

# Contents

<b>1</b>	<b>Tournament Formats</b>	<b>2</b>
<b>2</b>	<b>Brackets</b>	<b>3</b>
2.1	Bracket Signatures . . . . .	6
2.2	Proper Brackets . . . . .	11
2.3	Ordered Brackets I . . . . .	18
2.4	Ordered Brackets II . . . . .	24
2.5	Multiple Elimination . . . . .	30

# 1 Tournament Formats

## Definition 1.0.1: Gameplay Function

A *gameplay function*  $g$  on a list of teams  $\mathcal{T}$  is a nondeterministic function  $g : \mathcal{T} \times \mathcal{T} \rightarrow \mathcal{T}$  with the following properties:

- $\mathbb{P}[g(t_1, t_2) = t_1] + \mathbb{P}[g(t_1, t_2) = t_2] = 1$ .
- $\mathbb{P}[g(t_1, t_2) = t_1] = \mathbb{P}[g(t_2, t_1) = t_1]$ .

A gameplay function represents a process in which two teams compete in a game, with one of them emerging as the winner. This model simplifies away effects like home-field advantage or teams improving over the course of a tournament: a gameplay function is fully described by a single probability for each pair of teams in the list.

## Definition 1.0.2: Tournament Format

A *tournament format* is an algorithm that takes as input a list of teams  $\mathcal{T}$  and a gameplay function  $g$  and outputs a champion  $t \in \mathcal{T}$ .

## Definition 1.0.3: Playing, Winning, and Losing

If a tournament format queries  $g$  on input  $(t_1, t_2)$  we say that  $t_1$  and  $t_2$  *played a game*. We say that the team that got outputted by  $g$  *won*, and the team that did not *lost*.

(This chapter will be fleshed out but I'm including the important definitions here for the sake of the next chapter.)

## 2 Brackets

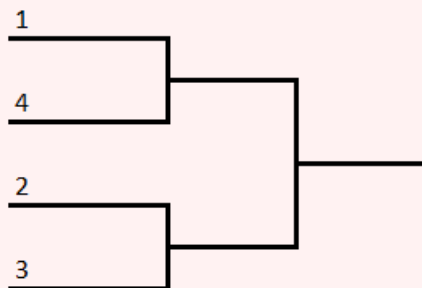
### Definition 2.0.1: Bracket

A *bracket* is a tournament format in which:

- Teams don't play any games after their first loss,
- Games are played until only one team has no losses, and that team is crowned champion, and
- The matchups between teams that have not yet lost are determined based on the ordering of the teams in  $\mathcal{T}$  in advance of the outcomes of any games.

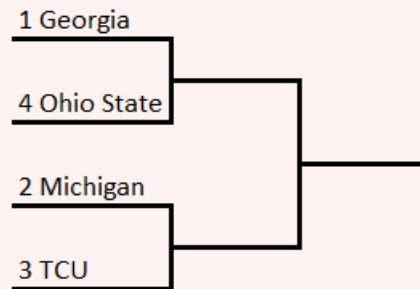
We can draw a bracket as a tree-like structure in the following way:

**Figure 2.0.2: The 2023 College Football Playoff**



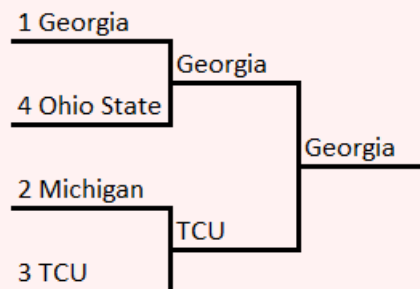
The numbers 1, 2, 3, and 4 indicate where the first, second, third and fourth team in  $\mathcal{T}$  are placed to start. In the actual 2023 College Football Playoff, the list of teams  $\mathcal{T}$  was Georgia, Michigan, TCU, and Ohio State, in that order, so the bracket was filled in like so:

**Figure 2.0.3: The 2023 CFP After Team Placement**



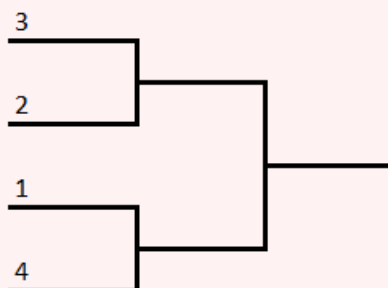
As games are played, we write the name of the winning teams on the corresponding lines. This bracket tells us that Georgia played Ohio State, and Michigan played TCU. Georgia and TCU won their respective games, and then Georgia beat TCU, winning the tournament.

**Figure 2.0.4: The 2023 CFP After Completion**



Rearranging the way the bracket is pictured, if it doesn't affect any of the matchups, does not create a new bracket. For example, Figure 2.0.5 is just another way to draw the same 2023 CFP Bracket.

Figure 2.0.5: Alternative Drawing of the 2023 CFP



One key piece of bracket vocabulary is the *round*.

**Definition 2.0.6: Round**

A *round* is a set of games such that the winners of each of those games have the same number of games remaining to win the tournament.

For example, the 2023 CFP has two rounds. The first round included the games Georgia vs Ohio State and Michigan vs TCU, and the second round was just a single game: Georgia vs TCU.

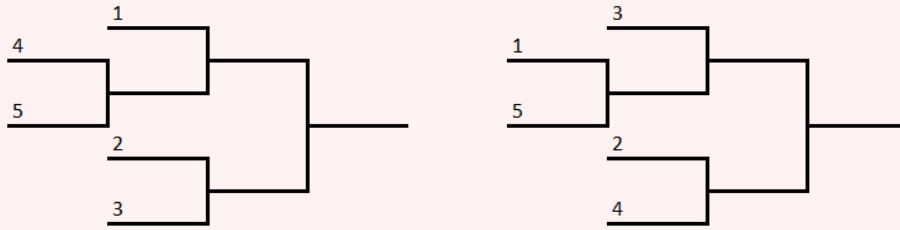
## 2.1 Bracket Signatures

### Definition 2.1.1: Shape

The *shape* of a bracket is the tree that underlies it.

For example, the following two brackets have the same shape:

Figure 2.1.2: Two Brackets with the Same Shape



### Definition 2.1.3: Bye

A team has a *bye* in round  $r$  if it plays no games in round  $r$  or before.

One way to describe the shape of a bracket is its signature.

### Definition 2.1.4: Bracket Signature

The *signature*  $[[\mathbf{a}_0; \dots; \mathbf{a}_r]]$  of an  $r$ -team bracket  $\mathcal{A}$  is a list of natural numbers, such that  $a_i$  is the number of teams with  $i$  byes.

The signature of a bracket is defined only by its shape: the two brackets in Figure 2.1.2 have the same shape, so they also have the same signature.

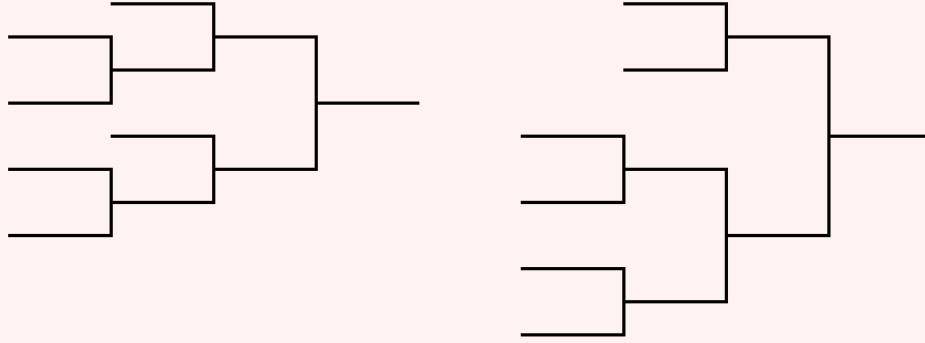
The signatures of the brackets discussed in this section are shown in Figure 2.1.5. It's worth verifying the signatures we've seen so far and trying to draw brackets with the signatures we haven't yet before moving on.

**Figure 2.1.5: The Signatures of Some Brackets**

Bracket	Signature
2023 College Football Playoff	$[[4; 0; 0]]$
The brackets in Figure 2.1.2	$[[2; 3; 0; 0]]$
The brackets in Figure 2.1.6	$[[4; 2; 0; 0]]$
2023 WCC Men's Basketball Tournament	$[[4; 2; 2; 2; 0; 0]]$

Two brackets with the same shape must have the same signature, but the converse is not true: two brackets with different shapes can have the same signature. For example, both bracket shapes depicted in Figure 2.1.6 have the signature  $[[4; 2; 0; 0]]$ .

