The Most Important Situations in Tennis – and in R&D Competition*

Avinash Dixit, Princeton University

and

Susan Skeath, Wellesley College

Most good tennis players have a special shot in their armory – an exceptionally fast or well-placed serve, or volley, or passing shot. It has a much higher probability of success than their usual play, but it cannot be deployed on a routine basis, because then their opponents will practice to counter it, and it will not succeed so well. It has to be deployed rarely, on the most important situations in a match. What constitutes an important situation? One where winning versus losing that point makes the biggest difference to one's probability of overall success in the match. But to calculate the difference, one must look further ahead to the situation that will ensue on winning and losing this point. That in turn depends on where those outcomes will lead, and so on. Therefore the correct calculation must perform all this look-ahead, and reason back to the current situation. In formal terms, backward induction or rollback analysis must be used. Here we offer such a sample calculation; readers can then adapt and apply it to their own context. The context does not have to be tennis, or any sport. Similar, and far more consequential, strategic decisions of this kind arise in industrial competitions, especially R&D races.

To win a game in tennis, you have to win four or more points (absolute achievement) with a margin of at least two over the opponent (relative achievement). The analogy with R&D is remarkably close. For example, a pharmaceutical company developing a new drug wants it to be effective (absolute achievement), and to be sufficiently better than other similar drugs in order to be successful in the market (achievement relative to that of rivals). Research is a lengthy process, and as each company proceeds through the stages of trying various compounds, patenting them, taking the promising ones through animal trials and then human trials, getting approval from the regulating authorities like the Food and Drug Administration, and so on, it not only thinks about its own progress but also watches how its competitors are doing. Its resources are limited and have

---

* This material was in the first edition of our introductory textbook on game theory, *Games of Strategy* (W.W. Norton, 1999). It turned out to be a somewhat complicated digression in that context, so we omitted it from subsequent editions. We are resurrecting it here as a fun example of probability calculations and backward induction for general readership.

other uses such as refinement and marketing of existing products. Its ability to expand its R&D division sufficiently fast is also limited. It has to decide when it should intensify the R&D efforts. The answer is: when the race with the other firm is at some especially crucial juncture. Identifying that is the same strategic problem as that of identifying important points in tennis.[†]

We analyze this well-known game, find results that accord well with intuition, and then translate the findings back to the context of business competition.

Tennis

Just in case you didn't know, "game" is a technical term in tennis. It consists of a sequence of "points." In each point, one player, the server, puts the ball in play, and the other player, the returner, hits it back. They go on hitting the ball back and forth until one player hits it out of the field of play or into the net; the other then wins the point.[‡] To win a single "game" of tennis, you have to win four or more points and you have to win by a margin of at least two. The scoring is peculiar: the first point in a game is called 15, the second 30, the third 40.[§] Thus if the server has won two points and the returner one point, the score is 30-15 in favor of the server. If the score reaches 40-40, the situation is called "deuce". Then the game continues until one player is ahead by two points. Being ahead by one point at this juncture is called having the "advantage", or simply "Ad". In principle a game could go on for ever, as the players switch back and forth between deuce, and advantage to the one or the other. But in practice all games played so far have ended in finite time.

We consider just one game. Actually, a game is part of a "set." In a set, the players take turns serving alternate games. You win a set by winning six or more games, with a margin of at least two games. If a set goes to 6 games each, a further game called the "tie-break" is played under somewhat different rules, and its winner wins the set 7-6. A set, in turn, is part of a complete

---

[†] For a more sophisticated mathematical model of dynamic R&D competition, see Christopher Harris and John Vickers, "Racing with Uncertainty," *Review of Economic Studies*, Volume 54, Issue 1, January 1987, pp. 1–21.

[‡] We are leaving out some complications; for example, a serve has to land in a limited part of the whole field for it to be legal, the server is allowed two attempts (serves) at putting the ball into play, and so on. These can be incorporated into the analysis, but the calculations are a little more involved.

[§] This is presumably because scores used to be kept on a clock-like dial by rotating a pointer like the minute-hand through a quarter circle for each point. Then 40 would be a misnomer for 45.

