# How much math do you need to win your March Madness pool?

March 16 2016, by Bryan Clair, Saint Louis University



The University of Dayton Arena, where March Madness will kick off again this year. Credit: Greenstrat

Deciding which teams to pick in your NCAA basketball pool? Then you're faced with a classic decision problem – and here, science can help.

On one hand, you want to pick good teams, the "favorites," because those teams seem more likely to win. On the other hand, you want to

pick some weaker teams, the "underdogs," so your bracket will stand out from the rest and win the pool. These two opposing forces make for an interesting math problem, because somewhere in the middle is an optimal solution.

In my heart, I always know which teams will win, or at least which teams I want to root for. As an academic, though, I'd rather squeeze all the fun out of it by overanalyzing the situation. Let's do that here!

## Estimating the likelihood of winning

To find the best way to build our own brackets, we need to first build a mathematical model for simulating the [tournament](tournament).

Suppose we model the tournament by replacing basketball games with coin flips, except with coins that don't land evenly heads or tails but rather are weighted to reflect each game's actual odds. For example, when Baylor plays Yale on Thursday, instead of playing the game, we just flip a coin that gives the higher-seeded Baylor a greater chance of winning. We'd need to flip one of these coins for every first-round game, every potential second-round game, and for each possible matchup in the tournament. Each coin must be weighted in a way that models the actual game, so its probabilities must be determined by the specific matchup.
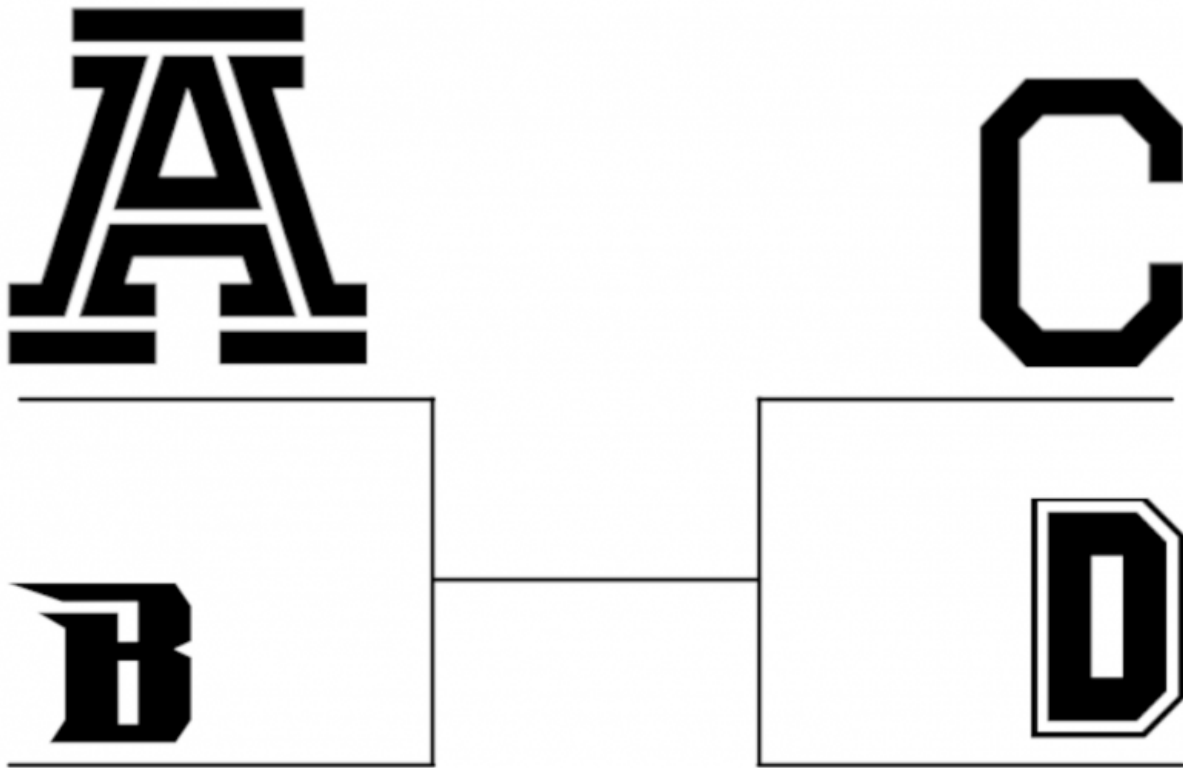
An example bracket for a four-team tournament.