**Figure 2.1.6: Two Shapes with the Signature  $[[4; 2; 0; 0]]$**



Despite this, bracket signatures are a useful way to talk about the shape of a bracket. Communicating a bracket's signature is a lot easier than communicating its shape, and much of the important information (such as how many games each team must win in order to win the tournament) is contained in the signature.

Bracket signatures have one more important property.

**Theorem 2.1.7**

Let  $\mathcal{A} = [[\mathbf{a}_0; \dots; \mathbf{a}_r]]$  be a list of natural numbers. Then  $\mathcal{A}$  is a bracket

signature if and only if

$$\sum_{i=0}^r a_i \cdot \left(\frac{1}{2}\right)^{r-i} = 1.$$

*Proof.* Let  $\mathcal{A}$  be the signature for some bracket. Assume that every game in the bracket was a coin flip, and consider each team's probability of winning the tournament. A team that plays their first game in the  $i$ th round must win  $r - i$  games to win the tournament, and so will do so with probability  $\left(\frac{1}{2}\right)^{r-i}$ . For each  $i \in \{0, \dots, r\}$ , there are  $a_i$  teams that play their first game in the  $i$ th round, so (because any two teams winning are mutually exclusive)

$$\sum_{i=0}^r a_i \cdot \left(\frac{1}{2}\right)^{r-i}$$

is the probability that one of the teams wins, which is 1.

We prove the other direction by induction on  $r$ . If  $r = 0$ , then the only list with the desired property is  $[[1]]$ , which is the signature for the unique one-team bracket. For any other  $r$ , first note that  $a_0$  must be even: if it were odd, then

$$\begin{aligned} \sum_{i=0}^r a_i \cdot \left(\frac{1}{2}\right)^{r-i} &= \frac{1}{2^r} \cdot \sum_{i=0}^r a_i \cdot 2^i \\ &= \frac{1}{2^r} \cdot \left(a_0 + 2 \sum_{i=1}^r a_i \cdot 2^{i-1}\right) \\ &= k/2^r && \text{for some odd } k \\ &\neq 1. \end{aligned}$$

Now, consider the signature  $\mathcal{B} = [[\mathbf{a}_1 + \mathbf{a}_0/2; \mathbf{a}_2; \dots; \mathbf{a}_r]]$ . By induction, there exists a bracket with signature  $\mathcal{B}$ . But if we take that bracket and replace  $a_0/2$  of the teams with no byes with a game whose winner gets placed on that line, we get a new bracket with signature  $\mathcal{A}$ .  $\square$

In the next few sections, we will use the language and properties of bracket



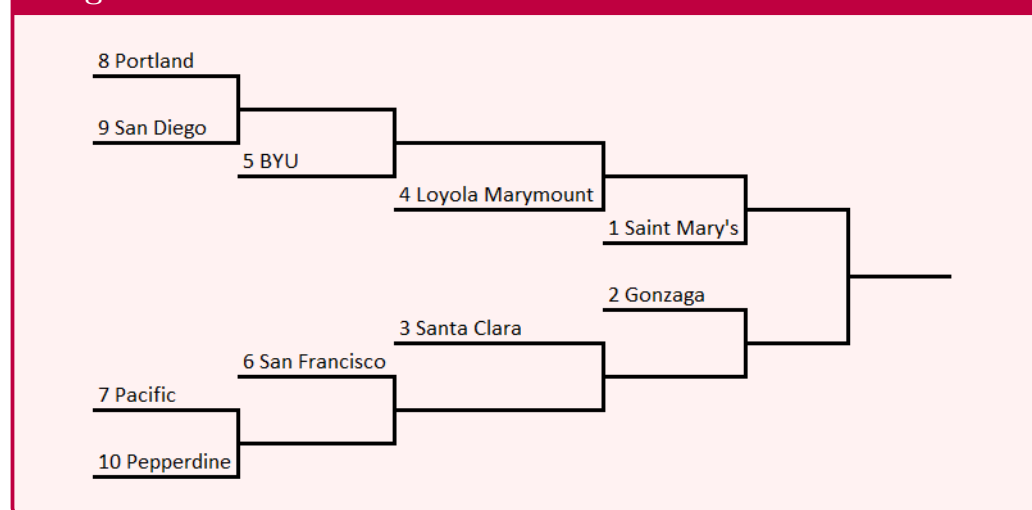
signatures to describe the brackets that we work with. For now though, let's return to the 2023 College Football Playoff. The bracket used in the 2023 CFP has a special property that not all brackets have: it is *balanced*.

### Definition 2.1.8: Balanced Bracket

A *balanced bracket* is a bracket in which none of the teams have byes.

The 2023 West Coast Conference Men's Basketball Tournament, on the other hand, is unbalanced:

**Figure 2.1.9: The 2023 WCC Men's Basketball Tournament**



Saint Mary's and Gonzaga each have three byes and so only need to win two games to win the tournament, while Portland, San Diego, Pacific, and Pepperdine need to win five. Unsurprisingly, this format conveys a massive advantage to Saint Mary's and Gonzaga, but this was intentional: those two teams were being rewarded for doing the best during the regular season.

In many cases, however, it is undesirable to grant advantages to certain teams over others. One might hope, for any  $n$ , to be able to construct a balanced bracket for  $n$  teams, but unfortunately this is rarely possible.

### Theorem 2.1.10

There exists an  $n$ -team balanced bracket if and only if  $n$  is a power of two.

*Proof.* A bracket is balanced if no teams have byes, which is true exactly when its signature is of the form  $\mathcal{A} = [[\mathbf{n}; \mathbf{0}; \dots; \mathbf{0}]]$  where  $n$  is the number of teams in the bracket. If  $n$  is a power of two, then by Theorem 2.1.7  $\mathcal{A}$  is indeed a bracket signature and so points to a balanced bracket for  $n$  teams. If  $n$  is not a power of two, however, then Theorem 2.1.7 tells us that  $\mathcal{A}$  is not a bracket signature, and so no balanced brackets exist for  $n$  teams.  $\square$

Given this, brackets are not a great option when we want to avoid giving some teams advantages over others unless we have a power of two teams. They are a fantastic tool, however, if doling out advantages is the goal, perhaps after some teams did better during the regular season and ought to be rewarded with an easier path in the bracket.

## 2.2 Proper Brackets

### Definition 2.2.1: Seeding

The *seeding* of an  $n$ -team bracket is the arrangement of the numbers 1 through  $n$  in the bracket.

Together, the shape and seeding fully specify a bracket.

### Definition 2.2.2: $i$ -seed

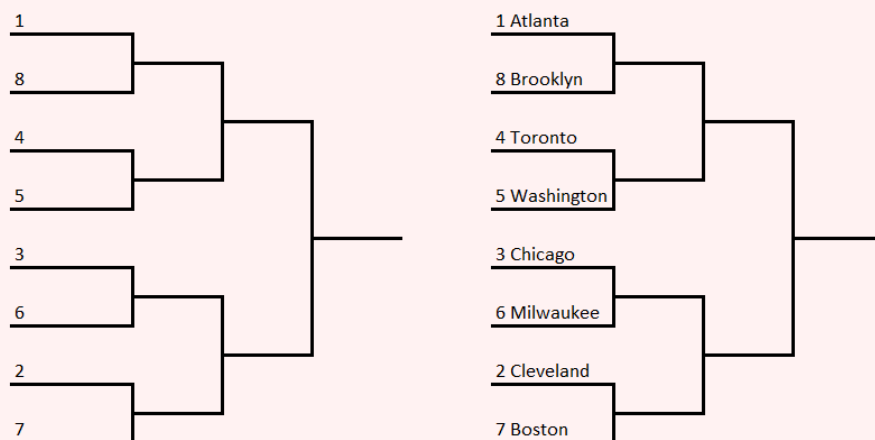
The  $i$ -seed is  $i$ th team in list of teams inputted to a bracket. Sometimes we use the notation  $t_i$  to refer to the  $i$ -seed in a bracket.

### Definition 2.2.3: Higher and Lower Seeds

Somewhat confusingly, convention is that lower numbers are the *higher seeds*, and higher numbers are the *lower seeds*.

Seeding is typically used to reward better and more deserving teams. As an example, on the left is the 8-team bracket used in the 2015 NBA Eastern Conference Playoffs. At the end of the regular season, the top eight teams in the Eastern Conference were ranked and placed into the bracket as shown on the right.

