Contents

1 Tournament Formats

1.1 Definitions

Definition 1.1.1: Gameplay Function

A *gameplay function* g on a list of teams $\mathcal{T} = [t_1, ..., t_n]$ is a nondeterministic function $g : \mathcal{T} \times \mathcal{T} \to \mathcal{T}$ with the following properties:

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

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 1.1.2: Playing, Winning, and Losing

When g is queried on input (t_i, t_j) we say that t_i and t_j *played a game*. We say that the team that got outputted by g *won*, and the team that did not *lost*.

The information in a gameplay function can be encoded into a *matchup table*.

Definition 1.1.3: Matchup Table

The *matchup table* implied by a gameplay function g on a list of teams $\mathcal{T}$ of length n is a n-by-n matrix $\mathbf{M}$ such that $\mathbf{M}_{ij} = \mathbb{P}[t_i \text{ beats } t_j]$.

For example, let $\mathcal{T}$ =[Favorites, Rock, Paper, Scissors, Conceders], and g be such that the Conceders concede every game they play, the Favorites are 70% favorites against Rock, Paper, and Scissors, and Rock, Paper, and Scissors match up with each other as their name implies. Then the matchup table would look like so:

Figure 1.1.4: The Matchup Table for $(\mathcal{T}, g)$

	Favorites	Rock	Paper	Scissors	Conceders
Favorites	0.5	0.7	0.7	0.7	1.0
Rock	0.3	0.5	0.0	1.0	1.0
Paper	0.3	1.0	0.5	0.0	1.0
Scissors	0.3	0.0	1.0	0.5	1.0
Conceders	0.0	0.0	0.0	0.0	0.5

Theorem 1.1.5

If $\mathbf{M}$ is the matchup table for $(\mathcal{T}, g)$, then $\mathbf{M} + \mathbf{M}^T$ is the matrix of all ones.

Proof. $(\mathbf{M} + \mathbf{M}^T)_{ij} = \mathbf{M}_{ij} + \mathbf{M}_{ji} = \mathbb{P}[t_i \text{ beats } t_j] + \mathbb{P}[t_j \text{ beats } t_i] = 1.$ $\square$

Definition 1.1.6: 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 $t \in \mathcal{T}$.

We use a gameplay function rather than a matchup table in the definition of a tournament format because a tournament format cannot simply look at the matchup table itself in order to decide which teams are best. Instead, formats query the gameplay function (have teams play games) in order to gather information about the teams. That said, matchup tables will often be useful in our *analysis* of tournament formats.

We also introduce some shorthand to help make notation more concise.

Definition 1.1.7: $\mathbb{W}_{\mathcal{A}}(t, \mathcal{T})$

$\mathbb{W}_{\mathcal{A}}(t, \mathcal{T})$ is the probability that team $t \in \mathcal{T}$ wins tournament format $\mathcal{A}$ when it is run on the list of teams $\mathcal{T}$.

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

2 Brackets

2.1 Brackets and Rounds

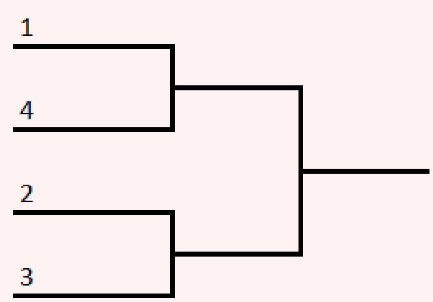
Definition 2.1.1: Bracket

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 are determined based on the ordering of the teams in $\mathcal{T}$ in advance of the outcomes of any games.

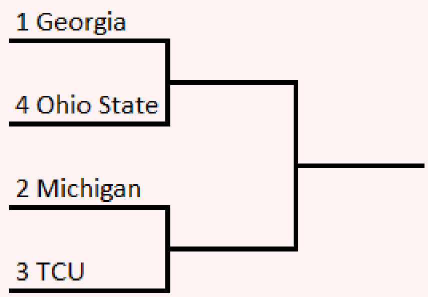
We can draw a bracket as a tree-like structure in the following way:

Figure 2.1.2: The 2023 College Football Playoff



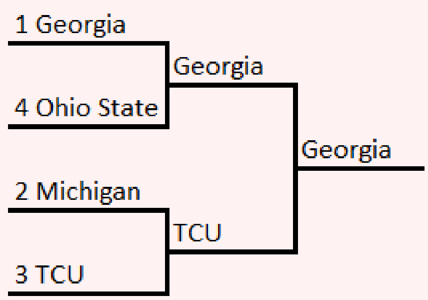
The numbers 1, 2, 3, and 4 indicate where t_1, t_2, t_3, and t_4 in $\mathcal{T}$ are placed to start. In the actual 2023 College Football Playoff, the list of teams $\mathcal{T}$ was Georgia, Michigan, TCU, and Ohio State, in that order, so the bracket was filled in like so:

Figure 2.1.3: The 2023 CFP After Team Placement



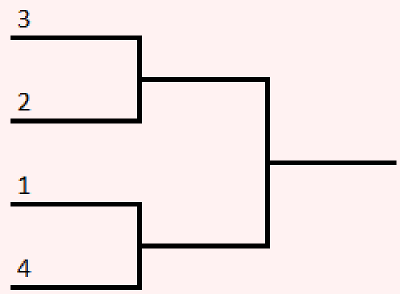
As games are played, we write the name of the winning teams on the corresponding lines. This bracket tells us that Georgia played Ohio State, and Michigan played TCU. Georgia and TCU won their respective games, and then Georgia beat TCU, winning the tournament.

Figure 2.1.4: 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 2.1.5 is just another way to draw the same 2023 CFP Bracket.

Figure 2.1.5: Alternative Drawing of the 2023 CFP



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

Definition 2.1.6: 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.

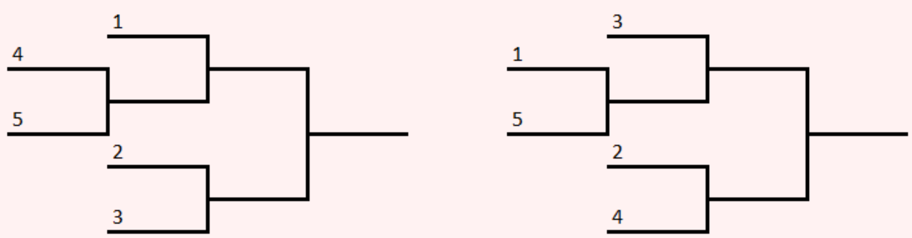
2.2 Bracket Signatures

Definition 2.2.1: Shape

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

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

Figure 2.2.2: Two Brackets with the Same Shape



Definition 2.2.3: 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 its signature.

Definition 2.2.4: Bracket Signature

The *signature* $[[\mathbf{a_0}; \ldots; \mathbf{a_r}]]$ of an r-round bracket $\mathcal{A}$ is list 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 2.2.2 have the same shape, so they also have the same signature.

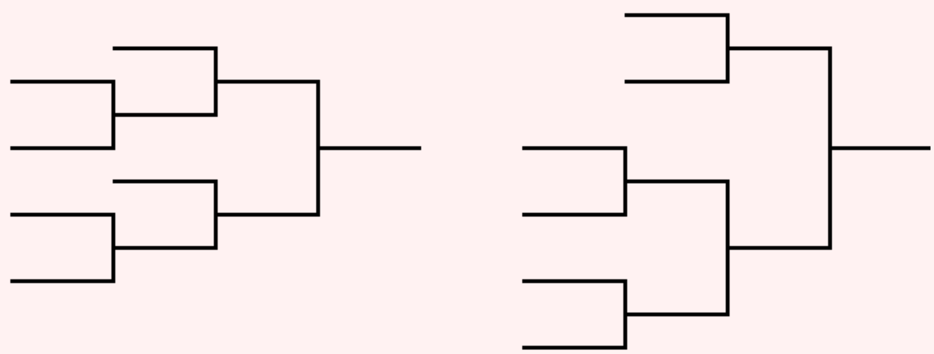
The signatures of the brackets discussed in this section are shown in Figure 2.2.5. It's worth verifying the signatures we've seen so far and trying to draw brackets with the signatures we haven't yet before moving on.

Figure 2.2.5: The Signatures of Some Brackets

Bracket	Signature
2023 College Football Playoff	$[[\mathbf{4};\mathbf{0};\mathbf{0}]]$
The brackets in Figure 2.2.2	$[[\mathbf{2};\mathbf{3};\mathbf{0};\mathbf{0}]]$
The brackets in Figure 2.2.6	$[[\mathbf{4};\mathbf{2};\mathbf{0};\mathbf{0}]]$
2023 WCC Men's Basketball Tournament	$[[\mathbf{4};\mathbf{2};\mathbf{2};\mathbf{2};\mathbf{0};\mathbf{0}]]$

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 2.2.6 have the signature $[[\mathbf{4};\mathbf{2};\mathbf{0};\mathbf{0}]]$.

Figure 2.2.6: 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 2.2.7

Let $\mathcal{A} = [[\mathbf{a_0};...;\mathbf{a_r}]]$ 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 has i byes 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 $i \in \{0, ..., r\}$, there are a_i teams with i byes, 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 $[[\mathbf{1}]]$, 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 $\mathcal{B} = [[\mathbf{a_1} + \mathbf{a_0}/\mathbf{2}; \mathbf{a_2}; ...; \mathbf{a_r}]]$. 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 $\mathcal{A}$. $\square$

The operation of transforming a bracket signature $\mathcal{A} = [[\mathbf{a_0}; ...; \mathbf{a_r}]]$ into a bracket signature with one fewer round $\mathcal{B} = [[\mathbf{a_1} + \mathbf{a_0}/\mathbf{2}; \mathbf{a_2}; ...; \mathbf{a_r}]]$ that we used at the end of the proof of Theorem 2.2.7 will become somewhat frequent, as we often induct on the number of rounds in a bracket, so it has a name:

Definition 2.2.8: The Successor Signature

If $\mathcal{A} = [[\mathbf{a_0}; ...; \mathbf{a_r}]]$, then the successor signature

$$\mathfrak{S}(\mathcal{A}) = [[\mathbf{a_1} + \mathbf{a_0/2}; \mathbf{a_2}; ...; \mathbf{a_r}]].$$

(The successor signature of zero-round signatures is undefined.)

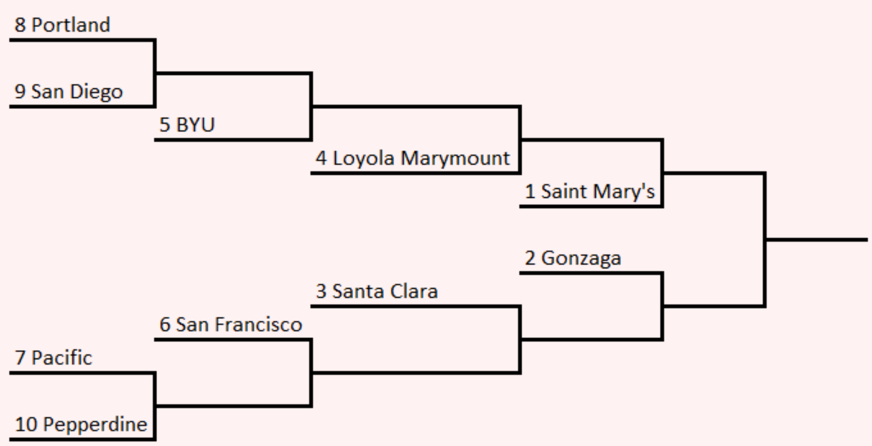
In the next few sections, we will use the language and properties of bracket signatures to describe the brackets that we work with. For now though, let's return to the 2023 College Football Playoff. The bracket used in the 2023 CFP has a special property that not all brackets have: it is *balanced*.

Definition 2.2.9: 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 2.2.10: 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 2.2.11

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 $\mathcal{A} = [[\mathbf{n}; \mathbf{0}; ...; \mathbf{0}]]$ where n is the number of teams in the bracket. If n is a power of two, then by Theorem 2.2.7 $\mathcal{A}$ 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 2.2.7 tells us that $\mathcal{A}$ is not a bracket signature, and so no balanced brackets exist for n teams. □

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.

2.3 Proper Brackets

Definition 2.3.1: 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 2.3.2: i-seed

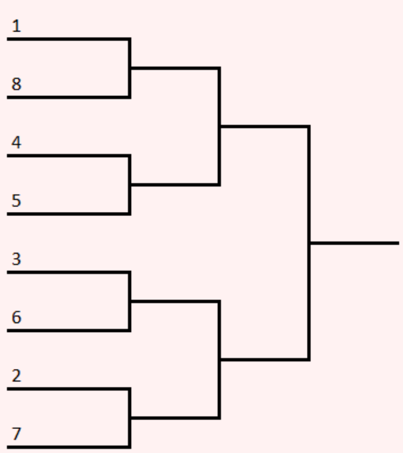
In a list of teams $\mathcal{T} = [t_1, ..., t_n]$, we refer to t_i as the i-seed.

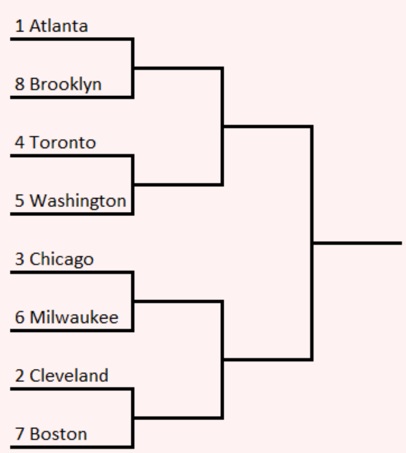
Definition 2.3.3: Higher and Lower Seeds

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

Seeding is typically used to reward better and more deserving teams. As an example, on the left is the eight-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 2.3.4: 2015 NBA Eastern Conference Playoffs

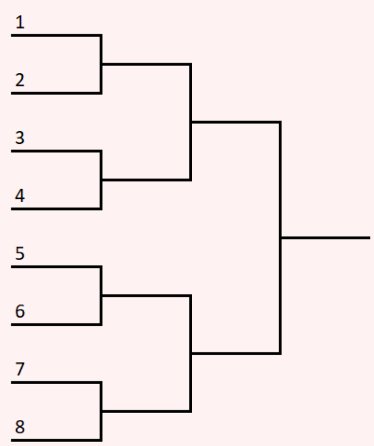




Despite this bracket being balanced, the higher seeds are still at advantage: they have an easier set of opponents. Compare 1-seed Atlanta, whose first two rounds are versus 8-seed Brooklyn and then (most likely) 4-seed Toronto, versus 7-seed Boston, whose first two rounds are versus 2-seed Cleveland 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 2.3.5: 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 2.3.6: Chalk

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

Definition 2.3.7: 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 2.3.8: 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 2.3.9

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). □

Lemma 2.3.10

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_ℓ. They beat t_ℓ, 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_ℓ, violating condition (2). Thus, the bracket is not proper. □

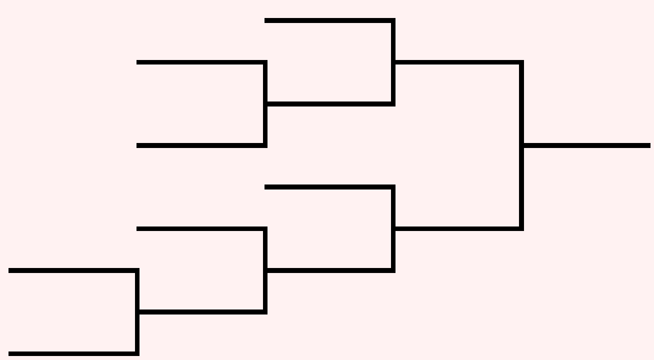
Lemma 2.3.11

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 $i \in \{1, ..., k\}$.

Proof. In the given round, there are $m+2k$ teams remaining. Theorem 2.3.10 tells us that (if the bracket goes chalk) those teams must be seeds 1 through $m+2k$. Theorem 2.3.9 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 $i \in \{1, ..., k\}$. □

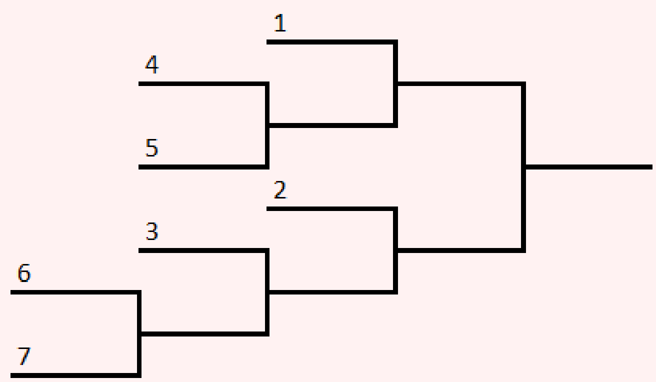
We can use Lemmas 2.3.9 through 2.3.11 to properly seed various bracket shapes. For example, consider the following seven-team shape:

Figure 2.3.12: A Seven-Team Bracket Shape



Lemma 2.3.9 tells us that the first-round matchup must be between the 6-seed and the 7-seed. Lemma 2.3.11 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 2.3.11 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:

Figure 2.3.13: A Seven-Team Bracket, Properly Seeded

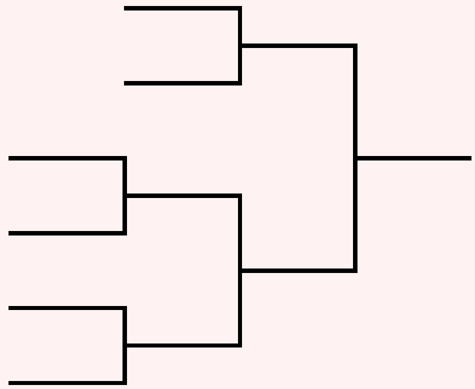


We can also quickly simulate the bracket going chalk to verify Lemma 2.3.10.

