

Fair Play

by
Edward A. Bender

How can we decide if a game with two competitors is fair? One simple answer is:

If the roles of the competitors are reversed,
their probability of winning does not change. (1)

Isn't that always true? No. For example, going first may give a player an advantage or disadvantage. Using (1) as our standard, we'll look at the problem of fairness for tennis and baseball.

1 Is a Tennis Match Fair?

A good tennis player is more likely to win the point when serving than when receiving. To counteract this, being the server switches back and forth between the players after each point.

Still, the person serving first apparently has an advantage because she has a better chance of winning that point. Is this true? That depends on how a winner is chosen. We'll look at three possibilities, the third being the way the winner of a game is actually chosen in tennis. Call the players First and Second to indicate when they serve.

1.1: Reach a Score One possible way to select a winner is to play until one player gets N points for some N . How can we see this? One way is to calculate the probability of winning for each player, but that could be complicated. Another way is to look at a particular case where computations are easy and show that whoever serves first has an advantage.

- Computations are easy if the chances of winning the point are 50:50, regardless of who serves. In this case, it's like flipping a coin until either heads or tails has appeared N times. Clearly the split is 50:50.
- We need a simple situation where the server has an advantage. Again, computations are easy if the server *always* wins the point. In this case, the players are tied after Second has served because both players have served the same number of times. On the other hand, First is ahead by one point after she serves because she will have served one more time than Second. Thus First always wins because she gets to N points first, namely, she gets her N th point just after both players have $N - 1$ points. This is not fair.

Of course, in an actual tennis game the true odds are somewhere between these two extremes. We can expect a smooth change from no unfairness at 50:50 odds of winning the point when serving to total unfairness at 100:0 odds.

1.2: Win Two Points in a Row Another way would be to play two points. If one player wins both, that player is the winner. If each player has won one point, start over. This is fair — the only way a player can win is by winning a point when serving and also when receiving.¹

1.3: Generalized Tennis Yet another way to play is to play until one player has at least N points and is ahead by at least two points. (When $N = 4$, we have the rule for a tennis game.) We claim it is fair if we ignore the first condition and just consider being ahead by two points. Look at the first two serves of the game. Either one player wins both points and so wins the game, or each player wins one point and the players are tied. If the players are tied, how many points they've won doesn't matter since the only way to win is to lead by two points. Thus, we're in the situation in the previous paragraph. What happens when we add back in the restriction that the winner must have at least N points? Before doing that, it will be useful to look at the game without the restriction in another way.

¹This is not completely true: First must win while serving and then when receiving whereas Second must win while receiving and then while serving. If there are psychological factors that enter because a player is ahead or behind by one point, it can matter who serves first. Players tiring could also be a factor. Can you see how who served first might matter in that situation? Anyway, we're ignoring such problems.

Suppose we do not have the “at least N points” restriction and that the winner has K points. Then the loser has $K - 2$. Hence there have been $2K - 2$ serves and so each player has had $K - 1$ serves. If the players are equally matched, then the probability that First has won K of these equals the probability that Second has won K .

Now suppose we require that the winner be ahead by $2p$ points. If the winner has K , the loser has $K - 2p$ and so there were $2K - 2p$ serves, each player had $K - p$ serves and we can reason as in the previous paragraph.

Let’s go back to the full rules (win at least N points and lead by at least two points) and add one more rule: The winner must lead by an even number of points. By the idea previous two paragraphs, this is fair.

Finally, we return to the full rules. We want to use the previous paragraph, but we can’t if the winner has N points and the loser has $N - t$ for some odd number t . We use a trick in this case. Since the winner must lead by at least two points and t is odd, we have $t \geq 3$. Here’s the trick: Play one more point. If the winner wins it, the scores are now $N + 1$ and $N - t$. If the loser wins it, the scores are now N and $N + 1 - t$. In either case, the winner is still ahead by at least two points and leads by an even number of points. Since the lead is an even number of points, this is fair. We showed:

- The game in which the winner must lead by an at least two points and have at least N points is equivalent to the game in which we add the rule that the winner must lead by an even number of points.
- The modified game is fair because both players served for exactly half the points.