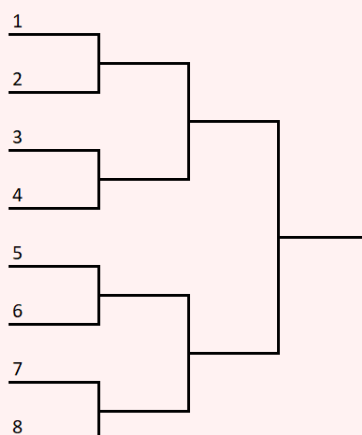
Figure 2.2.4: 2015 NBA Eastern Conference Playoffs



Despite this bracket being balanced, the higher seeds are still at advantage: they have an easier set of opponents. Compare 1-seed Atlanta, whose first two rounds are versus 8-seed Brooklyn and then (most likely) 4-seed Toronto, versus 7-seed Boston, whose first two rounds are versus 2-seed Cleveland and then (most likely) 3-seed Chicago. Atlanta's schedule is far easier: despite them having the same number of games to win as Boston, Atlanta is expected to play lower seeds in each round than Boston will.

Thus, we've identified two ways in which brackets can convey an advantage onto certain teams: by giving them more byes, and by giving them easier (expected) opponents. Not every seeding of a bracket does this: for example, consider the following alternative seeding for the 2015 NBA Eastern Conference Playoffs.

**Figure 2.2.5: An Alternative Seeding of the 2015 NBA Eastern Conference Playoffs**



This seeding does a very poor job of rewarding the higher-seeded teams: the 1- and 2-seeds are matched up in the first round, while the easiest road is given to the 7-seed, who plays the 8-seed in the first round and then (most likely) the 5-seed in the second. Since the whole point of seeding is to give the higher-seeded teams an advantage, we introduce the concept of a *proper seeding*.

### Definition 2.2.6: Chalk

We say a tournament *went chalk* if the higher-seeded team won every game during the tournament.

### Definition 2.2.7: Proper Seeding

A *proper seeding* of a bracket is one such that if the bracket goes chalk, in every round it is better to be a higher-seeded team than a lower-seeded one, where:

- (1) It is better to have a bye than to play a game.
- (2) It is better to play a lower seed than to play a higher seed.

### Definition 2.2.8: Proper Bracket

A *proper bracket* is a bracket that has been properly seeded.

It is clear that the actual 2015 NBA Eastern Conference Playoffs was properly seeded, while our alternative seeding was not.

A few quick lemmas about proper brackets:

### Lemma 2.2.9

In a proper bracket, if  $m$  teams have a bye in a given round, those teams must be seeds 1 through  $m$ .

*Proof.* If they did not, the seeding would be in violation of condition (1).  $\square$

### Lemma 2.2.10

If a proper bracket goes chalk, then after each round the  $m$  teams remaining will be the top  $m$  seeds.

*Proof.* We will prove the contrapositive. Assume that for some  $i < j$ , after some round,  $t_i$  has been eliminated but  $t_j$  is still alive. Let  $k$  be the seed of the team that  $t_i$  lost to. Because the bracket went chalk,

$k < i$ . Now consider what  $t_j$  did in that round. If they had a bye, then the bracket violates condition (1). Assume instead they played  $t_\ell$ . They beat  $t_\ell$ , so  $j < \ell$ , giving,

$$k < i < j < \ell.$$

In the round that  $t_i$  was eliminated,  $t_i$  played  $t_k$ , while  $t_j$  played  $t_\ell$ , violating condition (2). Thus, the bracket is not proper.  $\square$

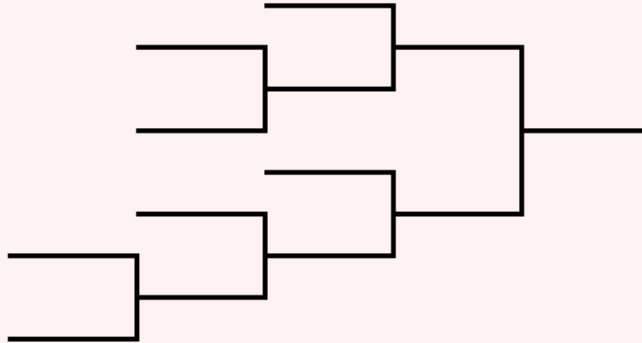
### Lemma 2.2.11

In a proper bracket, if  $m$  teams have a bye and  $k$  games are being played in a given round, then if the bracket goes chalk those matchups will be seed  $m + i$  vs seed  $(m + 2k + 1) - i$  for  $i \in \{1, \dots, k\}$ .

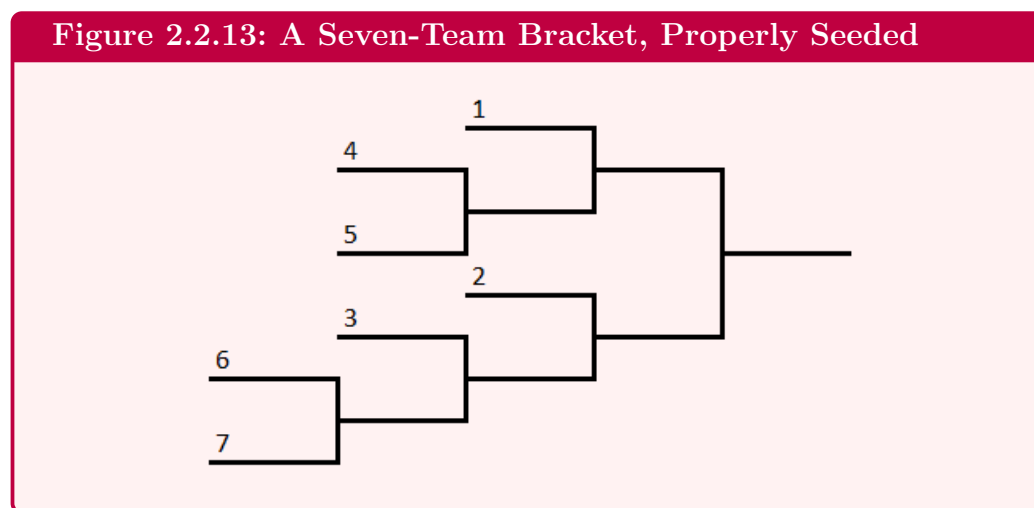
*Proof.* In the given round, there are  $m + 2k$  teams remaining. Theorem 2.2.10 tells us that (if the bracket goes chalk) those teams must be seeds 1 through  $m + 2k$ . Theorem 2.2.9 tells us that seeds 1 through  $m$  must have a bye, so the teams playing must be seeds  $m + 1$  through  $m + 2k$ . Then condition (2) tells us that the matchups must be exactly  $m + i$  vs seed  $(m + 2k + 1) - i$  for  $i \in \{1, \dots, k\}$ .  $\square$

We can use Lemmas 2.2.9 through 2.2.11 to properly seed various bracket shapes. For example, consider the following seven-team shape:

Figure 2.2.12: A Seven-Team Bracket Shape



Lemma 2.2.9 tells us that the first-round matchup must be between the 6-seed and the 7-seed. Lemma 2.2.11 tells us that if the bracket goes chalk, the second-round matchups must be 3v6 and 4v5, so the 3-seed play the winner of the first-round matchup. Finally, we can apply Lemma 2.2.11 again to the semifinals to find that the 1-seed should play the winner of the 4v5 matchup, while the 2-seed should play the winner of the 3v(6v7) matchup. In total, our proper seeding looks like:

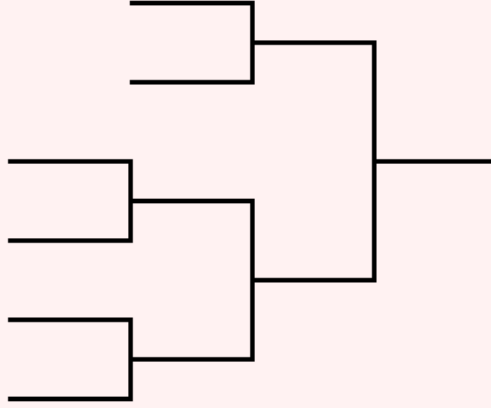


We can also quickly simulate the bracket going chalk to verify Lemma 2.2.10.

Lemmas 2.2.9 through 2.2.11 are quite powerful. It is not a coincidence that we managed to specify exactly what a proper seeding of the above bracket must look like with no room for variation: soon we will prove that the proper seeding for a particular bracket shape is unique.

But not every shape admits even this one proper seeding. Consider the following six-team shape:

**Figure 2.2.14: A Six-Team Bracket Shape**



This shape admits no proper seedings. Lemma 2.2.9 requires that the two teams getting byes be the 1- and 2-seed, but this violates Lemma 2.2.11 which requires that in the second round the 1- and 2-seeds do not play each other. So how can we think about which shapes admit proper seedings?

**Theorem 2.2.15: The Fundamental Theorem of Brackets**

There is exactly one proper bracket with each bracket signature.

