde: IZA, P.O Box 7240, 53072 Bonn, Germany. Email: sunde@iza.org. [seedmansci.tex]

# 1 Introduction

Kentucky and Arizona, the highest-ranked teams to reach the Final Four during the 2003 National Collegiate Athletic Association (NCAA) Basketball March Madness, were on the same bracket and therefore could meet only in the semifinal. Not for the first time, an emotional debate began: should the Final Four teams be reseeded after the regional finals, placing the two top teams in separate national semifinals with the highest ranked team facing the lowest ranked?[1]

The present paper offers a simple game-theoretic model of an elimination tournament and an analysis of the effect of seedings on several performance criteria. For example, our model predicts that the probability of a final among the two top-ranked teams without reseeding (i.e., as occurring under random seeding) is in fact equal to the probability of such a final if reseeding is done according to the method described above.[2] But, we also show that careful reseeding will increase the probability that the top-ranked team actually wins the tournament, and we show that there exists a reseeding method - different from the one described above - which increases the probability of a final among the two top-ranked teams. Finally, we compare our theoretical results with historical data from the NCAA basketball tournament.

In single elimination (or knockout) tournaments teams or individual players play pair-wise matches. The winner advances to the next round while the loser is eliminated from the competition. Many sportive events (or their respective final stages, sometimes called playoffs) are organized in such a way. Examples are the ATP tennis tournaments, professional playoffs in US-basketball, -football, -baseball and -hockey, NCAA college basketball, the FIFA (soccer) world-championship playoffs, the UEFA champions' league, Olympic disciplines such as fencing, boxing and wrestling, and top-

[1] For example, a similar event occurred in 1996 where the top ranked Kentucky and Massachusetts also met in a semifinal that was thought to be the "real" final. For the full story see USA Today, March 25, 2003. We assume that all US readers are experts in the mechanics of this tournament. Ignorant individuals (this group previously included the present authors) can find useful information at http://www.sportsline.com/collegebasketball/.

[2] I.e., where the top team meets the lowest ranked team in one semifinal, while the second and third ranked teams meet in the other.

level bridge and chess tournaments. There are also numerous elimination tournaments among students that solve scientific problems, and even tournaments among robots or algorithms that perform certain tasks. Less rigidly structured variants of elimination tournaments are also used within firms, for promotions or budgeting decisions, and by committees who need to choose among several alternatives.

A widely observed procedure in elimination tournaments is to rank competitors based on some historically observed performance, and then to match them according to their ranks: the team or player that is historically considered to be best (or ranks first after some previous stage of the tournament) meets the lowest ranked player, the second best team meets the second lowest team and so on. In the second round, the winner of the highest ranked vs. lowest ranked match meets the (expected) lowest ranked winner from the first round, and so on[3]. The above design logic is deeply ingrained in our mind. For example, Webster's College Dictionary defines the relevant meaning of the verb "to seed" as:

> "**a.** to rank (players or teams) by past performance in arranging tournament pairings, so that the most highly ranked competitors will not play each other until later rounds. **b.** to arrange (pairings or a tournament) by means of such a ranking."

As the above quotation makes clear, the raison d'être of seeding is to protect top teams from early elimination: two teams ranked among the top $2^N$ should not meet until the field has been reduced to $2^N$ or fewer teams. In particular, the two highest ranked teams can meet only in the final, and, with the above seeding method, indeed meet there if there are no surprises along the way. Presumably, this delivers the most thrilling match in the final **!**. An outcome where these teams meet in an earlier round greatly reduces further interest in the tournament and probably does not make financial sense[4]

[3]This design is used, for example, in the professional basketball (NBA)- and ice hockey (NHL)-playoffs.

[4]See also Chan, Courty and Li (2007) who analyze incentives in a dynamic contest among two players under a preference for close outcomes.

$>$From the large literature on contests, however, we know that expected effort and win-probabilities in any two-player contest do not solely depend on the absolute strength (win valuations) of the respective players, but also on their relative strength (see for example Baye et al. (1993)). For example, if the difference in strength between the best and second-best team is larger than the difference between the second and the third, a final between the second and third best teams may induce both more effort and "thrill" (in the sense of more symmetric expected probabilities of winning) than a final between the two strongest teams. Consequently, there might be (at least theoretically) rationales for various seedings.

There are many possible seedings in an elimination tournament. The reader may amuse herself/himself by calculating that, with $2^N$ players, there are $\frac{(2^N)!}{2^{(2^N-1)}}$ different seedings. This yields 3 seedings for 4 players, 315 seedings for 8 players, 638,512,875 seedings for 16 players and $1.2253\times10^{26}$ seedings for 32 players.

There is a significant statistical literature on the design of various forms of elimination tournaments. The pioneering paper[5] is David (1959) who considered the win probability of the top player in a four player tournament with a random seeding. This literature assumes that, for each encounter among players $i$ and $j$, there is a fixed, exogenously given probability that $i$ beats $j$. In particular, this probability does not depend on the stage of the tournament where the particular match takes place, and does not depend on the identity of the expected opponent at the next stage[6]. Most results in that literature offer formulas for computing overall probabilities with which various players will win the tournament. For specific numerical examples it has been noted that the seeding where best meets worst, etc...yields for the top ranked player a higher probability of winning than a random seeding. Several papers (see for example, Hwang (1982), Horen and Reizman (1985) and Schwenk (2000)) consider various optimality criteria for choosing seedings. Given the sheer number of possible seedings and match outcomes, there are no general results for tournaments with more than four

[5] See also Glenn (1960) and Searles (1963) for early contributions.

[6] Additional assumptions are that $i's$ probability to win against $j$ is larger than vice-versa if $i$ is higher ranked than $j$ (and thus it is at least 50%), and that the win probability decreases if the opponent's rank is increased. Probability matrices satisfying these conditions are called *doubly-monotone*.

players. In particular, the optimal seeding for a given criterion may depend on the particular matrix of win probabilities (see Horen and Reizman (1985) who consider general, fixed win probabilities and analyze tournaments with four and eight players).

In contrast to the above mentioned literature, we consider here a tournament model where forward looking agents exert effort in order to win a match and advance to the next stage. We assume that players have different, common knowledge valuations for winning, and we model each match among two players as an all-pay auction: the prize for the winner of a particular match is either the tournament's prize if that match was the final, or else the right to compete at the next round. As a result, win probabilities in each match become *endogenous* - they result from mixed equilibrium strategies, and are positively correlated to win valuations. Moreover, the win probabilities depend on the stage of the tournament where the match takes place, and on the identity of the future expected opponents (which are determined in other parallel matches). Thus, in order to determine the tournament's outcome, we need to compute a dynamically intertwined set of pair-wise equilibria for each seeding. We provide here full analytic solutions for the case of four players which yields three different seedings in the semifinals.

The players' ranking can be used by the designer in order to determine the tournament's seeding structure, and we look for the optimal seeding from the designer's point of view. In reality there are many possible designer's goals, tailored to the role and importance of the competition, to local idiosyncracies (such as fan support for a home team), to commercial contracts with large sponsors (that may be also related to prominent competitors), or with media companies. We consider here three separate optimality criteria, and, additionally, a "fairness" criterion:

1. Find the seeding(s) that maximizes the probability of a final among the two highest ranked players.

2. Find the seeding(s) that maximizes the win probability of the highest ranked player.

3. Find the seeding(s) that maximizes total expected effort in the tournament.

4. Find the seeding(s) with the property that higher ranked players have a higher

probability of winning the tournament.

