Introduction

Alice laughed: 'There's no use trying,' she said; 'one can't believe impossible things.'

'I daresay you haven't had much practice,' said the Queen. 'When I was younger, I always did it for half an hour a day. Why, sometimes I've believed as many as six impossible things before breakfast.'

'Where shall I begin,' she asked.

'Begin at the beginning,' the king said, 'and stop when you get to an end.'

Lewis Carroll

It does not take a student of mathematics long to discover results which are surprising or clever or both and for which the explanations themselves might enjoy those same virtues. In the author's case it is probable that in the long past the 'coin rolling around a coin' puzzle provided Carroll's beginning and a welcome, if temporary, release from the dry challenges of elementary algebra:

> Two identical coins of equal radius are placed side by side, with one of them fixed. Starting head up and without slipping, rotate one about the other until it is on the other side of the fixed coin, as shown in figure 1.
>
> Is the rotated coin now head up or head down?

Within a random group of people both answers are likely to be proffered as being 'obviously true', yet one of them is false and a quiet experiment with two coins quickly reveals which. We must prove the fact though, and too much knowledge is dangerous here: fix on a point on the circumference of the moving circle and we have an epicycloid to consider (or, more precisely, a cardioid) – and there could be hard mathematics to deal with.

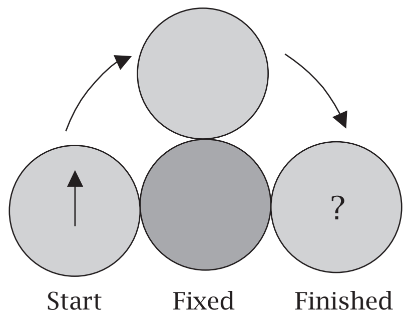

Figure 1. A coin rolling around another fixed coin.

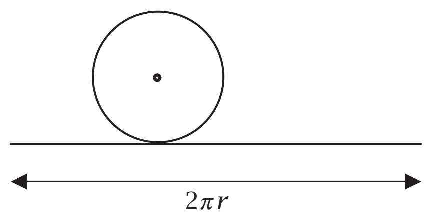

Figure 2. The situation simplified.

Alternatively, concentrate on the path of the centre of the moving coin and let us suppose that the common radii of the coins are r. During the motion, the path traced out by this centre is a semicircle, whose centre is itself the centre of the fixed coin and whose radius is $2r$; the motion will cause the centre to move a distance $\pi(2r) = 2\pi r$.

Now simplify matters and consider the moving coin rotating without slipping along a straight line of length $2\pi r$, the distance moved by its centre, as shown in figure 2. It is perfectly clear that it will have turned through 360° – and so be the right way up.

When it is first seen, the result is indeed surprising – and the solution clever.

It is a suitable preliminary example as this book chronicles a miscellany of the surprising, with a nod towards the clever, at least in the judgement of its author. The choice of what to include or, more painfully, what to exclude has been justly difficult to make and a balance has been found which recognizes the diversity of the surprising as well as the large role played by probability and statistics in bringing about surprise: it is they and the infinite which abound in the counterintuitive; other areas of mathematics dally with it. To reflect all of this, the fourteen chapters which constitute the book are divided evenly and alternate between results which intrinsically depend on probability and statistics and those which arise in other, widely diverse, areas; one such is the infinite. To reflect these tensions further, this is the first of two such books, the second providing the opportunity to embrace what the reader may have considered as unfortunate omissions. Wherever it has been possible, the provenance of the result in question has been discussed, with a considerable emphasis placed on historical context; no mathematics grows like Topsy, someone at some time has developed it.

Apart from chapter 13 (and where else could that material be placed?), the level of mathematics increases as the book progresses, but none of it is beyond a committed senior high school student: looking hard is not at all the same as being hard. It is hoped that the reader, young or not-so-young, will find something in the pages that follow to inform or remind him or her of the frailty of the intuition we routinely employ to guide us through our everyday lives, but which is so easily confounded – only to be replaced by the uncompromising reason of mathematical argument.