*Proof.* Let  $\mathcal{A} = [[\mathbf{a}_0; \dots; \mathbf{a}_r]]$  be an  $n$ -team bracket signature. We proceed by induction on  $r$ . If  $r = 0$ , then the only possible bracket signature is  $[[1]]$ , and it points to the unique one-team bracket, which is indeed proper.

For any other  $r$ , the first-round matchups of a proper bracket with signature  $\mathcal{A}$  are defined by Theorem 2.2.11. Then if those matchups go chalk, we are left with a proper bracket with signature  $\mathcal{B} = [[\mathbf{a}_1 + \mathbf{a}_0/2; \mathbf{a}_2; \dots; \mathbf{a}_r]]$ , which induction tells us exists admits exactly one proper bracket.

Thus both the first-round matchups and the rest of the bracket are determined, and by combining them we get a proper bracket with signature  $\mathcal{A}$ , there is exactly one proper bracket with signature  $\mathcal{A}$ .



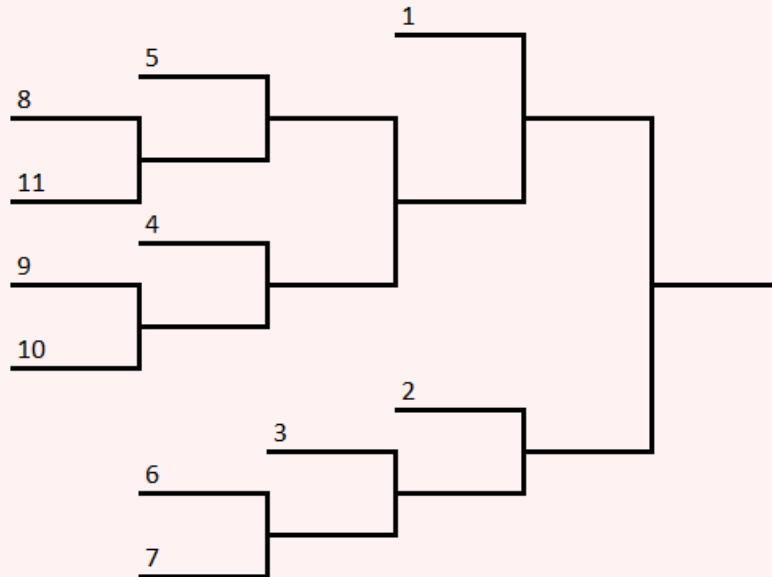
□

The fundamental theorem of brackets means that we can refer to the proper bracket  $\mathcal{A} = [[\mathbf{a}_0; \dots; \mathbf{a}_r]]$  in a well-defined way, as long as

$$\sum_{i=0}^r a_i \cdot \left(\frac{1}{2}\right)^{r-i} = 1.$$

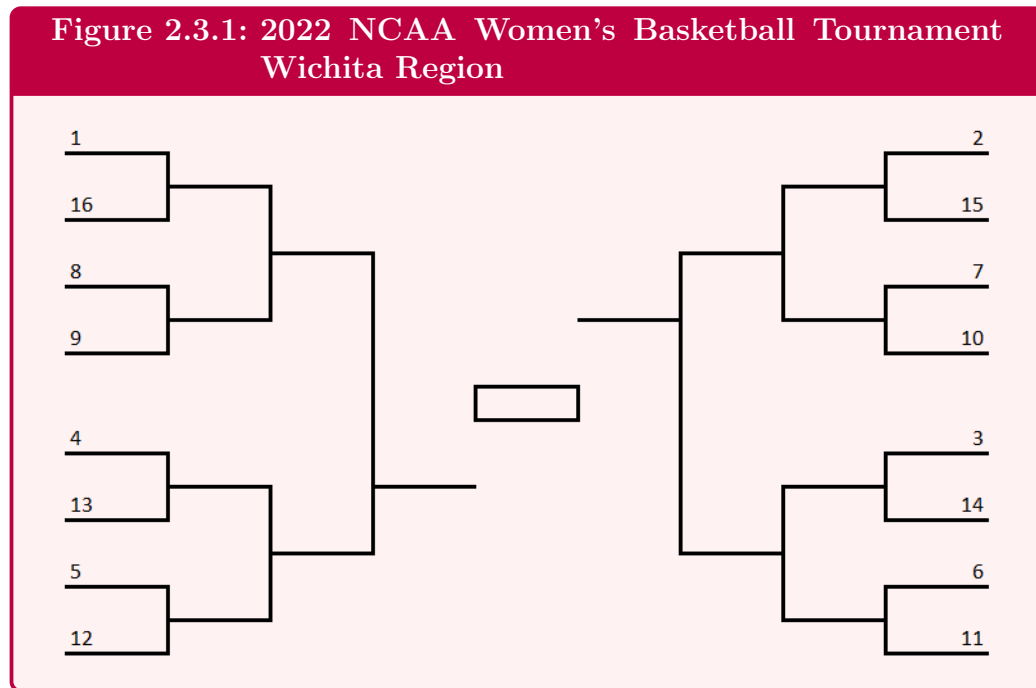
In practice, virtually every sports league that uses a traditional bracket uses a proper one: while different leagues take very different approaches to how many byes to give teams (compare the 2023 West Coast Conference Men's Basketball Tournament with the 2015 NBA Eastern Conference Play-offs), they are almost all proper. This makes bracket signatures a convenient labeling system for the set of brackets that we might reasonably encounter. They also are a powerful tool for specifying new brackets: if you are interested in (say) an eleven-team bracket where four teams get no byes, four teams get one bye, one team gets two byes and two teams get three byes, we can describe the proper bracket with those specs as  $[[4; 4; 1; 2; 0; 0]]$  and use Lemmas 2.2.9 through 2.2.11 to draw it with ease:

**Figure 2.2.16: The Proper Bracket of Signature  $[[4; 4; 1; 2; 0; 0]]$**



## 2.3 Ordered Brackets I

Let's consider the proper bracket  $[[16; 0; 0; 0; 0]]$ , which was used in the 2022 NCAA Women's Basketball Tournament Wichita Region, and is shown below. (Sometimes brackets are drawn in the manner below, with teams starting on both sides and the winner of each side playing in the championship game.)



The definition of a proper seeding ensures that as long as the bracket goes chalk (that is, higher seeds always beat lower seeds), it will always be better to be a higher seed than a lower seed. But what if it doesn't go chalk?

One counter-intuitive fact about the NCAA Basketball Tournament is that it is probably better to be a 10-seed than a 9-seed. (This doesn't violate the proper seeding property because 9-seeds have an easier first-round matchup than 10-seeds, and for further rounds, proper seedings only care about what happens if the bracket goes chalk, which would eliminate both the 9-seed and 10-seed in the first round.) Why? Let's look at whom each seed-line matchups against in the first two rounds:

**Figure 2.3.2: NCAA Basketball Tournament 9- and 10-seed Schedules**

Seed	First Round	Second Round
9	8	1
10	7	2

The 9-seed has an easier first-round matchup, while the 10-seed has an easier second-round matchup. However, this isn't quite symmetrical. Because the teams are probably drawn from a roughly normal distribution, the expected difference in skill between the 1- and 2-seeds is far greater than the expected difference between the 7- and 8-seeds, implying that the 10-seed does in fact have an easier route than the 9-seed.

Nate Silver investigated this matter in full, finding that in the NCAA Basketball Tournament, seed-lines 10 through 15 give teams better odds of winning the region than seed-lines 8 and 9. Of course this does not mean that the 11-seed (say) has a better chance of winning a given region than the 8-seed does, as the 8-seed is a much better team than the 11-seed. But it does mean that the 8-seed would love to swap places with the 11-seed, and that doing so would increase their odds to win the region.

This is not a great state of affairs: the whole point of seeding is confer an advantage to higher-seeded teams, and the proper bracket  $[[16; 0; 0; 0; 0]]$  is failing to do that. Not to mention that giving lower-seeded teams an easier route than higher-seeded ones can incentivize teams to lose during the regular season in order to try to get a lower but more advantageous seed.

To fix this, we need a stronger notion of what makes a bracket effective than properness. The issue with proper seedings is the false assumption that higher-seeded teams will always beat lower-seeded teams. A more nuanced assumption might look like this:

**Definition 2.3.3: Strongly Stochastically Transitive**

A list of teams is *strongly stochastically transitive* if for each  $i, j, k$  such that  $j < k$ ,

$$\mathbb{P}(t_i \text{ beats } t_j) \leq \mathbb{P}(t_i \text{ beats } t_k).$$

A list of teams being strongly stochastically transitive (or SST) captures the intuition that each team ought to do better against lower-seeded teams

than against higher-seeded teams. A few quick implications of this definition are:

**Corollary 2.3.4**

If a list of teams is SST, then for each  $i < j$ ,  $\mathbb{P}(t_i \text{ beats } t_j) \geq 0.5$ .