Lemmas 2.3.9 through 2.3.11 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 2.3.14: A Six-Team Bracket Shape



This shape admits no proper seedings. Lemma 2.3.9 requires that the two teams getting byes be the 1- and 2-seed, but this violates Lemma 2.3.11 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 2.3.15: The Fundamental Theorem of Brackets

There is exactly one proper bracket with each bracket signature.

Proof. Let $\mathcal{A}$ be an r-round bracket signature. We proceed by induction on r. If $r = 0$, then the only possible bracket signature is $[[\mathbf{1}]]$, 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 Lemma 2.3.11. Then if those matchups go chalk, we are left with a proper bracket of signature $\mathfrak{S}(\mathcal{A})$, which induction tells us exists admits exactly one proper bracket.

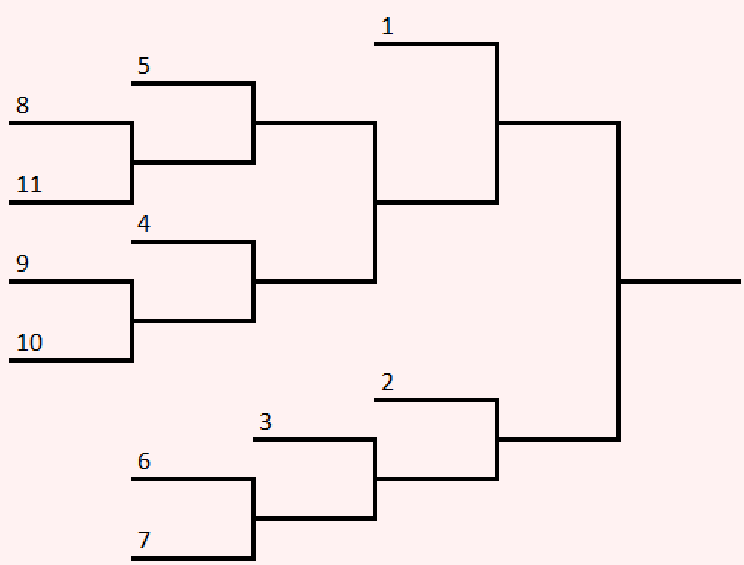
Thus both the first-round matchups and the rest of the bracket are determined, and by combining them we get a proper bracket with signature $\mathcal{A}$, so there is exactly one proper bracket with signature $\mathcal{A}$. $\square$

The fundamental theorem of brackets means that we can refer to the proper bracket $\mathcal{A} = [[\mathbf{a_0}; ...; \mathbf{a_r}]]$ 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 Playoffs), 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 $[[\mathbf{4}; \mathbf{4}; \mathbf{1}; \mathbf{2}; \mathbf{0}; \mathbf{0}]]$ and use Lemmas 2.3.9 through 2.3.11 to draw it with ease:

Figure 2.3.16: The Proper Bracket of Signature $[[4; 4; 1; 2; 0; 0]]$

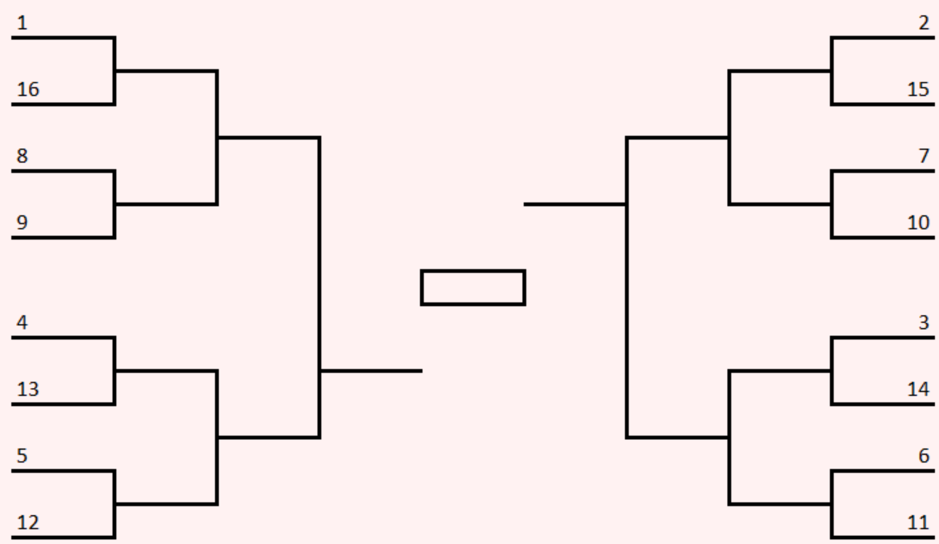


Due to these properties, we will almost exclusively discuss proper bracket from here on out: unless stated otherwise, assume all brackets are proper.

2.4 Ordered Brackets

Consider the proper bracket $[[\mathbf{16}; \mathbf{0}; \mathbf{0}; \mathbf{0}; \mathbf{0}]]$, which was used in the 2021 NCAA Men's Basketball Tournament South 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.)

Figure 2.4.1: 2021 NCAA Men's Basketball Tournament South Region



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 2.4.2: 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 [10] 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 $[[\mathbf{16}; \mathbf{0}; \mathbf{0}; \mathbf{0}; \mathbf{0}]]$ 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, initially proposed by H.A. David [4], might look like this:

Definition 2.4.3: Strongly Stochastically Transitive

A list of teams $\mathcal{T}$ 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 (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 2.4.4

If $\mathcal{T}$ is SST, then for each $i < j$, $\mathbb{P}[t_i \text{ beats } t_j] \geq 0.5$.

Corollary 2.4.5

If $\mathcal{T}$ is SST, then for each i, j, k, ℓ such that $i < j$ and $k < \ell$,

$$\mathbb{P}[t_i \text{ beats } t_\ell] \geq \mathbb{P}[t_j \text{ beats } t_k].$$

Corollary 2.4.6

If $\mathcal{T}$ is SST, then the matchup table $\mathbf{M}$ is monotonically increasing along each row and monotonically decreasing along each column.

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*, first defined by Chen and Hwang [3] (though we use the name proposed by Edwards [5]).

Definition 2.4.7: Ordered

A tournament format $\mathcal{A}$ is *ordered* if, for any SST list of teams $\mathcal{T}$, if $i < j$, then $\mathbb{W}_{\mathcal{A}}(t_i, \mathcal{T}) \geq \mathbb{W}_{\mathcal{A}}(t_j, \mathcal{T})$

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 2.4.8

Every ordered bracket is proper.

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

Let $\mathcal{T}$ be SST with matchup table $\mathbf{M}$ where $\mathbf{M}_{ij} = 0.5$. A team that plays their first game in the ith round will win the tournament with probability $(0.5)^{r-i}$, 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 2.3.11: 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 $i \in \{1, ..., k\}$. 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 the bracket goes chalk, t_{m+j} will face off against seed t_ℓ in the given round. Consider the following SST matchup table: every game is a coin flip, except for games involving a team seeded ℓ 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_ℓ 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. $\square$

With Theorem 2.4.8, 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 2.4.9: The One-Team Bracket [[1]]

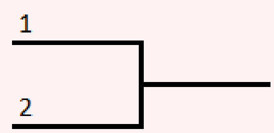
Theorem 2.4.10

The one-team bracket $[[\mathbf{1}]]$ is ordered.

Proof. Since there is only team, the ordered bracket condition is vacuously true. □

Next we look at the unique two-team bracket.

Figure 2.4.11: The Two-Team Bracket $[[2;0]]$



Theorem 2.4.12

The two-team bracket $[[\mathbf{2};\mathbf{0}]]$ is ordered.

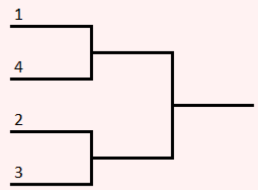
Proof. Let $\mathcal{A} = [[\mathbf{2};\mathbf{0}]]$. Then,

$$\mathbb{W}_{\mathcal{A}}(t_1, \mathcal{T}) = \mathbb{P}[t_1 \text{ beats } t_2] \geq 0.5 \geq \mathbb{P}[t_2 \text{ beats } t_1] = \mathbb{W}_{\mathcal{A}}(t_2, \mathcal{T})$$

so $\mathcal{A}$ is ordered. □

And thirdly, we show that the balanced four-team bracket is ordered, first proved by Horen and Riezman [6].

Figure 2.4.13: The Four-Team Bracket $[[4; 0; 0]]$



Theorem 2.4.14

The four-team bracket $[[\mathbf{4}; \mathbf{0}; \mathbf{0}]]$ is ordered.

Proof. Let $\mathcal{A} = [[\mathbf{4}; \mathbf{0}; \mathbf{0}]]$ and let $p_{ij} = \mathbb{P}[t_i \text{ beats } t_j]$. Then,

$$\begin{aligned}
\mathbb{W}_{\mathcal{A}}(t_1, \mathcal{T}) &= p_{14} \cdot (p_{23}p_{12} + p_{32}p_{13}) \\
&= p_{14}p_{23}p_{12} + p_{14}p_{32}p_{13} \\
&\geq p_{14}p_{23}p_{21} + p_{24}p_{41}p_{23} \\
&= p_{23} \cdot (p_{14}p_{21} + p_{41}p_{24}) \\
&= \mathbb{W}_{\mathcal{A}}(t_2, \mathcal{T})
\end{aligned}$$

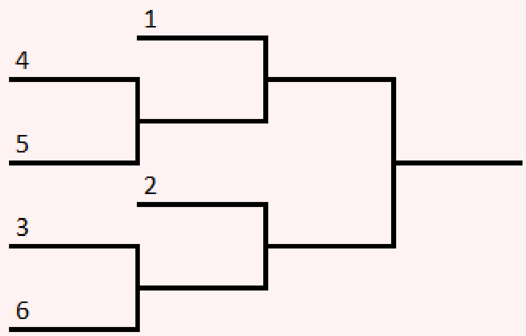
$$\begin{aligned}
\mathbb{W}_{\mathcal{A}}(t_2, \mathcal{T}) &= p_{23} \cdot (p_{14}p_{21} + p_{41}p_{24}) \\
&\geq p_{32} \cdot (p_{14}p_{31} + p_{41}p_{34}) \\
&= \mathbb{W}_{\mathcal{A}}(t_3, \mathcal{T})
\end{aligned}$$

$$\begin{aligned}
\mathbb{W}_{\mathcal{A}}(t_3, \mathcal{T}) &= p_{32} \cdot (p_{14}p_{31} + p_{41}p_{34}) \\
&= p_{32}p_{14}p_{31} + p_{32}p_{41}p_{34} \\
&\geq p_{42}p_{23}p_{41} + p_{32}p_{41}p_{43} \\
&= p_{41} \cdot (p_{23}p_{42} + p_{32}p_{43}) \\
&= \mathbb{W}_{\mathcal{A}}(t_4, \mathcal{T})
\end{aligned}$$

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

However, not every proper bracket is ordered. One particularly important example of a non-ordered proper bracket is $[[\mathbf{4}; \mathbf{2}; \mathbf{0}; \mathbf{0}]]$

Figure 2.4.15: The Six-Team Bracket $[[\mathbf{4};\mathbf{2};\mathbf{0};\mathbf{0}]]$



Theorem 2.4.16

The six-team bracket $[[\mathbf{4};\mathbf{2};\mathbf{0};\mathbf{0}]]$ is not ordered.

Proof. Let $\mathcal{A} = [[\mathbf{4};\mathbf{2};\mathbf{0};\mathbf{0}]]$, and let $\mathcal{T}$ have the following matchup table:

	t_1	t_2	t_3	t_4	t_5	t_6
t_1	0.5	0.5	0.5	0.5	0.5	1
t_2	0.5	0.5	0.5	0.5	0.5	1
t_3	0.5	0.5	0.5	0.5	0.5	0.5
t_4	0.5	0.5	0.5	0.5	0.5	0.5
t_5	0.5	0.5	0.5	0.5	0.5	0.5
t_6	0	0	0.5	0.5	0.5	0.5

Then

$$\mathbb{W}_{\mathcal{A}}(t_1, \mathcal{T}) = 0.5 \cdot 0.5 = 0.25,$$

but

$$\mathbb{W}_{\mathcal{A}}(t_2, \mathcal{T}) = (0.5 \cdot 0.5 + 0.5 \cdot 1) \cdot 0.5 = 0.375.$$

Thus $\mathcal{A}$ is not ordered. $\square$

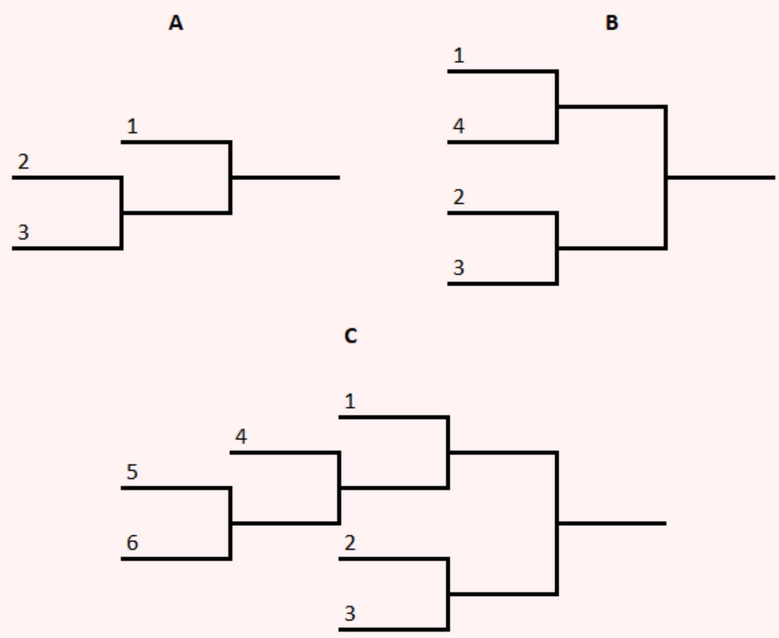
In the next section, we move on from describing particular ordered and non-ordered brackets in favor of a more general result.

2.5 Edwards's Theorem

We now attempt to completely classify the set of ordered brackets. Edwards [5] originally accomplished this without access to the machinery of bracket signatures or proper brackets: we present a quicker proof that makes use of the fundamental theorem of brackets and develop two nice lemmas along the way.

We begin with the stapling lemma, which allows us to combine two smaller ordered brackets 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 2.5.1.

Figure 2.5.1: Setup of the Stapling Lemma with $\mathcal{A} = [[\mathbf{2}; \mathbf{1}; \mathbf{0}]]$, $\mathcal{B} = [[\mathbf{4}; \mathbf{0}; \mathbf{0}]]$, and $\mathcal{C} = [[\mathbf{2}; \mathbf{1}; \mathbf{3}; \mathbf{0}; \mathbf{0}]]$



Lemma 2.5.2: The Stapling Lemma

If $\mathcal{A} = [[\mathbf{a_0}; ...; \mathbf{a_r}]]$ and $\mathcal{B} = [[\mathbf{b_0}; ...; \mathbf{b_s}]]$ are ordered brackets, then $\mathcal{C} = [[\mathbf{a_0}; ...; \mathbf{a_r} + \mathbf{b_0} - \mathbf{1}; ...; \mathbf{b_s}]]$ is an ordered bracket as well.

Proof. Let $\mathcal{A}, \mathcal{B}$, and $\mathcal{C}$ be as specified. Let $\mathcal{T}$ be an SST list of teams $n+m-1$ teams, and let $\mathcal{R}, \mathcal{S} \subset \mathcal{T}$ be the lowest n and the highest $m-1$ seeds of $\mathcal{T}$ respectively. We divide proving that $\mathcal{C}$ is ordered into proving three sub-statements:

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

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

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

$$\begin{aligned}
\mathbb{W}_{\mathcal{C}}(t_i, \mathcal{T}) &= \sum_{k=m}^{n+m-1} \mathbb{W}_{\mathcal{A}}(t_k, \mathcal{R}) \cdot \mathbb{W}_{\mathcal{B}}(t_i, \mathcal{S} \cup \{t_k\}) \\
&\geq \sum_{k=m}^{n+m-1} \mathbb{W}_{\mathcal{A}}(t_k, \mathcal{R}) \cdot \mathbb{W}_{\mathcal{B}}(t_j, \mathcal{S} \cup \{t_k\}) \\
&= \mathbb{W}_{\mathcal{C}}(t_j, \mathcal{T})
\end{aligned}$$

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.

$$\begin{aligned}
\mathbb{W}_{\mathcal{C}}(t_{m-1}, \mathcal{T}) &= \sum_{k=m}^{n+m-1} \mathbb{W}_{\mathcal{A}}(t_k, \mathcal{R}) \cdot \mathbb{W}_{\mathcal{B}}(t_{m-1}, \mathcal{S} \cup \{t_k\}) \\
&\geq \mathbb{W}_{\mathcal{A}}(t_m, \mathcal{R}) \cdot \mathbb{W}_{\mathcal{B}}(t_{m-1}, \mathcal{S} \cup \{t_m\}) \\
&\geq \mathbb{W}_{\mathcal{A}}(t_m, \mathcal{R}) \cdot \mathbb{W}_{\mathcal{B}}(t_m, \mathcal{S} \cup \{t_m\}) \\
&= \mathbb{W}_{\mathcal{C}}(t_m, \mathcal{T})
\end{aligned}$$

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

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

$$\begin{aligned}
\mathbb{W}_{\mathcal{C}}(t_i, \mathcal{T}) &= \mathbb{W}_{\mathcal{A}}(t_i, \mathcal{R}) \cdot \mathbb{W}_{\mathcal{B}}(t_i, \mathcal{S} \cup \{t_i\}) \\
&\geq \mathbb{W}_{\mathcal{A}}(t_j, \mathcal{R}) \cdot \mathbb{W}_{\mathcal{B}}(t_i, \mathcal{S} \cup \{t_i\}) \\
&\geq \mathbb{W}_{\mathcal{A}}(t_j, \mathcal{R}) \cdot \mathbb{W}_{\mathcal{B}}(t_j, \mathcal{S} \cup \{t_j\}) \\
&= \mathbb{W}_{\mathcal{C}}(t_j, \mathcal{T})
\end{aligned}$$

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. □

Now, if we begin with the set of brackets $\{[[\mathbf{1}]], [[\mathbf{2}; \mathbf{0}]], [[\mathbf{4}; \mathbf{0}; \mathbf{0}]]\}$ and then repeatedly apply the stapling lemma, we can construct a set of brackets that we know are ordered. In other words,

Corollary 2.5.3

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 2.5.3 uses the tools that we have developed so far to identify a set of ordered brackets. Somewhat surprisingly, this set is complete: any bracket not reachable using the process in Corollary 2.5.3 is not ordered. To prove this we first need to show the containment lemma.