Chapter 1

THREE TENNIS PARADOXES

> So that as tennis is a game of no use in itself, but of great use in respect it maketh a quick eye and a body ready to put itself into all postures; so in the mathematics, that use which is collateral and intervenient is no less worthy than that which is principal and intended.
>
> Roger Bacon

In this first chapter we will look at three examples of sport-related counterintuitive phenomena: the first two couched in terms of tennis, the third intrinsically connected with it.

Winning a Tournament

The late Leo Moser posed this first problem during his long association with the University of Alberta. Suppose that there are three members of a club who decide to embark on a private tournament: a new member M, his friend F (who is a better player) and the club's top player T.

M is encouraged by F and by the offer of a prize if M wins at least two games in a row, played alternately against himself and T.

It would seem sensible for M to choose to play more against his friend F than the top player T, but if we look at the probabilities

Table 1.1. The situation if the new member plays his friend twice.

F	T	F	Probability
W	W	W	ftf
W	W	L	$ft(1-f)$
L	W	W	$(1-f)tf$

Table 1.2. The situation if the new member plays the club's top player twice.

T	F	T	Probability
W	W	W	tft
W	W	L	$tf(1-t)$
L	W	W	$(1-t)ft$

associated with the two alternative sequences of play, FTF and TFT, matters take on a very different look. Suppose that we write f as the probability of M beating F and t as the probability of M beating T (and assume independence).

If M does choose to play F twice, we have table 1.1, which lists the chances of winning the prize.

This gives a total probability of winning the prize of

$$\begin{aligned} P_{\mathrm{F}} &= ftf + ft(1-f) + (1-f)tf \\ &= ft(2-f). \end{aligned}$$

Now suppose that M chooses the seemingly worse alternative of playing T twice, then table 1.2 gives the corresponding probabilities, and the total probability of winning the prize becomes

$$\begin{aligned} P_{\mathrm{T}} &= tft + tf(1-t) + (1-t)ft \\ &= ft(2-t). \end{aligned}$$

Since the top player is a better player than the friend, $t < f$ and so $2-t > 2-f$, which makes $ft(2-t) > ft(2-f)$ and $P_{\mathrm{T}} > P_{\mathrm{F}}$. Therefore, playing the top player twice is, in fact, the better option.

Table 1.3. Outcome of an all-plays-all tournament between the various teams.

		T_{B}: 10	2	3	7	5
	1	B	B	B	B	B
	8	B	W	W	W	W
T_{W}	9	B	W	W	W	W
	6	B	W	W	B	W
	4	B	W	W	B	B

Logical calm is restored if we look at the expected number of wins. With FTF it is

$$\begin{aligned} E_{\mathrm{F}} = {} & 0 \times (1-f)(1-t)(1-f) \\ & + 1 \times \{f(1-t)(1-f) + (1-f)t(1-f) + (1-f)(1-t)f\} \\ & + 2 \times \{ft(1-f) + f(1-t)f + (1-f)tf\} + 3 \times ftf \\ = {} & 2f + t \end{aligned}$$

and a similar calculation for TFT yields $E_{\mathrm{T}} = 2t + f$.

Since $f > t$, $2f - f > 2t - t$ and so $2f + t > 2t + f$, which means that $E_{\mathrm{F}} > E_{\mathrm{T}}$ – and that we *would* expect!

Forming a Team

Now let us address a hidden pitfall in team selection.

A selection of 10 tennis players is made, ranked 1 (the worst player, W) to 10 (the best player, B). Suppose now that W challenges B to a competition of all-plays-all in which he can chose the two best remaining players and B, to make it fair, must choose the two worst remaining players.

The challenge accepted, W's team is $T_{\mathrm{W}} = \{1, 8, 9\}$ and B's team is $T_{\mathrm{B}} = \{10, 2, 3\}$. Table 1.3 shows the (presumed) inevitable outcome of the tournament; at this stage we are interested only in the upper left corner. We can see that W's disadvantage has not been overcome since T_{B} beats T_{W} 5 games to 4.

Table 1.4. The average rankings of each of the three pairs of teams.