Thus

tennis games are fair.

We’ve looked at a single game; however, we need to look at an entire match. A tennis match consists of several “sets”. Whoever wins the most sets is the winner. A set consists of games and the rules for winning a set are like the rules for a game. A player must win at least six games and lead by two; however, if both players have won six games, a tie-breaker game is played to determine the winner of the set. Do these detailed rules matter in determining fairness? No. Matches and sets are built out of games and we’ve

proved that games are fair. Therefore, regardless of the rules for winning a match,

tennis matches are fair.

2 Is the World Series Fair?

The baseball World Series consists of up to seven games. A team must win four games to win the series. Care is taken to make the series fair:

- Many carefully-chosen umpires are used.
- Being the first team at bat is rotated.
- The games are rotated between home fields as follows, where A stands for the American League team's home field and N stands for the National League team's home field.

one year: A A N N N A A next year: N N A A A N N.

Is this fair?

There may be psychological advantage or disadvantage to being first at bat. We won't explore this. Instead, we'll limit attention to the "home-field advantage." It is claimed that playing on one's home field gives a team an advantage, presumably due to familiarity with the field. Here are some natural questions to ask:

- Is there really a home-field advantage?
- If so, how does it affect the fairness of the World Series?
- If the World Series is unfair, how can we make it fair?

We'll look at them one by one.

Is There a Home-Field Advantage? It is frequently claimed that there is a home field advantage. How can we test this claim? But first, what does it mean? One interpretation is that batters do better in home games. Another

is that teams are more likely to win an home game. There are statistics indicating batters do better in home games, but we're interested in teams winning and I haven't seen a statistic on that. Nevertheless, it should be a simple matter to gather statistics and check it out. For each game played, it is either won by the home team or the visiting team. Let H be the number of games won by the home team and A the number won by the visiting team. If H is significantly larger than A , there is an at home advantage. Significance can be measured by computing the signage

$$\sigma = \frac{H - A}{\sqrt{H + A}}.$$

A σ of 1 or less means the data is unconvincing, a σ of 3 or more means the data is significant. The probability that a random team will win an at-home game can be estimated by $p = \frac{H}{H+A}$. If we already know that the team has some probability P of winning a game and half its games are at-home, then the probability it will win an at-home game is $2pP$.

Suppose there are two equally matched teams in the World Series and each team has probability p of winning a home game. The following table gives the probability that the team starting on its home field will win the series for various values of p .

p	0.50	0.52	0.54	0.56	0.58	0.60
win	0.500	0.506	0.513	0.519	0.525	0.532

It's clear that starting in one's home field gives a slight advantage in the World Series; however, even with a large home-field advantage, the advantage in the series is slight. Considering the many other random factors that enter, there is probably no point in trying to balance out the home field advantage. If one wanted to make such an adjustment, it could be done by insisting that the winning team must have played at least as many away games as home games; however, this could lead to a World Series that went on for many games.

3 The Chances of Winning in Tennis

Before proceeding, we need a little background in probability. Then we'll look at simplified tennis. Finally, we'll look at real tennis.

3.1 Probability

We could think of probability as the fraction of time we expect an event to occur. For example, if we toss a coin, we expect it to land heads half the time. Thus we say “the probability of heads is one-half.” In mathematical notation, $\Pr(\text{heads}) = 1/2$. In both cases, “heads” is an abbreviation for “when a fair coin is tossed, it lands heads.” Suppose we roll two fair dice. The probability that the sum is 6 is $5/36$. Why is this? There are 36 possible ways the two dice could land, all of which are equally likely. Of these, there are five of them with sum 6:

$$1 + 5, \quad 2 + 4, \quad 3 + 3, \quad 4 + 2 \quad \text{and} \quad 5 + 1.$$

What about a unique event? For example, what do we mean by

We think there is a 60% probability that Tiger Woods will win this tournament?