"match." Women play matches which are best-of-3-sets, men play best-of-5-sets in the major or "grand slam" tournaments and best-of-3 in most others. At Wimbledon and in the international Davis Cup matches, the last set does not have a tie-breaker, but must be played out until one player gets a two-game lead. The method we develop below for a game can easily be extended to analyze whole sets or matches using a computer; we leave that as an exercise for those readers who have the interest and the necessary programming ability.

We can show possible configurations that can arise in a tennis game using a two-dimensional grid as in Figure 1. (All figures are at the end of the paper.) The server's score is shown along the horizontal dimension (the x-axis) moving from left to right; the returner's score is along the vertical dimension (the y-axis) from the bottom to the top. Each crossing point of the grid represents a particular score: for example, the score at the bottom left corner is 0-0, that at the point three units to the right and two up from this corner is 40-30 to the server, and so on. The thick lines on the right side of the diagram represent points where the server wins the game; those on the top side represent points where the returner wins.

We can trace the progress of an actual game by a succession of horizontal and vertical moves along the grid lines starting at 0-0. For example, the server may win the first point, taking the score to 15-0, win the second (30-0), lose the next three (30-15, 30-30, and 30-40), and then win the next three, Deuce, Ad to the Server, and Game to the Server. One player wins if any such sequence reaches the thick line on his side of the diagram.

We begin with some simple calculations of the dynamics of a tennis game, without considering any choice of strategies at all. We want to know how it is likely to proceed – who is likely to win each point, and then who is likely to win the game. Generally, the server is more likely to win any given point. We will let x stand for the probability that the server wins any single point; the probability that the returner wins the point is then (1-x). For now, we will assume that we know x and that it is the same regardless of the current position in the game. If the server is more likely to win a point then $x > ½$. In top-level men's singles, for instance, a typical x might be 0.65 (65%). For convenience (and comparisons), we will also look at the case $x = ½$.

Now we ask: given x, can we calculate the probability that the server wins the whole game? A hard way to calculate this probability is to tot up all of the possible different ways the server can win, corresponding to all the zig-zag paths that can lead from 0-0 to the server's winning line on the right, find the probability of each, and add all these probabilities. The trouble is that there are infinitely many possible zig-zags between deuces and ads so the probabilities form a geometric series; summing it is a complicated and error-prone task. Rollback reasoning lets us cut through all that. We can begin by calculating the probabilities of the server winning from one of the later positions in the game (deuce or Ads) first, and then use this information to calculate the probabilities of the server winning starting from successively earlier positions, until we know the probability of winning from a 0-0 start.

To do this, we define a bit of notation. Let us write P(i,j) for the probability that the server wins the game, starting from (i,j), where i denotes the server's score and j the receiver's score within the game. For example, P(30,15) is the probability that the server wins the game, from being 30-15 up.* Then the probability that the server wins the game starting from any particular point equals the sum of two products, the probability that the server *wins* that point times the probability that the server wins the game after winning the point, *plus* the probability that the server *loses* the point times the probability that the server wins the game after losing the point. For example:

$$P(0,0) = x\ P(15,0) + (1 - x)\ P(0,15)$$

We emphasize two things about the formula just above. [1] All the P(i,j) are probabilities that the *server* wins the game eventually, so [1-P(I,j)] will be the probabilities for the returner's win. The different (i,j) are the different scores starting from which the server is supposed to win. [2] All the P(i,j) are unknowns thus far. We have to solve for them using such equations.

To solve for P(0,0), we would have to know P(15,0) and P(0,15). But to solve for those we would need to know P(30,0), P(15,15), and P(0,30). In turn, to get those we would need P(40,0) etc., etc. Thus we need to start at the *end* of the game and work our way back as we noted above; we need to use rollback. To start "at the end," we have to consider how a tennis game can end after it has gone to (40,40) or beyond. The end can occur in three possible ways. We use this fact and link three probabilities, all for *the server* to win the game, starting from each of the different end situations:

* We will only calculate the probabilities that the server wins the game. Of course, 1 - P(30,15) is the probability that the receiver wins the game starting from a score of (30,15) to the server, and similarly from any other starting point.