Our third optimality criterion is "conservative" in the sense that it treats all matches in the tournament symmetrically, and it does not a-priori bias the decision in favor of top players. The statistical literature did not analyze this criterion since there are no strategic decisions (e.g., about effort) in their models. In contrast, the other two optimality criteria have been discussed in the statistical literature, and seem to be prevalent in practice. The last criterion poses a constraint on the unfairness of a seeding by requiring that the overall win probabilities are naturally ordered according to the players' ranking. If this property does not hold for a given seeding, anticipating players have a perverse incentive to manipulate their ranking (e.g., by exerting less effort in a qualifying stage).

Our main findings are as follows: Let the four players be ranked in decreasing order of strength: 1,2,3,4. Seedings specify who meets whom in the semifinals. It turns out that the seeding most observed in practice[7], $A$:1-4,2-3, maximizes the win probability of the strongest player, and is the unique one with the property that strictly stronger players have a strictly higher probability of winning (criteria 3 and 4). On the other hand, seeding $B$:1-3,2-4 maximizes both total effort across the tournament and the probability of a final among the two top players (criteria 1 and 2). Seeding $C$:1-2,3-4, under which the two top players meet already in the semifinal, does not satisfy any of the optimality or fairness criteria, and the same holds for a random seeding[8].

Our results do not depend on cardinal differences between players' strength (win valuations), but only on the ordinal ranking specifying who is stronger than whom. It is this feature that allows us to compare the theoretical predictions to real-life tournaments where, in most cases, the remaining players in the semifinals need not be the a-priori highest ranked four players.

To understand the nature of such an empirical exercise, consider the four regional brackets of the NCAA basketball tournament. In each bracket 16 ranked teams play

[7] This is also the seeding method proposed for the Final Four.

[8] Schwenk (2000) argues for *cohort randomized* seeding based on three fairness criteria. In cohort randomized seeding players are first divided in several cohorts according to strength (say top, middle, bottom) and players in the same cohort are treated symmetrically in the randomization.

in an elimination tournament whose winner goes on to play in the national semifinals. Needless to say, the bracket semifinals are not necessarily played by the four originally highest ranked teams. For example, the 2002 Midwest semifinals where Kansas (1)-Illinois (4) and Oregon (2)-Texas (6). Since the top ranked team (Kansas) played against the third highest ranked team among the remaining ones (Illinois), this semifinal corresponds to our seeding $B$:1-3,2-4. The West semifinals were Oklahoma (2)-Arizona (3) and UCLA (8)-Missouri (12). Since the two top remaining teams (Oklahoma and Arizona) meet already in the semifinal, this corresponds to our seeding $C$:1-2,3-4 . The 2001 South semifinals were Tennessee (4)-N.Carolina (8) and Miami (6)-Tulsa (7). Since the highest ranked remaining team (Tennessee) plays against the lowest ranked remaining one (N.Carolina), this corresponds to seeding $A$. In this way, available data can generate observations for all three possible seedings, even if the initial method of seeding at the beginning of the tournament is fixed. We find that the data from college basketball tournaments is broadly in line with our theoretical predictions.

We conclude our Introduction by mentioning several related papers from the economics literature. In a classical piece, Rosen (1986) looks for the optimal prize structure in an elimination tournament with homogeneous players where the probability of winning a match is a stochastic function of players' efforts. In the symmetric equilibrium, the winner of every match is completely determined by the exogenous stochastic terms[9]. In Section IV he also considers an example with four players that can be either "strong" or "weak". Rosen finds (numerically) that a random seeding yields higher total effort than the seeding where strong players meet weak players in the semifinals. He did not consider the seeding strong/strong and weak/weak in the semifinals, but, in his numerical example, it turns out that this seeding (which corresponds then to our seeding $C$:1-2,3-4) yields the highest total effort.

As a by-product of our analysis, we show that total expected effort in the elimination tournament where the two strongest players meet in the final with probability

[9]His main result is that rewards in later stages must be higher than reward in earlier stages in order to sustain a non-decreasing effort along the tournament. Other works on allocation of resources in sequential contests are, among others, Konrad (2004), Warneryd (1998) and Klumpp and Polborn (2006).

one (seeding $B$:1-3,2-4) equals total effort in the all-pay auction where all players compete simultaneously. This should be contrasted with the main finding of Gradstein and Konrad (1999) who study a rent-seeking contest à la Tullock (with homogenous players). They found that simultaneous contests are strictly superior if the contest's rules are discriminatory enough (as in an all-pay auction). In a setting with heterogenous valuations, our analysis indicates that, for the Gradstein-Konrad result to hold, it is necessary that the multistage contest induces a positive probability that the two strongest players do *not* reach the final with probability one (e.g., our seedings $A$:1-4,2-3 and $C$:1-2,3-4)

Baye et al. (1993) look for the optimal set of contestants in an all-pay auction, and they find that it is sometimes advantageous to exclude the strongest player. These authors do not consider explicit mechanisms by which finalists are selected. Our analysis suggests that, given the rigid constraints imposed by the structure of an elimination tournament, it is not advantageous to exclude the strongest player from the final in our model.

The paper is organized as follows. We present the tournament model in Section 2. In Section 3 we present the optimality results, and briefly illustrate the employed techniques. In Section 4 we compare the theoretical results with historical data from the NCAA tournament. In Section 5 we gather several concluding remarks. All proofs are in an Appendix.

# 2 The Model

There are four players (or teams) $i = 1, ..., 4$ competing for a prize. The prize is allocated to the winner of a contest which is organized as an elimination tournament. First, two pairs of players simultaneously compete in two semifinals. The two winners (one in each semifinal) compete in the final, and the winner of the final obtains the prize. The losers of the semifinals do not compete further. We model each match among two players as an all-pay auction: both players exert effort, and the one exerting the higher effort wins.

Player $i$ values the prize at $v_i$, where $v_1 \geq v_2 \geq v_3 \geq v_4 > 0$. Valuations are common-knowledge. We assume that each finalist obtains a payment $k > 0$, independent from his performance in the final[10], and we consider the limit behavior as $k \to 0$. This technicality is required in order to ensure that all players have positive present values when competing in the semifinals - this is a necessary condition for the existence of equilibria in the semifinals.

In a final between players $i$ and $j$, the exerted efforts are $e_i^F$ , $e_j^F$. Net of $k$, the payoff for player $i$ is given by

$$u_i^F(e_i^F, e_j^F) = \begin{cases} -e_i^F & \text{if } e_i^F < e_j^F \\ \frac{v_i}{2} - e_i^F & \text{if } e_i^F = e_j^F \\ v_i - e_i^F & \text{if } e_i^F > e_j^F \end{cases} \tag{1}$$

and analogously for player $j$. Player $i's$ payoff in a semifinal between players $i$ and $j$ is given by

$$u_i^S(e_i^S, e_j^S) = \begin{cases} -e_i^S & \text{if } e_i^S < e_j^S \\ \frac{Eu_i^F + k}{2} - e_i^S & \text{if } e_i^S = e_j^S \\ Eu_i^F + k - e_i^S & \text{if } e_i^S > e_j^S. \end{cases} \tag{2}$$

and analogously for player $j$. Note that each player's payoff in a semifinal depends on the expected utility associated with a participation in the final. In turn, this expected utility depends on the expected opponent in the final. It is precisely this feature that can be "manipulated" by designing the seeding of the semifinals. The contest designer chooses the structure $s$ of the semifinals out of the set of feasible seedings $\{A, B, C\}$ , where: $A$:1-4,2-3, $B$:1-3,2-4 and $C$:1-2,3-4.