Definition 2.5.4: Containment

Let $\mathcal{A}$ and $\mathcal{B}$ be bracket signatures. We say $\mathcal{A}$ *contains* $\mathcal{B}$ if there exists some i such that

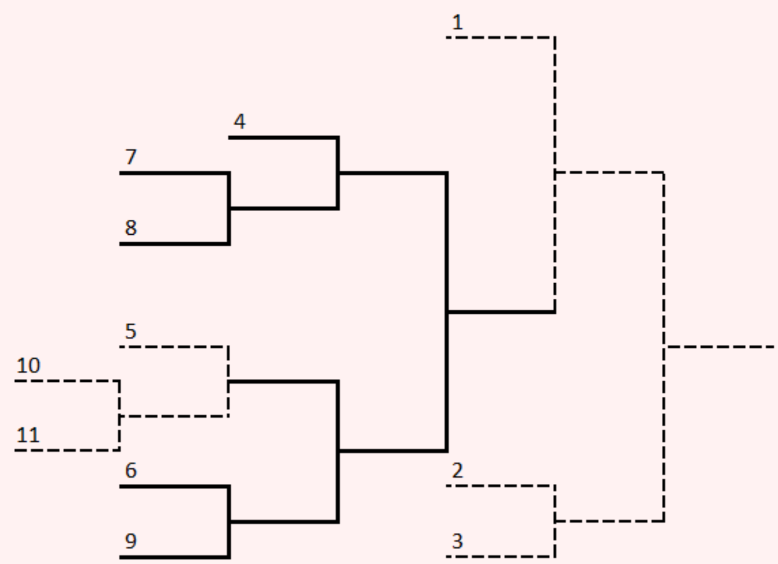
- At least as many games are played in the $(i+1)$th round of $\mathcal{A}$

as in the first round of $\mathcal{B}$, and

- For $j > 1$, there are exactly as many games played in the $(i+j)$th round of $\mathcal{A}$ as in the jth round of $\mathcal{B}$.

Intuitively, $\mathcal{A}$ containing $\mathcal{B}$ means that if $\mathcal{A}$ went chalk, and games within each round were played in order of largest seed-gap to smallest seed-gap, then at some point, there would be a bracket of shape $\mathcal{B}$ used to determine to identify the last team in the rest of bracket $\mathcal{A}$. Figure 2.5.5 shows $\mathcal{A} = [[\mathbf{2}; \mathbf{5}; \mathbf{1}; \mathbf{0}; \mathbf{3}; \mathbf{0}; \mathbf{0}]]$ containing $\mathcal{B} = [[\mathbf{4}; \mathbf{2}; \mathbf{0}; \mathbf{0}]]$. After the 10v11 game and the 5v(10v11) game, there is a bracket of shape $\mathcal{B}$ (the solid lines) that must be played to determine the last team in the rest of the bracket

Figure 2.5.5: Setup of the Containment Lemma with $\mathcal{A} = [[2; 5; 1; 0; 3; 0; 0]]$ and $\mathcal{B} = [[4; 2; 0; 0]]$.



Lemma 2.5.6: The Containment Lemma

If $\mathcal{A}$ contains $\mathcal{B}$, and $\mathcal{B}$ is not ordered, then neither is $\mathcal{A}$.

Proof. Let $\mathcal{A}$ be a bracket signature with r rounds and n teams, and let $\mathcal{B}$ have s round and m teams, such that $\mathcal{A}$ contains $\mathcal{B}$ and $\mathcal{B}$ is not ordered. Let k be the number of teams in $\mathcal{A}$ that get at least $s+i$ byes (where i is from the definition of contains).

$\mathcal{B}$ is not ordered, so let $\mathbf{M}$ be a matchup table that violates the orderedness condition, where none of the win probabilities are 0. (If we have an $\mathbf{M}$ that includes 0s, we can replace them with ϵ. For small enough ϵ, $\mathbf{M}$ will still violate the condition.) Let p be the minimum probability in $\mathbf{M}$. Let $\mathbf{P}$ be a matchup table in which the lower-seeded team wins with probability p, and let $\mathbf{Z}$ be a matchup table in which the lower-seeded team wins with probability 0.

Now, consider the following block matchup table on $\mathcal{T}$, a list of n teams:

	t_1 - t_k	t_{k+1} - t_{k+m}	t_{k+m+1} - t_n
t_1 - t_k	**P**	**P**	**Z**
t_{k+1} - t_{k+m}	**P**	**M**	**Z**
t_{k+m+1} - t_n	**Z**	**Z**	**Z**

Let $\mathcal{S} \subset \mathcal{T}$ be the sublist of teams seeded between $k+1$ and $k+m$. Then, for $t_j \in \mathcal{S}$,

$$\mathbb{W}_{\mathcal{A}}(t, \mathcal{T}) = \mathbb{W}_{\mathcal{B}}(t, \mathcal{S}) \cdot p^{r-s-i},$$

since t_j wins any games it might have to play in rounds i or before automatically, any games after $s+i$ with probability p, and any games in between according to $\mathbf{M}$.

However, $\mathbf{M}$ (and thus $\mathcal{S}$) violates the orderedness condition for $\mathcal{B}$, and so $\mathcal{T}$ does for $\mathcal{A}$. $\square$

With the containment lemma shown, we can proceed to the main theorem:

Theorem 2.5.7: Edwards's Theorem

The only ordered brackets are those described by Corollary 2.5.3.

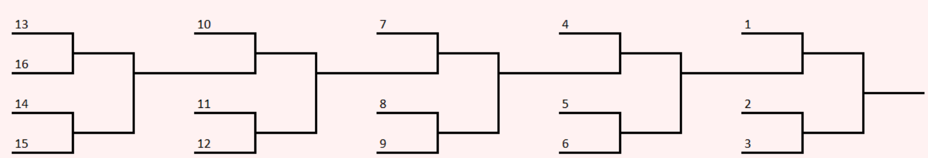
Proof. Let $\mathcal{A}$ be a proper bracket not described by Corollary 2.5.3. The corollary describes all proper brackets in which each round either has only game, or has two games but is immediately followed by a round with only one game. Thus $\mathcal{A}$ must include at least two successive rounds with two or more games each.

The final round in such a chain will be followed by a round with a single game, and so the final round must have only two games. Thus, $\mathcal{A}$ includes a sequence of three rounds, the first of which has at least two games, the second of which has exactly two games, and the third of which has one game.

Therefore, $\mathcal{A}$ contains $[[\mathbf{4}; \mathbf{2}; \mathbf{0}; \mathbf{0}]]$. But we know that $[[\mathbf{4}; \mathbf{2}; \mathbf{0}; \mathbf{0}]]$ is not ordered, and so by the containment lemma, neither is $\mathcal{A}$. $\square$

Edwards's Theorem is both exciting and disappointing. On one hand, it means that we can fully describe 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 the shortest proper bracket). If we want a bracket on many teams to be ordered, we risk forcing lower-seeded teams to play a large number of games, and we only permit the top-seeded teams to play a few. For example, the shortest ordered bracket that could've been used in the 2021 NCAA Basketball South Region is $[[\mathbf{4}; \mathbf{0}; \mathbf{3}; \mathbf{0}; \mathbf{3}; \mathbf{0}; \mathbf{3}; \mathbf{0}; \mathbf{3}; \mathbf{0}; \mathbf{0}]]$, which is played over a whopping ten rounds.

Figure 2.5.8: 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 $[[\mathbf{2}; \mathbf{1}; \mathbf{1}; \mathbf{1}; \mathbf{0}]]$, 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 $[[\mathbf{2}; \mathbf{1}; \mathbf{1}; \mathbf{1}; \mathbf{1}; \mathbf{0}]]$, we would have a case of an ordered bracket being used when a proper non-ordered bracket exists on the same number of teams.

2.6 Reseeding

Edwards's Theorem naturally raises the question: is there some bracket-like tournament format, one where undefeated teams face off until only one reamins, that expands the space of signatures that are ordered? *Reseeded* brackets are our first attempt at an answer.

Ultimately, the reason that proper brackets are not, in general, ordered, is that lower-seeded teams are treated, if they win, as the team that they beat for the rest of the format. Consider again the proper bracket analyzed by Silver: $[[\mathbf{16}; \mathbf{0}; \mathbf{0}; \mathbf{0}; \mathbf{0}]]$. If an 11-seed wins in the first round, they take on the schedule of a 6-seed for the rest of the tournament, while if the 9-seed wins, they take on the schedule of an 8-seed. Given that a 6-seed has an easier schedule than an 8-seed, it's not hard to see why it might be preferable to be an 11-seed rather than a 9-seed.

Reseeding (poorly named) fixes this by resorting the match-ups every round: if an 11-seed keeps winning, they will have to play teams according to their seed, rather than getting an effective upgrade to 6-seed status.

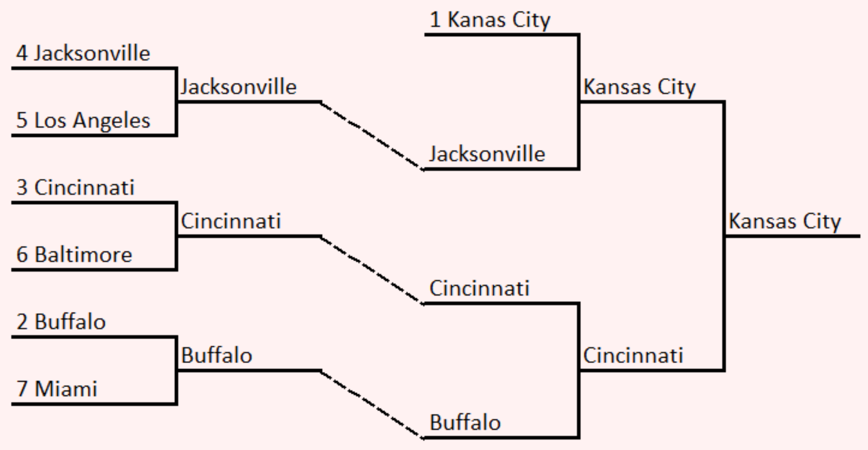
Definition 2.6.1: Reseeded Brackets

In a *reseeded* bracket, after each round, match up the highest-seeded team with the lowest-seeded team, second-highest vs second-lowest, etc.

Note that by Definition 2.1.1, a reseeded bracket is not a bracket at all, as matchups between teams that have not yet lost are not determined in advance of the outcomes of any games. However, because reseeded brackets act so similarly to traditional brackets, and because colloquially they are referred to as brackets, we opt to continue using the word "bracket" to describe them.

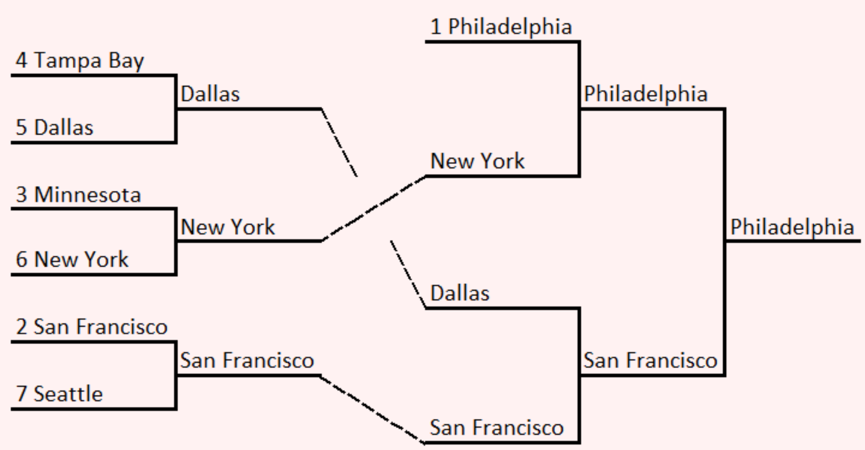
In 2023, both National Football League conferences use a reseeded bracket with signature $[[\mathbf{6}; \mathbf{1}; \mathbf{0}; \mathbf{0}]]^R$. (The superscript R indicates this is reseeded bracket.) If the first round of the bracket goes chalk, then it looks just like a normal bracket:

Figure 2.6.2: 2023 National Football League AFC Playoffs



The dotted lines are drawn after the first round of games has been played: if there are some first-round upsets, then the bracket is rearranged to ensure that it is still better to be a higher seed rather than a lower seed.

Figure 2.6.3: 2023 National Football League NFC Playoffs



In the NFC, 6-seed New York upset 3-seed Minnesota. Had a conventional bracket been used, the semifinal matchups would have been 1-seed vs 5-seed and 2-seed vs 6-seed: the 2-seed would have had an easier draw than the 1-seed, while the 6-seed would have an easier draw than the 5-seed. Reseeding fixes this by matching 6-seed New York with top-seed Philadelphia, and 2-

seed San Francisco with 5-seed Dallas.

Reseeding is a powerful technique. For one, the fundamental theorem still applies to reseeded brackets, allowing us to refer to reseeded brackets by their signatures as well.

Theorem 2.6.4

There is exactly one proper reseeded bracket with each bracket signature.

Proof. The definition of properness ensures that there is only one way byes can be distributed such that a reseeded bracket can be proper. Additionally, because reseeded brackets have no additional parameters beyond which seeds get how many byes, there is no more than one reseeded bracket with each signature that could be proper. Finally, that bracket is indeed proper: if the bracket goes to chalk, the matchups will be the exact same as a traditional bracket, which by the fundamental theorem is a proper set of matchups. □

But what about orderedness? It's intuitive to think that all proper reseeded are ordered: it feels like almost by definition, the higher-seeded teams have an easier path than the lower-seeded ones. Hwang [7] conjectured a weaker version of this.

Conjecture 2.6.5

All balanced proper reseeded brackets are ordered.

Unfortunately, neither the stronger claim that all proper reseeded brackets are ordered, nor Hwang's weaker conjecture are true. Our classification of the ordered reseeded brackets takes the same route as our proof of Edwards's Theorem did: we first examine the orderedness of certain important brackets, and then we use the stapling and containment lemmas to specify the complete set of ordered reseeded brackets.

Note that the proofs of the stapling and containment lemmas for reseeded brackets, as well as the fact that all ordered reseeded brackets are proper, are so similar to the corresponding proofs for traditional brackets that we just state them without proof.

Theorem 2.6.6

All ordered reseeded brackets are proper.

Lemma 2.6.7: The Stapling Lemma for Reseeding

If $\mathcal{A} = [[\mathbf{a_0}; ...; \mathbf{a_r}]]^R$ and $\mathcal{B} = [[\mathbf{b_0}; ...; \mathbf{b_s}]]^R$ are ordered reseeded brackets, then $\mathcal{C} = [[\mathbf{a_0}; ...; \mathbf{a_r} + \mathbf{b_0} - \mathbf{1}; ...; \mathbf{b_s}]]^R$ is an ordered reseeded bracket as well.

Lemma 2.6.8: The Containment Lemma for Reseeding

If $\mathcal{A}$ and $\mathcal{B}$ are reseeded brackets, $\mathcal{A}$ contains $\mathcal{B}$, and $\mathcal{B}$ is not ordered, then neither is $\mathcal{A}$.

We now examine particular brackets.

Theorem 2.6.9

$[[\mathbf{1}]]^R$, $[[\mathbf{2}; \mathbf{0}]]^R$, and $[[\mathbf{4}; \mathbf{0}; \mathbf{0}]]^R$ are ordered.

Proof. Since no reseeding is done in a bracket of two or fewer rounds, and since the traditional brackets of these signatures are ordered, so are the reseeded brackets. □

Our primary example of a reseeded bracket that is ordered despite the traditional bracket of the same signature not being ordered is $[[\mathbf{4}; \mathbf{2}; \mathbf{0}; \mathbf{0}]]^R$.

Theorem 2.6.10

$[[\mathbf{4}; \mathbf{2}; \mathbf{0}; \mathbf{0}]]^R$ is ordered.

