
# Soccermatics


## MATHEMATICAL ADVENTURES IN THE BEAUTIFUL GAME


DAVID SUMPTER

BLOOMSBURY

# A NOTE ON THE AUTHOR

**David Sumpter** is Professor of Applied Mathematics at the University of Uppsala, Sweden. Born in London but raised in Scotland, he completed his doctorate in Mathematics at Manchester, and was a Royal Society Research Fellow in Oxford before heading to Sweden.

David's research has shown how mathematics can be applied to anything and everything, and in particular to social behaviour. An incomplete list of his research projects includes: pigeons flying in pairs over Oxford; clapping undergraduate students in the north of England; swarms of locusts traveling across the Sahara; disease spread in remote Ugandan villages; the gaze of London commuters; and the tubular structures built by Japanese amoebae.

In his spare time, he exploits his mathematical expertise in training a successful under-tens football team, Uppsala IF P05. David is a Liverpool supporter with a lifelong affection for Dunfermline Athletic.

# SOCCERMATICS

# MATHEMATICAL ADVENTURES IN THE BEAUTIFUL GAME

PRO-EDITION

David Sumpter

BLOOMSBURY sigma

# Contents


The Kick-off

*Part I On the Pitch*

Chapter 1: I Never Predict Anything and I Never Will
Chapter 2: How Slime Moulds Built Barcelona
Chapter 3: Check My Flow
Chapter 4: Statistical Brilliance
Chapter 5: Zlatan Ibrah Rocket Science

*Part II In the Dugout*

Chapter 6: Three Points for the Bird-brained Manager
Chapter 7: The Tactical Map
Chapter 8: Total Cyber Dynamo
Chapter 9: The World in Motion

*Part III From the Crowd*

Chapter 10: You'll Never Walk Alone
Chapter 11: Bet Against the Masses
Chapter 12: Putting My Money Where My Mouth Is
Chapter 13: The Results Are In

*Part IV The Analysts*

Chapter 14: Finding the Talent
Chapter 15: Football's Intelligent Future
The Full-Time Whistle

Notes
Acknowledgements
Index

# The Kick-off

Mathematics can't compete with football. Football captures the hopes and dreams of nations. It brings us together in admiration of ability and commitment. It has superstars and tactics, entertainment and excitement. Football covers the back pages of newspapers and fills our Twitter feed. Tens of thousands of fans cram into grounds, and billions of people watch the World Cup on TV. Compare this with mathematics. Obscure academic journals lie unread in empty libraries. Seminars are attended by two gently snoring professors and a small group of bored PhD students. Football and mathematics? There is no competition.

If mathematics could compete with football, then we would be ready to pay £40 a month for a subscription to Sky Mathematics. Instead of spending Wednesday evening in front of the Champions League, we would load up Khan Academy and brush up on linear inequalities. If mathematics could compete with football, we would spend our November afternoons sitting on freezing plastic seats watching Marcus du Sautoy blackboard out that cocky Manc physicist from the TV. Arsenal one, Oldham Athletic nil. Instead of saying, 'It's a game of two halves', we'd say, 'It's a single division of the unit interval into sets of equal measure'. Instead of 'He gave 110%', the commentator would say . . . well, he'd say, 'He gave 100%.'

It is not as if mathematics wasn't given its chance. We all sat in school learning our times tables and tapping numbers into calculators. All those hours spent trying to remember whether 7 times 8 is 56 or 54, or whether the pi are squared or circled. Given all that time, and all that training, you would think that people would have noticed if mathematics was as exciting as football. But it seems that most people can't be so easily fooled. There may well be quite a few people who enjoy maths, but there are many, many more who totally love football.

I am one of those people who enjoy maths almost as much as I enjoy football. I am a mathematics professor, and I spend my day creating and understanding mathematical models. But even I wouldn't go as far as to claim that maths can compete with football. It can't. The numbers are against it.