The following well-known Lemma characterizes behavior in an all-pay auction among two heterogenous players.

**Lemma 1.** *Consider two players i and j with $0 < v_j \leq v_i$ that compete in an all-pay auction for a unique prize. In the unique Nash equilibrium both players randomize on the interval $[0, v_j]$. Player i's effort is uniformly distributed, while player j's effort is distributed according to the cumulative distribution function*[11] $G_j(e) = (v_i - v_j + e)/v_i$.

[10] There are many examples where such a feature is indeed present.

[11] Note that this distribution has an atom of size $(v_i - v_j)/v_i$ at $e = 0$.

*Given these mixed strategies, player $i'$s winning probability against $j$ is given by* $q_{ij} = 1 - \frac{v_j}{2v_i}$. *Player $i$'s expected effort is* $\frac{v_j}{2}$, *and player $j$'s expected effort is* $\frac{v_j^2}{2v_i}$. *Total expected effort is therefore* $\frac{v_j}{2}(1 + \frac{v_j}{v_i})$. *The respective expected payoffs are* $u_i = v_i - v_j$ *and* $u_j = 0$.

*Proof.* See Hillman and Riley (1989) and Baye et al. (1993). □

# 3 Semifinals Design

We provide here optimal seedings for the last crucial stage requiring design - the semi-finals. This Section has the following structure: We first verbally sketch the main intuition behind the results. We next offer an illustration for the simple special case where there are two equally strong and two equally weak players (this can be compared to Rosen's example mentioned in the Introduction). Finally, we present the general optimality results for the various criteria.

## 3.1 Intuition

Let us first look at design $A$:1-4,2-3. As $k$ goes to zero, player 1 reaches the final with almost certainty. This happens because player 4 expects a limit payoff of zero no matter which player (either 2 or 3) she meets in a final. A-priori, players 2 and 3 are not in a symmetric position: while both would obtain a limit payoff of zero in a final against player 1, their expected payoffs are positive but different in a final against player 4 (with 2 having the higher valuation). But, since the event of meeting 4 in a final has a zero limit probability, the position of 2 and 3 becomes symmetric: since both know that they are going to meet the stronger player 1 in the final, their limit expected valuation for the final is zero. Hence, both reach the final with probability one-half and meet there player 1.

In design $B$:1-3,2-4, player 4 has a limit expected utility of zero in *any* final (where he meets either 1 or 3 - both stronger players), whereas player 2 has a positive expected value stemming from the event where he meets player 3 in the final. In the limit, player 2 reaches the final with probability one. But then, player 3 does not expect a positive payoff in the final. Hence, player 1 reaches the final with probability one, and meets there player 2.[12]

Since the expected final in design $B$:1-3,2-4 (among players 1 and 2 ) is tighter than the expected final in design $A$:1-4,2-3 (where 1 meets either 2 or 3 , each with probability one-half), design $B$:1-3,2-4 dominates design $A$:1-4,2-3 with respect to total effort.

The comparison with respect to total effort between seedings $B$:1-3,2-4 and $C$:1-2,3-4 is more subtle. In design $C$:1-2,3-4 all four possible finals have a positive probability since both stronger players expect a positive payoff in a final, and both weak players expect a zero payoff. An important observation is that, in our elimination tournament, a semifinal among players 1 and 2 yields less total effort than a final among these players because in the semifinal both players anticipate that, in order to ultimately win the tournament, they need to exert an additional effort in the final. The decrease in effort caused by the fact that 1 and 2 meet already in the semifinal cannot be compensated by the additional effort in a final among one of the stronger players and one of the weaker players, and seeding $B$:1-3,2-4 also dominates seeding $C$:1-2,3-4 with respect to total effort.

Recall that in seeding $B$:1-3,2-4 there is a final among players 1 and 2 with limit probability one. Hence, player 1's overall win probability equals the probability with which he wins a final against player 2. In seeding $A$:1-4,2-3 player 1 also reaches the final with limit probability one, but meets there either player 2 or player 3 (with equal limit probabilities). Since player 1 is more likely to win a final against player 3 than a final against player 2, we obtain that seeding $A$:1-4,2-3 dominates seeding $B$:1-3,2-4 with respect to the top player's win probability.

The comparison between seedings $C$:1-2,3-4 and $A$:1-4,2-3 with respect to the top player's win probability is more subtle: player 1 is more likely to win the final in seeding $C$:1-2,3-4 (where he meets either player 3 or player 4) than in seeding $A$:1-4,2-3 (where he meets either 2 or 3) . But, in seeding $A$:1-4,2-3 player 1 makes it to the final for

[12] Of course, as $k$ gets small, neither 2 nor 4 have a positive valuation for the final where they meet player 1 for sure; but player 2's valuation converges to zero at a slower rate, confirming the above logic.

sure, while in seeding $C$ :1-2,3-4 only with some probability less than one (since he first has to win the semifinal against player 2). It turns out that this last handicap is significant, and it is always the case that seeding $A$:1-4,2-3 yields a higher overall win probability for player 1.

### 3.1.1 The Two-Type Case

We now briefly consider here the case where $v_1 = v_2 = v_H > v_L = v_3 = v_4$. Obviously, seedings $A$:1-4,2-3 and $B$:1-3,2-4 are here equivalent.

**Seedings A:1-4,2-3 and B:1-3,2-4.** Let $q_{ij}^S(k)$ denote the probability that $i$ beats $j$ in a semifinal among $i$ and $j$. Based on these probabilities we can compute expected values for the final, conditional on winning a semifinal.

Conditional on winning the semifinal, player 1 faces player 2 in the final with probability $q_{23}^S(k)$. This results in a payoff of zero for both finalists since they are of equal strength. Player 1 meets player 3 in the final with probability $1 - q_{23}^S(k)$. Since 3 has valuation $v_L < v_H$, player 1 expects a payoff of $v_H - v_L$ in that case. In any case, there is the additional payoff $k$ for making it to the final. Thus, player 1's expected value from winning the semifinal is given by

$$q_{23}^S(k) \cdot 0 + (1 - q_{23}^S(k))(v_H - v_L) + k = (1 - q_{23}^S(k))(v_H - v_L) + k. \tag{3}$$

Analogously, the expected value for player 2 is given by

$$(1 - q_{14}^S(k))(v_H - v_L) + k, \tag{4}$$

In the final player 4 faces player 2 with probability $q_{23}^S(k)$ and player 3 with probability $1 - q_{23}^S(k)$. Player 4's expected payoff is $k$ in both cases, and analogously for player 3.

Given the above computed values, Lemma 1 tells us that the winning probabilities $q_{14}^S(k)$ and $q_{23}^S(k)$ are determined by the following system of equations:

$$q_{14}^S(k) = 1 - \frac{k}{2[(1 - q_{23}^S(k))(v_H - v_L) + k]} \tag{5}$$

$$q_{23}^S(k) = 1 - \frac{k}{2[(1 - q_{14}^S(k))(v_H - v_L) + k]}. \tag{6}$$

Solving the above system (under the restriction $q \in [0,1]$) yields the symmetric solution

$$q_{14}^S(k) = q_{23}^S(k) = 1 + \frac{k}{2(v_H - v_L)} - \frac{1}{2(v_H - v_L)}\sqrt{(2(v_H - v_L) + k)k} \tag{7}$$

By Lemma 1, in each of the two semifinals the expected effort in each of the two semifinals is given by

$$\frac{1}{2}k + \frac{1}{2}\frac{k^2}{(1 - q^S(k))(v_H - v_L) + k} \tag{8}$$

where $q^S(k) \in \{q_{14}^S(k), q_{23}^S(k)\}$. Note that

$$\lim_{k \to 0} q_{14}^S(k) = \lim_{k \to 0} q_{23}^S(k) = 1 \tag{9}$$

Intuitively, the weak players have only a small chance to win the final, and hence exert almost no effort. This implies that the strong players do not have to exert a lot of effort in the semifinals either. Moreover, each of the strong players knows that he is going to meet the other strong player in the final (and thus that the payoff from the final will be low). This reduces the strong players' valuation for winning the semifinals.