Since this tournament is only played once and Tiger will either win or lose, we can't mean that $3/5$ of the time he'll win this tournament. There are various interpretations of what this means. Here are two:

- We think that a bet based on his winning $3/5$ of the time is an even bet; that is, we'd be equally willing to take either side. In such a bet, the person favoring Tiger puts \$2 into the pot and the person betting he'll lose puts \$3 into the pot. The winner gets the pot.
- If we could create a lot of near clones of our world, Tiger Woods would win the tournament in about 60% of them.

We won't discuss the problem of what probability means any further.

There are three important rules about probabilities that we need to know.²

Sum of Two If A and B are possible events that can occur, but we cannot have both A and B , then the probability that A or B occurs is the sum of the two separate probabilities:

$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) \quad \text{when } A \text{ and } B \text{ cannot both occur.}$$

²In “probability theory” these rules are built into the definitions of “probability” and “independence.”

Thus, if A stands for rolling six with two dice, which we know has $\Pr(A) = \frac{5}{36}$, and B stands for rolling twelve with two dice, which you should show has $\Pr(B) = \frac{1}{36}$, then $\Pr(A \text{ or } B) = \frac{5}{36} + \frac{1}{36} = \frac{1}{6}$.

Sum of All Something must occur. Hence the probability that something occurs is 1. If we add up the probabilities of all the things that could occur, the result will be 1. For example, if we toss a coin and assume it cannot land on its edge, then $\Pr(\text{heads}) + \Pr(\text{tails}) = 1$.

Product of Two If two events are independent of each other, then the probability of both events occurring is the product of the separate probabilities:

$$\Pr(A \text{ and } B) = \Pr(A) \Pr(B) \quad \text{when } A \text{ and } B \text{ are independent.}$$

For example, the probability that I'll roll six and Tiger Woods will win the tournament is $\frac{5}{36}(60\%) = \frac{5}{36} \frac{3}{5} = \frac{1}{12}$.

This concludes our background on probability.

3.2 Simplified Tennis

We will assume that the first player has a probability p_1 of winning the point when he serves against the second player and a probability q_1 of losing it. Similarly, we assume the second player's probability of winning and losing when serving are p_2 and q_2 . Since a player must either win or lose a point,

$$p_1 + q_1 = 1 \quad \text{and} \quad p_2 + q_2 = 1. \quad (2)$$

We assume that the probability of winning a point is independent of who has won points in the past — it depends only on who is serving. Thus we are ignoring the effects of thoughts such as “If I don't win this point, I'll lose the game” and “I've won the past three games, so I'm on a winning streak.”

Let's do a couple of examples. The probability that the first player wins the first two points is p_1q_2 . The probability that each player wins one of the first two points (a tie) is $p_1p_2 + q_1q_2$ because either the first player wins a point and then the second player wins (probability p_1p_2) or the first player loses and then the second loses (probability q_1q_2). In other words

$$\Pr(\text{tie after 2 points}) = p_1p_2 + q_1q_2. \quad (3)$$

We could get this another way:

$$1 = \Pr(\text{tie}) + \Pr(\text{player 1 wins both}) + \Pr(\text{player 2 wins both}),$$

and so

$$\Pr(\text{tie}) = 1 - \Pr(\text{player 1 wins both}) - \Pr(\text{player 2 wins both}). \quad (4)$$

We started this paragraph by noting that

$$\Pr(\text{player 1 wins both}) = p_1q_2.$$

Doing the same for the second player and using this in (4), we have

$$\Pr(\text{tie after 2 points}) = 1 - p_1q_2 + q_1p_2. \quad (5)$$

You should use (2) to show that (3) and (5) give the same answers.