P(AS) = probability that the server wins when initially he has the Ad,

P(D) = probability that the server wins when initially the game is at deuce,

P(AR) = probability that the server wins when initially the receiver has the Ad.

To win the game starting with having the Ad, the server either wins the next point and thereby the game (probability x), or loses the first point (probability 1-x) and then wins the game starting at Deuce (probability P(D)). Therefore

$$P(AS) = x + (1 - x) P(D)$$

Similar arguments starting at Deuce and Returner's Ad tell us that

$$P(D) = x P(AS) + (1 - x) P(AR),$$

and

$$P(AR) = x P(D).$$

To recapitulate the reasoning: From AS, if you (the server) win the point, you win the game, but if you lose the point, you are at deuce (D). From D, if you win the point, you are at AS, but if you lose the point, you are at AR. From AR, if you win the point, you are at D, but if you lose the point, you lose the game.

Now we have three linear equations in three unknowns, P(AS), P(D), and P(AR). These three can be solved simultaneously. A simple way is to start with the expressions for P(AR) and P(AS) (which only have P(D) in them), and plug them into the expression for P(D). This gives us just one equation in P(D). We solve it, and substitute back to get the other two. The answers are:

$$P(D) = x^2/(1 - 2x + 2x^2),$$

$$P(AR) = x^3/(1 - 2x + 2x^2).$$

$$P(AS) = x(1 - x + x^2)/(1 - 2x + 2x^2),$$

Once we know these three probabilities, we can proceed to find the probabilities that server wins from earlier points in game. For example, the by-now-familiar argument starting at 40-30 gives

$$P(40,30) = x + (1 - x) P(D),$$

and we know P(D) now, so can find P(40,30). Similarly we can find P(30,40); then we can do a similar solution for each preceding point, in turn.

The results of these probability calculations are shown schematically in Figure 2. The grid arrangement is exactly like that of Figure 1. At each grid point, we show the P(i,j) corresponding to that score. The figure shows two cases: the one in the upper grid is for $x = 0.50$; the lower grid is for $x = 0.65$. The latter is typical for the top professional level in the men's game. You can experiment with other values by plugging in different values of x in the probability expressions.

In the upper grid, where x = 0.50, we see a symmetry about the 45-degree line. If you (the server) are not favored to win any particular point, then you are not favored to win the game starting from any symmetric score either. And the probabilities of the server's winning the game from any two situations symmetrically on either side of the 45-degree line add to 1. For example, starting from 30-15 you (the server) will win 69 percent of the time, whereas starting from 15-30 you will win 31 percent of the time.

In the lower grid, the server is quite heavily favored to win any one point. Therefore the probabilities of the server winning the game starting from any symmetric initial score are also greater than 50 percent. These probabilities get bigger the closer we move along the 45-degree line to the start of the game: the probability of the server winning the game from deuce or 30-30 is 78 percent, from 15-15 it is 80 percent, and from 0-0 it is 83 percent. The reason is that the earlier the starting point, the more opportunities the server has for recovery from a setback.

Similarly, in the lower grid the server has good chances to win the game even from many initially unfavorable situations. For example, the server has an almost even chance (48 percent) of retrieving a game from 0-30 down. If you are surprised, keep track when next you watch a top-level men's match.

All the calculation up to now is not based on any strategic game or decision situation. Neither player is actually choosing anything; the probability x is just a given, based on an assessment of the relative skills of the two players. Now introduce strategy. Suppose the server has a special serve that is much more effective than his ordinary serve, but cannot use the special serve all the time, perhaps because it takes a lot of effort, but more likely because if it is overused, the other player will learn and practice how to counter it and its effectiveness will diminish.

Then the strategic question for the server entails determining when to use this special serve. The instinctive answer is of course that it should be used on the most "important" or "crucial" situations in the game. But that just raises another question – what makes one situation more important than another? This needs a little more thought, but the answer is simple. The ultimate objective here is to win the whole game. A particular point is crucial when winning versus losing *that* point makes the biggest difference to the chances of winning the full game.