Sometimes, when I look up at football and then back down at my maths books, I start to wonder what exactly I am doing with my life. Here I am, a professor of applied mathematics. I work on a wide range of different and interesting problems, with researchers from around the world. I have the opportunity to travel widely, to present my work at conferences in exotic locations, and to visit world-leading universities. All of this should be like playing for England. But it isn't, and I know it isn't. Being a mathematician is respectable, but it is nothing like succeeding in football.

The great footballers not only master technique and skill, they also achieve incredible levels of physical fitness. Footballers certainly aren't thick. On the contrary, the first thing football scouts look for in youngsters is 'intelligence', the ability to see quickly what is going on around them and to plan for all eventualities – something we academics might call spatial reasoning. Nor are footballers lazy. They are highly motivated, focused, driven individuals who decide from an early age that they want to succeed. Footballers are worshipped because they really have achieved greatness. The rest of us can only dream.

I'm the type of person who can't stop myself from dreaming. Despite now being 42 years old, and with two left feet and only a moderate interest in working out, I can't stop myself from believing that I can contribute to football. After all, planning and reasoning were also on that list of prerequisites for footballing success, weren't they? These are things I am good at. Maybe maths has something to offer football? And just maybe, football has something to offer maths?

There are good reasons to believe that my hard-earned modelling skills may prove useful after all. Numbers play an increasingly important role in football. Player and team rankings, assists and goals, possession and passing rate, tackle and interception frequency are just a few of the stats that feature in match reports. Detailed 'chalkboards' of corner angles, passing timelines and positional heat maps are displayed on managers' computer screens in post-match briefings. But these numbers are just a starting point. Mathematics is about putting statistics together in a way that allows us to see what is going on. Once we have numbers, mathematics gives us understanding.

There is a whole range of footballing questions that can be answered using mathematics. What is the probability of two last-minute goals in a Champions League final? Whatever Manchester United fans may say, this is a question about the nature of pure randomness. Why is Barcelona's tiki-taka passing so effective? This is a question of geometry and dynamics. Why do we give three points for a win in league games? This is a question of game theory and incentives. Who is best, Messi or Ronaldo? This is a question of large statistical deviations. What do heat maps and passing statistics actually tell us about a game? This is a question for big data and networked systems. How can bookies offer such attractive-looking spread bets? This is a question of combining probability and psychology. And why are these odds so hard to beat? This is a question of collective intelligence and averaging.

I will answer all these questions and more in this book, but my ambition stretches further. Soccermatics isn't simply about providing you with a few maths-related football facts that you can tell your friends down the pub; it's about changing the way you look at both maths and football. I believe that the two have a lot to offer each other and, while maths can't compete with football, both subjects can learn from each other. Maths can be used to understand football, and football helps to explain mathematics.

Football and maths start from the same point. Football starts with the 'laws of the game', the rules set out by the International Football Association Board. What football managers have to do is solve the problem of getting their team to win within the constraints imposed by these rules. Mathematics has its own set of rules, which the mathematician has to apply to get the right answer to the question posed. By following these rules, and with a little bit of inspiration, both the footballer and the mathematician seek to reach their goal. Management and mathematics both start with theory.

But the rules of the game aren't everything. It's one thing for the manager to explain the importance of the players holding their positions, but if a central defender picks up the ball in their own half, charges with confidence towards the opponents' goal and bangs it into the top-left corner, then not even Louis van Gaal will complain. Most of us are happy to accept that what happens in practice can be very different from what the theory says should happen. If everyone stuck to the theory, then football matches – and life in general – would be very boring indeed.

Exactly the same point holds for mathematics. Of course, when mathematical theories are proved, then they always remain true. Pythagoras's rule gives us a relationship between the lengths of the sides of a right-angled triangle, and this relationship always holds. But the real world isn't made of perfect triangles, and when mathematics meets the real world anything can happen. Sometimes our mathematical model of the real world is correct, but at other times we get it wrong. Sometimes, like football managers, we set up a beautiful theoretical idea, only to see our observations charge off in a completely different direction. Putting mathematics into practice is just as important as knowing the precise details of the theory.