**Corollary 2.3.5**

If a list of teams is SST, then for each  $i, j, k, \ell$  such that  $i < j$  and  $k < \ell$ ,

$$\mathbb{P}(t_i \text{ beats } t_\ell) \geq \mathbb{P}(t_j \text{ beats } t_k).$$

Note that not every set of teams can be seeded to be SST. Consider, for example, the game of rock-paper-scissors. Rock beats paper which beats scissors which beats rock, so no ordering of these “teams” will be SST. For our purposes, however, SST will work well enough.

Our new, nuanced alternative a proper bracket is an *ordered bracket*.

**Definition 2.3.6: Ordered Bracket**

A bracket and seeding are *ordered* if, for any SST list of teams, if  $i < j$ , then  $\mathbb{P}(t_i \text{ wins the tournament}) \geq \mathbb{P}(t_j \text{ wins the tournament})$ .

In an informal sense, a bracket being ordered is the strongest thing we can want without knowing more about why the tournament is being played. Depending on the situation, we might be interested in a format that almost always declares the most-skilled team as the winner, or in a format that gives each team roughly the same chance of winning, or anywhere in between. But certainly, better teams should win more, which is what the ordered bracket condition requires.

In particular, a bracket being ordered is a stronger claim than it being proper.

**Theorem 2.3.7**

Every ordered bracket is proper.

*Proof.* Let  $\mathcal{A}$  be an ordered  $n$ -team bracket with  $r$  rounds.

Consider a list of teams such that every team wins every game with probability 0.5. These teams are SST. A team that plays their first game in the  $i$ th round will win the tournament with probability  $(\frac{1}{2})^{r-i}$ , so teams that get more byes will have a higher probability to win the tournament than teams with fewer byes. This implies that higher-seeded teams must have more byes than lower-seeded teams, so in each round, the teams with byes must be the highest-seeded teams that are still alive. Thus, condition (1) is met.

We show that condition (2) is met by proving the stronger condition from Lemma 2.2.11: if  $m$  teams have a bye and  $k$  games are being played in round  $s$ , then if the bracket goes chalk, those matchups will be  $t_{m+i}$  vs  $t_{(m+2k+1)-i}$  for  $i \in \{1, \dots, k\}$ . We show this by strong induction on  $s$  and on  $i$ .

Assume that this is true for every round up until  $s$  and for all  $i < j$  for some  $j$ . Let  $\ell = (m + 2k + 1) - j$ . We want to show that if the bracket goes chalk,  $t_{m+j}$  will face off against seed  $t_\ell$  in the given round. Consider the following SST matchup table: every game is a coin flip, except for games involving a team seeded  $\ell$  or lower, in which case the higher seed always wins. Then, each team seeded between  $\ell - 1$  and  $m + j$  will win the tournament with probability  $(\frac{1}{2})^{r-s}$ , other than the team slated to play  $t_\ell$  in round  $s$  who wins with probability  $(\frac{1}{2})^{r-i-1}$ . In order for  $\mathcal{B}$  to be ordered, that team must be  $t_{m+j}$ .

Thus  $\mathcal{A}$  satisfies both conditions, and so is a proper bracket. □

With Theorem 2.3.7, we can use the language of bracket signatures to describe ordered brackets without worrying that two ordered brackets might share a signature. Now we examine three particularly important examples of ordered brackets.

We begin with the unique one-team bracket.

**Figure 2.3.8: The One-Team Bracket  $[[1]]$ .**



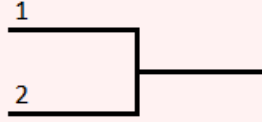
**Theorem 2.3.9**

The one-team bracket  $[[1]]$  is ordered.

*Proof.* Since there is only team, the ordered bracket condition is vacuously true.  $\square$

Next we look at the unique two-team bracket.

**Figure 2.3.10: The Two-Team Bracket  $[[2; 0]]$ .**



**Theorem 2.3.11**

The two-team bracket  $[[2; 0]]$  is ordered.

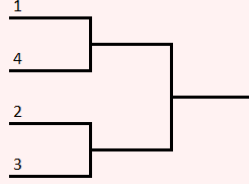
*Proof.* Let  $\mathcal{A} = [[2; 0]]$ . Then,

$$\mathbb{P}[t_1 \text{ wins } \mathcal{A}] = \mathbb{P}[t_1 \text{ beats } t_2] \geq 0.5 \geq \mathbb{P}[t_2 \text{ beats } t_1] = \mathbb{P}[t_2 \text{ wins } \mathcal{A}],$$

so  $\mathcal{A}$  is ordered.  $\square$

And thirdly, we show that the balanced four-team bracket is ordered.

**Figure 2.3.12: The Four-Team Bracket  $[[4; 0; 0]]$ .**



### Theorem 2.3.13

The four-team bracket  $[[4; 0; 0]]$  is ordered.

*Proof.* Let  $\mathcal{A} = [[2; 0]]$ . Let  $p_{ij}$  be the probability that the  $i$ -seed beats the  $j$ -seed. Then,

$$\begin{aligned}
 \mathbb{P}[t_1 \text{ wins } \mathcal{A}] &= p_{14} \cdot (p_{23}p_{12} + p_{32}p_{13}) \\
 &= p_{14}p_{23}p_{12} + p_{14}p_{32}p_{13} \\
 &\geq p_{14}p_{23}p_{21} + p_{24}p_{41}p_{23} \\
 &= p_{23} \cdot (p_{14}p_{21} + p_{41}p_{24}) \\
 &= \mathbb{P}[t_2 \text{ wins } \mathcal{A}]
 \end{aligned}$$

$$\begin{aligned}
 \mathbb{P}[t_2 \text{ wins } \mathcal{A}] &= p_{23} \cdot (p_{14}p_{21} + p_{41}p_{24}) \\
 &\geq p_{32} \cdot (p_{14}p_{31} + p_{41}p_{34}) \\
 &= \mathbb{P}[t_3 \text{ wins } \mathcal{A}]
 \end{aligned}$$

$$\begin{aligned}
 \mathbb{P}[t_3 \text{ wins } \mathcal{A}] &= p_{32} \cdot (p_{14}p_{31} + p_{41}p_{34}) \\
 &= p_{32}p_{14}p_{31} + p_{32}p_{41}p_{34} \\
 &\geq p_{42}p_{23}p_{41} + p_{32}p_{41}p_{43} \\
 &= p_{41} \cdot (p_{23}p_{42} + p_{32}p_{43}) \\
 &= \mathbb{P}[t_4 \text{ wins } \mathcal{A}]
 \end{aligned}$$

Thus the four-team bracket  $[[4; 0; 0]]$  is ordered. □

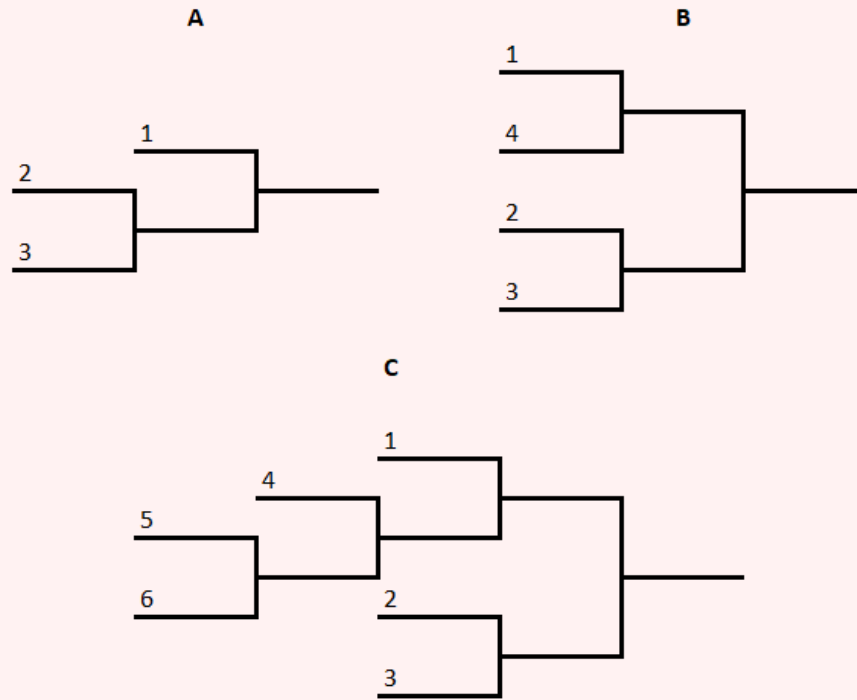
In the next section, we move on from describing particular ordered brackets in favor of a more general result.