Players 1 and 2 meet in the final with probability 1 (as $k$ tends to zero). Since both 1 and 2 have the same valuation $v_H$, total expected effort in the final is $v_H$.

**Seeding C:1-2,3-4**. The final will be between a player with valuation $v_H$ and a player with valuation $v_L$. Hence, by Lemma 1, expected effort in the final is $\frac{v_L}{2}\left(1 + \frac{v_L}{v_H}\right)$.

Consider first the semifinal between the strong players 1 and 2. Since the winner of this semifinal will meet a weak player in the final, both 1 and 2 expect a payoff $v_H - v_L + k$ in the final. By Lemma 1, total expected effort in this semifinal is $v_H - v_L + k$ (note that, for small $k$, this is less than total effort in a final among two strong players, which yields a total effort of $v_H$).

Consider now the semifinal between the weak players. Both have an expected payoff of $k$ in the final since this is the payoff in a final against a strong competitor with valuation $v_H$. Hence, total expected effort in this semifinal is also $k$.

Total limit effort in seeding $C$:1-2,3-4 is thus given by:

$$TE_C = \lim_{k \to 0}[\frac{1}{2}v_L\left(1 + \frac{v_L}{v_H}\right) + v_H - v_L + 2k] = v_H - \frac{1}{2}v_L\left(1 - \frac{v_L}{v_H}\right) \tag{10}$$

We can conclude that

$$TE_A = TE_B = v_H > v_H - \frac{1}{2}v_L\left(1 - \frac{v_L}{v_H}\right) = TE_C. \tag{11}$$

### 3.2 Total Effort

We now return to the general four-player case, and we assume first that the designer chooses the seeding $s$ in order to maximize total expected tournament effort $TE_s$ . Let $R$ be the set of all players, and let $\widetilde{F}(s)$ be the random set of players reaching the final for a given seeding $s$. The designer solves

$$\max_{s\in\{A,B,C\}}\{\sum_{i\in R}E(\widetilde{e}_i^S)+[\sum_{i>j}(E(\widetilde{e}_i^F)+E(\widetilde{e}_j^F))\cdot P_{rob}(i,j\in\widetilde{F}(s))]\} \tag{12}$$

**Proposition 1.** *For any valuations $v_1 > v_2 > v_3 > v_4$, the limit total tournament effort (as k goes to zero) is maximized in seeding B:1-3,2-4, where it equals $\frac{1}{2}(v_2+\frac{v_2^2}{v_1})$ .*

**Proof:** See Appendix.

### 3.3 Probability of a Final among the Two Top Players

We now assume that the designer chooses the seeding $s$ in order to maximize the probability of a final among the two top players. As already indicated in the previous section, seeding $B$:1-3,2-4 is again optimal.

**Proposition 2.** *For any valuations $v_1 > v_2 > v_3 > v_4$, a final among players 1 and 2 occurs with limit probability one in seeding B:1-3,2-4, and with limit probability of one-half in seeding A:1-4, 2-3.*[13]

**Proof:** See Lemmas 2, 3, 4 in the Appendix.

Interestingly, our model predicts that the probability of a final among the two top players under random seeding (roughly corresponding to the method now employed by the NCAA for the Final Four) is $\frac{1}{3}(1+\frac{1}{2}+0)=\frac{1}{2}$, which equals the probability of such a final under seeding $A$:1-4,2-3. Thus, reseeding according to $A$:1-4,2-3 is not likely to increase the probability of a final among the two top players, but, as we show in the next section, it will increase the probability that the top team wins the tournament.

Let us briefly discuss the relevance of the above finding for elimination tournaments among $2^N$ players where $N > 2$ is the number of rounds needed to produce a winner.

[13] The probability for seeding $C$:1-2,3-4 is obviously zero.

Order the agents by their valuations $v_1 \geq v_2 ... \geq v_{2^N}$ Let $M_{ij}$ denote a match among players $i$ and $j$, and let $M^w_{(ij)(hl)}$ denote a match among the winners in the matches $M_{ij}$ and $M_{hl}$.

**Definition 1.** *We say that seeding $s$ eliminates player $i$ in round $l < N$ if, as $k$ tends to zero, the probability that $i$ reaches stages $l+1$ (given that she reached stage $l$) tends to zero.*[14]

For example, recall our results for round $l = 1$ of a tournament with 4 players (see Lemmas 2, 3 and 4 in the Appendix): Seeding $C$:1-2,3-4 does not eliminate any player, and all four possible finals have positive probability. Seeding $A$ :1-4,2-3 eliminates only player 4 and the finals 1-2 and 1-3 have both positive probability. Finally, the optimal seeding $B$ :1-3,2-4 eliminates both players 3 and 4, and only the final 1-2 has positive probability (one).

It turns out that it is always possible to seed the players such that the two strongest players participate in the final with probability one (as $k$ tends to zero). Consider for example a tournament among 8 players with the following structure of matches: Round 1: $M_{18}, M_{27}, M_{36}, M_{45}$;Round 2: $M^w_{(18)(36)}, M^w_{(27)(45)}$; Round 3: final among winners in semifinals. It is easy to see that player 8 is eliminated at stage 1 from the same reason that player 4 is eliminated at stage 1 in seedings $A$ :1-4,2-3 and $B$ :1-3,2-4. Now, since player 8 does not reach the second stage for sure, player 7 is eliminated at the first stage from the same reason that player 3 is eliminated at the first stage in seeding $B$ :1-3,2-4. By induction, player 6 is eliminated at stage 1 as well. This seeding does not eliminate players 4 and 5 at stage 1 since they are in symmetric positions given their possible future opponents. Thus we obtain either the semifinals 1-3,2-4 or 1-3,2-5. By the logic of seeding $B$:1-3,2-4, the two respective weaker players get eliminated in stage 2, and we again obtain the desired final among the two best players.

It is important to note that, whereas in the four-player case seeding $B$:1-3,2-4 was the unique one with the property that it ensures a final among the two best players, there is more design freedom if there are $2^N > 4$ players.

[14] Obviously, at most $2^{N-l}$ players can be eliminated at stage $l$.

### 3.4 The Top Player's Win Probability

Seeding $B$:1-3,2-4 was found to be optimal for the previous two criteria. But recall that seeding $A$:1-4,2-3 is the one most often observed in real tournaments. It is reassuring to find that seeding $A$:1-4,2-3 is optimal with respect to the important criterion of maximizing the top player's win probability.

**Proposition 3.** *For any valuations $v_1 > v_2 > v_3 > v_4$, player 1's limit win probability (as $k$ goes to zero) is maximized in seeding A:1-4,2-3.*

**Proof:** See Appendix.

### 3.5 Fairness

We study now the win probabilities of all players, and check which seedings have the property that the probabilities to win the tournament are naturally ordered according to the players' ranking.