Now we'll get probabilities for the simplified game of tennis in which the first player to lead by two points wins. There will be some number (possibly zero) of ties after two points. Finally, either the first player wins by winning two points in a row or the second player wins two in a row. Let t be the probability of a tie as computed in (3) or (5). Then³

$$\begin{aligned} \Pr(\text{player 1 wins}) &= p_1q_2 + tp_1q_2 + ttp_1q_2 + ttt p_1q_2 + \cdots \\ &= (1 + t + t^2 + t^3 + \cdots)p_1q_2 \\ &= \frac{p_1q_2}{1 - t}, \end{aligned}$$

because $1 + t + t^2 + t^3 + \cdots$ is a geometric series and its sum is $\frac{1}{1-t}$. Similarly,

$$\Pr(\text{player 2 wins}) = \frac{q_1p_2}{1 - t}.$$

We can check our calculations as follows. Since some player must win, these numbers should sum to 1. Using (5),

$$\frac{p_1q_2}{1 - t} + \frac{q_1p_2}{1 - t} = \frac{p_1q_2 + q_1p_2}{1 - (1 - p_1q_2 + p_2q_1)} = 1.$$

³We sum over the events: no ties before winning, one tie before winning, etc. To compute, for example, the probability of two ties before winning, we have $\Pr(\text{tie}) \Pr(\text{tie}) \Pr(\text{win})$.

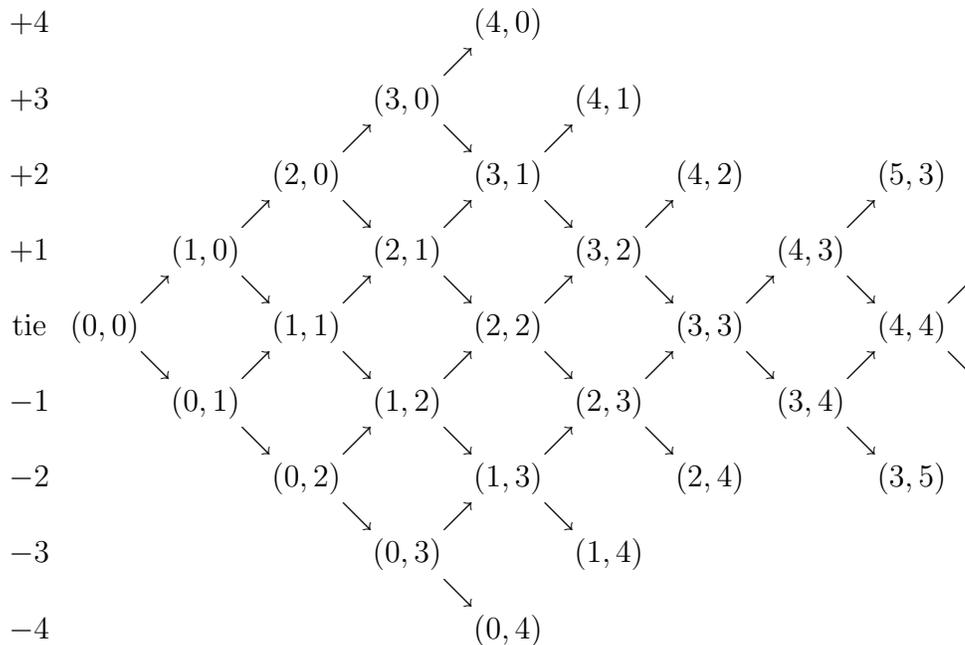


Figure 1: The possible points during the first part of a tennis game. A point is labeled (i, j) if the first player has won i points and the second player has won j . The left-hand column is the number of points by which the first player leads (or is behind, if negative) in that row. The symbol \nearrow (resp. \searrow) indicates a point won by the first (resp. second) player. A location with no arrow leading out is a win for one of the players.