## 2.4 Ordered Brackets II

We now attempt to completely specify the set of ordered brackets, culminating in Corollary 2.4.4 and Theorem 2.4.6.

We begin with a theorem that allows us to combine two smaller ordered brackets together into a larger ordered one by having the winner of one of the brackets be treated as the lowest seed in the other. This is depicted in Figure 2.4.1.

**Figure 2.4.1: The Setup of Theorem 2.4.2**



### Theorem 2.4.2

If  $\mathcal{A}$  and  $\mathcal{B}$  are  $n$ - and  $m$ -team ordered brackets, respectively, we can construct an  $(n + m - 1)$ -team ordered bracket by replacing the lowest seed in  $\mathcal{B}$  with the entire bracket  $\mathcal{A}$  (and lowering the seed of each team in  $\mathcal{A}$  by  $m - 1$ ).



*Proof.* Let  $\mathcal{C}$  be the bracket formed by merging  $\mathcal{A}$  and  $\mathcal{B}$ . We divide proving that  $\mathcal{C}$  is ordered into proving three substatements:

1. For  $i < j < m$ ,  $\mathbb{P}(t_i \text{ wins } \mathcal{C}) \geq \mathbb{P}(t_j \text{ wins } \mathcal{C})$
2.  $\mathbb{P}(t_{m-1} \text{ wins } \mathcal{C}) \geq \mathbb{P}(t_m \text{ wins } \mathcal{C})$
3. For  $m \leq i < j$ ,  $\mathbb{P}(t_i \text{ wins } \mathcal{C}) \geq \mathbb{P}(t_j \text{ wins } \mathcal{C})$

Together, these show that  $\mathcal{C}$  is ordered.

We begin with the first sub-statement. Let  $i < j < m$ . Then,

$$\begin{aligned}
 \mathbb{P}(t_i \text{ wins } \mathcal{C}) &= \mathbb{P}(t_i \text{ wins } \mathcal{B}) \\
 &= \sum_{k=n}^{n+m-1} \mathbb{P}(t_i \text{ wins } \mathcal{B} \mid t_k \text{ wins } \mathcal{A}) \cdot \mathbb{P}(t_k \text{ wins } \mathcal{A}) \\
 &\geq \sum_{k=n}^{n+m-1} \mathbb{P}(t_j \text{ wins } \mathcal{B} \mid t_k \text{ wins } \mathcal{A}) \cdot \mathbb{P}(t_k \text{ wins } \mathcal{A}) \\
 &= \mathbb{P}(t_j \text{ wins } \mathcal{B}) \\
 &= \mathbb{P}(t_j \text{ wins } \mathcal{C})
 \end{aligned}$$

The first and last equalities follow from the structure of  $\mathcal{C}$ , and the inequality follows from  $\mathcal{B}$  being ordered.

Now the second sub-statement.

$$\begin{aligned}
 \mathbb{P}(t_{m-1} \text{ wins } \mathcal{C}) &= \mathbb{P}(t_{m-1} \text{ wins } \mathcal{B}) \\
 &\geq \mathbb{P}(t_m \text{ wins } \mathcal{B} \mid t_m \text{ wins } \mathcal{A}) \\
 &\geq \mathbb{P}(t_m \text{ wins } \mathcal{B} \mid t_m \text{ wins } \mathcal{A}) \cdot \mathbb{P}(t_m \text{ wins } \mathcal{A}) \\
 &= \mathbb{P}(t_m \text{ wins } \mathcal{C})
 \end{aligned}$$

The equalities follow from the structure of  $\mathcal{C}$ , and the first inequality follows from  $\mathcal{B}$  being ordered.

Finally we show the third sub-statement. Let  $m \leq i < j$ . Then,

$$\begin{aligned}\mathbb{P}(t_i \text{ wins } \mathcal{C}) &= \mathbb{P}(t_i \text{ wins } \mathcal{B} \mid t_i \text{ wins } \mathcal{A}) \cdot \mathbb{P}(t_i \text{ wins } \mathcal{A}) \\ &\geq \mathbb{P}(t_i \text{ wins } \mathcal{B} \mid t_i \text{ wins } \mathcal{A}) \cdot \mathbb{P}(t_j \text{ wins } \mathcal{A}) \\ &\geq \mathbb{P}(t_j \text{ wins } \mathcal{B} \mid t_j \text{ wins } \mathcal{A}) \cdot \mathbb{P}(t_j \text{ wins } \mathcal{A}) \\ &= \mathbb{P}(t_j \text{ wins } \mathcal{C})\end{aligned}$$

The equalities follow from the structure of  $\mathcal{C}$ , the first inequality from  $\mathcal{A}$  being ordered, and the second inequality from the teams being SST.

We have shown all three sub-statements, and so  $\mathcal{C}$  is ordered.  $\square$

### Corollary 2.4.3

If  $\mathcal{A} = [[\mathbf{a}_0; \dots; \mathbf{a}_r]]$  and  $\mathcal{B} = [[\mathbf{b}_0; \dots; \mathbf{b}_s]]$  are ordered brackets, then  $\mathcal{C} = [[\mathbf{a}_0; \dots; \mathbf{a}_r + \mathbf{b}_0 - \mathbf{1}; \dots; \mathbf{b}_s]]$  is an ordered bracket as well.

We can then construct the set of brackets that we have thus far shown are ordered. We do this by starting with  $\{[[\mathbf{1}]], [[\mathbf{2}; \mathbf{0}]], [[\mathbf{4}; \mathbf{0}; \mathbf{0}]]\}$  and then repeatedly applying the above theorem on the set to expand it. In other words,

### Corollary 2.4.4

Any bracket signature formed by the following process is ordered:

1. Start with the list  $[[\mathbf{0}]]$  (note that this not yet a bracket signature).
2. As many times as desired, prepend the list with  $[[\mathbf{1}]]$  or  $[[\mathbf{3}; \mathbf{0}]]$ .
3. Then, add 1 to the first element in the list, turning it into a bracket signature.

Corollary 2.4.4 uses the tools that we have developed so far to demark a set of brackets as ordered. Somewhat surprisingly, this set is complete: any bracket not reachable using the process in Corollary 2.4.4 is not ordered. To prove this we first show a lemma.

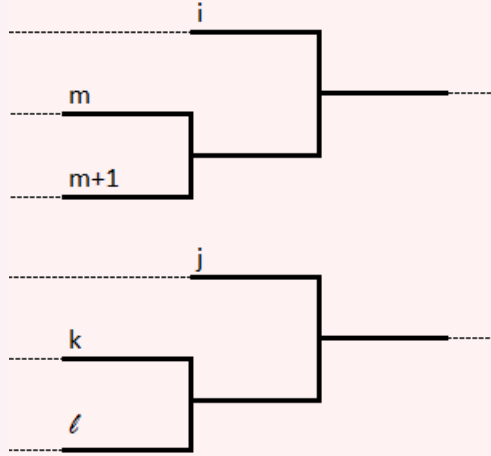
### Lemma 2.4.5

If  $\mathcal{A}$  is an  $r$ -round ordered bracket and two games are played in round  $s$ , then the winners of those games must play each other in round  $s+1$ . Further, this must be the only game of round  $s+1$ .

*Proof.* Let  $m$  be the number of teams that are still alive in  $\mathcal{A}$  after round  $s$ . Because  $\mathcal{A}$  is proper, we know that if the bracket goes chalk,  $t_m$  will play  $t_{m+1}$  in round  $s$ .

Now, assume for contradiction that two games are played in round  $s$  but the winners of those games do not play each other in the following round. So let  $k < \ell$  such that, if the bracket goes chalk,  $t_k$  plays  $t_\ell$  in round  $s$ , but the winner of that game doesn't play the winner of the  $t_m$  game in round  $s+1$ . In fact, let  $k$  be the lowest such seed. Finally, let  $i$  and  $j$  be such that if the bracket goes chalk, in round  $s+1$ ,  $t_i$  will play  $t_m$  and  $t_j$  will play  $t_k$ .

The situation so far (assuming all omitted games go chalk):



We can use  $\mathcal{A}$ 's properness to determine that  $i < j < k < m < m+1 < \ell$ .

Consider now the following SST set of matchups: games between two

teams seeded  $\ell + 1$  or higher are coin flips, games between  $t_\ell$  and teams seeded between  $t_{\ell+1}$  and  $t_k$  are coin flips, and all other games are always won by the higher seed.

Let's calculate the probability of  $t_i$  and  $t_j$  winning the tournament.  $t_i$  will auto-win any games the have prior to round  $s + 1$ , and then have to win  $r - s$  coin flips in order to win the tournament. This happens with probability  $0.5^{r-s}$ .