It is the combination of theory and practice that makes football the sport we love. You can dribble like Messi or bend it like Beckham, but if your team lacks structure you will never get the chance to show off your skills. You can sing your national anthem with pride and feeling, but 30 minutes later find yourself trailing 5–0 to a well-organised Germany. And you can know every formation in the book, but without all those hours of practice in the school playground and on the training pitch, you won't have mastered the touch you need to succeed. Football is more than tactics, it is more than mastery of the ball – and it is more than the feeling of winning.

While every football pundit knows that theory and tactics are just a small part of football, the same point is less widely acknowledged about mathematics. We hear about characters like Andrew Wiles, who locked himself in his office in Princeton, only to emerge seven years later with a proof of Fermat's Last Theorem. Films depict mathematicians as child prodigies, dusty professors covered in chalk or difficult geniuses without any friends. We are told that mathematics is an elaborate, ever-evolving game of chess that you have to study for years to learn the rules. It is almost the complete opposite of the fanatical world of football. Far too often, mathematics is admired for its purity and mathematicians for their dedication, but not for their impulsiveness or their imagination.

As beautiful as pure mathematics may be, that is not the type of mathematics that excites me. I have always aimed to put maths to use in unusual places. I have used networks to map out urban sprawl, railway networks and segregated neighbourhoods. I see equations in following the gazes of city commuters, the way students applaud after hearing a presentation and how heavy-metal fans jump around in mosh-pits. I have created models of fish swimming among coral in the Great Barrier Reef, democratic change in the Middle East, the traffic of Cuban leaf-cutter ants, swarms of locusts travelling across the Sahara, disease spreading in Ugandan villages, political decision-making by European politicians, dancing honeybees from Sydney, American stock-market investors, and the tubular structures built by Japanese slime mould. For me there is no limit to mathematical modelling. Everything can and should be modelled.

I realised early in my career that I was different from many of my mathematical colleagues, who specialised in specific equations and single areas of application. I wanted to get stuck into the data, and work together with biologists and sociologists. I love the abstract beauty of equations, but formulae are meaningless until they say something about reality. So, while much of my day is spent sitting in front of a computer or sketching ideas on a blackboard, at times I can be found building a racetrack for locusts, talking to government ministries about tackling social problems, traipsing through the forest counting ants, or handing out tablets to school classes so I can study how they play interactive math-games. I don't let logic alone tell me what problems to study – I indulge my emotions, my feelings and my sense of humour. I play mathematics just like I play football, only much, much better.

There has always been a unifying rationale behind all my seemingly random projects. I see very different parts of the world as being related to one another, and I use mathematics to create links between them. I use a mathematics that isn't scared to get dirty, to change tactics at half-time or to involve players from all backgrounds and all over the world in a gigantic kickabout. It is a mathematics that aims to entertain as well as to impress, and where we celebrate the team just as much as the individual. It is this approach that is Soccermatics.

In this book, I use Soccermatics to attack a whole load of very different problems. Football is always the starting point, but I don't stop there. Each chapter is a story about how football and maths can work together to create powerful analogies. I show that managers use the same tactics to fight over points as birds use to fight over a worm, and cancer cells use to fight over our bodies. I decompose the network structure of Champions League teams, I show how the spread of football chants can explain everything from polite audience applause and transfer rumours to disease in the poorest parts of Africa. I show that, while the Mexican wave may be fun for fans, it is life and death for fish. These stories link together the physical, biological, social and footballing worlds.

Beneath these individual stories lies a deeper message. The Soccermatics philosophy is about a more accessible and creative style of mathematics. It is about a mathematics that crosses boundaries, one that creates links and analogies. It is about a mathematics that can be applied to anything. I use footballing analogies to explain other parts of the world, and I use other parts of the world to explain football. These analogies become possible because mathematical models provide a powerful way of seeing connections. When you work as a mathematical modeller, you see relationships that other people have missed.

Just like football, anyone can play at modelling. If you are the sort of person who sees things more clearly through footballing analogies, through sporting analogies, weather analogies, analogies in film and music, analogies to nature or any other type of analogy, then you are already one step closer to becoming a mathematical modeller. If you can make good analogies, then you can make good mathematical models. Being a modeller is first of all about using your imagination, and then homing in on the problem. It is a creative activity, but one that is subject to rules and procedures. I want to show you how to think in this way, and hopefully help you understand more about your own life and the world around you. Mathematics is a way of seeing problems and finding solutions.