To summarize,

$$\Pr(\text{player 1 wins}) = \frac{p_1 q_2}{p_1 p_2 + q_1 q_2} \tag{6}$$

$$\Pr(\text{player 2 wins}) = \frac{q_1 p_2}{p_1 p_2 + q_1 q_2}.$$

3.3 Winning a Tennis Game

We now turn to the rules for an actual tennis game: The winner must lead by at least two points and must have won at least four points. We'll keep the same notation (i.e., p_1, p_2, q_1, q_2) and assumptions as in the simplified case.

Figure 1 is a picture of the situation for the first eight points — if the

game lasts that long. Notice that if a game reaches six points without a win, then it is tied at three points each. After this tie, the game proceeds like a simplified tennis game. Let $r_{i,j}$ be the probability that the game reaches the point (i, j) in Figure 1. By adding up all winning events in Figure 1 with at most six points total and adding to it $r_{3,3}$ times the probability of winning the simplified game, we have the probability of winning. Using (6), this gives us

$$\begin{aligned} \Pr(\text{player 1 wins}) &= r_{4,0} + r_{4,1} + r_{4,2} + r_{3,3} \frac{p_1 q_2}{p_1 q_2 + p_2 q_1} \\ \Pr(\text{player 2 wins}) &= r_{0,4} + r_{1,4} + r_{2,4} + r_{3,3} \frac{q_1 p_2}{p_1 q_2 + p_2 q_1}. \end{aligned} \tag{7}$$

To use these equations, we need to compute the various $r_{i,j}$. One way is to simply list all the sequences of who wins which point that leads to those scores in Figure 1. It turns out that there are 20 ways to reach $(3, 3)$ and another 15 ways to reach the other three scores with $i > j$. Thus we need to list $20 + 15 + 15 = 50$ things and compute their probabilities.

There is an easier way. The idea is to work across Figure 1 from left to right, calculating probabilities of reaching score (i, j) as we go. In other words, going from one point played to two points played to three points played and so on. Let's do it, but first we'll make an observation that reduces work. Notice that the win-loss "routes" to (i, j) and (j, i) are closely related — just switch the winner and loser at each point along the way. What does this do to probabilities? It simply interchanges the symbols p and q , without changing their subscripts. Thus we have the *symmetry principle*:

$$r_{j,i} \text{ is } r_{i,j} \text{ with the symbols } p \text{ and } q \text{ interchanged.}$$

This also provides a check on the value of $r_{i,i}$.

Now we're ready. We'll list the results by number of points played.

1. $r_{1,0} = p_1$ and $r_{0,1} = q_1$ because the first player is serving.
2. Remember that now we are starting with one point having been played and moving to two points. To reach $(2, 0)$ we must first reach $(1, 0)$ and then the first player must win a point when the second player serves. Thus $r_{2,0} = r_{1,0} q_2 = p_1 q_2$. By our symmetry principle, $r_{0,2} = q_1 p_2$.

There are two distinct ways to reach $(1, 1)$, namely from $(1, 0)$ and from $(0, 1)$. Thus

$$r_{1,1} = r_{1,0}p_2 + r_{0,1}q_2 = p_1p_2 + q_1q_2$$

To summarize,

$$r_{2,0} = p_1q_2, \quad r_{0,2} = q_1p_2 \quad \text{and} \quad r_{1,1} = t, \quad \text{where} \quad t = p_1p_2 + q_1q_2.$$

Note that t is unchanged when we interchange p and q .

3. Working the way we did for two points, we get

$$r_{3,0} = r_{2,0}p_1 = p_1^2q_2, \quad r_{2,1} = r_{2,0}q_1 + r_{1,1}p_1 = p_1(q_1q_2 + t)$$

and $r_{0,3}$ and $r_{1,2}$ are found by symmetry.