Proof. This can be shown by computing the probability of each team winning the format and then applying the SST conditions to establish the inequalities, as we did in Theorem 2.4.14. In the interest of brevity, however, we instead give an intuitive argument.

$\mathbb{W}_A(t_1, \mathcal{T}) \geq \mathbb{W}_A(t_2, \mathcal{T})$ because from those two teams perspectives, this format is just $[[\mathbf{4}; \mathbf{0}; \mathbf{0}]]^R$. $\mathbb{W}_A(t_2, \mathcal{T}) \geq \mathbb{W}_A(t_3, \mathcal{T})$ because t_2

has better odds if t_3 wins in the first round and they meet in the semifinals, and certainly has better odds if t_3 loses in the first round. $\mathbb{W}_A(t_4, \mathcal{T}) \geq \mathbb{W}_A(t_5, \mathcal{T})$ because t_4 is at least as likely to win the first-round matchup, and then their paths would be identical.

$\mathbb{W}_A(t_3, \mathcal{T}) \geq \mathbb{W}_A(t_4, \mathcal{T})$ holds because if both teams win the first round then t_3 has better odds in the remaining $[[\mathbf{4}; \mathbf{0}; \mathbf{0}]]^R$ bracket. Meanwhile if only one does, then t_3 will be joined by t_5 while t_4 will be joined by t_6, and so t_3 is more likely to dodge playing t_1 in the finals. The same argument applies to show that $\mathbb{W}_A(t_5, \mathcal{T}) \geq \mathbb{W}_A(t_6, \mathcal{T})$ as well. $\square$

Unfortunately, that is where the power of reseeding to convert non-ordered signatures into ordered ones ends. The following two signatures are not ordered:

Theorem 2.6.11

$[[\mathbf{6}; \mathbf{1}; \mathbf{0}; \mathbf{0}]]^R$ is not ordered.

Proof. Let $\mathcal{A} = [[\mathbf{6}; \mathbf{1}; \mathbf{0}; \mathbf{0}]]^R$, and let $\mathcal{T}$ have the following matchup table:

	t_1	t_2	t_3	t_4	t_5	t_6	t_7
t_1	0.5	$1-p$	$1-p$	$1-p$	$1-p$	$1-p$	$1-p$
t_2	p	0.5	$1-p$	$1-p$	$1-p$	$1-p$	$1-p$
t_3	p	p	0.5	0.5	0.5	$1-p$	$1-p$
t_4	p	p	0.5	0.5	0.5	0.5	0.5
t_5	p	p	0.5	0.5	0.5	0.5	0.5
t_6	p	p	p	0.5	0.5	0.5	0.5
t_7	p	p	p	0.5	0.5	0.5	0.5

Then

$$\mathbb{W}_{\mathcal{A}}(t_6, \mathcal{T}) = O(p^3),$$

but

$$\mathbb{W}_{\mathcal{A}}(t_7, \mathcal{T}) = 0.25p^2 + O(p^3).$$

Thus, for small enough p, $\mathbb{W}_{\mathcal{A}}(t_6, \mathcal{T}) < \mathbb{W}_{\mathcal{A}}(t_7, \mathcal{T})$, so $\mathcal{A}$ is not ordered. $\square$

Theorem 2.6.12

$[[\mathbf{4};\mathbf{2};\mathbf{2};\mathbf{0};\mathbf{0}]]^R$ is not ordered.

Proof. Let $\mathcal{A} = [[\mathbf{4};\mathbf{2};\mathbf{2};\mathbf{0};\mathbf{0}]]^R$, and let $\mathcal{T}$ have the following matchup table:

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
t_1	0.5	$1-p^2$	$1-p^2$	$1-p^2$	$1-p^2$	$1-p^2$	$1-p^2$	$1-p^2$
t_2	p^2	0.5	0.5	0.5	$1-p$	$1-p$	$1-p^2$	$1-p^2$
t_3	p^2	0.5	0.5	0.5	$1-p$	$1-p$	$1-p$	$1-p$
t_4	p^2	0.5	0.5	0.5	0.5	$1-p$	$1-p$	$1-p$
t_5	p^2	p	p	0.5	0.5	$1-p$	$1-p$	$1-p$
t_6	p^2	p	p	p	p	0.5	$1-p$	$1-p$
t_7	p^2	p^2	p	p	p	p	0.5	0.5
t_8	p^2	p^2	p	p	p	p	0.5	0.5

Then

$$\mathbb{W}_{\mathcal{A}}(t_7, \mathcal{T}) = 0.25p^5 + O(p^6)$$

but

$$\mathbb{W}_{\mathcal{A}}(t_8, \mathcal{T}) = 0.5p^5 + O(p^6).$$

Thus, for small enough p, $\mathbb{W}_{\mathcal{A}}(t_7, \mathcal{T}) < \mathbb{W}_{\mathcal{A}}(t_8, \mathcal{T})$, so $\mathcal{A}$ is not ordered. □

Recapping,

Figure 2.6.13: Which Proper Reseeded Brackets are Ordered

Ordered	Not Ordered
$[[\mathbf{1}]]^R$	$[[\mathbf{6};\mathbf{1};\mathbf{0};\mathbf{0}]]^R$
$[[\mathbf{2};\mathbf{0}]]^R$	$[[\mathbf{4};\mathbf{2};\mathbf{2};\mathbf{0};\mathbf{0}]]^R$
$[[\mathbf{4};\mathbf{0};\mathbf{0}]]^R$	
$[[\mathbf{4};\mathbf{2};\mathbf{0};\mathbf{0}]]^R$	

Finally, we apply the stapling and containment lemmas to complete the theorem.

Theorem 2.6.14

The ordered reseeded brackets are exactly those corresponding to signatures that can be generated in the following way:

1. Start with the list $[[\mathbf{0}]]^R$ (note that this not yet a bracket signature).

2. As many times as desired, prepend the list with $[[\mathbf{1}]]$, $[[\mathbf{3};\mathbf{0}]]$, or $[[\mathbf{3};\mathbf{2};\mathbf{0}]]$.

3. Then, add 1 to the first element in the list, turning it into a bracket signature.

Proof. The stapling lemma, combined with the fact that $[[\mathbf{1}]]^R$, $[[\mathbf{2};\mathbf{0}]]^R$, $[[\mathbf{4};\mathbf{0};\mathbf{0}]]^R$, and $[[\mathbf{4};\mathbf{2};\mathbf{0};\mathbf{0}]]^R$ are ordered, ensure that any reseeded brackets generated by the above procedure is indeed ordered. Left is to use the containment lemma to ensure that these are the only ones.

Let $\mathcal{A}$ be a bracket signature that cannot be generated by the procedure. Then, either there is a round in which three or more games are to be played, or there is a round in which exactly two games are played and the next two rounds each have exactly two games played as well.

Let i be the latest such round. If round i is the first of three rounds with two games each, then round $i+3$ must have only one game played (otherwise i would not be the latest such round). But then $\mathcal{A}$ contains $[[\mathbf{4};\mathbf{2};\mathbf{2};\mathbf{0};\mathbf{0}]]^R$, and so is not ordered.

If round i has three or more games, then round $i+1$ must contain exactly two games (any less and not every winner would have a game, any more and i would not be the latest such round.) Then, if round $i+2$ has one game, then $\mathcal{A}$ contains $[[\mathbf{6};\mathbf{1};\mathbf{0};\mathbf{0}]]^R$, and if it has two, then $\mathcal{A}$ contains $[[\mathbf{4};\mathbf{2};\mathbf{2};\mathbf{0};\mathbf{0}]]^R$. In either case, $\mathcal{A}$ is not ordered.

Thus, the ordered reseeded brackets are exactly those generated by the procedure. □

So, the space of ordered reseeded brackets is slightly larger than the space of ordered traditional brackets, although perhaps this is not quite as much of an expansion as we would've liked or expected. Despite this, reseeded brackets definitely *feel* more ordered than traditional brackets of the same signature, even if neither is ordered in the definitional sense.

Conjecture 2.6.15

There is some reasonable restriction on a set of teams that is stronger than SST under which all reseeded brackets ordered.

In the meantime, reseeding remains an important tool in our tournament design toolkit. But it is not without its drawbacks, as discussed by Baumann, Matheson, and Howe [1].

In a reseeded bracket, teams and spectators alike don't know who they will play or where their next game will be until the entire previous round is complete. This can be an especially big issue if parts of the bracket are being played in different locations on short turnarounds: in the NCAA Basketball Tournament, the first two rounds are played over a weekend at various predetermined locations. It would cause problems if teams had to pack up and travel across the country because they got reseeded and their opponent and thus location changed.

In addition, part of what makes the NCAA Basketball Tournament (affectionately known as "March Madness") such a fun spectator experience is the fact that these matchups are known ahead of time. In "bracket pools," groups of fans each fill out their own brackets, predicting who will win each game and getting points based on how many they get right. If it wasn't clear where in the bracket the winner of a given game was supposed to go, this experience would be diminished.

Finally, reseeding gives the top seed(s) an even greater advantage than they already have: instead of playing against merely the *expected* lowest-seeded team(s) each round, they would get to play against the *actual* lowest-seeded team(s). In March Madness, "Cinderella Stories," that is, deep runs by low seeds, would become much less common.

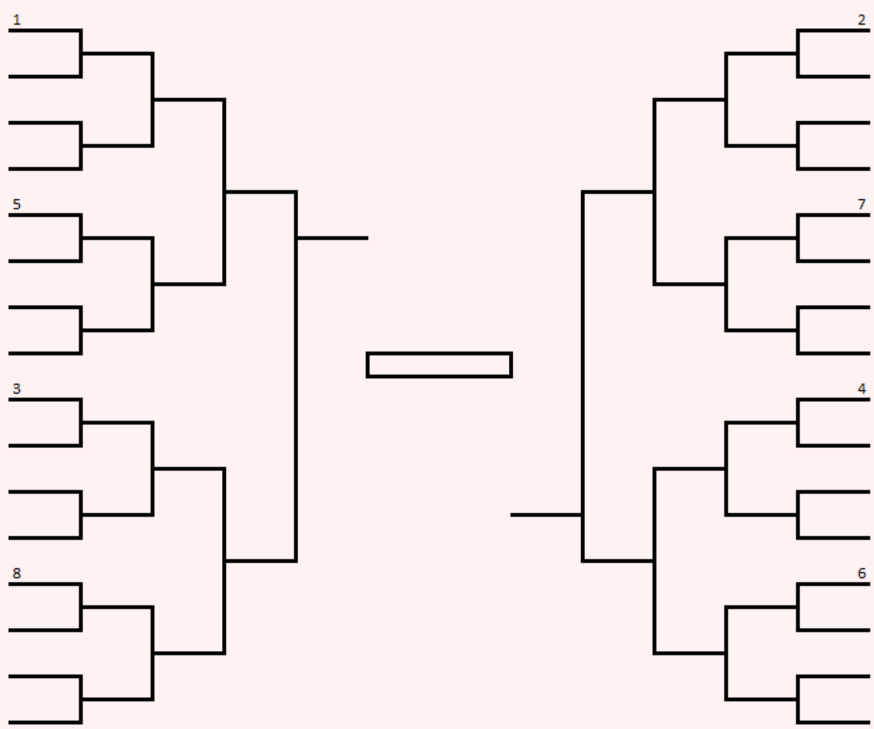
In many ways, the NFL conference playoffs are a perfect place to use a reseeded bracket: games are played once a week, giving plenty of time for travel; only seven teams make the playoffs in each, so a huge March Madness-style bracket challenge is unlikely; as a professional league, the focus is far more on having the best team win and protecting Cinderella Stories isn't as important; and because the bracket is only three rounds long, reseeding is only required once. Somewhat ironically, the NFL conference playoffs used to use the format $[[\mathbf{4};\mathbf{2};\mathbf{0};\mathbf{0}]]^R$ which is ordered, but have since allowed a seventh team from each conference into the playoffs and changed to the non-ordered $[[\mathbf{6};\mathbf{1};\mathbf{0};\mathbf{0}]]^R$.

Other leagues with similar structures might consider adopting forms of reseeding to protect their incentives and competitive balance (looking at you, Major League Baseball), but in many cases, the traditional bracket structure is too appealing to adopt a reseeded one.

2.7 Cohort Randomized Brackets

Consider the 2022 Wimbledon Mixed Double's Championship, whose bracket is depicted in Figure 2.7.1.

Figure 2.7.1: 2022 Wimbledon Mixed Double's Championship



The 2022 Wimbledon Mixed Double's Championship used a balanced bracket of signature $[[\mathbf{32}; \mathbf{0}; \mathbf{0}; \mathbf{0}; \mathbf{0}; \mathbf{0}]]$. However its seeding is certainly not proper, and in fact, looks quite strange.

For one thing, 24 of the 32 starting lines in the bracket are not seeded at all. What does this mean? In Wimbledon, only the top 25% of teams are seeded. The other 75% of teams are randomly placed in the remaining starting lines.

But secondly, even the top 8 teams are not seeded properly: if all games go chalk, the quarterfinals will be 1v5, 2v7, 3v5, and 4v6, and the semifinal will be 1v3 and 2v4. How did Wimbledon come up with such a bizarre seeding? Randomly.

Before each Wimbledon Mixed Double's Championship, the seeding is constructed in the following way: the 1- and 2-seeds are placed normally. Then half the 3- and 4-seeds are placed normally, and half the time they are swapped. Then the 5- through 8-seeds are randomly placed into the four starting lines that are normally assigned to those four seeds. Finally, the remaining 24 teams are placed randomly into the remaining spots.

Thus the bracket in Figure 2.7.1 is only one of the forty eight different possible brackets that could be used in the Wimbledon Mixed Double's Championship: it happens to be the one used in 2022.

Why does Wimbledon use such a strange format? Wimbledon is one of many tournaments on the ATP Tour (the set of tournaments played by the professional tennis players) that all use almost identical formats: large balanced brackets. Additionally, the seeding for these tournaments is set by the ATP rankings, which tend to be slow to update. As a result, if every ATP Tour tournament used a proper seeding, the 6-seed and 27-seed would play each other in the first round at every tournament until one of them moved up or moved down. These rematches were deemed undesirable and so this randomization procedure was introduced: The 1-seed's quarterfinals matchup (if everything goes chalk) is now randomly drawn from the 5- through 8-seeds, instead of always being the 8-seed.

Additionally, this particular way of grouping seeds ensures that after round s for $s > 1$, if the format goes chalk, the top 2^{r-s} out of 2^r teams remain. Finally, it might even allow a balanced bracket on more than four teams to be ordered, a feat that proper and reseeded brackets were both unable to accomplish.

The format used in the Wimbledon Mixed Double's Championship is a particular example of a class of formats called *cohort randomized brackets* (a generalization of Schwenk's cohort randomized seeding [9]).

Definition 2.7.2: Composition

A *composition* $\mathcal{B} = (\mathbf{b_1} + ... + \mathbf{b_n})$ of a natural number n is a way of writing n as the sum of a sequence of natural numbers.

We will use compositions to describe which seeds are to be randomized with which other seeds (each of these groups is called a cohort). In the particular case of the Wimbledon Mixed Double's Championship, the bottom 24 teams are randomized, as are the next 4 and 2 after that, and then the 2-seed and 1-seed aren't randomized at all. So the composition of the Wimbledon Mixed Double's Championship is

$$\mathcal{B} = (\mathbf{24} + \mathbf{4} + \mathbf{2} + \mathbf{1} + \mathbf{1}).$$

Definition 2.7.3: Cohort Randomized Bracket

A *cohort randomized bracket* is an n-team tournament format parameterized $(\mathcal{A}, \mathcal{B})$, where $\mathcal{A}$ is an n-team bracket signature and $\mathcal{B} = (\mathbf{b_1} + \ldots + \mathbf{b_k})$ is a composition of n. The proper bracket of signature $\mathcal{A}$ is constructed, but, for each j, seeds

$$n + 1 - \sum_{i=1}^{j} b_i$$

through

$$n - \sum_{i=1}^{j-1} b_i$$

are shuffled randomly. The resulting bracket is played out normally.

Thus the 2022 Wimbledon Mixed Double's Championship employs the cohort randomized bracket parameterized by

$$([[\mathbf{32}; \mathbf{0}; \mathbf{0}; \mathbf{0}; \mathbf{0}; \mathbf{0}]], (\mathbf{24} + \mathbf{4} + \mathbf{2} + \mathbf{1} + \mathbf{1})).$$

But it is certainly not the simplest example of a cohort randomized bracket: a format that totally ignores seeding and just randomly places all the teams into a bracket is also cohort randomized.

Definition 2.7.4: Totally Randomized Bracket

The *totally randomized bracket* of the n-team signature $\mathcal{A}$ is the cohort randomized bracket parameterized by $(\mathcal{A}, (\mathbf{n}))$.

Chung and Hwang had a conjecture about totally randomized brackets [3].

Conjecture 2.7.5

All totally randomized brackets are ordered.

Certainly once the randomization is been complete and starting lines have been set this is not true: the resulting bracket will either be improper, in which case it is certainly not ordered, or proper, in which case it is ordered only when it satisfies the condition of Edwards's Theorem. However, in expectation *over the randomization*, it is natural to hope that all totally randomized brackets are ordered: there is no advantage to being one seed or another, so certainly the better teams should win more?

Unfortunately, as was the case in the previous section, the sweeping conjecture ultimately falls short, this time due to a counterexample given by Israel [8].