By thinking Soccermatically, you will see players, teams, managers and fans in a new light. You'll see why Bastian Schweinsteiger is a whirlwind, why Bayern Munich defenders are lionesses and why the Barcelona team of 2015 is a jet fighter. You'll learn how to motivate a team by making them work like ants, and how to discourage slackers by changing incentives. You'll see why betting is like trying to build a communication cable into the future, understand why punters who know almost nothing about the game can make smart predictions together, and realise why you should never trust the experts. You may even find out how to make a few quid at the bookies.

This book explains the important models through words, computer simulations and pictures. Instead of filling pages with obscure symbols, I'll do meaningful calculations that reveal the inner workings of a football team. You won't need to dig out your graphical calculator, because I'll use my laptop to process vast quantities of match data. And while we will still need our trusty blackboard, it will be used to sketch diagrams and create intuitive pictures. I do not assume an in-depth knowledge of maths, but for those who want to learn more I have included details in the endnotes. I'll show that making mathematical models is about seeing patterns and making analogies. By the end of the book, you'll be able to find mathematics everywhere.

Neither do I assume that you have an in-depth knowledge of football. And I'm also going to be honest with you from the start. While I guarantee that this book will give you a unique perspective on football, I know my limitations: I'm not, of course, a world-famous manager, just a reasonably successful academic. Before I started researching this book I was like many other British men. I watch football, I read about it, I play it with friends and I spend my spare time training a (very talented) team of 10-year olds. Those friends who have seen me play will laugh out loud when they find out that I've written a book on the subject.

Instead of pretending to be an expert, I give a different perspective. Mathematicians and economists have dabbled in football writing before. They have argued that if teams could just change the way they take corners or throw-ins, then they would score more goals. They give the world's leading players advice about how to take a penalty kick. They provide confident statistical arguments for why England will win the next World Cup – or why they will never win it again. Some of these mathematically inspired suggestions make sense, but others don't. I'll look at how we can evaluate these claims and create our own modelling arguments.

The same is true when we see football presented on TV. Studios are now equipped with advanced technology for showing and analysing highlights and tactics. The question for us viewers is, what useful information do these displays provide? An animation of player positions may look good with Jamie Carragher standing in front of it, but it is Jamie Carragher who understands football, not the programmer who generated the graphics. When we see new ways of displaying data, we should be careful not to confuse presentation with substance. For this, we need to understand the maths that lies behind the big fancy screens.

This overload of data and statistics is not unique to football. Mathematics is now used to tackle problems throughout science and society. House prices, project scheduling, Facebook friend networks, viral marketing, artificial intelligence, online poker, economic growth, genetic engineering, computational biology, crowd and disaster planning, and most of the rest of modern life are all impacted by mathematics. So even if you are not particularly interested in watching 22 people kick a ball around a pitch, you can't stay outside the mathematical world. You need to understand how applied mathematics works and how we mathematicians think.

Football offers a way to understand the connection between maths and the modern world. Soccermatics is about how analogy is used to understand science, society and football. So forget those boring rules for sines and cosines – I'll show you how mathematical modelling is about thinking freely and broadly. We'll start on the pitch, then move into the dugout, and finally we'll find ourselves out in the crowd and trying to outwit the online bookies. We're about to set off on a mathematical adventure through the beautiful game.

PART I

# On the Pitch

CHAPTER ONE

# I Never Predict Anything and I Never Will

The England midfielder Paul Gascoigne once said 'I never predict anything and I never will.' For me, this statement contains just as much genius as his Euro '96 goal against Scotland. In eight words it shows just why predictions are unavoidable: after four words he was wrong about the past and the present, and after the next four he was wrong about the future too. But even though he was so wrong, Gazza still told us something profound. He summed up a very deep fact about life: there are patterns to be found in everything.