**Proposition 4.** *Assume that $v_1 > v_2 > v_3 > v_4$. Let $p_i(s)$ denote the probability that player $i$ wins the tournament for a given seeding $s \in \{A, B, C\}$. We have:*

1. $p_1(A) > p_2(A) > p_3(A) > p_4(A) = 0$.
2. $p_1(B) > p_2(B) > p_3(B) = p_4(B) = 0$.
3. In seeding $C$ :1-2,3-4 it may happen that $v_i > v_j$ but $p_i(s) < p_j(s)$.

**Proof:** See Appendix.

## 4 Some Empirical Observations

In order to investigate the empirical relevance of our theory, we confront the theoretical predictions with data from semi-finals and finals from 100 regional NCAA college basketball tournaments. We use data from the four regional elimination tournaments (East, West, Midwest, South) over the period 1979 to 2003. All 16 participating teams in a region are ranked (during the selection process) according to their previous performance.[15] The rankings in each region *at the beginning* of the respective tournament allows us to determine the *relative* ranking of the four teams playing in the semi-finals and the resulting seeding. Moreover, seedings should be random. In the NCAA tournaments, two rounds have already been played prior to the semi-finals. To rule out any selection effects, we exclude semi-finals that are implied by a particular initial seeding.[16] Then, the constructed seedings for the semi-finals are presumably random and unaffected by the initial seeding.[17] We observe all seedings although $A$:1-4,2-3 (25 observations) and $B$:1-3,2-4 (25 observations) are much more common than $C$:1-2,3-4 (14 observations).

The raw frequencies of winner types by seeding provide support for the prediction of Proposition 4 that the probability of winning the tournament is positively correlated with the contestants' ranking. Higher ranked teams or players are more likely to win a tournament in all seedings. Proposition 3 predicts that the winning probability of the strongest team is maximized by seeding $A$:1-4,2-3. The raw NCAA data support this hypothesis. Under seeding $A$:1-4,2-3, the best team wins the tournament in 22

[15] Data and relevant links can be found on the internet under http://old.sportsline.com/u/madness/2002/history/index.html. During the selection process, a committee determines the 64 "best" college teams of the respective season and allocates them to the regional conferences. This selection is based on several measures reflecting the teams' recent performance. Roughly speaking, teams are split into groups of four according to rankings. The four highest ranked teams are allocated to the four regions with one team per region, followed by next group of four teams, etc., until all teams are allocated. The aim of this procedure is to balance the brackets in all regions. The regional brackets are initially seedings of type $A$ for the respective 16 teams. A team's regional allocation may change from year to year and bears no relation to geography. More details can be found under http://www.ncaa.org/library/handbooks/basketball/2003.

[16] As a consequence of the initial seeding, which in the the case of the NCAA regional conferences is type $A$, particular pairings of cardinal rankings are not possible (e.g. a semi-final among the two strongest regional teams). We also analyzed all semi-finals without modification to test the robustness of the results and to rule out any potential biases due to selection effects stemming from initial seeding. The results are qualitatively identical to those obtained from the full sample.

[17] We make the identifying assumption that during the first two rounds of the tournament there is sufficient noise in match outcomes in order to generate a random selection of teams, and thus random (ordinal) seedings at the level of the semi-final. Note that winning streaks, or "hot hand" should work against finding evidence for the effects of the seeding according to initial strength.

out of 25 cases, under seeding $B$:1-3,2-4 only in 17 out of 25 cases. In 9 out of 14 cases, the strongest team wins the tournament under seeding $C$:1-2,3-4, constituting the smallest percentage. To investigate this result in more depth, we estimate a logit model of the probability that the team ranked 1 prior to the tournament wins the final, on a measure of ranks based on the expected round of elimination of a team as well as on the respective seed.[18] Compared to seeding $A$:1-4, 2-3, both seedings $B$:1-3,2-4 and $C$:1-2,3-4 entail a significantly smaller probability of the strongest team winning the NCAA basketball tournament. The respective coefficients for seedings $B$ and $C$ are significant and negative on the 10 percent level. Alternatively, compared to any other seeding, seeding $A$:1-4, 2-3 exhibits a significantly higher probability for the best team winning, again on the ten percent level. In the raw data, this effect is even stronger and significant on the 5 percent level. The data also do not allow us to reject the prediction of Proposition 2 that the probability of a final among the two strongest teams is higher in seeding $B$:1-3,2-4 than in $A$:1-4, 2-3 (while it is zero by definition in $C$:1-2,3-4). We refrain from testing predictions about effort exertion as in Lemma 1 and Proposition 1, because it is difficult to find a good and observable measure for effort in basketball data. To sum up, data from the NCAA college basketball tournaments are broadly in line with the theoretical predictions. We find similar results for data from the semi-finals and finals of 12 tournaments for male tennis professionals for the years 1990 until 2002.[19] We interpret these empirical observations as corroborating evidence for the relevance of the problem and the applicability of the theory.

# 5 Concluding Remarks

We have analyzed optimal seedings in an elimination tournament where players have to exert effort in order to advance to the next stage. We established that seedings involving a delayed encounter among the top players are optimal for a variety of criteria. We have also exhibited the effects of switching the ranks of the opponents that play against the top players in the semifinals. In principle, it is possible to generalize the analysis

[18] The rank of a team $i$, $r_i$, is given by $r_i = 2 - log_2(rank_i)$, see Klaassen and Magnus (2003).

[19] See the discussion paper version for more detailed results and robustness checks.

conducted here to tournaments with more players (and possibly more prizes). But, the exponentially growing number of seedings, and the complexity of the fixed-point arguments suggest that analytic solutions are difficult to come by.

Our model and results offer a wealth of testable hypotheses. We have compared its predictions to the results of NCAA tournaments. The data do not refute the theory and provide some support for theoretical predictions. A-priori, it seems possible that other, more complex models (e.g, where some exogenous noise is added, where some dynamic "exhaustion" effects are introduced, or where the probability of winning with a higher effort is not equal to one) may better reflect some aspects of real-life tournaments. The crucial difference between such models and ours will mainly consist of the win-probabilities they produce for each match. These probabilities ultimately drive the results, and are easily measurable (in contrast to effort). In this context, it is important to note that Horen and Riezman (1985) looked at general four-player tournaments where the only requirement on the (exogenous and fixed) matrix of win-probabilities is that the entries are naturally ordered by the strength of the teams[20]. These authors showed that, for any such matrix, seeding $A$ maximizes the strongest competitor's probability of winning, and it is the only "fair" one. Moreover, if $\frac{p_{14}}{p_{13}} \leq [\geq] \frac{p_{24}}{p_{23}}$ (where $p_{ij}$ denotes the probability that $i$ beats $j$) , the probability that the two strongest players meet in the final is maximized by seeding $B$ $[A]$. Thus, any theoretical model that yields natural and minimal monotonicity requirements on the win-probabilities in each match will display results that are very similar to ours, the only exception being the possibility that seeding $A$ also maximizes the probability of a final among the top players for some values of the parameters.