Theorem 2.7.6

Let $\mathcal{A} = [[\mathbf{16}; \mathbf{0}; \mathbf{0}; \mathbf{0}; \mathbf{1}; \mathbf{0}]]$. The totally randomized bracket $\mathcal{B} = (\mathcal{A}, (\mathbf{17}))$ is not ordered.

Proof. □

However, again analogously to the previous section, orderedness can be rescued for a smaller set of brackets, this time by Chen and Hwang [2].

Theorem 2.7.7

All totally randomized balanced brackets are ordered.

Proof. □

Chen and Hwang conjectured that this theorem could be extended.

Definition 2.7.8: Nearly Balanced

A bracket is *nearly balanced* if no team receives more than one bye.

Conjecture 2.7.9

All totally randomized nearly balanced brackets are ordered.

Conjecture 2.7.9 remains open, but would be very powerful if true: unlike balanced brackets, there is a nearly balanced bracket for every number of teams, if the conjecture held, then total randomization would be a format that can provide orderedness to arbitrary numbers of teams n without having $O(n)$ rounds, as the traditional and reseeded options do.

Of course, this orderedness does not come without drawbacks. For one, the randomization makes the orderedness feel a bit cheap: once the randomization is complete, before any games have even been played, the orderedness is lost. (Compare to the proper and reseeded ordered brackets, which maintain their orderedness throughout the whole tournament.)

But secondly, total randomness has the undesirable property that it might make for some very lopsided and anti-climatic brackets. It could be that top-two teams, whom everyone wants to see face off in the championship game, are set to play each other in the first round!

To fix this, we define a new class of cohort randomized brackets: *chalk randomized brackets*.

Definition 2.7.10: Chalk Randomized Brackets

The *chalk randomized bracket* of the n-team signature $\mathcal{A} = [[\mathbf{a_0}; ...; \mathbf{a_r}]]$ is the cohort randomized bracket parameterized by $(\mathcal{A}, (\mathbf{b_1} + \mathbf{b_2} + ... + \mathbf{b_r} + \mathbf{1}))$, where b_i is the number of games being played in round i of $\mathcal{A}$.

Chalk randomization solves the second issue with total randomness: that good teams might meet earlier than desired.

Theorem 2.7.11

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

Proof. We proceed by induction on r. If $r = 0$, then $\mathcal{A} = [[\mathbf{1}]]$, and so the theorem holds. For any other r, in the first round, a_0 teams are playing $b_1 = a_0/2$ games. However, the b_1 lowest seeds in the tournament are shuffled between themselves, so none of them can play each other. Because the tournament goes chalk, they will all lose, so after the first round, the $n - b_1$ remaining teams are the top $n - b_1$

top seeds. Further, we are left with the chalk randomized bracket of signature $\mathfrak{S}(\mathcal{A})$, so the theorem holds by induction. □

The chalk randomizations of balanced brackets have particularly nice forms: the chalk randomization of an r-round balanced bracket is parameterized by

$$([[\mathbf{2^r}; \mathbf{0}; ...; \mathbf{0}]], (\mathbf{2^{r-1}} + \mathbf{2^{r-2}} + ... + \mathbf{2^0} + \mathbf{1})).$$

This is very similar to the Wimbledon format, which employs a chalk randomized balanced bracket but merges the first two cohorts. Wimbledon even uses this scheme in their single's tournaments, which are on 128 teams and are parameterized by

$$([[\mathbf{128}; \mathbf{0}; ...; \mathbf{0}]], (\mathbf{96} + \mathbf{16} + \mathbf{8} + \mathbf{4} + \mathbf{2} + \mathbf{1})).$$

Chalk randomized balanced brackets are of a nice enough form that we can show their orderedness:

Theorem 2.7.12

All chalk randomized balanced brackets are ordered.

Proof. □

However, the natural extension, that all chalk randomized brackets are ordered, remains open.

Conjecture 2.7.13

All chalk randomized brackets are ordered.

Finally, the Wimbledon formats are not technically chalk randomized, as they merge the first two cohorts. However, we conjecture that they too are ordered.

Conjecture 2.7.14

The Wimbledon formats are ordered.

Ultimately, cohort randomization is a useful tool that uses randomness to generate ordered formats for a given signature (balanced ones in particular), when traditional and reseeded brackets were unable to do so. As we discussed, however, this orderedness can feel a bit fake, as the format is orderedness only over the randomness in the starting line selection: as soon as the bracket has been set, the format is no longer ordered.

Cohort randomization can still be a valuable tool in cases where many tournaments are being played and rematches would like to be avoided (as in the ATP), cases where a random distribution of advantages might be preferred to a disordered distribution of advantages (as in March Madness), or cases when the true seeding is unknown (and so total randomization is a more accurate model of the format).

But it is certainly not a completely satisfying tool for generating ordered brackets from arbitrary signatures. In fact, the question of whether such a tool exists at all remains an open one, and we will conclude this chapter without an answer. Instead, we will explore other kinds of formats all together, such as *multibrackets*, *round robins*, and *pools*.

3 Multibrackets

3.1 Semibrackets

In this chapter, we will develop the theory of a new kind of format called a *multibracket*. The multibracket is a very versatile class of formats, unifying many seemingly different formats that are used in practice and allowing us to study them all more effectively.

Amongst these formats are:

1. Third-place games, as in the 2015 AFL Asian Cup.

2. Assigning bids for a future tournament, as in the Ultimate Frisbee Sectional and Regional Tournaments.

3. Swiss systems, as in the 2023 League of Legends World Championships.

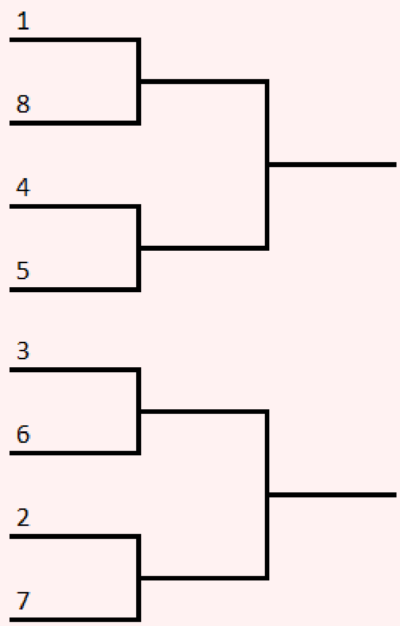
4. Double elimination tournaments, as in the 2005 Women's College World Series.

But before we can develop the concept of a multibracket or see how any of these tournaments are instances of it, we need to first investigate a class of formats called *semibrackets* of which multibrackets are composed.

Consider the following tournament design problem: we are tasked with designing an eight-team tournament to select the top two teams who will go on to compete as a part of the national tournament. The catch: here's only enough time for two rounds: perhaps due to field space or team fatigue, each team can only play two games. What design should we use?

The most natural answer to this question is to simply use a traditional eight-team bracket, but to leave the championship game unplayed. This format is displayed in the figure below.

Figure 3.1.1: $[[8; 0; 0; 0]]$ with no Championship Game



The format in Figure 3.1.1 does exactly what we need. The championship game being left unplayed is not a bug but a feature: each team plays a maximum of two games, and the two teams that advance to the national tournament are clear.

While it would be reasonable to describe the format in Figure 3.1.1 as two brackets that run side-by-side, it would be nice to be able to describe it as a single format: a bracket in which the championship game is left unplayed.

Definition 3.1.2: Semibracket

A *semibracket* is a tournament format in which:

- Teams don't play any games after their first loss.
- The matchups between teams are determined based on the ordering of the teams in $\mathcal{T}$ in advance of the outcomes of any games.

All teams that finish a semibracket with no losses are declared co-champions.

Recall that a bracket is a tournament format in which:

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

Semibrackets are formats that adhere to the first and third requirement, but not (necessarily) the second one. Multiple teams can finish a semibracket undefeated, and they are each winners of the semibracket.

Figure 3.1.3 describes which properties various bracket-like formats require. The row for multibrackets is included even though we won't formally define a multibracket until the next section.

Figure 3.1.3: Properties of Bracket-like Formats

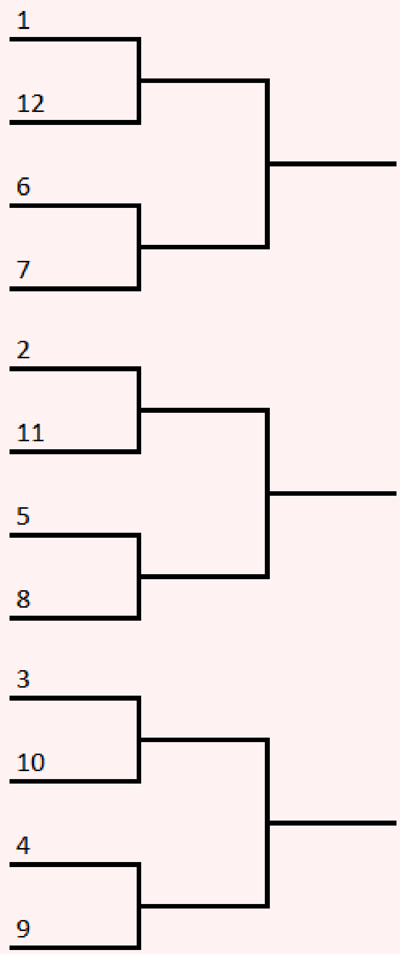
Format	No games after first loss	Only one team finishes undefeated	Matchups determined in advance
Traditional Bracket	✔	✔	✔
Semibracket	✔	✘	✔
Multibracket	✘	✘	✔
Reseeded Bracket	✔	✔	✘

Definition 3.1.2 implies that traditional brackets are a subset of semibrackets. Of course, not all semibrackets are traditional brackets: the format in Figure 3.1.1 is one such example.

That said, this format is not a particularly exciting example of a semibracket: after all, it is just a traditional bracket minus one game. Are there any examples of semibrackets that are not traditional brackets with some rounds left uncompleted?

Indeed there are. Let's modify the original problem so that we need to pick a top three teams out of twelve. Again, no team can play more then two games. The natural choice is shown below in Figure 3.1.4.

Figure 3.1.4: A More Exciting Semibracket



There is no potential for the format in Figure 3.1.4 to be completed into a traditional bracket, the next round would include three teams: and odd number. But as a semibracket, this is still a viable format, one that perfectly solves the tournament design problem that we were given.

Definition 3.1.5: Rank of a Semibracket

If semibracket $\mathcal{A}$ has m co-champions, then $\mathrm{Rank}(\mathcal{A}) = m$. We say $\mathcal{A}$ has rank m or that $\mathcal{A}$ *ranks* m *teams*.

Thus, traditional brackets are exactly the semibrackets that rank one team. The formats in Figures 3.1.1 and 3.1.4 rank two and three teams, respectively.

We can adapt the concept of a bracket signature to semibrackets.

Definition 3.1.6: Semibracket Signature

The *signature* $[[\mathbf{a_0};...;\mathbf{a_r}]]_m$ of an r-round semibracket $\mathcal{A}$ is list such that a_i is the number of teams with i byes and $m = \text{Rank}(\mathcal{A})$. (In the case where $m = \text{Rank}(\mathcal{A}) = 1$, it can be omitted.)

Thus the signature of traditional brackets are the same as when they are viewed as semibrackets that rank one team. The signatures of the formats in Figures 3.1.1 and 3.1.4 are $[[\mathbf{8};\mathbf{0};\mathbf{0}]]_2$ and $[[\mathbf{12};\mathbf{0};\mathbf{0}]]_3$, respectively.

In analogy with traditional bracket signature's Theorem 2.2.7, we have

Theorem 3.1.7

Let $\mathcal{A} = [[\mathbf{a_0};...;\mathbf{a_r}]]_m$ be a list of natural numbers. Then $\mathcal{A}$ is a semibracket signature if and only if

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

The proof is almost identical to that of Theorem 2.2.7 so we leave it out for brevity. Likewise, the fundamental theorem still applies, again with almost the exact same proof (also left out for brevity).

Theorem 3.1.8

There is exactly one proper semibracket with each semibracket signature.

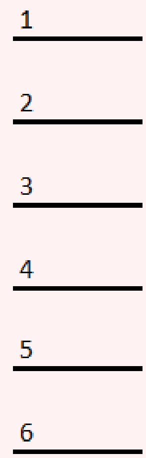
Finally, the space of semibrackets contains some slightly bizarre formats.

Definition 3.1.9: Trivial Semibrackest

We say semibrackets is *trivial* if it has signature $[[\mathbf{n}]]_n$ for some n.

No games are played in a trivial semibracket: all teams that enter one exit having been declared champion. The only trivial traditional bracket is $[[\mathbf{1}]]$, but there is one trivial semibracket with each rank. Trivial semibrackets look a bit strange when drawn.

Figure 3.1.10: $[[\mathbf{6}]]_6$



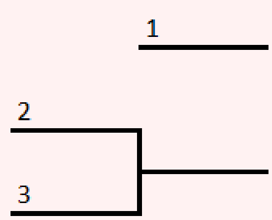
Arguably stranger than trivial semibrackets are *semitrivial* semibrackets.

Definition 3.1.11: Semitrivial Semibracket

We say a semibracket $\mathcal{A} = [[\mathbf{a_0}; ...; \mathbf{a_r}]]_m$ is *semitrivial* if $r \geq 1$ and $a_r \neq 0$.

In a semitrivial semibracket, some teams are declared champion without playing any games, while other have games on their schedule. The simplest example of a semitrivial semibracket is $[[\mathbf{2}; \mathbf{1}]]_2$.

Figure 3.1.12: $[[\mathbf{2}; \mathbf{1}]]_2$



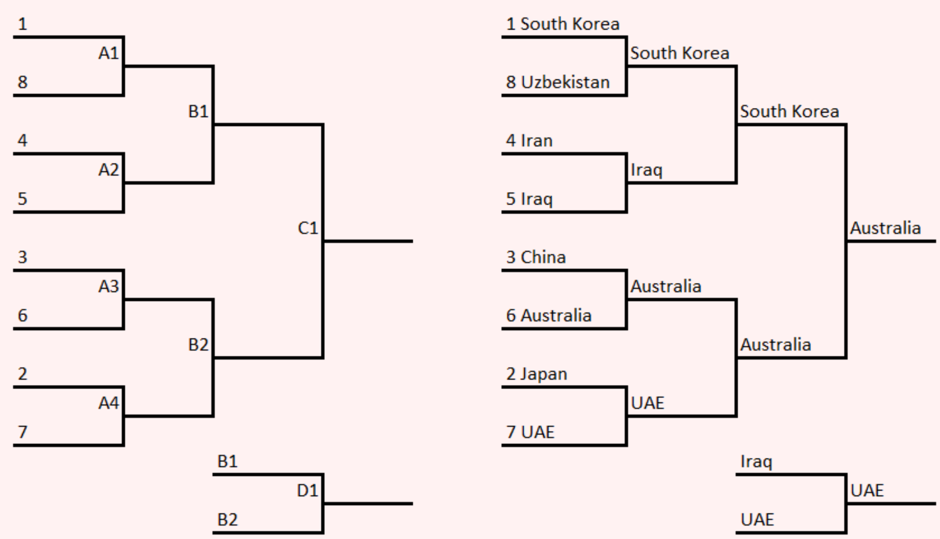
There are no semitrivial traditional brackets: if a team wins a traditional bracket without playing any games, they must be the only team in the bracket. Semitrivial semibrackets are pretty unintuitive: luckily, we will soon see that semitrivial semibracket aren't required to develop the theory of multibrackets, and so we won't have to worry about them.

With the idea of a semibracket developed and fleshed out, we can now move on to the meat of the chapter: multibrackets.

3.2 Third-Place Games

Consider the format used in the 2015 Asian Football Confederation Asian Cup: a bracket of signature $[[\mathbf{8}; \mathbf{0}; \mathbf{0}; \mathbf{0}]]$, plus a third-place game.

Figure 3.2.1: 2015 AFL Asian Cup



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 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 winners of those games always continue on in the original bracket, so such labels only refer to the losers.

The 2015 AFL Asian Cup is a *multibracket*: a sequence of brackets (or semibrackets) in which teams that lose in earlier brackets fall into later brackets instead of being eliminated outright, and teams finish in a place dependent on which bracket they win. Formally,

Definition 3.2.2: Multibracket

A *multibracket* $\mathcal{A}$ is a sequence of semibrackets $\mathcal{A}_1 \to \ldots \to \mathcal{A}_k$ where some of the starting lines in some of the semibrackets are assigned to teams that lost certain games in other semibrackets, subject to the following conditions:

1. No game sends its loser to multiple locations.
2. If the loser of a game in $\mathcal{A}_i$ is sent to $\mathcal{A}_j$, then $i < j$.
3. If the loser of a game in round r of $\mathcal{A}_i$ is sent to $\mathcal{A}_j$, then the loser of each game in round s of $\mathcal{A}_i$ for $r \leq s$ is sent to a $\mathcal{A}_\ell$ for $i < \ell \leq j$.
4. If the loser of a game in $\mathcal{A}_i$ is sent to $\mathcal{A}_j$, then the each loser of each game in $\mathcal{A}_\ell$ for $i < \ell < j$ is sent to a $\mathcal{A}_m$ for $\ell < m \leq j$.
5. If the i-seed starts in $\mathcal{A}_j$, then the loser of each game in $\mathcal{A}_\ell$ for $\ell < j$ is sent to $\mathcal{A}_m$ for $\ell < m \leq j$.
6. If the i-seed starts in $\mathcal{A}_j$, and $\ell < i$, then the ℓ-seed starts in $\mathcal{A}_m$ for $m < j$.
7. None of $\mathcal{A}_i$ are semitrivial.