Average ranking of T_B	5	$5\frac{1}{2}$	$5\frac{2}{5}$
Average ranking of T_W	6	6	$5\frac{3}{5}$

The remaining players are $\{4, 5, 6, 7\}$ and W reissues the challenge, telling B that he can add to his team one of the remaining players and then he would do the same from the remainder; of course, both B and W choose the best remaining players, who are ranked 7 and 6 respectively. The teams are now $T_W = \{1, 8, 9, 6\}$ and $T_B = \{10, 2, 3, 7\}$ and the extended table 1.3 now shows that, in spite of B adding the better player to his team, the result is worse for him, with an 8–8 tie.

Finally, the challenge is reissued under the same conditions and the teams finally become $T_W = \{1, 8, 9, 6, 4\}$ and $T_B = \{10, 2, 3, 7, 5\}$ and this time the full table 1.3 shows that T_W now beats T_B 13–12.

A losing team has become a winning team by adding in worse players than the opposition.

Table 1.4 shows, in each of the three cases, the average ranking of the two teams. We can see that in each case the T_B team has an average ranking less than that of the T_W team and that the average ranking is increasing for T_B and decreasing (or staying steady) for T_W as new members join. This has resonances with the simple (but significant) paradox known as the *Will Rogers Phenomenon*.

Interstate migration brought about by the American Great Depression of the 1930s caused Will Rogers, the wisecracking, lariat-throwing people's philosopher, to remark that

> When the Okies left Oklahoma and moved to California, they raised the intellectual level in both states.

Rogers, an 'Okie' (native of Oklahoma), was making a quip, of course, but if we take the theoretical case that the migration was from the ranks of the least intelligent of Oklahoma, all of whom were more intelligent than the native Californians(!), then what he quipped would obviously be true. The result is more subtle, though. For example, if we consider the two sets $A = \{1, 2, 3, 4\}$ and $B = \{5, 6, 7, 8, 9\}$, supposedly ranked by intelligence level (1 low, 9 high), the average ranking of A is 2.5 and that of B is 7. However, if we move the 5 ranking from B to A we have that $A = \{1, 2, 3, 4, 5\}$ and $B = \{6, 7, 8, 9\}$ and the average ranking of A is now 3 and that of B is 7.5: both average intelligence levels have risen.

If we move from theoretical intelligence levels to real-world matters of the state of health of individuals, we approach the medical concept of *stage migration* and a realistic example of the Will Rogers phenomenon. In medical stage migration, improved detection of illness leads to the fast reclassification of people from those who are healthy to those who are unhealthy. When they are reclassified as not healthy, the average lifespan of those who remain classified as healthy increases, as does that of those who are classified as unhealthy some of whose health has been poor for longer. In short, the phenomenon could cause an imaginary improvement in survival rates between two different groups. Recent examples of this have been recorded (for example) in the detection of prostate cancer (I. M. Thompson, E. Canby-Hagino and M. Scott Lucia (2005), 'Stage migration and grade inflation in prostate cancer: Will Rogers meets Garrison Keillor', *Journal of the National Cancer Institute* 97:1236–37) and breast cancer (W. A. Woodward et al. (2003), 'Changes in the 2003 American Joint Committee on cancer staging for breast cancer dramatically affect stage-specific survival', *Journal of Clinical Oncology* 21:3244–48).

Winning on the Serve

Finally, we revert to lighter matters of tennis scoring and look at a situation in which an anomaly in the scoring system can, in theory, be exposed.

The scoring system in lawn tennis is arcane and based on the positions of the hands of a clock. For any particular game it is as follows.

> If a player wins his first point, the score is called 15 for that player; on winning his second point, the score is called

> 30 for that player; on winning his third point, the score is called 40 for that player, and the fourth point won by a player causes the player to win, unless both players have won three points, in which case the score is called deuce; and the next point won by a player is scored 'advantage' for that player. If the same player wins the next point, he wins the game; if the other player wins the next point the score is again called deuce. This continues until a player wins the two points immediately following the score at deuce, when that player wins.

The great tennis players of the past and present might be surprised to learn that, with this scoring system, a *high quality tennis player serving at 40–30 or 30–15 to an equal opponent has less chance of winning the game than at its start.*

We will quantify the players being evenly matched by assigning a fixed probability p of either of them winning a point as the server (and $q = 1 - p$ of losing it); for a high quality player, p will be close to 1. The notation $P(a, b)$ will be used to mean the probability of the server winning the game when he has a points and the receiver b points; we need to calculate $P(40, 30)$ and $P(30, 15)$ and compare each of these with $P(0, 0)$, which we will see will take some doing!