Where should we get these probabilities? The NCAA provides you with a handy little number next to each team, the team's seed. For the first few rounds, each game has a favorite, and that choice was made by people with a tremendous amount of basketball knowledge. You could look back over history and observe that when a #5 seed plays a #12 seed, the #5 seed wins 65 percent of the time.

But there are plenty of other methods: Las Vegas betting odds give a point spread for each game, and based on those teams' scoring averages, you can convert the point spread into a probability of winning. Computer rating systems abound, and you can convert these ratings into probabilities by considering the ratings difference between two teams – a

method known as the [Bradley-Terry model](#). Some more sophisticated systems can even produce a probability custom fit to the two teams in the game.

## Know the odds? It's still not easy

So, pick your favorite method. Even then, things aren't as simple as they seem. The most likely outcome of the tournament is not necessarily that all favorites win. Look at this example:

Imagine a four-team tournament with teams A, B, C and D as shown. Assume that A always beats B, and C beats D with probability 0.6. Finally, A always beats D, but has only 0.5 probability of beating C. The only possible outcomes are: A wins over C (probability 0.3), C wins over A (probability 0.3) and A wins over D (probability 0.4). The most likely outcome contains the upset D beats C.

Further complicating the situation, the rules of your office or friends' pool probably mean that picking correctly in later-round games earns more points than early-round picks. How do you pick a bracket that gets you those crucial late-round points?

In one of the first analytic papers on this subject, [Kaplan and Garstka](#) gave an algorithm for deciding which picks are expected to score the highest. Their method builds a list of 64 brackets backwards, round-by-round, starting each one with a different team as the winner. For example, Duke's bracket starts with just Duke, and adds one round at a time, doubling in size but always keeping Duke as the winner. In the end, the algorithm selects the best from each of the 64 team-specific brackets.

You won't win this trophy, but math can help you you pick correctly the team that does. Credit: randomduck, CC BY-SA

This doesn't sound like something a human would do, and in fact it is best implemented by a computer. The brackets produced tend to be "chalk" – in which higher-ranked teams are most likely to win – but do not always select the higher seed. And Kaplan and Garstka did observe that their algorithm did better than just automatically picking the high seeds.

## It's about winning, not just scoring

To this point our model is ignoring an important fact: the goal of picking your bracket is not to achieve a high score, but to win a pool against other people. And people behave irrationally.

In a psychological experiment, McCrea and Hirt found evidence that pool participants pursue "probability matching": if a collection of games (say, the 5-12 matchups) has historically produced an upset one-third of the time, people will attempt to predict upsets in about one-third of those games in their brackets. In fact, people do no better than random chance at making such predictions, and so hurt their overall chances in the pool.

On the other hand, when choosing the tournament winner, people flock to the favorites. Every year, ESPN Tournament Challenge publishes data on its 11 million entries. Last year, 48 percent of their players had selected prohibitive favorite Kentucky as champion. Picking the correct champion is important, but if everyone else has the same opinion then you need to pick a bunch of other games well, too.

This brings us back to what makes this problem interesting: you need to

pick teams that win, but not the same teams as everyone else – so you come out on top in your pool.

To improve your odds in your pool, you need to model the other players you're up against. Each year, large, free, Internet pools publish data on player behavior, and they publish it before your brackets are due on Thursday morning.

Let's assume people make their picks the same way we modeled the games, by flipping biased coins for each game in the bracket. The national Internet pools give exactly the data you need to properly bias the coins. Nobody I know actually picks their bracket this way, but it turns out that real (human-picked) brackets and randomized brackets have nearly the same score distribution.

## Playing the odds means a long, long wait

In my own research, we used this model to calculate optimal picks. The brackets produced tend to be very conservative in the first two rounds, include one or two surprises in the Final Four, and a strong but not heavily favored champion. They never, ever, pick an upset in a 5-12 game. According to the computers, these picks increase the chances of winning a big Internet pool by a factor of 100 to 1,000.

This sounds great. It is great! But there's a catch: the NCAA basketball tournament happens only once a year. And your probability of winning is very low indeed – even with a boost from math and computer analytics. It will likely take thousands of years before the strategy pays off.

And that's the beautiful thing about scientific studies of the NCAA tournament. Serious modeling and data analysis quail before the absurdity of predicting such a notoriously unpredictable event. After a decade of study, the only things we really know are that the tournament

is madness and that your friend whose picks are based on mascots will probably win your pool.

*This article was originally published on* [The Conversation](link)*. Read the* [original article](link)*.*

Source: The Conversation