Then, teams that win semibracket $\mathcal{A}_j$ finish in place

$$\sum_{i=1}^{j} \mathrm{Rank}(\mathcal{A}_i).$$

The first requirement ensures that teams are not playing in multiple semibrackets simultaneously, and the last requirement allows us to avoid having to consider somewhat pathological semitrivial semibrackets. The other five requirements ensure that it is always better to win games than to lose them: losing games should never put you in an earlier bracket than winning them, even accounting for the fact that future games in earlier brackets will likely be against better teams then future games in later brackets.

Note that the definition the multibracket allows multiple teams to finish in the same place. For example, the multibracket of signature $[[\mathbf{8}; \mathbf{0}; \mathbf{0}; \mathbf{0}]] \to$

$[[\mathbf{3}]]_3$ crowns one team champion and grants fourth-place to three teams. However,

Theorem 3.2.3

No more than m teams in a multibracket finish in the top-m.

Proof. This follows from how placed finishes are defined in Definition 3.2.2. $\square$

Colloquially, we use the terms *higher* and *lower* to refer to certain semibrackets in a multibracket.

Definition 3.2.4: Higher and Lower Semibrackets

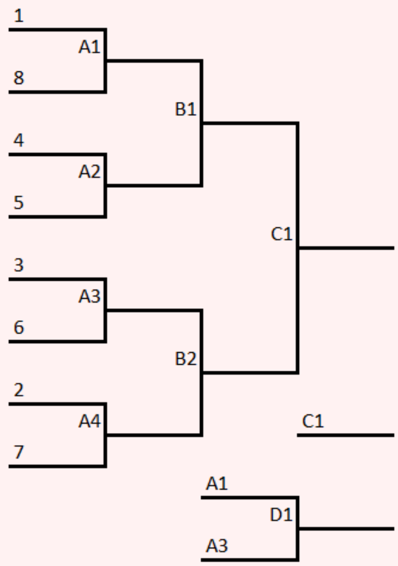
If $\mathcal{A}_i$ and $\mathcal{A}_j$ are two semibrackets in a multibracket $\mathcal{A}$ such that $i < j$, we say $\mathcal{A}_i$ is the *higher semibracket* and $\mathcal{A}_j$ is the *lower semibracket*.

The notion of higher and lower semibrackets fits with the intuitive idea of teams falling down the multibracket as they lose.

So, how does the 2015 AFL Asian Cup fit into this schema? Though Figure 3.2.1 seems to indicate that it ought to be a sequence of two brackets, this doesn't quite work. For one, the multibracket rule (3) prevents the losers of games **B1** and **B2** from falling into the second bracket without the loser of **C1** being placed anywhere. Additionally, if the format had only two brackets, the winner of the game between **B1** and **B2** would be awarded second place, rather than third.

However, both of these issues can be fixed if we think of the 2015 AFL Asian Cup as a sequence of three brackets, the second of which has signature $[[\mathbf{1}]]$.

Figure 3.2.5: $[[8; 0; 0; 0]] \to [[1]] \to [[2; 0]]$



This format satisfies all of the requirements of the multibracket, and correctly assigns first, second, and third place. Thus, we say that the 2015 AFL Asian Cup is a multibracket of signature $[[\mathbf{8}; \mathbf{0}; \mathbf{0}; \mathbf{0}]] \to [[\mathbf{1}]] \to [[\mathbf{2}; \mathbf{0}]]$.

Definition 3.2.6: Rank of a Multibracket

If $\mathcal{A} = \mathcal{A}_1 \to \ldots \to \mathcal{A}_k$ is a multibracket, then

$$\text{Rank}(\mathcal{A}) = \sum_{i=1}^{k} \text{Rank}(\mathcal{A}_i).$$

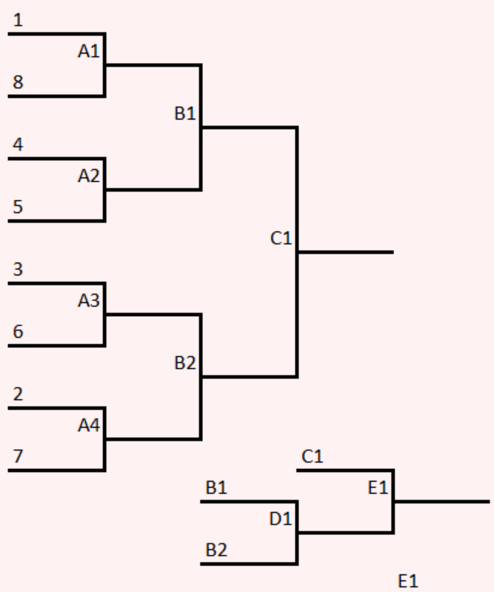
We say $\mathcal{A}$ has rank $\text{Rank}(\mathcal{A})$ or that $\mathcal{A}$ *ranks* $\text{Rank}(\mathcal{A})$ teams.

The 2015 AFL Asian Cup determines a top-three, so the multibracket $[[\mathbf{8}; \mathbf{0}; \mathbf{0}; \mathbf{0}]] \to [[\mathbf{1}]] \to [[\mathbf{2}; \mathbf{0}]]$ has rank three. But this multibracket is far from the only multibracket of rank three that the AFL could have used to dole 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 arguing: South Korea and we both 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 $[[\mathbf{8}; \mathbf{0}; \mathbf{0}; \mathbf{0}]] \to [[\mathbf{2}; \mathbf{1}; \mathbf{0}]] \to [[\mathbf{1}]]$.

Figure 3.2.7: $[[8; 0; 0; 0]] \to [[2; 1; 0]] \to [[1]]$



If the AFL used the format in Figure 3.2.7 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 Figure 3.2.7 isn't feasible.

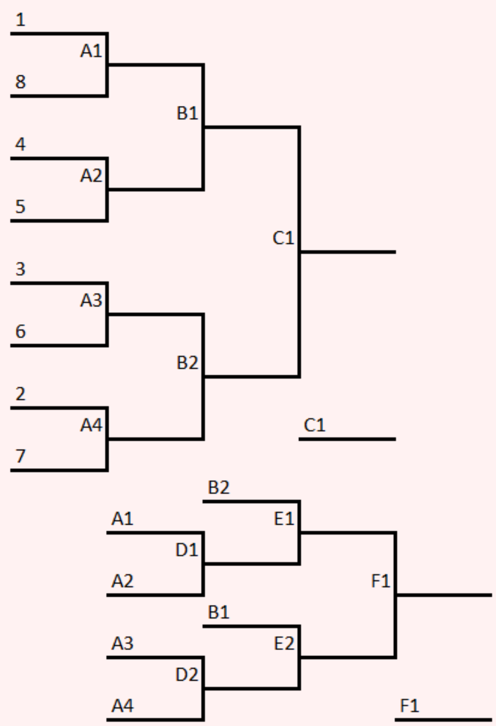
Another concern: what if Iraq had beaten the UAE when they played in game **D1**? Then the two teams with a claim to second place would have been South Korea and Iraq, except South Korea already beat Iraq! In this world, South Korea being given second place without having to win a rematch with Iraq seems 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 a rematch, although this would no longer be a multibracket and is a bit out of scope.

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 a 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, 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. (Formally, the multibracket $[[\mathbf{8};\mathbf{0};\mathbf{0};\mathbf{0}]] \to [[\mathbf{1}]] \to [[\mathbf{2};\mathbf{0}]]$ only ranks three teams but it could easily be extended to the following multibracket that ranks four, $[[\mathbf{8};\mathbf{0};\mathbf{0};\mathbf{0}]] \to [[\mathbf{1}]] \to [[\mathbf{2};\mathbf{0}]] \to [[\mathbf{1}]]$, which has the property mentioned above.)

A better format for selecting the top four teams might look like this:

Figure 3.2.8: $[[8; 0; 0; 0]] \to [[1]] \to [[4; 2; 0; 0]] \to [[1]]$



The multibracket in Figure 3.2.8 selects a top four without having the selection be determined only by the first-round games. In fact, $[[\mathbf{8}; \mathbf{0}; \mathbf{0}; \mathbf{0}]] \to [[\mathbf{1}]] \to [[\mathbf{4}; \mathbf{2}; \mathbf{0}; \mathbf{0}]] \to [[\mathbf{1}]]$ has the attractive property that a team will finish in the top four if and only if it wins two of its first three games.

With the general format of the multibracket established, the next few sections will examine particular classes of the design.

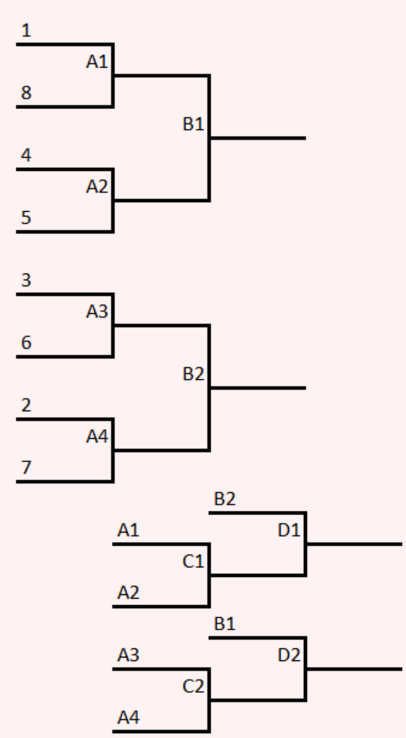
3.3 Efficient Multibrackets

At the end of the previous section, we considered the format in Figure 3.2.8: a multibracket of signature $[[\mathbf{8}; \mathbf{0}; \mathbf{0}; \mathbf{0}]] \rightarrow [[\mathbf{1}]] \rightarrow [[\mathbf{4}; \mathbf{2}; \mathbf{0}; \mathbf{0}]] \rightarrow [[\mathbf{1}]]$, which categorizes the top four teams. If it's important to rank the teams one trough four then that format works well enough.

But if all we care about is which of the teams are in top four, and not the ranking among them, then some of the games are unnecessary. In particular, games **C1** and **F1** could be left unplayed, as both the winner and loser of each of those games finish in the top four.

The resulting format is shown below.

Figure 3.3.1: An Efficient Format for Selecting a Top Four



The format in Figure 3.3.1 is still a multibracket of rank four. But instead of being composed of four traditional brackets, it is composed of two semibrackets each of which have rank two: one with the **A** and **B** round games, and one with the **C** and **D** round games. And, as desired, there no games played between two teams such that both the winner and loser of each of those games finish in the top four.

This format has signature $[[\mathbf{8};\mathbf{0};\mathbf{0}]]_2 \to [[\mathbf{4};\mathbf{2}]]_2$ and we say that it is *weakly efficient*.

Definition 3.3.2: Weakly Efficient

A multibracket is *weakly efficient* if there are no games played within it such that both the winner and loser of that game are guaranteed to be ranked by the multibracket.

Identifying whether a multibracket is weakly efficient can be done just by looking at its signature.

Lemma 3.3.3

In a multibracket $\mathcal{A}$, if the loser of game $\mathbf{G}$ goes to bracket $\mathcal{A}_j$, then the winner of game $\mathbf{G}$ will either:

1. Win a bracket $\mathcal{A}_i$ for $i \leq j$, or

2. Lose a game in $\mathcal{A}_j$.

Proof. Let $\mathcal{A}_i$ be a semibracket in $\mathcal{A}$, and $\mathbf{G}$ be a game in $\mathcal{A}_i$ such that the loser of $\mathbf{G}$ goes to bracket $\mathcal{A}_j$. Let t be the team that won $\mathbf{G}$. Assume that t does not win any bracket $\mathcal{A}_i$ for $i \leq j$. Thus, t must have lost at least one game after playing $\mathbf{G}$. Upon losing this game, by multibracket rule (3), they must fall into bracket $\mathcal{A}_\ell$ for $i < \ell \leq j$. If they fall into $\mathcal{A}_\ell$ for $\ell < j$, then again they must lose and again by multibracket rule (4) fall into $\mathcal{A}_m$ for $\ell < m \leq j$. At some point, then, t must fall into $\mathcal{A}_j$. And since t does not win bracket $\mathcal{A}_j$ either, they must lose in $\mathcal{A}_j$ as well. $\square$

Theorem 3.3.4

A multibracket $\mathcal{A} = \mathcal{A}_1 \to \ldots \to \mathcal{A}_k$ is weakly efficient if and only if there is some j such that all brackets $\mathcal{A}_i$ for $i \leq j$ are trivial and all brackets $\mathcal{A}_i$ for $i > j$ are not.

Proof. Let $\mathcal{A}$ be a multibracket.

Assume no such j exists, and let i be the first trivial bracket that follows a nontrivial one. Thus there is at least one game $\mathbf{G}$ such that the loser drops into $\mathcal{A}_i$. Because $\mathcal{A}_i$ is trivial, the loser of $\mathbf{G}$ wins $\mathcal{A}_i$. Applying Lemma 3.3.3, we see that the winner of $\mathbf{G}$ will either win a semibracket as well, or lose in $\mathcal{A}_i$. But $\mathcal{A}_i$ is trivial, so they must win a semibracket and thus get ranked. Therefore, $\mathcal{A}$ is not weakly efficient.

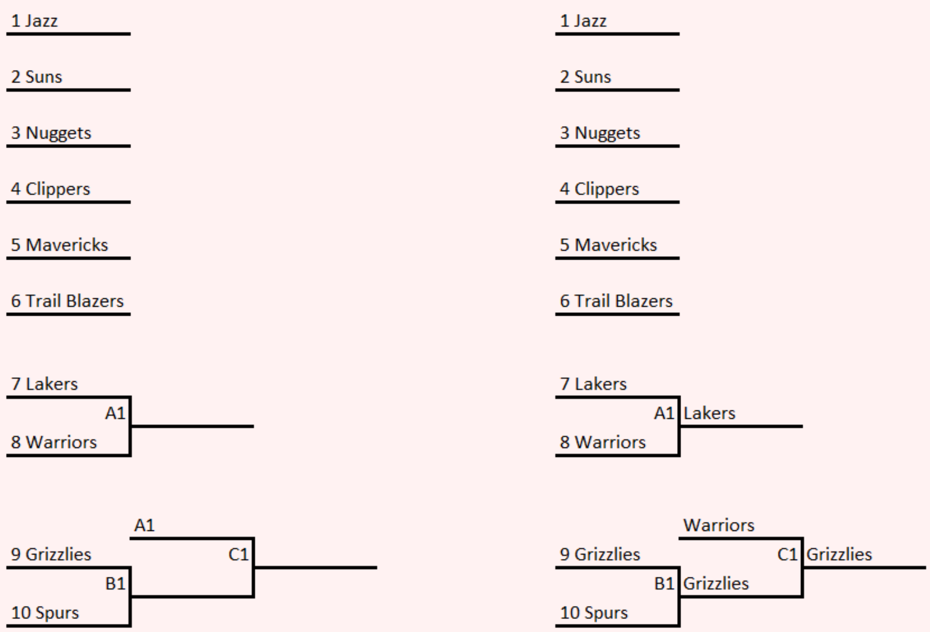
Now assume that such a j exists. We will show by inducting on the semibrackets in $\mathcal{A}$ in reverse that none of the semibrackets contain a game that violates the efficiency condition. Firstly, $\mathcal{A}_k$ upholds the condition because any team that loses a game in $\mathcal{A}_k$ doesn't fall into another semibracket, much less have a chance to win one.

Now we must show that if all of the semibrackets from $\mathcal{A}_{i+1}$ to $\mathcal{A}_k$ uphold the condition, then $\mathcal{A}_i$ does as well. If $i \leq j$, then $\mathcal{A}_i$ is trivial so there are no games to violate the condition with. Otherwise, let $\mathbf{G}$ be a game in $\mathcal{A}_i$. If the loser of $\mathbf{G}$ does not fall into another semibracket, then we are done. If they do, then because that bracket is not trivial, they will play another game. However, by induction, the loser of this game is not guaranteed to end up ranked by the format. Thus neither is the loser of $\mathbf{G}$.

So by induction, if such a j exists, then $\mathcal{A}$ is weakly efficient. Thus we have proved the theorem. □

Another example of a weakly efficient multibracket is the 2021 NBA Western Conference Play-in Tournament, which was a ten-team multibracket with order eight and the following signature: $[[\mathbf{6}]]_6 \to [[\mathbf{2};\mathbf{0}]]_1 \to [[\mathbf{2};\mathbf{1};\mathbf{0}]]_1$. The play-in tournament was used to whittle the top ten teams in the conference down to eight teams who would qualify for the playoffs.

Figure 3.3.5: 2021 NBA Western Conference Play-in



Finally, the USA Ultimate Manual of Championship Series Tournament Formats [11], which is used to determine the formats to be used at the various sectional and regional tournaments in the sport of ultimate frisbee, contains a host of weakly efficient multibrackets for selecting the top m teams out of a list of n for m and n ranging from 1 to 24.

We note two things about the notion of weak efficiency presented above. First, Theorem 3.3.4 implies that a weakly efficient multibracket can begin with a long string of trivial semibrackets before the nontrivial ones begin. While this is sufficient for avoiding playing unnecessary games, it does not completely remove unnecessary semibrackets: the set of leading trivial semibrackets $[[\mathbf{m_1}]]_{m_1} \to ... \to [[\mathbf{m_j}]]_{m_j}$ of a weakly efficient multibracket can be combined into a single trivial semibracket $[[\mathbf{m_1} + ... + \mathbf{m_j}]]_{(m_1+...+m_j)}$ without affecting which teams end up ranked.