We can construct quantitative measures of this importance from our knowledge of the probabilities P(i,j), and show how to construct it for any one point (i,j), say 30-15 in favor of the server. If the server wins the 30-15 point, the probability of his winning the game becomes P(40,15). If the server loses the 30-15 point, the probability of winning the game from then on is P(30,30). Therefore the *difference* made to the probability of the server winning the game by winning *versus* losing the 30-15 point is just P(40,15) - P(30,30). This is what we define as the measure of importance of the 30-15 point. Using the symbol I(30,15) for it, we therefore write

$$I(30,15) = P(40,15) - P(30,30).$$

Consider the case of x = 0.5, shown in the upper grid of Figure 2. Using the actual numbers there, P(40,15) = 88 percent and P(30,30) = 50 percent; therefore I(30,15) = 88 - 50 = 38 percent.

Similar calculations can be done for all scores, and Figure 3 shows the resulting measures of importance of all positions for our two x values. The upper grid shows the case of x = 0.50, and the lower grid shows the case of x = 0.65. In the upper grid, for example, the number at the 30-30 location is 38, as we calculated just above. You can verify all the other values similarly, using the numbers in Figure 2 and doing the appropriate subtractions.

What about the importance of a point to the returner? The probability that the returner wins starting from any score is one minus the probability that the server wins starting from the same score. Consider starting at 30-15 as we did above. If the returner wins it, the score goes to 30-30; if he loses it, the score goes to 40-15 in favor of the server. Therefore the importance from the returner's perspective is

$$[\ 1 - P(30,30\ ] - [\ 1 - P(40,15)\ ] = P(40,15) - P(30,30) = I(30,15).$$

This is exactly the same as the importance of 30-15 to the server! The two have exactly the opposite objectives, but they are in full agreement as to which situations are more important.[1]

These measures of importance have several noteworthy properties:

1. The importance of points with equal initial scores rises steadily as we move northeast along any 45-degree line. Why? As we get closer to the end of a game, there is less opportunity to retrieve the loss of a point. Therefore it is more vital to win rather than lose the point.

2. The importance is highest along a central 45-degree line and falls off to each side (northwest or southeast) of it. Why? If you are very far ahead (bottom right of diagram), you are quite likely to win even if you lose this point. If you are very far behind (upper left of diagram),

---

[1] This is because the game is zero-sum; for non-zero-sum games the importance of any situation can differ for the two players.

the game may be lost even if you win this point. In neither case does winning *versus* losing this point make much of a difference. Only when the initial situation is reasonably balanced does winning or losing this point make a big difference.

3. The highest-importance line is truly central, namely the line of equality, for x = 0.50, but higher up for x = 0.65. Why? If you are favored to win any single point, then you have good prospects of coming from behind to win the game. Losing one point when you are initially equal does not matter all that much, so the line of exact equality is less crucial than the line where you are a little behind. In the x = 0.65 case, the most crucial point in the game is when the server is 30-40 or ad down. This conforms to the accepted wisdom of tennis commentators, and also to the practice of the top men players. You see them use their best serves in these situations. Pete Sampras at his best went a level beyond that. His probability of winning a single point on serve was even higher than 65 percent. Therefore the most important points to him were when he was serving 15-40 or even 0-40 down; that is when he delivered in succession two or even three very special unreturnable serves.

R&D competition

We asked you to regard the tennis game as a metaphor for business competition, especially an R&D race for a new product, where the objective is to combine some absolute standard of achievement and some relative level of superiority over your rivals. Now we can offer valuable insights into some aspects of this competition. The most crucial stages are when the rivals are reasonably close to each other and both products are reasonably close to completion – that is when one misstep will be difficult to correct in time. And therefore that is when both firms should exert maximum effort. If the two firms are unequal in their innate strengths, then the most crucial situation comes when, through some combination of luck or skill, the weaker firm has stolen a lead over the stronger one. If Intel is dominant in the current generation of computer chips, and AMD forges ahead to a leading position in the race for the next generation, this is when AMD will work hardest to protect its lead, and Intel will also fight hardest to retain its dominance. You should be able to think of several other examples of this principle.