There are patterns in how long it takes us to get to work in the morning rush hour. There are patterns in our networks of friends and how often we meet up with them. There are patterns in what we eat for dinner each night and what we buy in the supermarket. And there are, of course, patterns in football. The challenge lies in finding these patterns and understanding them. Once we've seen the patterns, we can start to make predictions.

## Random Subbuteo

I can trace my own fascination with patterns back to a large orange hardback full of football statistics I was given for Christmas when I was eight. I would sit for hours looking at pages full of numbers. I loved the tables that had the names of the teams along the top and down the left, and the entries that were the scores of the matches between them in a season. I'd scan down the table, adding up all the goals scored and looking for matches with strange results. 4–3 was a favourite, and 5–2 also sounded good.

I don't get quite as much time to read football almanacs nowadays, but luckily it takes just a few seconds to find all the results and tables on the internet. If you do this, you can get a feeling for the unpredictability that Gascoigne was talking about. The 2012/13 Premier League season is a good one – there were some pretty exciting matches and unexpected outcomes. Liverpool won 5–0 twice and 6–0 another time, but still failed to qualify for Europe. The season ended with the retirement of Alex Ferguson, the king of unexpected last-minute changes in fortune. His last match as Manchester United manager was no exception: a 5–5 draw in which West Bromwich Albion scored three goals in the last ten minutes. 'Football, bloody hell!', as Fergie once put it.

These results are the exciting exceptions, the season's most memorable matches. There were also a fair number of boring 0–0 draws, forgotten by the fans, perhaps, but not by the season's statistics. But if we want to understand the underlying pattern, we need to include these in our analysis, too. Figure 1.1 is a histogram of the number of goals scored in all Premier League games in the 2012/13 season. The average number of goals scored was a bit less than three per match, 2.79 to be exact.

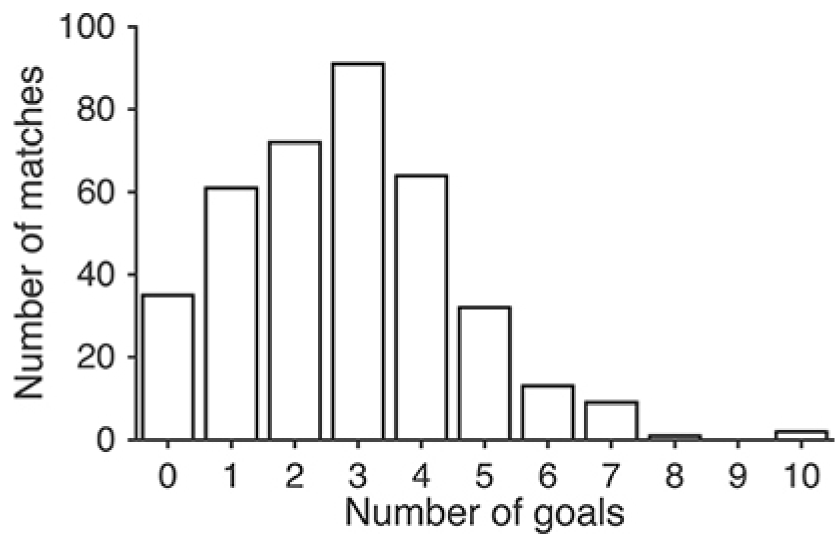



**Figure 1.1** Histogram of number of goals scored during the English Premier League season 2012/13.

The histogram shows how often various scorelines occurred. In all there were 35 0–0 draws, which is the first bar on the histogram. Ferguson's last match was one of two that season which ended with 10 goals being scored, as can be seen over on the right. In the middle, the most common goal tally was three, and in the majority of those games the final score was 2–1. A pattern is already starting to appear. The next step is to see if we can understand where this pattern comes from, and for this we need a mathematical model.

I've been interested in mathematical modelling for almost as long as I've been interested in statistics. My other big hobby around the time I was reading big orange football almanacs was playing Subbuteo table football.[1] With my friend David Paterson, I set up a Subbuteo league. We played every day after school, fitting in five or six matches before dinner, and making a note of each score. But we never had time to complete all the 380 fixtures that make up a league tournament (20 teams each playing 19 home games makes 20 × 19 = 380 matches). There just weren't enough hours in the day.