# 6 Appendix

All results are based on the three basic lemmas that determine equilibrium behavior for each seeding. Let $q_{ij}^{S}(k)$ denote the limit probability that $i$ beats $j$ if they meet in

[20] This means the following: 1) If competitor $i$ is stronger than $j$, then $i$ has at least a 50% chance to beat $j$ in a match. 2) If $i$ is stronger than $j$, then any other competitor $h$ has a higher probability to win against $j$ than against $i$.

a semifinal, and let $q^S_{ij} = \lim_{k\to 0} q^S_{ij}(k)$. $TE^S_{ij}(k)$ denotes total equilibrium effort in a semifinal among $i$ and $j$; $TE^F_s(k)$ denotes total equilibrium effort in a final resulting from seeding $s$; $TE_s(k)$ denotes total equilibrium effort in all three matches of seeding $s$, and define $TE_s = \lim_{k\to 0} TE_s(k)$;

**Lemma 2.** *Consider seeding $A$ :1-4,2-3 . In the limit, as $k \to 0$, player 1 reaches the final with probability one, while players 2 and 3 reach the final with probability one-half each. In addition the following hold:*

$$\lim_{k\to 0} \left(TE^S_{14}(k) + TE^S_{23}(k)\right) = 0 \tag{13}$$

*and*

$$TE_A = \lim_{k\to 0} TE_A(k) = \lim_{k\to 0} TE^F_A(k) = \frac{1}{4}\left(v_2 + \frac{v_2^2}{v_1} + v_3 + \frac{v_3^2}{v_1}\right) \tag{14}$$

**Proof:** By Lemma 1, player 1's valuation for the semifinal is $(1 - q^S_{32}(k))(v_1 - v_2 + k) + q^S_{32}(k)(v_1 - v_3 + k)$ and player 4's valuation is $k$. By Lemma 1, we know that the total expected effort in this semifinal is given by

$$TE^S_{14}(k) = \frac{1}{2}k + \frac{1}{2}\frac{k^2}{(1 - q^S_{32}(k))(v_1 - v_2 + k) + q^S_{32}(k)(v_1 - v_3 + k)}. \tag{15}$$

Player's 4 probability of winning is given by

$$q^S_{41}(k) = \frac{1}{2}\frac{k}{(1 - q^S_{32}(k))(v_1 - v_2 + k) + q^S_{32}(k)(v_1 - v_3 + k)} \tag{16}$$

Players 2 and 3 play in the other semifinal. Their valuations for the semifinal are $q^S_{41}(k)(v_j - v_4 + k) + (1 - q^S_{41}(k))k,\ j = 2, 3$ , and expected total efforts in this semifinal is given by

$$\begin{aligned} TE^S_{23}(k) \quad &= \quad \frac{1}{2}[q^S_{41}(k)(v_3 - v_4 + k) + (1 - q^S_{41}(k))k] + \\ &\quad\quad \frac{1}{2}\frac{(q^S_{41}(k)(v_3 - v_4 + k) + (1 - q^S_{41}(k))k)^2}{q^S_{41}(k)(v_2 - v_4 + k) + (1 - q^S_{41}(k))k} \end{aligned} \tag{17}$$

Player's 3 probability of winning is given by

$$q^S_{32}(k) = \frac{1}{2}\frac{q^S_{41}(k)(v_3 - v_4 + k) + (1 - q^S_{41}(k))k}{q^S_{41}(k)(v_2 - v_4 + k) + (1 - q^S_{41}(k))k}. \tag{18}$$

In the limit, as $k \to 0$, the unique fixed point is $q_{41}^S = 0$ and $q_{32}^S = 1/2$. We have then

$$TE_A^F = \frac{1}{4}\left(v_2 + \frac{v_2^2}{v_1} + v_3 + \frac{v_3^2}{v_1}\right)$$

*Q.E.D.*

**Lemma 3.** *Consider seeding B:1-3,2-4. In the limit, as $k \to 0$, the final takes place among players 1 and 2 with probability one. In addition, the following hold:*

$$\lim_{k\to 0}\left(TE_{13}^S(k) + TE_{24}^S(k)\right) = 0 \tag{19}$$

*and*

$$TE_B = \lim_{k\to 0} TE_B(k) = \lim_{k\to 0} TE_B^F(k) = \frac{1}{2}(v_2 + \frac{v_2^2}{v_1}) \tag{20}$$

**Proof:** Player 1's valuation for the semifinal is $(1-q_{42}^S(k))(v_1-v_2+k)+q_{42}^S(k)(v_1-v_4+k)$ and player 3's valuation is $(1-q_{42}^S(k))k + q_{42}^S(k)(v_3-v_4+k)$. Total expected effort in this semifinal is given by

$$\begin{aligned} TE_{13}^S(k) &= \frac{1}{2}((1-q_{42}^S(k))k + q_{42}^S(k)(v_3-v_4+k)) + \\ &\quad \frac{1}{2}\frac{((1-q_{42}^S(k))k + q_{42}^S(k)(v_3-v_4+k))^2}{(1-q_{42}^S(k))(v_1-v_2+k)+q_{42}^S(k)(v_1-v_4+k)} \end{aligned} \tag{21}$$

Player's 3 probability of winning is given by

$$q_{31}^S(k) = \frac{1}{2}\frac{(1-q_{42}^S(k))k + q_{42}^S(k)(v_3-v_4+k)}{(1-q_{42}^S(k))(v_1-v_2+k)+q_{42}^S(k)(v_1-v_4+k)} \tag{22}$$

Players 2 and 4 play in the other semifinal. Their valuations for the semifinal are $q_{31}^S(k)(v_2-v_3+k)+(1-q_{31}^S(k))k$ for player 2 and $k$ for player 4. Expected total effort in this semifinal is given by

$$TE_{24}^S(k) = \frac{1}{2}k + \frac{1}{2}\frac{k^2}{q_{31}^S(k)(v_2-v_3+k)+(1-q_{31}^S(k))k}. \tag{23}$$

Player's 4 probability of winning the semifinal is given by

$$q_{42}(k) = \frac{1}{2}\frac{k}{q_{31}(k)(v_2-v_3+k)+(1-q_{31}(k))k}. \tag{24}$$

Solving for $q_{31}^S$ by combining equations (22) and (24) yields

$$q_{31}^S(k) = \frac{-k[2(v_1 - v_2 + k) + (v_3 - v_4)]}{4(v_2 - v_3)(v_1 - v_2 + k)} + \sqrt{\left(\frac{k[2(v_1 - v_2 + k) + (v_3 - v_4)]}{2(v_2 - v_3)(v_1 - v_2 + k)}\right)^2 + k(2k + (v_2 - v_4))}. \quad (25)$$

Taking the limit in equation (25), we obtain $\lim_{k\to 0} q_{31}^S(k) = 0$.

By equation (24) we also have

$$q_{42}^S(k) = \frac{1}{2}\frac{k}{q_{31}^S(k)(v_2 - v_3) + k}. \quad (26)$$

Note that $q_{31}(k)$ converges faster to zero than $k$, due to the square root. But then, $q_{42}^S \to 0$ as $k \to 0$. To see this, we apply l'Hospital's rule. Denote:

$$g(k) = \frac{-k[2(v_1 - v_2 + k) + (v_3 - v_4)]}{4(v_2 - v_3)(v_1 - v_2 + k)},$$

$$z(k) = \left(\frac{k[2(v_1 - v_2 + k) + (v_3 - v_4)]}{2(v_2 - v_3)(v_1 - v_2 + k)}\right)^2$$

and

$$f(k) = k(2k + (v_2 - v_4)).$$

Taking the derivatives of the numerator and denominator in the expression for $q_{31}^S(k)$ yields the expression

$$\frac{2\sqrt{z(k) + f(k)}}{2\sqrt{z(k) + f(k)} + g'(k)(v_2 - v_3)2\sqrt{z(k) + f(k)} + z'(k) + f'(k)}. \quad (27)$$

Note that $z(0) = f(0) = z'(0) = 0$, but that $f'(0) > 0$. Hence $\lim_{k\to 0} q_{42}^S(k) = 0$.