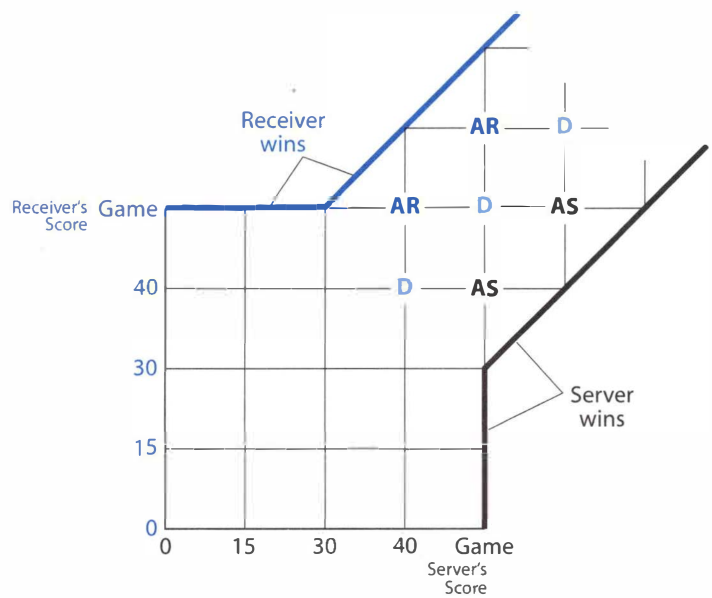

Figure 1: Grid showing possible positions in one tennis game