Constrained by parents who seemed to think we had to do things like eat and sleep, Patzi and I had to find a different way to complete the league. The answer was dice. Patzi would throw one dice for one team and I would throw for the other. We'd then take away one from the result on each dice to get the result. So if Arsenal played Manchester City, he rolled a red dice and I rolled a blue one. If the red dice showed 5 and the blue one showed 3, then Arsenal won 4–2. This model can generate games with from 0 to 10 goals, just as in the Premier League histogram.

After a lot of dice-throwing, and some small adjustments to the advantage of our favourite clubs, we'd fill in their scores based on the numbers that came up. We'd compile leagues and statistics and write it all out neatly on lined paper. I think I was always destined to become a mathematician (and the other David is now a successful accountant).

Dice throwing is a very simple example of a mathematical model, but there are a few problems with it. Chelsea beat Aston Villa 8–0 just before Christmas 2012, which simply couldn't occur in our dice-rolling model. Another problem is that 0–0 draws happen a lot in real football. For the dice, a 0–0 result is just as likely as 5–5, but in the histogram 0 goals is almost 20 times more likely than 10 goals. The model doesn't work. Football games aren't random, like a dice throw.

But football games *are* random, in another way. What makes football and other team sports exciting is their unpredictability. If you are watching a match and look away for a few seconds, you can miss an important build-up and a sudden goal. As a modeller, this tells me something important. A goal is just as likely to occur at any time during the match. While there are all sorts of factors determining the rate at which teams score, the timing of goals is more or less random.

We can turn this assumption into a simulation. Imagine a football game as 90 single-minute slots, in each of which a goal is equally likely. With an average of 2.79 goals per match, the probability that there is a goal in any one of these slots is 2.79/90 = 0.031. This means that the chance of us seeing a goal in any randomly chosen minute is about 1 in 32. Not a very large chance, but enough to make sure you keep watching.

Using this model we can run a computer simulation over 90 minutes, where in each simulated minute a goal is scored with probability of 0.031. If we keep on running the simulation over lots and lots of matches, we can find out what a typical season looks like. This simulated season is plotted in Figure 1.2 as a solid line, superimposed on the histogram for the real Premier League 2012/13 season.

The correspondence between model and reality is very good. Remember all the complexity at play here. All the shouting by the manager from the touchline. The fans trying to rally their team or, more often than not, telling them how useless they are. The thoughts in the heads of the players as they tell themselves that now is their chance to score. None of these factors seem to affect the distribution of goals scored. On the contrary, it is all these factors acting together that generate the type of randomness assumed in the model. The more factors involved, the greater the randomness in the goals, and the better the match of our simulated histogram with reality.

The solid line in Figure 1.2, generated by my simulation, is known as a Poisson distribution. This kind of distribution arises whenever the timing of previous events has no effect on future events. This is exactly what I assumed in my simulation, and it's what actually happens in football: neither the number of goals scored so far, nor the amount of time played, influences the probability of another goal being scored. The resulting Poisson distribution succeeds remarkably well in capturing the overall shape of the goal histogram.[2] The events make each minute of the football match unpredictable, and out pops the Poisson distribution. It's a pattern that arises from pure randomness.