*Q.E.D*

**Lemma 4.** *Consider seeding C:1-2,3-4. In the limit, as $k \to 0$, players 3 and 4 reach the final with probability one-half each. Player 2 reaches the final with probability $\frac{1}{2}\frac{2v_2 - v_3 - v_4}{2v_1 - v_3 - v_4}$. In addition the following hold:*

$$\lim_{k\to 0} TE_{34}^S(k) = 0 \quad (28)$$

$$\lim_{k\to 0} TE_{12}^S(k) = \frac{1}{4}[(2v_2 - v_3 - v_4) + \frac{(2v_2 - v_3 - v_4)^2}{2v_1 - v_3 - v_4}] \tag{29}$$

$$TE_C = \lim_{k\to 0} TE_C(k) = \frac{1}{2}v_2 + \frac{1}{4}\left(\frac{v_4^2 + v_3^2}{v_1}\right) + \frac{1}{4}\frac{(2v_2 - v_3 - v_4)}{(2v_1 - v_3 - v_4)}\left(2v_2 - v_3 - v_4 + \frac{v_4^2 + v_3^2}{v_2} - \frac{v_4^2 + v_3^2}{v_1}\right) \tag{30}$$

**Proof:** Player 1's valuation for the semifinal is $q_{43}^S(k)(v_1 - v_4 + k) + (1 - q_{43}^S(k))(v_1 - v_3 + k)$ and player 2's valuation is $q_{43}^S(k)(v_2 - v_4 + k) + (1 - q_{43}^S(k))(v_2 - v_3 + k)$. The probability of winning for player 2 is

$$q_{21}^S(k) = \frac{1}{2}\frac{q_{43}^S(k)(v_2 - v_4 + k) + (1 - q_{43}^S(k))(v_2 - v_3 + k)}{q_{43}^S(k)(v_1 - v_4 + k) + (1 - q_{43}^S(k))(v_1 - v_3 + k)} \tag{31}$$

By Lemma 1 we know that total expected effort in this semifinal is

$$TE_{12}^S(k) = \frac{1}{2}(q_{43}^S(k)(v_2 - v_4 + k) + (1 - q_{43}^S(k))(v_2 - v_3 + k)) + \frac{1}{2}\frac{(q_{43}^S(k)(v_2 - v_4 + k) + (1 - q_{43}^S(k))(v_2 - v_3 + k))^2}{q_{43}^S(k)(v_1 - v_4 + k) + (1 - q_{43}^S(k))(v_1 - v_3 + k)}. \tag{32}$$

Players 3 and 4 play in the other semifinal. Their valuations for the semifinal are $k$ and expected total effort is also $TE_{34}^S(k) = k$. The respective probabilities of winning are $q_{43}^S = 1 - q_{43}^S = \frac{1}{2}$

The expected effort in the final is given by

$$TE_C^F(k) = \frac{1}{2}[q_{43}^S q_{21}^S\left(v_4 + \frac{v_4^2}{v_2}\right) + (1 - q_{43}^S)q_{21}^S\left(v_3 + \frac{v_3^2}{v_2}\right) + \tag{33}$$

$$+ q_{43}^S(1 - q_{21}^S)\left(v_4 + \frac{v_4^2}{v_1}\right) + (1 - q_{43}^S)(1 - q_{21}^S)\left(v_3 + \frac{v_3^2}{v_1}\right)] \tag{34}$$

Total expected effort is given by

$$TE_C(k) = TE_{12}^S(k) + TE_{34}^S(k) + TE_C^F(k). \tag{35}$$

Note that

$$\lim_{k\to 0} TE_{12}^S(k) = \frac{1}{4}[(2v_2 - v_3 - v_4)) + \frac{(2v_2 - v_3 - v_4)^2}{2v_1 - v_3 - v_4}]$$

Since $q_{43}^S = 1/2$ we also obtain

$$\lim_{k \to 0} q_{21}(k) = \frac{1}{2} \frac{2v_2 - v_3 - v_4}{2v_1 - v_3 - v_4}. \tag{36}$$

Combining all pieces gives the desired formula for $TE_C$. *Q.E.D.*

## 6.1 Proof of Proposition 1

We compare the total efforts in each seeding. Since $v_3 \leq v_2$ we immediately obtain that

$TE_A = \frac{1}{4}\left(v_2 + \frac{v_2^2}{v_1} + v_3 + \frac{v_3^2}{v_1}\right) \leq TE_B = \frac{1}{2}(v_2 + \frac{v_2^2}{v_1})$ , with strict inequality for $v_3 < v_2$.

Hence, in order to find the optimal seeding, it remains to compare

$$TE_B = \frac{1}{2}v_2 + \frac{v_2^2}{2v_1} \tag{37}$$

with

$$\begin{aligned} TE_C \;&=\; \frac{1}{2}v_2 + \frac{1}{4}\left(\frac{v_4^2 + v_3^2}{v_1}\right) + \\ &\frac{1}{2}\frac{(2v_2 - v_3 - v_4)}{(2v_1 - v_3 - v_4)}\left(v_2 - \frac{v_3}{2} - \frac{v_4}{2} + \frac{1}{2}\frac{v_4^2 + v_3^2}{v_2} - \frac{1}{2}\frac{v_4^2 + v_3^2}{v_1}\right) \end{aligned} \tag{38}$$

The idea is to look for values $v_i^*, i = 3, 4$ which maximize $TE_C(v_1, v_2, v_3, v_4)$ under the restriction $0 \leq v_4 \leq v_3 \leq v_2 \leq v_1$, while treating $v_1$ and $v_2$ as exogenous parameters.

We write $TE_C(v_3, v_4)$ for fixed $v_2$ and $v_1$. Note that $TE_C(v_3, v_4)$ is symmetric in its variables. It can be shown that its maximizers must be symmetric, that is, $v_3^* = v_4^*$. Hence we set $v_4 = v_3$ and look at total effort as a function of $v_3$ only:

$$TE_C(v_3) = \frac{1}{2}\left(v_2 + \frac{(v_2 - v_3)^2}{(v_1 - v_3)} + \frac{1}{2}\frac{(v_2 - v_3)}{(v_1 - v_3)}\frac{v_3^2}{v_2} + \frac{1}{2}\frac{(v_1 - v_2)}{(v_1 - v_3)}\frac{v_3^2}{v_1}\right). \tag{39}$$

In the remaining part of the proof we show that the function $TE_C$ is strictly convex for all $v_3 \in [v_2, v_4]$ , and that it achieves maxima at $v_3^* = v_2$ and at $v_3^* = v_4$ only if $v_4 = 0$.

The following facts are used:

$(i)$

$$TE'_C(v_3)_{v_3=v_2} = \frac{v_2}{v_1} > 0 \tag{40}$$

$(ii)$

$$TE'_C(v_3)_{v_3=0} = \frac{v_2(2v_2 - 4v_1)}{(-v_1)^2} < 0. \tag{41}$$

$(iii)$ The numerator of $TE_C(v_3)$ is a polynomial of third degree in $v_3$ . Hence it can have at most two roots between 0 and $v_1$, where its first derivative is equal to zero.

$(iv)$ $TE_C(0) = TE_C(v_2) = \frac{1}{2}v_2 + \frac{1}{2}\frac{v_2^2}{v_1}$.