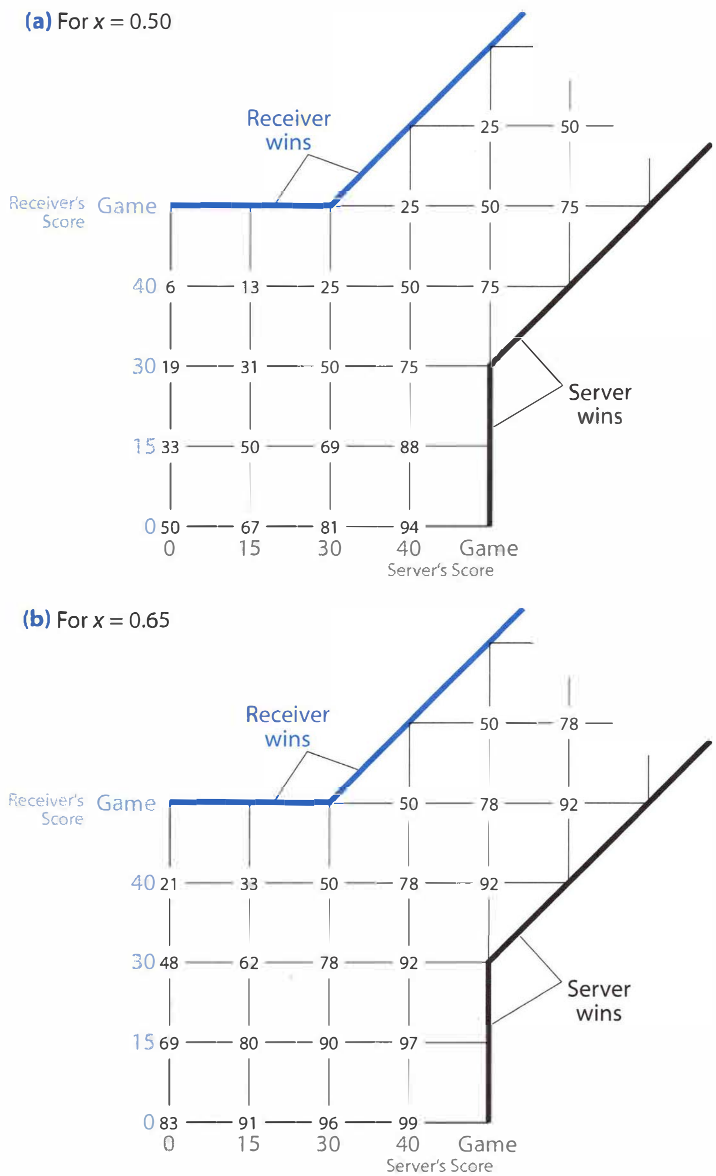

Figure 2: P-values

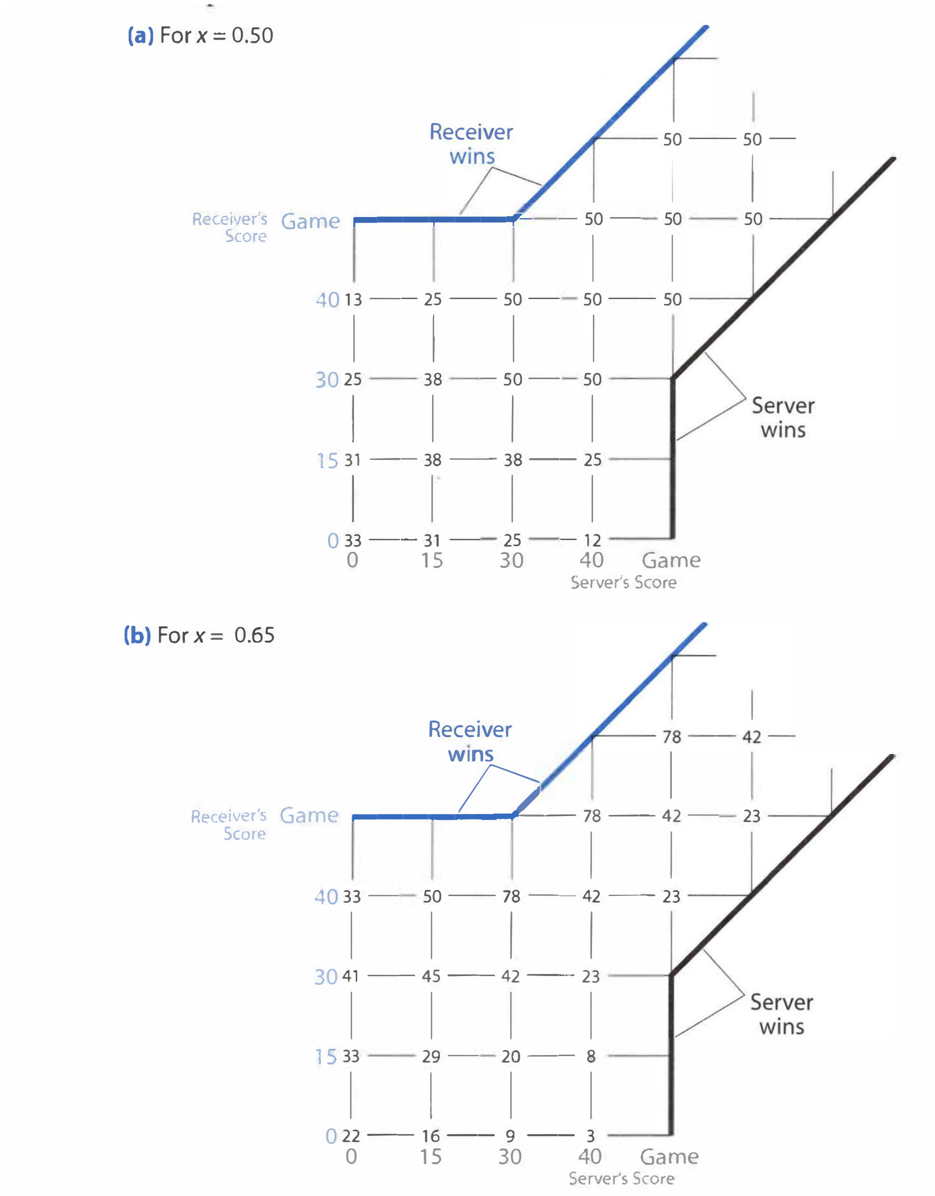

Figure 3: I-values