First, notice that the position at 'advantage' is the same as that at $(40, 30)$, which means that the situation at deuce, when divided into winning or losing the next point, is given by

$$P(40, 40) = pP(40, 30) + qP(30, 40),$$

also, using the same logic, we have

$$P(30, 40) = pP(40, 40) \quad \text{and} \quad P(40, 30) = p + qP(40, 40).$$

If we put these equations together, we get

$$P(40, 40) = p(p + qP(40, 40)) + q(pP(40, 40))$$

and so

$$P(40, 40) = \frac{p^2}{1 - 2pq}.$$

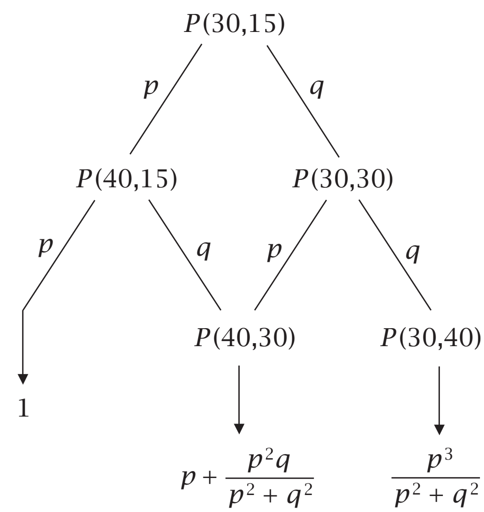

Figure 1.1. Finding $P(30,15)$.

Using the identity $1 - 2pq = (p+q)^2 - 2pq = p^2 + q^2$ we have the more symmetric form for the situation at deuce,

$$P(40,40) = \frac{p^2}{p^2+q^2},$$

and this makes

$$P(30,40) = pP(40,40) = \frac{p^3}{p^2+q^2}$$

and the first of the expressions in which we have interest is then

$$\boxed{P(40,30) = p + \frac{p^2 q}{p^2+q^2}}$$

Now we will find the expression for $P(30,15)$, which takes a bit more work, made easier by the use of a tree diagram which divides up the possible routes to success and ends with known probabilities, as shown in figure 1.1.