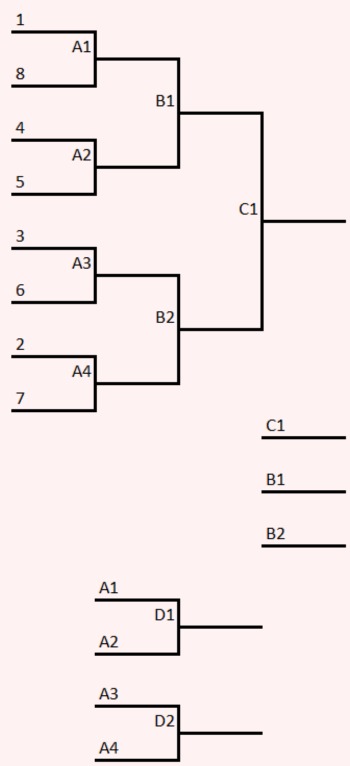
To patch this, we strengthen the notion of weak efficiency into just *efficiency.*

Definition 3.3.6: Efficient

We say a multibracket is *efficient* if it has no trivial semibrackets, or if only its leading semibracket is trivial.

Second, efficiency makes a lot of sense if we are only interested in the top-m, where $m = \text{Rank}(\mathcal{A})$, and not in the ranks of of the teams with them. However, sometimes we might be interested in the intermediate ranks as well. For example, let's see we want to design an eight-team tournament format in which the top team receives the grand prizes, while the second- through sixth-place each get equivalent consolation prizes, and seventh and eighth each get nothing. While not weakly efficient, the following format assigns the desired places without playing any games between teams that are guaranteed to receive the same prize.

Figure 3.3.7: $[[8;0;0;0]] \to [[3]]_3 \to [[4;0]]_2$



To account for this, we introduce the notion of a *prize structure*.

Definition 3.3.8: Prize Structure

A *prize structure* $\mathcal{P}$ is a sequence $(\mathbf{p_1}, ..., \mathbf{p_m})$ indicating that the top p_1 teams in a format receive some prize, the next p_2 receive some smaller prize, etc. Any teams finishing in place $1 + \sum_{i=1}^{m} p_i$ or worse receive no prize.

Then,

Definition 3.3.9: Efficient with Respect to a Prize Struture

We say a multibracket $\mathcal{A} = \mathcal{A}_1 \to ...\mathcal{A}_k$ is *efficient with respect to a prize structure* $\mathcal{P} = (\mathbf{p_1}, ..., \mathbf{p_m})$ if $\mathcal{A}_j$ being trivial implies that for some $\ell < m$,

$$\sum_{i=1}^{j-1} \text{Rank}(\mathcal{A}_j) = \sum_{i=1}^{\ell} p_i.$$

So the multibracket $[[\mathbf{8}; \mathbf{0}; \mathbf{0}; \mathbf{0}]] \to [[\mathbf{3}]]_3 \to [[\mathbf{4}; \mathbf{0}]]_2$ is efficient with respect to the prize structure $(\mathbf{1}, \mathbf{5})$. A multibracket $\mathcal{A}$ being efficient is the same as it being efficient with respect the prize structure $(\text{Rank}(\mathcal{A}))$.

Efficient multibrackets are great tournament designs for tournaments whose primary goal is to select the top m teams to move on to the next stage of the competitions, as discussed in the beginning of this section. They do so excitingly, with each spot in the top m being awarded as the winner of a particular game; efficiently, with no games being played between teams who will receive the same prize; and fairly, as the multibracket rules ensure that winning is always better than losing. They will become an important class of multibracket as our general analysis of tournament formats progresses.

3.4 Swiss Systems

Many tournaments, particularly those in which there are many teams each looking to play a similar number of games against teams of similar skill, use a set of formats referred to as *swiss systems.* In particular, swiss systems or near-variants are commonly used in board game tournaments, such as chess or Magic: The Gathering.

The idea behind a swiss system is to play a fixed number of rounds, and in each round have each matchup be between teams with the same record. This gives every team a bunch of games, while ensuring that teams are paired with teams that are probably similarly skilled. We can formally describe a swiss system in the language of multibrackets.

Definition 3.4.1: Swiss System

A *swiss system* is a weakly respectful standard multibracket signature in which

1. Each matchup is between teams of the same record,
2. All teams play the same number of games,
3. All teams start in the primary semibracket, and
4. The multibracket ranks every team.

The first two requirements come from the intuitive notion of a swiss systems, while the last two requirements are technical detail that ensures we don't double count systems: for example once where teams that finish with no wins drop into a trivial last semibracket, and one where they don't drop into any semibracket at all.

Definition 3.4.2: r-Round Swiss

We say a swiss system is an *r-round swiss system* if each team plays r games.

We begin our analysis by noting a key structural fact about Swiss systems.

Theorem 3.4.3

All r-round swiss systems are on $m \cdot 2^r$ teams for some m. Further, m divides the rank of each its semibrackets.

Proof. In order for a multibracket to be swiss system, its primary semibracket must be balanced. (Otherwise, teams that get byes will play fewer games than teams that don't.) Additionally, since each winner of the primary semibracket will play all of their games in that bracket, it must be exactly r rounds long. A balanced semibracket that is r rounds long has signature $[[\mathbf{m \cdot 2^r}; \mathbf{0}; ...; \mathbf{0}]]$ for some m. Thus, since every team starts in the primary semibracket, there must be $m \cdot 2^r$ teams participating for some m.

To prove the second half of the theorem, we will show by induction on s that after each team has played s games, the number of teams in each semibracket is divisible by $m \cdot 2^{r-s}$. Then the case of $r = s$ shows that the number of teams that win each semibracket is divisible by m, and so m divides the rank of each semibracket.

For $s = 0$, all $m \cdot 2^r$ teams are in the primary semibracket and no teams are in any of the others, so the statement holds. Now assume the statement holds for $s - 1$. Let t_i be the number of teams in the ith semibracket after each team has played $s - 1$ games. By induction $m \cdot 2^{r-s+1}$ divides t_i for all i. After each team plays their sth game, the ith semibracket contains $t_i/2$ teams that just won, and $t_i/2$ teams just lost and are dropped into another semibracket. Thus, each semibracket now has $\sum_{i \in S} t_i/2$ teams in it, for some set S. However, each $t_i/2$ is divisible by $m \cdot 2^{r-s}$, so the inductive case holds.

Therefore by induction, m divides the order of each its semibrackets. $\square$

Theorem 3.4.3 indicates that r-round swiss systems on $m \cdot 2^r$ teams are, in a sense, actually m different simultaneous and identical swiss tournaments each operating on 2^r teams. Because of this, it is useful to study just the swiss systems that operate on 2^r teams, as this will give us strong insights into the full space of swiss systems.

Definition 3.4.4: Compact

We say a swiss system is *compact* if its primary semibracket has order one.

Theorem 3.4.3 guarantees that r-round compact swiss systems are on exactly 2^r teams.

We omit the proofs until the next section, but there are unique compact 0, 1, and 2-round swiss systems, and exactly two 3-round ones. To give a more concrete sense of what exactly a swiss system looks like, we exhibit the signatures a sample brackets of those five formats below.

Figure 3.4.5: The Compact r-Round Swiss Systems for $r \leq 3$

Name	Rounds	Signature
$\mathcal{S}_0$	0	$[[\mathbf{1}]]$
$\mathcal{S}_1$	1	$[[\mathbf{2};\mathbf{0}]] \to [[\mathbf{1}]]$
$\mathcal{S}_2$	2	$[[\mathbf{4};\mathbf{0};\mathbf{0}]] \to [[\mathbf{1}]] \to [[\mathbf{2};\mathbf{0}]] \to [[\mathbf{1}]]$
$\mathcal{S}_3$	3	$[[\mathbf{8};\mathbf{0};\mathbf{0};\mathbf{0}]] \to [[\mathbf{1}]] \to [[\mathbf{2};\mathbf{0}]] \to [[\mathbf{1}]] \to [[\mathbf{4};\mathbf{0};\mathbf{0}]] \to [[\mathbf{1}]] \to [[\mathbf{2};\mathbf{0}]] \to [[\mathbf{1}]]$
$\mathcal{T}_3$	3	$[[\mathbf{8};\mathbf{0};\mathbf{0};\mathbf{0}]] \to [[\mathbf{1}]] \to [[\mathbf{4};\mathbf{2};\mathbf{0}]]_2 \to [[\mathbf{2}]]_2 \to [[\mathbf{2};\mathbf{0}]] \to [[\mathbf{1}]]$

Figure 3.4.6: The Unique Compact 0-, 1-, and 2-Round Swiss Systems

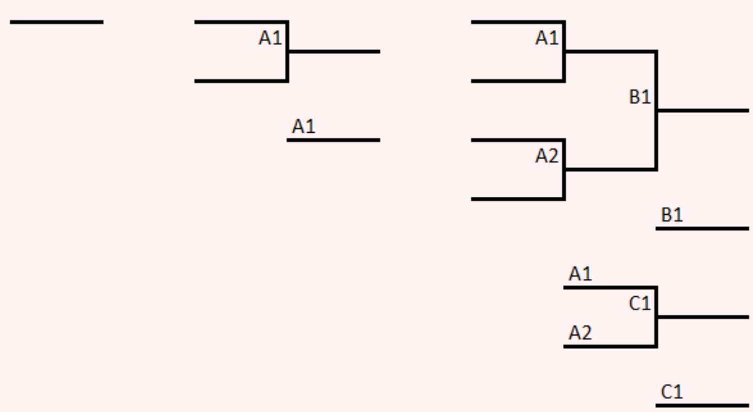
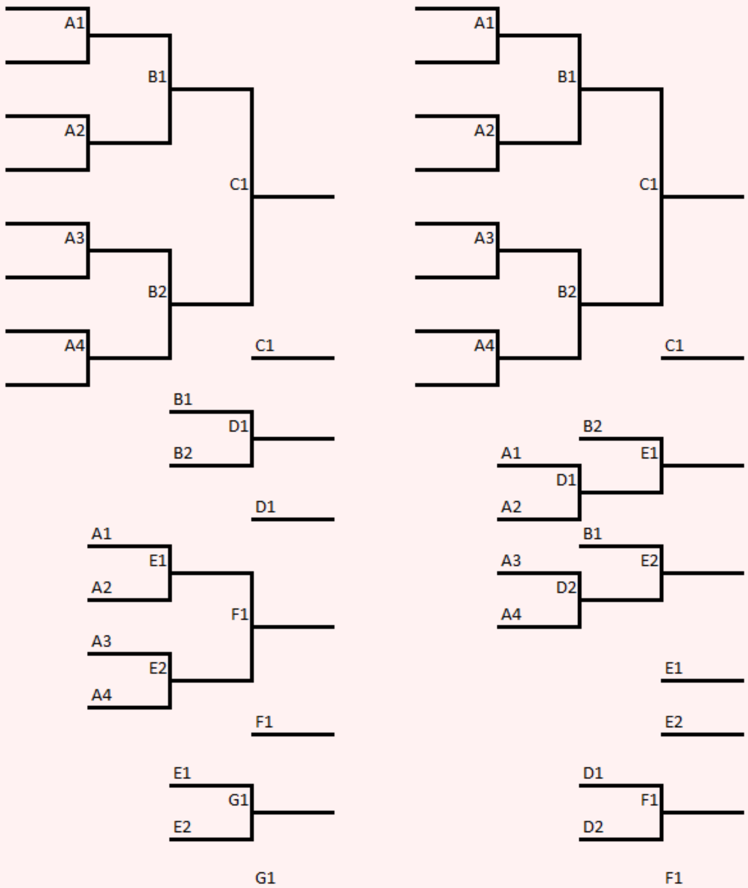


Figure 3.4.7: The Two Compact 3-Round Swiss Systems



We will prove that these are the only compact r-round systems for $r \leq 3$, discuss the advantages and disadvantages of $\mathcal{S}_3$ and $\mathcal{T}_3$, an generalize to $r > 3$ in the next section, but before then, we prove a few key structural results about general swiss systems.

Theorem 3.4.8

After each team in an r-round compact swiss system has played s games, for each i, $2^{r-s} \cdot \binom{s}{i}$ teams will have i wins.

Proof. We show this by induction on s. For $s = 0$, no games have yet been played, so all $2^r = 2^{r-0} \cdot \binom{0}{0}$ teams have no wins.

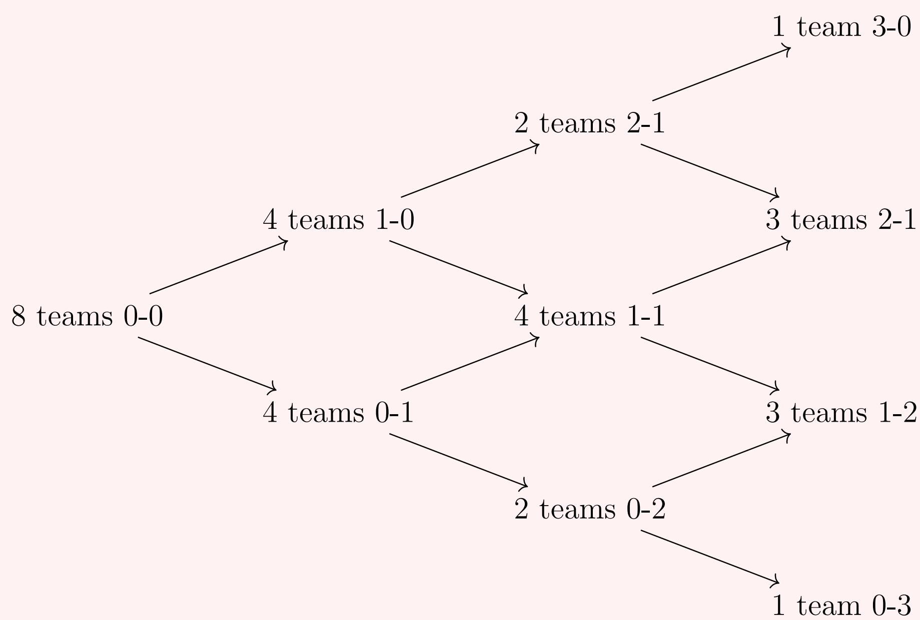
Assume the theorem holds for $s - 1$ and fix i. A team that after s games has i wins will, after $s-1$ games, have had either i or $i-1$ wins. In fact, half of the teams with i wins after $s - 1$ games will have lost and still have i wins, and half of the teams that had $i - 1$ wins will have won and now have i wins. Thus, the number of teams that have i wins after s games is

$$\begin{aligned}
&\frac{1}{2}\left(2^{r-s+1} \cdot \binom{s-1}{i} + 2^{r-s+1} \cdot \binom{s-1}{i-1}\right) \\
&= 2^{r-s}\left(\binom{s-1}{i} + \binom{s-1}{i-1}\right) \\
&= 2^{r-s} \cdot \binom{s}{i}.
\end{aligned}$$

$\square$

Figure 3.4.9 visualizes how Theorem 3.4.8 applies to compact 3-round swiss systems.

Figure 3.4.9: Theorem 3.4.8 with $r = 3$.



Theorem 3.4.10

The loser of the championship game of a compact swiss system will come in second.

Proof. The team that loses the championship a compact swiss system will have played all r of their games, so they must fall into a trivial bracket, and by the definition of a multibracket, it must be the second semibracket. Finally, because every other team that lost in the primary bracket has more games to play, they cannot have fallen into the second bracket. Thus the loser of the championship game of a compact swiss system will come in second. $\square$

Theorem 3.4.11

If team t_1 finishes ahead of t_2 in a swiss system, then there exists some round s such that after round s, t_1 has more wins than t_2.

Proof. We will show that for all s, if after every round $u < s$, t_2 had at least as many wins as t_1, then in round s, t_2 will be playing in as least as high of a semibracket as t_1. The theorem follows.

We show this by induction on s. For $s = 1$, every team has zero wins and is playing the same multibracket. Now assume the statement holds for s and that after every round $u < s + 1$, t_2 had at least as many wins as t_1.

By the inductive hypothesis, in round s, t_2 played at least as high a semibracket as t_1. Now because t_2 had at least as many wins as t_1 after both round $s - 1$ and round s, there are three things that could have happened in round s:

1. Both t_1 and t_2 won, in which case they each remain in the same semibracket as before, and so t_2 is still playing in at least as high of a semibracket as t_1 is.
2. t_1 lost and t_2 won. t_2 remains in their semibracket, while t_1 drops down, and so t_2 is in a higher semibracket as t_1.
3. Both t_1 and t_2 lost. If they were in the same semibracket in round s, then they will still be in the same semibracket in round $s + 1$ by multibracket rule (3). If t_2 was in a higher semibracket in round s, then t_2 will be in at least as high of a semibracket in round $s + 1$ by multibracket rule (4).

Thus, for all s, if after every round $u < s$, t_2 had at least as many wins as t_1, then in round s, t_2 will be playing in as least as high of a semibracket as t_1.

$\square$

4 Round Robins

4.1 Tiebreakers

Definition 4.1.1: Round Robin

A *round robin* is a tournament format in which each team plays each other team once, and then teams are ranked according to how many games they won.

Round robins, or close variants, are used in many leagues across many sports, especially during the regular season or qualifying rounds. For example, the 2014 Ivy League Football Regular Season was structured as round robin. At the conclusion of a round robin, a league table can be used to display the results and rank the teams.