4. Now

$$\begin{aligned} r_{4,0} &= r_{3,0}q_2 = (p_1q_2)^2, & (8) \\ r_{3,1} &= r_{3,0}p_2 + r_{2,1}q_2 = 2p_1q_2t, \\ r_{2,2} &= r_{2,1}p_2 + r_{1,2}q_2 = p_1p_2(q_1q_2 + t) + q_1q_2(p_1p_2 + t) \\ &= 2p_1p_2q_1q_2 + t^2. \end{aligned}$$

5. Since $(4, 0)$ is a winning score, it does not contribute to $(4, 1)$ and so we have

$$\begin{aligned} r_{4,1} &= r_{3,1}p_1 = 2p_1^2q_2t, & (9) \\ r_{3,2} &= r_{2,2}p_1 + r_{3,1}q_1 = p_1t^2 + 2p_1q_1q_2(t + p_1p_2). \end{aligned}$$

6. Now we have

$$r_{4,2} = r_{3,2}q_2 = p_1q_2t^2 + 2p_1q_1q_2^2(t + p_1p_2), \quad (10)$$

$$r_{3,3} = r_{3,2}p_2 + r_{2,3}q_2 = (t^2 + 6p_1p_2q_1q_2)t, \quad (11)$$

after some algebra.

	0.500	0.550	0.600	0.650	0.700
0.500	0.500	0.438	0.376	0.316	0.258
0.550	0.562	0.500	0.437	0.373	0.311
0.600	0.624	0.563	0.500	0.435	0.369
0.650	0.684	0.627	0.565	0.500	0.432
0.700	0.742	0.689	0.631	0.568	0.500

Table 1: The left column (resp. top row) gives probability of the first (resp. second) player winning the point when serving. The other entries are the probabilities of the first player winning the tennis game.

Finally, we can put (8) through (11) into (7):

$$\begin{aligned}
\Pr(\text{player 2 wins}) &= (p_1q_2)^2 + 2p_1^2q_2t \\
&\quad + (p_1q_2t^2 + 2p_1q_1q_2^2(t + p_1p_2)) \\
&\quad + (t^2 + 6p_1p_2q_1q_2)t \frac{p_1q_2}{p_1q_2 + p_2q_1}
\end{aligned} \tag{12}$$

Using (12), we computed the probabilities in Table 1.⁴ For comparison, Table 2 gives the probabilities of winning the simplified game which requires only that the winner lead by two points. Notice that the better player's chances of winning are slightly better for a true tennis game.

4 Winning a Tennis Set and Match

What about the probability of winning a set? Recall that a set is won by either winning at least six games and then leading by two games or, if both players have at least six games, leading by one game. At first, this may seem more involved because the winner must have at least six games as opposed to just four points. Actually, the calculations are simpler because a game is fair. Instead of p_1 and p_2 , we now have just p , the probability that the first player will win a game. The first player wins in one of three ways:

⁴The table provides a check on our formula: As explained in the table's caption the entry for row p_1 and column p_2 is the probability that first player wins. By fairness, the entry for row p_2 and column p_1 must be the probability that the second player wins. Since someone wins, the sum of these two probabilities should be one.

	0.500	0.550	0.600	0.650	0.700
0.500	0.500	0.450	0.400	0.350	0.300
0.550	0.550	0.500	0.449	0.397	0.344
0.600	0.600	0.551	0.500	0.447	0.391
0.650	0.650	0.603	0.553	0.500	0.443
0.700	0.700	0.656	0.609	0.557	0.500

Table 2: Similar to Table 1, but the entries now give the probability of winning a simplified game; that is, one where the winner must lead by two points.

- The final score is $(6, k)$ where $0 \leq k \leq 4$. To reach such a win, the last point must have been one by the first player. It turns out that the probability of such an event is

$$\Pr((6, k) \text{ win}) = \frac{(k+1)(k+2)\cdots(k+5)}{1 \cdot 2 \cdots 5} p^6 q^k \quad \text{where } q = 1 - p. \quad (13)$$

It would take us to far afield into simple counting problems to derive this result, so you'll just have to accept it.