$t_j$  will also win all of its games prior to round  $s + 1$ , also has to win a coin flip for each games in round  $s + 2$  or later. For round  $s + 1$ , however, half the time  $t_j$  will be matched up with  $t_k$ , which is a coin flip, but half the time they will be matched with  $t_k$ , which is an auto-win. Thus,  $t_j$  will win the tournament with probability  $0.75(0.5)^{r-s-1}$ .

Therefore,  $t_j$  has a better chance of winning the tournament than  $t_i$  does, despite  $t_i$  being higher seeded, so  $\mathcal{A}$  is not ordered, completing the contradiction. Thus if two games are played in round  $s$ , the winners of those games must play each other in round  $s + 1$ .

This immediately implies that at most two games can be played in each round of an ordered bracket.

Finally, we know that this can be the only game played in round  $s + 1$  because if there was another game, it would have to be between two teams that are higher-seeded than the two teams who won in the previous round, meaning that round violated condition (2).  $\square$

And now the main theorem,

#### Theorem 2.4.6

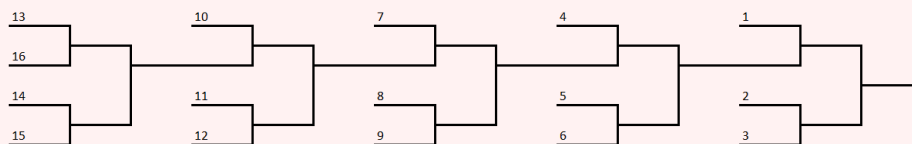
The only ordered brackets are those described by Corollary 2.4.4.

*Proof.* Corollary 2.4.4 describes all proper brackets in which each round either has only game, or has two games but is immediately fol-

lowed by a round with only one game. A proper bracket not described by Corollary 2.4.4 would thus have to include a round with two or more games followed by another round with two or more games, violating Lemma 2.4.5.  $\square$

Theorem 2.4.6 is both exciting and disappointing. On one hand, it means that Corollary 2.4.4 fully describes the set of ordered brackets, making it easy to check whether a given bracket is ordered or not. On the other hand, it means that in an ordered bracket at most three teams can be introduced each round, so the length of the shortest ordered bracket on  $n$  teams grows linearly with  $n$  (rather than logarithmically as is the case for proper brackets). If we want a bracket on many teams to be ordered, we risk forcing lower-seeded teams to play large numbers of games, and we only permit the top seeded teams to play a few. For example, if the shortest ordered bracket that could've been used in the 2022 NCAA Women's Basketball Wichita Region is  $[[4; 0; 3; 0; 3; 0; 3; 0; 3; 0; 0]]$ , which is played over a whopping ten rounds.

**Figure 2.4.7: The Shortest Sixteen-Team Ordered Bracket**



Because of this, few leagues use ordered brackets, and those who do usually have so few teams that every proper bracket is ordered (the 2023 College Football Playoffs, for example). Even the Korean Baseball Organization League, which uses a somewhat unconventional  $[[2; 1; 1; 1; 0]]$ , only sends five teams to the playoffs, and again every five-team proper bracket is ordered. If the KBO League ever expanded to the six-team bracket  $[[2; 1; 1; 1; 1; 0]]$ , we would have a case of an ordered bracket being used instead of a proper non-ordered bracket on the same number of teams.

## 2.5 Multiple Elimination

While the bracket is a very powerful and important tournament design, it has two intimately related shortcomings. First, brackets lead teams to play wildly different numbers of games: in the bracket  $[[8; 0; 0; 0]]$ , for example, two teams will play three game, two teams will play two, and four teams will play only one.

And second, brackets tend to do a very poor job of ranking teams beyond just selecting a winner. Again considering the bracket  $[[8; 0; 0; 0]]$ , the first-place finish is obviously the winner of the bracket, and we can easily grant the loser of the championship game second-place, but the two semifinals losers both might have a claim to the third-place, and sorting through the fifth-through eight-place finishes is even trickier.

These problems are reflections of each other: the reason that ranking the lower-placing teams is so hard is because they play so few games. It's easier to differentiate between first and second because both teams have played three games, but differentiating between the four teams that have played only a single game is nigh impossible.

In some cases, these problems do not cause concern: perhaps we are only interested in crowning a champion and don't care about exactly who came in third, or maybe this bracket is being played at the conclusion of a long season and so teams playing variable numbers of games is not a big deal. But the interconnected nature of the two problems lets us solve them together, leveraging the extra games that lower-ranked teams have left in order to rank them.

We consider class of formats called *simple multibrackets*.

### Definition 2.5.1: Simple Multibracket

A simple multibracket is a sequence of brackets in which the losers of certain games in the upper brackets fall into the lower brackets rather than being eliminated outright, and team place based on which bracket they won.

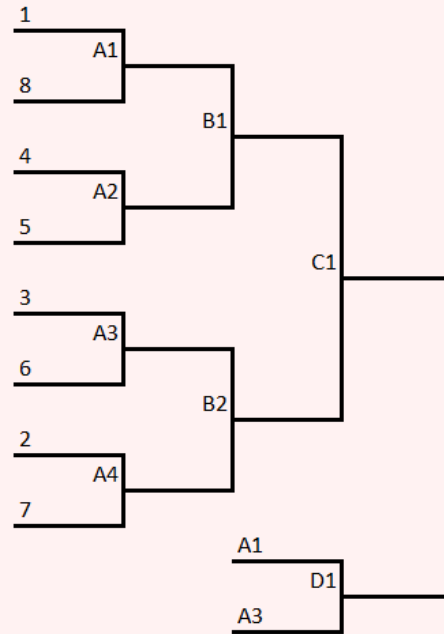
The simplest example of a simple multibracket is the third-place game, in which the losers of the two semifinal games play each other for third place. The 2015 Asian Football Confederation Asian Cup, whose bracket is of signature  $[[8; 0; 0; 0]]$ , employs a third-place game.



Each game in this figure is labeled. In the primary bracket, first-round games are **A1** through **A4**, while the semifinals are **B1** and **B2**, and the finals is game **C1**. The actual labeling scheme is arbitrary, but we adopt the system used in Figure 2.5.2, in which each round is given a letter and then games are numbered from the top to bottom of the bracket within each round. Note that the third-place game is labeled **D1**: even though it could be played concurrently to the championship game, it is part of a different bracket and so we label it as a different round.

We indicate that the third-place game is to be played in between the losers of games **B1** and **B2** by labeling the starting lines in the third-place game with those games. This is not ambiguous because the winner of those games always just continues on in the original bracket, so such labels always refer to the losers. In theory, the Asian Cup could have had any two teams play for third.

Figure 2.5.3: 2015 AFL Asian Cup Alternative Format



In this alternative, the third-place game is played between the loser of two of **A1** and **A3** instead of between the two semifinal losers. This is probably a bad idea: the losers of **B1** and **B1** are clearly more deserving than the losers of **A1** and **A3**, and so they ought to be the teams playing the third-place game.

We can use this intuition to define what it means for a simple multibracket to be proper.

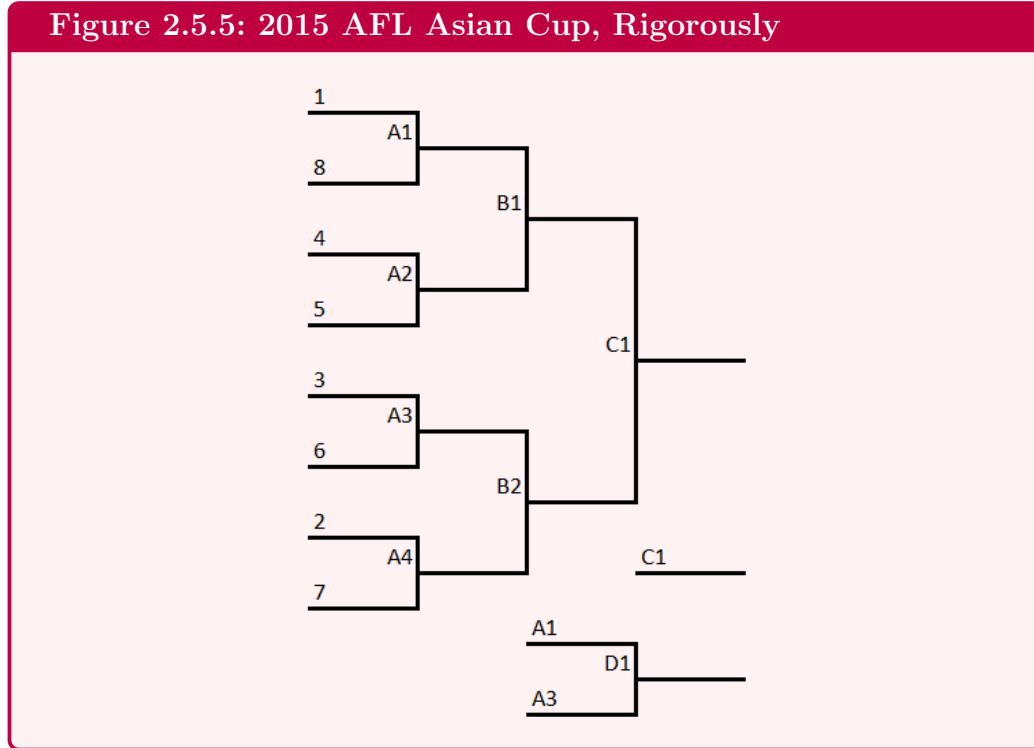
#### Definition 2.5.4: Proper Simple Multibracket