Figure 4.1.2: 2014 Ivy League Football Regular Season

Rank	Team	Games	Wins	Losses
1	Harvard	7	7	0
2	Dartmouth	7	6	1
3	Yale	7	5	2
4	Princeton	7	4	3
5	Brown	7	3	4
6	Penn	7	2	5
7	Cornell	7	1	6
8	Columbia	7	0	7

At the end of an n-team round robin, each team has played each other team once, for a total of $n-1$ games. There are n possible records a team could have after playing $n-1$ games, so it is possible for each team to end the tournament with a different record: the 2014 Ivy League Football Regular Season has this property.

However, this is far from guaranteed: consider the 2019 Big 12 Football Regular Season (strangely enough, in 2019 the Big 12 had only ten teams).

Figure 4.1.3: 2019 Big 12 Football Regular Season

Rank	Team	Games	Wins	Losses
1	Oklahoma	9	8	1
2	Baylor	9	8	1
3	Texas	9	5	4
4	Oklahoma State	9	5	4
5	Kansas State	9	5	4
6	Iowa State	9	5	4
7	West Virginia	9	3	6
8	TCU	9	3	6
9	Texas Tech	9	2	7
10	Kansas	9	1	8

Nearly every team, including the two leaders Oklahoma and Baylor, ended the season tied with at least one other team. To rank teams with the same record, a *tiebreaking algorithm* is used.

Definition 4.1.4: Tiebreaking Algorithm

A *tiebreaking algorithm* is an algorithm for ranking teams that finish with the same record at the conclusion of a round robin.

Since every team in the 2014 Ivy League Football Regular Season ended with a different record, the tiebreaking algorithm wasn't employed: no matter what algorithm the Ivy League had prescribed, the ranking would have been the same. Not so for the 2019 Big 12 Football Regular Season, where many teams ended the season with identical records: different tiebreaking algorithms might have resulted in different rankings of the tied teams. (Though of course, Oklahoma and Baylor would always be first and second in some order, Texas and the three States would always be third through sixth in some order, etc.)

Definition 4.1.5: Tiebreaker

A *tiebreaker* is a single statistic that can be used to compare teams that finish with the same record at the conclusion of a round robin. Tiebreakers need not be able to successfully generate an order for any given set of tied teams.

Most tiebreaking algorithms are composed of a sequence of individual *tiebreakers*. These tiebreaker are applied one-by-one: if the first tiebreaker successfully breaks the tie, then the algorithm is complete. Otherwise, we proceed to the next tiebreaker.

Although individual tiebreakers are not required to be able to break all possible ties, the tiebreaking algorithm is. Thus, that last tiebreaker (and only the last tiebreaker) in a tiebreaking algorithm must be *terminal*.

Definition 4.1.6: Terminal Tiebreaker

A tiebreaker is *terminal* if it is guaranteed to generate an order for a set of tied teams.

Definition 4.1.7: `PointsScored`

The `PointsScored` tiebreaker ranks teams by how many points they scored over the course of the round robin: the more points, the better.

`PointsScored` is not terminal: any number of tied teams might have scored the same number of points over the course of the round robin and thus would remain tied.

Definition 4.1.8: `Random`

The `Random` tiebreaker ranks teams randomly: with each ordering being equally likely.

`Random` is terminal, and in fact, `Random` is used as the last tiebreaker in the tiebreaking algorithm of many leagues.

On the other hand, many leagues' first tiebreaker is `HeadToHead`.

Definition 4.1.9: `HeadToHead`

The `HeadToHead` tiebreaker ranks teams by their record against the other tied teams: the more wins, the better.

Theorem 4.1.10

`HeadToHead` is terminal as a two-team tiebreaker, but not terminal for more than two teams.

Proof. If only two teams are tied, then their record against each other must be 1-0 and 0-1, in some order, and so `HeadToHead` will successfully break the tie.

If n teams are tied for $n \geq 3$, then let t_1, t_2, and t_3 be three of the tied teams, and consider the situation where t_1 beat t_2, t_2 beat t_3, t_3 beat t_1, and all three of t_1, t_2, and t_3 beat every other tied team. Then each of t_1, t_2, and t_3 will have $(n-2)$ wins and one loss against tied teams, so `HeadToHead` cannot break their tie and thus is not terminal. $\square$

4.2 Faithfulness

As we saw in the last section, round robins can vary in their tiebreaking algorithms, and there are a number of different tiebreakers with various properties to chose from. But what can we say about the results of an n-team round robin without knowing what the tiebreaking algorithm is? For one thing,

Theorem 4.2.1

If a team goes undefeated in a round robin, they will come in first.

Proof. The team that goes undefeated will finish with $n-1$ wins, while every other team lost to the undefeated team and so can finish with at most $n-2$ wins. $\square$

But this is not a property particularly unique to round robins: it is also true for every bracket and many multibrackets. We will now strengthen the claim in Theorem 4.2.1 in two different ways, each of which that are indeed somewhat unique to round robins.

We begin with the first strengthening.

Theorem 4.2.2

If round robin on a set of teams $\mathcal{T}$ ends such that for some subset of teams $\mathcal{S} \subset \mathcal{T}$, for every team $s \in S$ and $t \in \mathcal{T} \setminus \mathcal{S}$, s beat t, then every team in $\mathcal{S}$ will finish in the top-$|\mathcal{S}|$.

Proof. Let $n = |\mathcal{T}|$ and $m = |\mathcal{S}|$. Each team in $\mathcal{S}$ will beat every team in $\mathcal{T} \setminus \mathcal{S}$, and so will finish with at least $n-m$ wins. Each team in $\mathcal{T} \setminus \mathcal{S}$ will lose to every team in $\mathcal{S}$ and so finish with at least m losses and thus at most $n-m-1$ wins. Thus every team in $\mathcal{S}$ will finish ahead of every team in $\mathcal{T} \setminus \mathcal{S}$, and thus in the top-$m$. $\square$

In the case where $|S| = 1$, this reduces to Theorem 4.2.1. Theorem 4.2.2 is a nice result, ensuring that if the competing teams can be cleanly divided into a "better group" and "worse group," the members of the better group of teams will finish ahead of the teams in the worse group. However, for a team to be able to use this theorem to guarantee themselves a spot in the top-m for some m, they would need to rely on the results of many games that they did participate in (except in the trivial case where $|S| = 1$.) The second strengthening of Theorem 4.2.1 allows a team to guarantee themselves a spot in the top m for various m based only on their own performance.

Theorem 4.2.3

If a team finishes a round robin with ℓ losses, they are guaranteed to finish in the top $2\ell + 1$.

Proof. Assume for contradication that team $t \in \mathcal{T}$ finishes with ℓ losses but is not in the top $2\ell + 1$. Then there are at least $2\ell + 1$ teams that finish ahead of t, each of which who must have ℓ losses or fewer. Let $\mathcal{S}$ be such a set of $2\ell + 1$ teams. Then the set of teams $\mathcal{S} \cup \{t\}$ has a combined $(2\ell + 2) \cdot \ell = 2\ell^2 + 2\ell$ losses or fewer. However, there are

$$\binom{2\ell + 2}{2} = \frac{(2\ell + 2)(2\ell + 1)}{2} = 2\ell^2 + 3\ell + 1$$

games played between them, and each one must end in a loss. Contradiction! $\square$

Once again, for $\ell = 0$, this reduces to Theorem 4.2.1.

We can think of Theorems 4.2.2 and 4.2.3 as guarantees that round robins make to competing teams: if you are a member of an m-team set that beats every other team, or if you lose no more than $(m - 1)/2$ games, then you are guaranteed a spot in the top-m. These are important guarantees, and we generalize them onto formats where not every team plays every other team, (or where two teams might play more than once,) with the notion of faithfulness.

Definition 4.2.4: Dominating Set

We say a subset of teams $\mathcal{S} \subset \mathcal{T}$ is a *dominating set* if for every $s \in \mathcal{S}$ and $t \in \mathcal{T} \setminus \mathcal{S}$,

$$\mathbb{P}[s \text{ beats } t] = 1.$$

Definition 4.2.5: Faithful

We say an n-team tournament format $\mathcal{A}$ is *faithful to its top-m* if, for any set of teams $\mathcal{T}$ the following two conditions hold:

- If $\mathcal{S} \subset \mathcal{T}$ is a dominating set and $|\mathcal{S}| \leq m$ then
$$\forall s \in \mathcal{S}\ \mathbb{W}^m_{\mathcal{A}}(s, \mathcal{T}) = 1.$$
- If there is a team $t \in \mathcal{T}$ and a set of teams $\mathcal{S} \subset (\mathcal{T} \setminus \{t\})$ such that
$$|\mathcal{T} \setminus \{t\}| - |\mathcal{S}| \leq (m-1)/2$$
and
$$\forall s \in \mathcal{S}\ \mathbb{P}[t \text{ beats } s] = 1,$$
then
$$\mathbb{W}^m_{\mathcal{A}}(t, \mathcal{T}) = 1.$$

Theorems 4.2.2 and 4.2.3 imply that

Corollary 4.2.6

For all $m < n$, every n-team round robin is faithful to its top-m.

Additionally, since every team ends up in the top-n no matter what,

Theorem 4.2.7

Every n-team format is faithful to its top-n.

Faithfulness is an important property for formats that aim to select a top-m, as it protects teams from getting unfairly unlucky with whom they are matched up against. For example, if a format is faithful to its top-three, teams know that one bad matchup will not eliminate their chance of medaling, even if they get an unlucky draw, as long as they take care of business against the rest of their opponents.

Many multibrackets (and formats in general) are faithful to their top-1.

Theorem 4.2.8

A multibracket is faithful to its top-1 if and only if every competing team starts in its primary primary semibracket and the primary semibracket ranks one team.

Proof. If there is a team that does not start in the primary semibracket, then even if that team always beats every other team, they will not come in first. And if the primary semibracket ranks more than one team, then no team will come in first. But if the primary semibracket has rank one and if every team starts in it, then any team that always beats every other team is guarenteed to win that semibracket and come in first. □

In general, however, multibrackets struggle to be faithful to their top-m for $1 < m < n$. For example,

Theorem 4.2.9

A compact r-round swiss systems is faithful to its top-m if and only if:

- $m = 1$,
- $m = 2^r$, or
- $r = 2$ and $m = 3$.

Proof. We note that the first two cases are covered by Theorems 4.2.8 and 4.2.7 respectively. Left is to show the specific third case as well as to prove that these are the only cases.

Consider first the case when $r \geq 2$ and $m = 2$. Let $\{t_1, t_2\} \subset \mathcal{T}$ be a dominating set of teams. By Theorem 3.4.10, the team that comes in second-place in a compact swiss system is the loser of the championship game. Thus if t_1 and t_2 are matched up against each other in the first round, one of them will not finish in the top-two, so for $r \geq 2$, r-round compact swiss systems are not faithful to their top two.

Now fix r and m, let

$$\ell = \min(\lceil m/2 \rceil - 1, r),$$

and consider a team t that loses its first ℓ games and then wins its last $(r - \ell)$. By Theorem 3.4.11, such a team cannot finish ahead of any team that finishes with the same record or better, and by Theorem 3.4.8, there are

$$k = \sum_{i=0}^{\ell} \binom{r}{i}$$

such teams (including team t). Team t can finish at best in kth place, despite losing only ℓ games, so if $k > m$, then compact r-round swiss systems are not faithful to their top-m. This inequality holds when $r = 3$ and $m \in \{3, 5, 6, 7\}$ or when $r > 3$ and $2 < m < 2^r$.

We are left with two uncovered cases, when $r = 2$ and $m = 3$, and when $r = 3$ and $m = 4$.

Recall that there is only one compact 2-round swiss system:

$$[[\mathbf{4}; \mathbf{0}; \mathbf{0}]] \to [[\mathbf{1}]] \to [[\mathbf{2}; \mathbf{0}]] \to [[\mathbf{1}]].$$

Any three-team dominating set will necessarily force the fourth-team into fourth place, securing a top-three finish for the teams in the set, and any team with one bad matchup will win one of their two games, again guaranteeing themselves a spot in the top three. Thus, 2-round swiss systems are faithful to their top-3.

Finally, in the case when $r = 3$ and $m = 4$, consider a dominating set of size four, such that teams in the set are matched up against each other in first round and the losers of those two games are matched up against each other in the second round. By Theorems 3.4.11 and 3.4.8, the loser of that second round game will finish in seventh place at best, so 3-round swiss systems are not faithful to their top-4. $\square$

Round robin's faithfulness is very impressive: many multibrackets only faithful to their top-m for a few values of m, but round robins are simultaneously faithful top their top-m for all m. In some sense, this is because the definition of faithfulness is designed to express to what extent other formats are able to successfully simulate round robin results, but it is still a very powerful contract between round robins and the teams competing in them.

Round robins are able to honor this contract because they call for so many games: in an n-team round robin, $O(n^2)$ games. Compare this to compact swiss systems, which call for $O(n \log n)$ games, or brackets, which take only $O(n)$. In fact, we conjecture that asymptotically, round robins are the best we can do.

Conjecture 4.2.10

The fastest n-team format that is faithful to its top-m for all m takes $O(n^2)$ games to complete.

However not all hope of faithful yet relatively fast formats is lost. In practice, rather than needing to be faithful to the top-m for any m, there is often a particular m, or list $(m_1, ..., m_k)$, that we want to be faithful to. In the next section, we will combine round robins and multibrackets into a larger format that is faithful to its top-m for particular m that we might care about.

4.3 Pools

Consider the 2023 FIFA Women's World Cup. 32 teams qualified for the World Cup, and they were divided into eight pools of four teams each. Each pool played out a self-contained round robin. Then, the bottom two teams in each pool were eliminated, and the remaining sixteen teams (two from each pool) played out the multibracket $[[\mathbf{16}; \mathbf{0}; \mathbf{0}; \mathbf{0}; \mathbf{0}]] \to [[\mathbf{1}]] \to [[\mathbf{2}; \mathbf{0}]]$. The entire format as it played out is displayed in Figure 4.3.1.

Figure 4.3.1: The 2023 FIFA Women's World Cup

Note that the seeds in the primary semibracket $[[\mathbf{16}; \mathbf{0}; \mathbf{0}; \mathbf{0}; \mathbf{0}]]$ are not just the numbers 1 through 16, but instead letter-number combination. **B2** indicates that the second-place team from pool **B** should be placed in that location. To avoid confusion, we start at I when labeling the rounds. In alignment with the discussion in Section XYZ, the 2023 FIFA Women's World Cup attempts to avoid rematches, ensuring that the two teams that qualify to bracket from each pool can only meet in the finals.

The format in Figure 4.3.1 combines the two genres of format we've discussed thus far: the first half is a (bunch of separate) round robins, and the second half is a multibracket. How teams perform in the round robin phase of the format affects how they are seeded (and whether they qualify at all) in the second half. We call such formats *pool to bracket formats.*

Definition 4.3.2: Pool to Bracket

A *pool to bracket* is a tournament format in which teams are divided amongst a number of pools, each pool plays a round robin, and then the results of the round robins are used to place (some) of the teams into a multibracket which is then played out to get the final placements.

Pool to bracket formats make use to the strengths of both round robins and multibrackets in order to cover for each others weaknesses. Round robins are maximally faithful, ensuring that unlucky draws don't screw teams out of top-ms for any m, but they take $O(n^2)$ games to complete, and they aren't very exciting to watch: there is no championship game, and often times games are played in which one or both teams have already been eliminated from placing (or have already secured a particular finish). Meanwhile, multibrackets can be completed in $O(n \log n)$ games or even $O(n)$ games, and are often incredible exciting, with every game in an (efficient) multibracket being played for stakes, but struggle to be faithful: a poor first-round matchup can significantly hinder an otherwise deserving teams ability to finish highly.

Pool to bracket formats use ...

Because of this, pool to bracket formats are nearly ubiquitous across sporting leagues. Both the Men's and Women's FIFA World Cup use pool to bracket formats, as well as many olympic sports. Most club sport regional and national tournaments also use pool to bracket formats. Even many NCAA sports use a format very similar to pool to bracket: colleges are divided into conferences, play something resembling a round robin with the teams in their conferences, and then proceed to the post-season which is a bracket that mixes together teams from various conferences. Professional leagues such as the NFL can even be looked at through this lense as well.

5 References

[1] Robert Baumann, Victor Matheson, and Cara Howe. Anomalies in tournament design: The madness of march madness. *Journal of Quantitative Analysis in Sports*, 6(4), 2010.

[2] Robert Chen and F. K. Hwang. Stronger players win more balanced knockout tournaments. *Graphs and Combinatorics*, 4, 1988.

[3] F. R. K. Chung and F. K. Hwang. Do stronger players win more knockout tournaments? *Journal of the American Statistical Association*, 73(363), 1978.

[4] H.A. David. *The method of paired comparisons*, volume 12. London, 1963.

[5] Christopher Edwards. *The Combinatorial Theory of Single-Elimination Tournaments*. PhD thesis, Montana State University, 1991.

[6] Jeff Horen and Raymond Riezman. Comparing draws for single elimination tournaments. *Operations Research*, 33(2), 1985.

[7] F. K. Hwang. New concepts in seeding knockout tournaments. *The American Mathematical Monthly*, 89(4), 1982.

[8] Robert B. Israel. Stronger players need not win more knockout tournaments. *Journal of the American Statistical Association*, 76(376), 1981.

[9] Allen J. Schwenk. What is the correct way to seed a knockout tournament? *The American Mathematical Monthly*, 107(2), 2000.

[10] Nate Silver. When 15th is better than 8th: The math shows the bracket is backward, 2011.

[11] Eric Simon. The upa manual of championship series tournament formats, 2008.