- The final score is $(7, 5)$. To reach such a win, the score $(5, 5)$ must be reached and then the first player must win two points. (This is the case because the any other way to $(7, 5)$ would require the first player to have six points at a time when the second has less than five.) Thus $\Pr((7, 5)) = \Pr((5, 5))p^2$. It turns out (again, just accept it) that

$$\Pr((5, 5)) = \frac{6 \cdot 7 \cdots 10}{1 \cdot 2 \cdots 5} p^5 q^5. \quad (14)$$

- The final score is $(7, 6)$ The only way is to reach $(5, 5)$, then have each player win a point and then have the first player win a point. There are two ways for each player to win a point, so this event has probability $pq + qp = 2pq$. Putting it all together gives

$$\Pr((7, 6) \text{ win}) = \Pr((5, 5))(2pq)p \quad (15)$$

To get the probability of the first player winning, add up (13) for $0 \leq k \leq 4$, use (14) and (15) to add in $\Pr((7, 5) \text{ win})$ and $\Pr((7, 6) \text{ win})$. After some

	0.500	0.550	0.600	0.650	0.700
0.500	0.500	0.328	0.186	0.088	0.034
0.550	0.672	0.500	0.326	0.181	0.082
0.600	0.814	0.674	0.500	0.321	0.171
0.650	0.912	0.819	0.679	0.500	0.313
0.700	0.966	0.918	0.829	0.687	0.500

Table 3: Similar to Table 1, but the entries now give the probability of winning a tennis set.

algebra, it turns out that

$$\begin{aligned} \Pr(\text{win set}) = & p^6 \left(1 + 6q \left(1 + \frac{7q}{2} \left(1 + \frac{8q}{3} \left(1 + \frac{9q}{4} \right. \right. \right. \right. \\ & \left. \left. \left. \left. \times (1 + 2q(p + 2pq)) \right) \right) \right) \right) \right) \quad \text{where } q = 1 - p. \end{aligned} \tag{16}$$

We finally come to tennis matches. To win a match, all one needs to do is win a majority of the sets; however, the number of sets in a tennis match varies and may be one, three or five.⁵ To compute probabilities, one can list all possible ways to win, compute the probability of each way, and add them up. Of course, shortcuts are possible, but we won't go into them. For one set, simply use Table 3. For three or five sets, use the probability from Table 3 and interpolate as described in Table 4. For example, suppose Amy and Brenda are playing three sets. When they play each other, we expect Amy to win the point 65% of the time when serving and expect Brenda to win the point 60% of the time when she serves. What is the probability that Alice will win? By Table 1, the probability that Alice will win a game is 56.5%; however, what we need is the probability of her winning a set. By Table 3, this is 67.9%. Since this lies between 0.65 and 0.70, we interpolate between the two row entries of 0.718 and 0.784 to obtain a 75.7% probability of winning. Thus, although Alice and Brenda appear rather closely matched, Alice will win about 3/4 of the matches she plays against Brenda.

⁵There is no reason to have an even number of sets: If a player wins more than half of $2K$ sets, then she is ahead by at least two games and would therefore still be ahead if one more set were played. Hence they may just as well have played $2K + 1$ sets. If the players are tied, a tie-breaking set must be played and so $2K + 1$ sets must be played. Thus, the probability of winning is the same for $2K$ sets with tie breakers as it is for $2K + 1$ sets.

	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950
3	0.575	0.648	0.718	0.784	0.844	0.896	0.939	0.972	0.993
5	0.593	0.683	0.765	0.837	0.896	0.942	0.973	0.991	0.999

Table 4: Get the probability p of the better player winning a set from Table 3. Find the two columns that lie on either side of p , interpolate the corresponding entries in the 3-set or 5-set row, depending on the number of sets per match.