Combining facts$(i)$, $(ii)$ and $(iv)$, it must be that $TE'_C$ has exactly one root in $[0, v_2]$ and it must be a minimum. Therefore $TE_C$ is strictly convex on the interval $[v_2, v_4]$ and it satisfies

$$TE_C(v_3) \leq TE_C(0) = TE_C(v_2) = \frac{1}{2}v_2 + \frac{1}{2}\frac{v_2^2}{v_1} = TE_B$$

*Q.E.D.*

## 6.2 Proof of Proposition 3

Let $q_{ij}^F$denote the limit probability (as $k$ goes to zero) that player $i$ beats player $j$ if these players meet in the final. Recall that $q_{ij}^S$ denote the limit probability that $i$ beats $j$ if they meet in a semifinal. Moreover, let $p_1(s), s \in \{A, B, C\}$ denote the limit probability that the strongest player, player 1, wins the tournament with a given seeding. We have

$$p_1(A) = q_{14}^S(q_{12}^F q_{23}^F + q_{13}^F q_{32}^F) \tag{42}$$

$$p_1(B) = q_{13}^S(q_{12}^F q_{24}^F + q_{14}^F q_{42}^F) \tag{43}$$

$$p_1(C) = q_{12}^S(q_{13}^F q_{34}^F + q_{14}^F q_{43}^F) \tag{44}$$

By Lemmas 1, 2, 3 and 4 we obtain that:

$$p_1(A) = \frac{1}{2}(1 - \frac{1}{2}\frac{v_2}{v_1} + 1 - \frac{1}{2}\frac{v_3}{v_1}) \tag{45}$$

$$p_1(B) = 1 - \frac{1}{2}\frac{v_2}{v_1} \tag{46}$$

$$p_1(C) = \frac{1}{2}\left(1 - \frac{1}{2}\frac{v_2 - \frac{1}{2}(v_3+v_4)}{v_1 - \frac{1}{2}(v_3+v_4)}\right)\left(1 - \frac{1}{2}\frac{v_3}{v_1} + 1 - \frac{1}{2}\frac{v_4}{v_1}\right) \tag{47}$$

We clearly have $p_1(A) > p_1(B)$. For the other inequality, note that $p_1(C)$ can be written as an expression of the sum $t \equiv v_3 + v_4$:

$$p_1(C) = p_1(C,t) = \frac{1}{2}\left(1 - \frac{1}{2}\frac{v_2 - \frac{t}{2}}{v_1 - \frac{t}{2}}\right)\left(2 - \frac{1}{2}\frac{t}{v_1}\right). \tag{48}$$

We first show that, on the interval of definition $[0, 2v_2]$ , $p_1(C,t)$ attains a maximum at $t = 2v_2$. Note that for all $t \in [0, 2v_2]$

$$\frac{\partial^2 p_1(C,t)}{\partial t^2} = -\frac{v_1 - v_2}{(t - 2v_1)^3} > 0 \tag{49}$$

Hence, $p_1(C,t)$ is strictly convex in $t$ and its maximum can not be interior. We also get

$$p_1(C, 2v_2) = p_1(C, 0) = 1 - \frac{1}{2}\frac{v_2}{v_1} \tag{50}$$

Thus, for any $v_3$ and $v_4$, we obtain that

$$p_1(C) \leq 1 - \frac{1}{2}\frac{v_2}{v_1}. \tag{51}$$

On the other hand, since $p_1(A)$ is strictly decreasing in $v_3$ , we obtain for all $v_3 \leq v_2$ that

$$p_1(A) \geq 1 - \frac{1}{2}\frac{v_2}{v_1}. \tag{52}$$

For all $v_3$ and $v_4$ such that $v_4 < v_3 < v_2$ we obtain $p_1(A) > p_1(C)$. *Q.E.D.*

## 6.3 Proof of Proposition 4

The results for seedings $A$:1-4,2-3 and $B$:1-3,2-4 follow by Lemmas 2 , 3 in the Appendix. We now display an example where $p_3(C) > p_4(C) > p_2(C)$.

By Lemma 4, we have:

$$p_2(C) = \frac{1}{2}\left(\frac{1}{2}\frac{v_2 - \frac{1}{2}(v_3+v_4)}{v_1 - \frac{1}{2}(v_3+v_4)}\right)\left(1 - \frac{1}{2}\frac{v_3}{v_2} + 1 - \frac{1}{2}\frac{v_4}{v_2}\right). \tag{53}$$

and for $j = 3, 4$,

$$p_j(C) = \frac{1}{2}\left[\left(1 - \frac{1}{2}\frac{v_2 - \frac{1}{2}(v_3+v_4)}{v_1 - \frac{1}{2}(v_3+v_4)}\right)\left(\frac{1}{2}\frac{v_j}{v_1}\right) + \left(\frac{1}{2}\frac{v_2 - \frac{1}{2}(v_3+v_4)}{v_1 - \frac{1}{2}(v_3+v_4)}\right)\left(\frac{1}{2}\frac{v_j}{v_2}\right)\right] \tag{54}$$

For $v_1 = 15, v_2 = 13, v_3 = 11, v_4 = 10$, we obtain $p_2(C) = 0.166 < 0.174 = p_4(C) < 0.324 = p_3(C)$. *Q.E.D.*

# References


Baye, M. and Kovenock, D. and C. de Vries (1996): "The All-Pay Auction with Complete Information", *Economic Theory* **8**, 291-305.

Baye, M. and Kovenock, D. and C. de Vries (1993): "Rigging the Lobbying Process", *American Economic Review* **83**, 289-294.

Chan, W., Courty, P. and H. Li (2007): "Suspense", *CEPR* Discussion paper No. 3251.

David, H. (1959): "Tournaments and Paired Comparisons", *Biometrika* **46**, 139-149.

Ehrenberg, R. and Bognanno, M. (1990): "Do Tournaments Have Incentive Effects?", *Journal of Labor Economics* **98**, 1307-1324

Glenn, W. (1960): "A Comparison of the Effectiveness of Tournaments", *Biometrika* **47**, 253-262.

Gradstein, M. and K.Konrad (1999): "Orchestrating Rent Seeking Contests", *The Economic Journal* **109**, 536-545.

Harbaugh, N. and T.Klumpp (2005): "Early Round Upsets and Championship Blowouts", *Economic Inquiry* **43**, 316-329.

Hillman, A. and J.Riley (1989): "Politically Contestable Rents and Transfers", *Economics and Politics* **1**, 17-40.

Horen, J. and Riezman, R. (1985): "Comparing Draws for Single Elimination Tournaments", *Operations Research* **3**(2), 249-262.

Hwang, F. (1982): "New Concepts in Seeding Knockout Tournaments", *American Mathematical Monthly* **89**, 235-239.

Klaassen, F. and Magnus, J. (2003): "Forecasting the Winner of a Tennis Match", *European Journal of Operational Research* **148**, 257-267.

Klumpp, Tilman, and Mattias K. Polborn (2006): "Primaries and the New Hampshire effect", *Journal of Public Economics,* 90, 1073-1114.

Konrad, Kai A. (2004): "Bidding in hierarchies", *European Economic Review*, 48, 1301-1308.

Rosen, S. (1986): "Prizes and Incentives in Elimination Tournaments", *American Economic Review* **74**, 701-715.

Searles, D. (1963): ”On the Probability of Winning with Different Tournament Procedures”, *Journal of the American Statistical Association* **58**, 1064-1081.

Schwenk, A. (2000): ”What is the Correct Way to Seed a Knockout Tournament”, *American Mathematical Monthly* **107**, 140-150.

Sunde, U. (2003): ”Potential, Prizes and Performance: Testing Tournament Theory with Professional Tennis Data”, *IZA Discussion Paper* **947**.

*USA Today* (2003): ”Weighted Brackets in NCAA draw fire”, March 26.

Warneryd, Karl (1998): "Distributional conflict and jurisdictional organization”, J*ournal of Public Economics*, 69, 435-450.