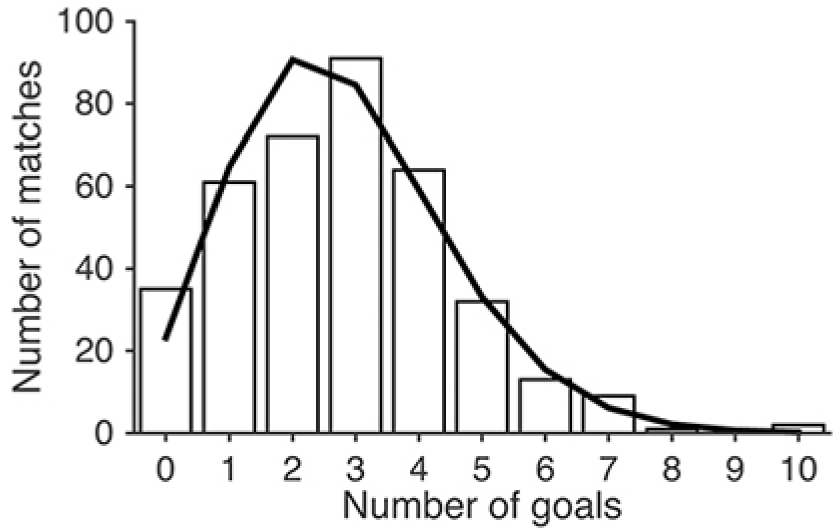



**Figure 1.2** The histogram of goals during the Premier League season 2012/13 (histogram boxes) compared to the Poisson distribution (solid line).

I didn't choose to look at Premier League football because I knew beforehand that it would follow the Poisson model. It just happens that I'm into football. I could have chosen pretty much any sport in which goals can be scored at any time. To make sure of this, I looked at all the results for NHL ice hockey games in the 2012/13 season. There was an average of 5.2 goals during the 60 minutes of normal time. Figure 1.3 shows a histogram of the number of goals in the 720 games that made up the season. The solid line is the corresponding Poisson distribution.

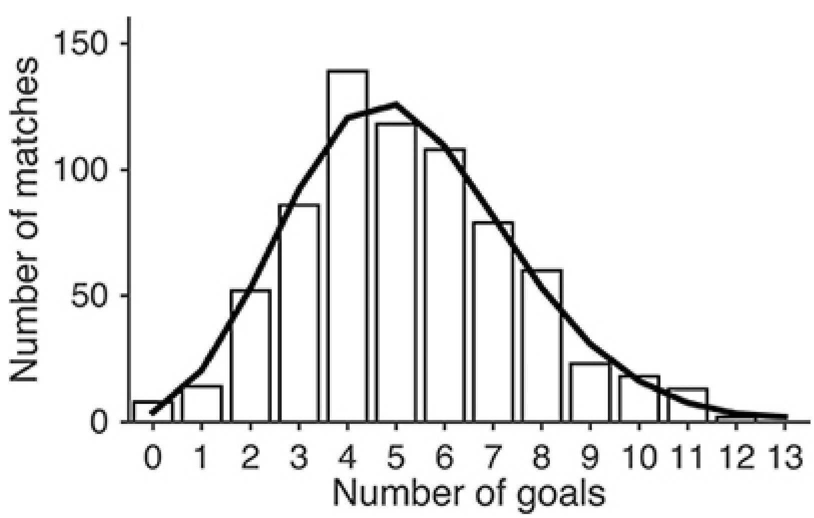


**Figure 1.3** Histogram of number of goals scored during the NHL ice hockey in season 2012/13 (boxes) compared to the Poisson distribution (solid line).

The higher average number of goals shifts the peak in the histogram to the right, but the simulation again fits the data. The data and the model are not very different at all, and the small difference in terms of four-goal matches can be put down to fluctuations from one season to another.[3] Goals go in more often in ice hockey, but no more or no less randomly than in football.

## Kicked by the Horse

If you start thinking in terms of random simulations and the Poisson distribution, then you see it everywhere. In undergraduate statistics courses, the lecturer's best (and only) joke is that bus arrivals are Poisson-distributed. The bus company starts off with timetables, but a lot of random factors get in the way: an old man takes his time getting on the bus, or a cyclist hogs the middle of the bus lane. Another classic example is the number of light bulbs you have to change in your house every year. Every time you switch on the light there is a small chance that the element will break. Add up all the breakages and you get a Poisson distribution.