A simple multibracket is *proper* if: NEED BETTER DEFINITION

Note that as a simple multibracket with only two brackets, the Asian Cup is actually not quite proper: remember that because a team's final place is based on which bracket it wins. In order for the Asian Cup to have a top three, it needs to have three brackets in its multibracket. Otherwise, the winner of game **D1** would be given second place, leaving the loser of **C1** out in the cold and violating the properness condition.



In reality there is an implied one-team bracket in between the primary bracket and the third-place game. Rigorously, the Asian Cup multibracket looks like this:



However for clarity, when drawing figures, we will omit brackets in a multibracket of signature  $[[1]]$ . We restrict our study of simple multibrackets to proper ones, so if the loser of the championship of a bracket doesn't appear in the next bracket, then there is an implied bracket of signature  $[[1]]$  between them.

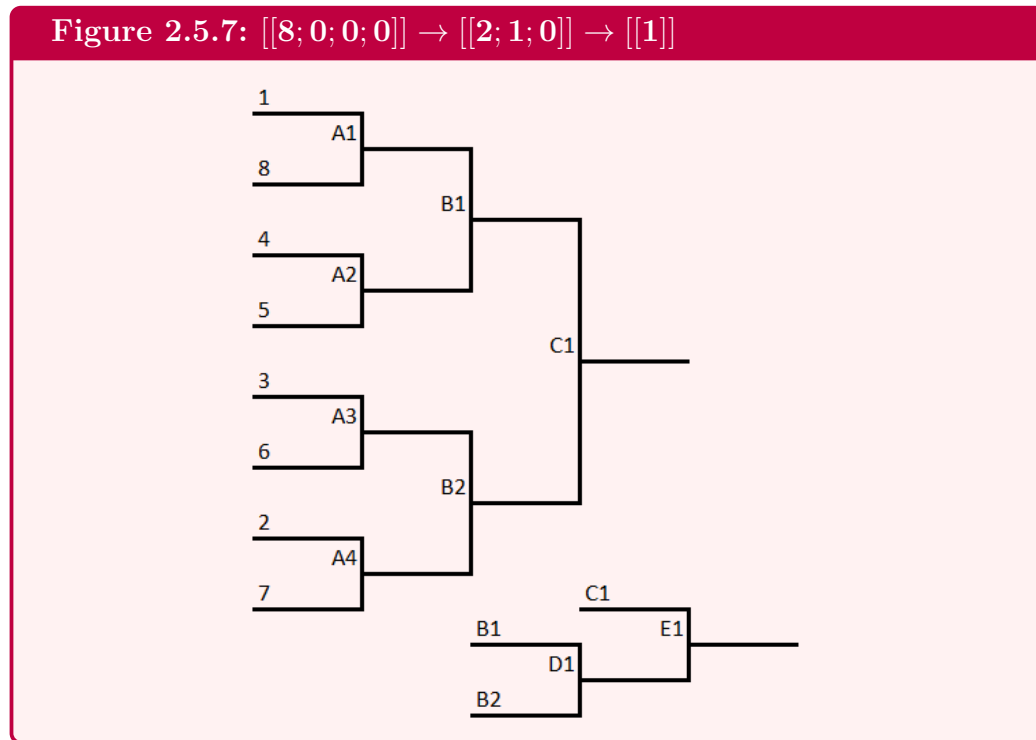
#### Definition 2.5.6: Multibracket Signature

The *signature* of a multibracket is simply the list of signatures of its brackets.

So the 2015 AFL Asian Cup has signature  $[[8; 0; 0; 0]] \rightarrow [[1]] \rightarrow [[2; 0]]$ . But the multibracket with this signature is far from the only multibracket that the AFL could have used to give out gold, silver, and bronze. In fact, it's not clear the loser of **C1**, who comes in second place, is really more deserving

than the winner of **D1**, who comes in third. One could imagine the UAE saying: both South Korea and us finished with two wins and one loss – a first round win, a win against Iraq, and a loss against Australia. The only reason that South Korea came in second and we came in third was because South Korea lucked out by having Australia on the other half of the bracket as them. That’s not fair!

If the AFL took this complaint seriously, they could modify their format to have signature  $[[8; 0; 0; 0]] \rightarrow [[2; 1; 0]] \rightarrow [[1]]$ .



Again, the third bracket of signature  $[[1]]$  is implied and left out of the figure.

If the AFL used the format in Figure 2.5.7 in 2015, then South Korea and the UAE would have played each other for second place after all of the other games were completed. In some sense, this is a more equitable format than the one used in reality: we have the same data about the UAE and South Korea and so we ought to let them play for second place instead of having decided almost randomly.

However, swapping formats doesn’t come without costs. For one thing,

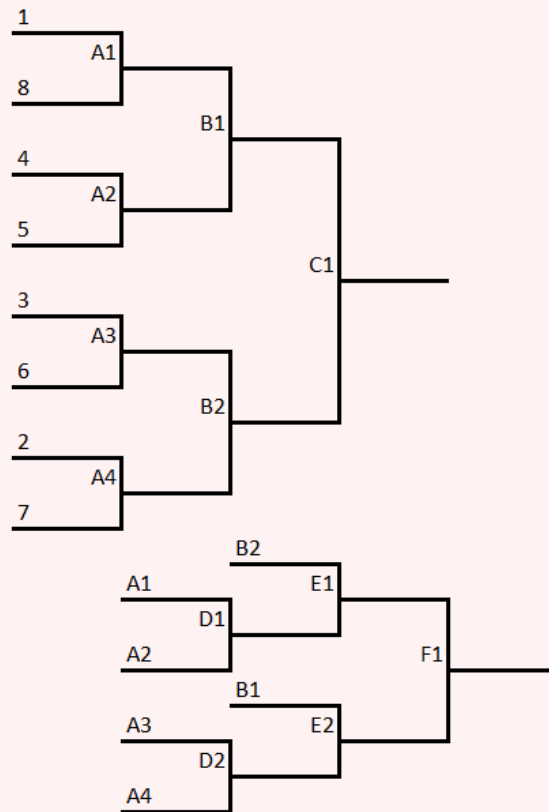
South Korea and the UAE would've had to play a fourth game: if the AFL had only three days to put on the tournament and teams can play at most one game a day, then the format in 2.5.7 isn't feasible.

Another concern: what if Iraq had beaten the UAE when they played in game **D1**? Then the two teams with a claim to second-place would have been South Korea and Iraq, except South Korea already beat Iraq! In this world, South Korea being given second place without having to win a rematch with Iraq seems more equitable than giving Iraq a second chance to win. To address this, one could imagine a format in which game **E1** is played only if it is not rematch, although this would no longer be just a multibracket.

Ultimately, whether including game **E1** is worth it depends on the goal of the format. If there is a huge difference between the prizes for coming in second and third, for example, if the top two finishing teams in the Asian Cup qualified for the World Cup, then **E1** is quite important. If, on the other hand, this is self-contained format played purely for bragging rights, **E1** could probably be left out. In reality, the 2015 AFL Asian Cup qualified only its winner to another tournament (the 2017 Confederations Cup), and gave medals to its top three, and so game **E1** which distinguishes between second and third place, is probably unnecessary.

Let's imagine, however, that instead of just the champion, the top four teams from the Asian Cup advanced to the Confederations Cup. In this case, the format used in 2015 would be quite poor, as teams finish in the top-four based only on the result of their first-round game: the rest of the games don't even have to be played. A better format for selecting the top four teams might look like this:

**Figure 2.5.8:**  $[[8; 0; 0; 0]] \rightarrow [[1]] \rightarrow [[4; 2; 0; 0]] \rightarrow [[1]]$



The multibracket in Figure 2.5.8 has the really attractive property that a team will finish in the top four if and only if it wins two of its first three games. In fact,