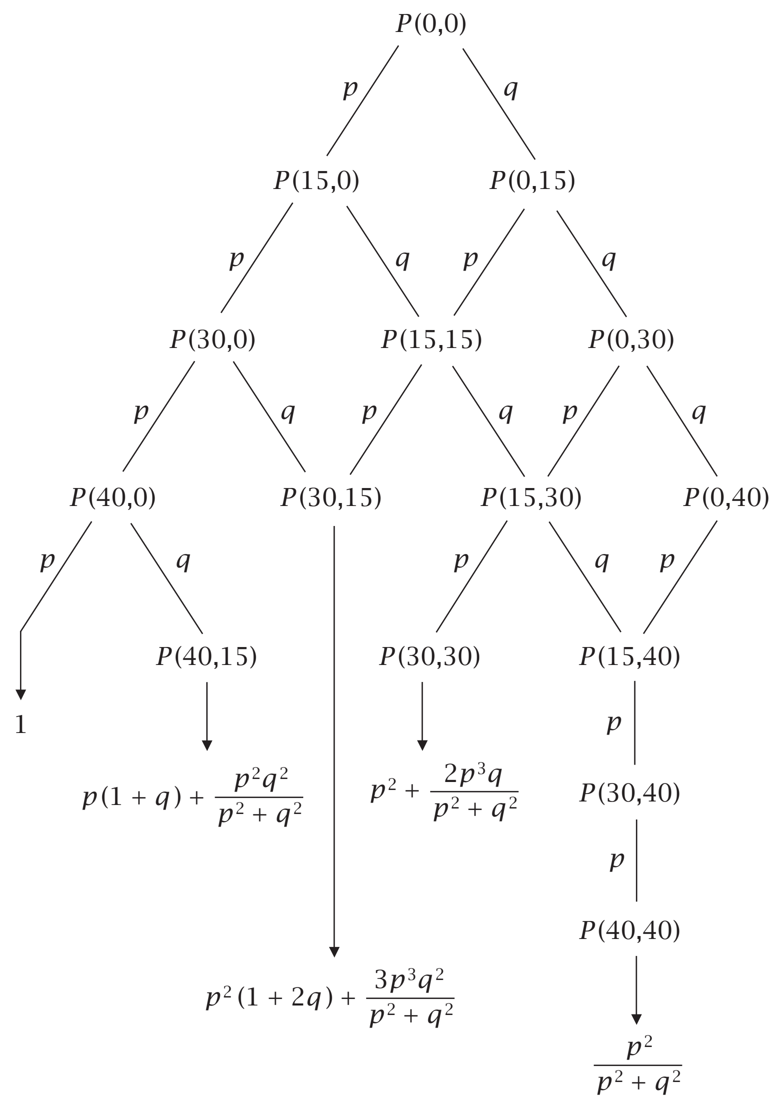

Figure 1.2. Finding $P(0,0)$.

Every descending route is counted to give

$$P(30,15) = p^2 + 2pq\left(p + \frac{p^2 q}{p^2+q^2}\right) + q^2\left(\frac{p^3}{p^2+q^2}\right)$$
$$= p^2(1+2q) + \frac{3p^3q^2}{p^2+q^2}$$

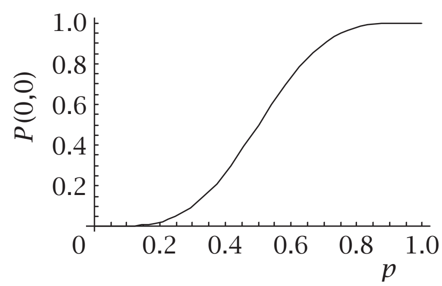

Figure 1.3. $P(0,0)$ plotted against p.

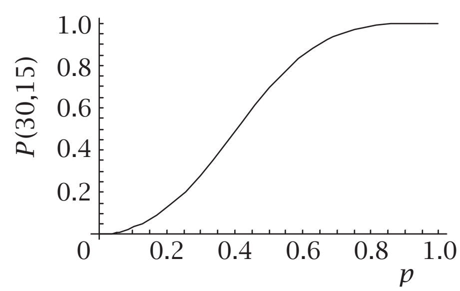

Figure 1.4. $P(30,15)$ plotted against p.

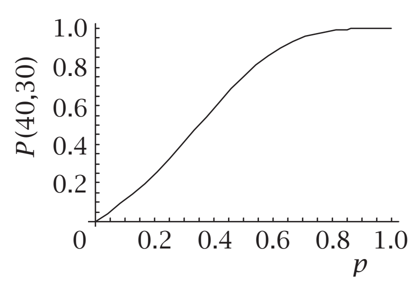

Figure 1.5. $P(40,30)$ plotted against p.

and so we have found the second of our expressions

$$\boxed{P(30,15) = p^2(1+2q) + \frac{3p^3q^2}{p^2+q^2}}$$

We only need the starting probability $P(0,0)$, which is by far the hardest goal, and to reach it without getting lost we will make use of the more complex tree diagram in figure 1.2, which again

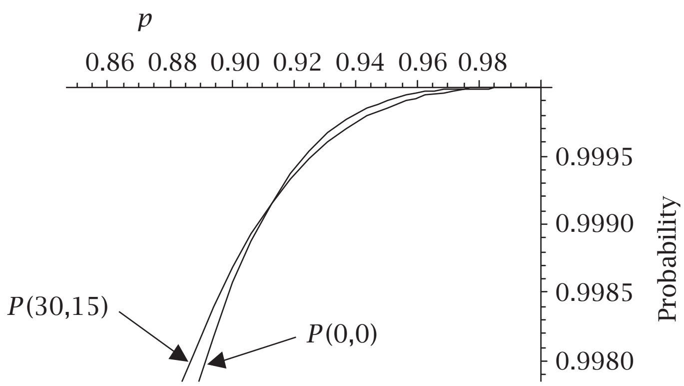

Figure 1.6. The intersection of $P(0,0)$ with $P(30,15)$.

shows the ways in which the situations divide until a known probability is reached. We then have

$$\begin{aligned}P(0,0) = p^4 + p^3q\left(p(1+q) + \frac{p^2q^2}{p^2+q^2}\right)\\ + 3p^2q\left(p^2(1+2q) + \frac{3p^3q^2}{p^2+q^2}\right)\\ + 3p^2q^2\left(p^2 + \frac{2p^3q}{p^2+q^2}\right) + 4p^3q^3\left(\frac{p^2}{p^2+q^2}\right)\\ = p^4(1+4q+10q^2) + \frac{20p^5q^3}{p^2+q^2},\end{aligned}$$

and the final expression needed is

$$\boxed{P(0,0) = p^4(1+4q+10q^2) + \frac{20p^5q^3}{p^2+q^2}}$$

Plots of the three probabilities, shown in figures 1.3–1.5, for all values of p (remembering that $q = 1 - p$) show that they have very similar behaviour to one another, but there are intersections and if we plot the pairs $\{P(0,0), P(30,15)\}$ and $\{P(0,0), P(40,30)\}$ on the same axes for large p we can see them. This is accomplished in figures 1.6 and 1.7.

Of course, to find those intersections we need to do some algebra.

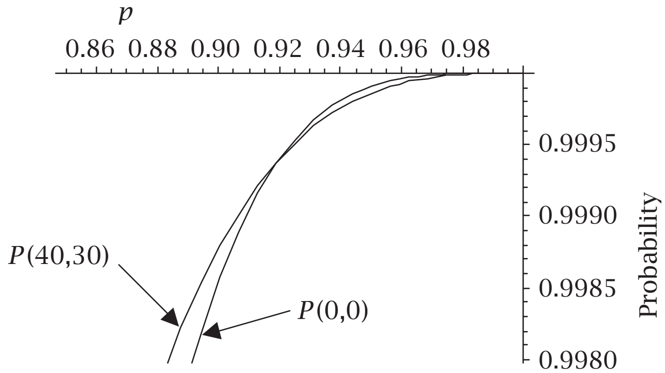

Figure 1.7. The intersection of $P(0,0)$ with $P(40,30)$.

The Intersection of $P(30,15)$ *and* $P(0,0)$

To find the point of intersection we need to solve the formidable equation

$$p^2(1+2q) + \frac{3p^3q^2}{p^2+q^2} = p^4(1+4q+10q^2) + \frac{20p^5q^3}{p^2+q^2},$$

again remembering that $q = 1 - p$.

Patience (or good mathematical software) leads to the equation in p,

$$p^2(1-p)^3(8p^2-4p-3) = 0,$$

which has repeated trivial roots of $p = 0, 1$ as well as the roots of the quadratic equation $8p^2 - 4p - 3 = 0$.

The only positive root is $p = \frac{1}{4}(1+\sqrt{7}) = 0.911\,437\ldots$ and for any $p > 0.911\,437\ldots$ we will have $P(0,0) > P(30,15)$ and the result for this case is established.

The Intersection of $P(40,30)$ *and* $P(0,0)$

This time the equation to be solved is

$$p + \frac{p^2q}{p^2+q^2} = p^4(1+4q+10q^2) + \frac{20p^5q^3}{p^2+q^2}$$

and, after a similarly extravagant dose of algebra, this reduces to

$$p(1-p)^3(8p^3-4p^2-2p-1)=0,$$

which again has trivial roots of $p = 0, 1$.

The remaining cubic equation $8p^3 - 4p^2 - 2p - 1 = 0$ has the single real root,

$$p = \tfrac{1}{6} + \tfrac{1}{24}\sqrt[3]{1216 - 192\sqrt{33}} + \tfrac{1}{6}\sqrt[3]{19 + 3\sqrt{33}},$$

which evaluates to $p = 0.919\,643\ldots$.

Again, for any $p > 0.919\,6\ldots$, we will have $P(0,0) > P(40,30)$, with the paradox once again established.

In conclusion, two equal players who are good enough to win the point on their serve just over 90% of the time are better off at the game's start than they are when the score is either 30–15 or 40–30 in their favour.