The Poisson distribution is named after Siméon Denis Poisson, the Frenchman who first described it in the early nineteenth century. However, his description emphasised the mathematical equations behind the distribution and not how it could be used to model the real world. It was first applied in the way I use it here by a Pole, Ladislaus Bortkiewicz, working in Germany in 1898.[4] He investigated two data sets. The first was a macabre set of statistics he found on the number of children under the age of 10 who committed suicide over a 24-year period. The second data set, only slightly less disturbing, concerned soldiers who had died after being accidentally kicked or otherwise hit by a horse. Bortkiewicz looked at 14 different regiments over 20 years, plotting how many soldiers had been killed in this way. Obviously, he hadn't realised that only a few years earlier the English Football League had been established. This could have provided him with all the data he needed, without having to delve into the German death statistics.

In both of his data sets, Bortkiewicz found a good correspondence with the Poisson distribution. Deaths caused by kicking horses were uncommon. Of the 280 regiments he studied, there had been no deaths at all in 144. But in two unlucky regiments there were four deaths in a single year. By fitting the Poisson distribution, Bortkiewicz could show that these regiments were not necessarily treating their horses any worse than any other regiment: they had just had bad luck that year. Football may or may not be more important than life and death, but all three are governed by the same rules.

Comparison with a Poisson distribution is one of the first things I do when I'm presented with new data. Sometimes a colleague will come into my office with some newly collected experimental results. 'It's weird,' he'll say. 'Most of the fish never swim past a predator, but there was one fish that swam past it four times! It must have a bold personality type or something.' Three minutes later, I'll be plotting the Poisson distribution and overlaying it on my colleague's data. 'No, your fish wasn't particularly bold,' I'll say. 'It was just a statistical necessity.' Getting chased by a predator over and over again is like getting a 5–0 drubbing. It looks bad when it happens, but it could happen to anyone.

The Poisson distribution is our first example of a mathematical analogy. It works in so many contexts. It works for football matches, it works for light bulbs and it works for death by horse. Whenever it is reasonable to assume that events can happen unexpectedly, at any time, independently of how many events have happened prior to the next one, then it is reasonable to expect a Poisson distribution.

Away from football, most modern applications of the Poisson distribution continue the tradition started by Bortkiewicz. Statisticians seem to have a perverse fascination with death, injuries and accidents. Or maybe it's just that we pay them to work out the bad things that could happen to us, so that we don't have to think about it. Whatever the reason for their interest in misfortune, statisticians have found the Poisson distribution in car crashes, truck collisions, head injuries, aeroplane engine shutdowns, bankruptcies, suicides, murders, work-related accidents and the numbers of hazardous building sites.[5] They have even found it in the number of wars starting between 1480 and 1940. And when they are done with death and injury, they find the Poisson distribution in printing errors, manufacturing defects, network failures, computer virus attacks and divorces. Wherever there is death or destruction, misfortune or mistakes, then the same pattern of randomness can be found.

In 2015, Cristian Tomasetti, an applied mathematician, and Bert Vogelstein, a medical doctor, used a statistical argument to show that two-thirds of cases of cancer were due to 'bad luck'.[6] While certain cancers can be linked to lifestyle choices, for example lung cancer and smoking, this was only part of the story. The more important part lies in the unavoidable cell divisions that take place in our bodies. Every time a cell divides, there is a very small chance of a genetic mutation that could cause cancer. What Cristian and Bert found was that parts of the body where cells divide faster are more likely to develop cancer, and they concluded that cancer is primarily explained by these random mutations.

This study caused some controversy. If cancer is just random, then why should we spend so much money researching its causes? To justify using the term 'bad luck' and in order to better explain their conclusions, Cristian and Bert used an analogy with car accidents. The more time you spend in the car driving around, they said, the more likely you are to be in an accident. How you drive the car is a factor, but your time at the wheel is important too.

A footballing analogy works just as well, if not a little bit better. You can think of every cell division in your body as akin to a single minute of a football match. When a cell divides, there is a (very) tiny chance of a random cancerous mutation, just as there is a (much bigger) chance of conceding a goal in a football match. It is in this sense that cancer can be thought of as bad luck. Sometimes our team gets through a match without conceding a goal, and hopefully we get through our life without getting cancer. And while sometimes we lose because the opposition were good, no one can deny that luck plays an important part in any particular match. Our health is like a Saturday afternoon watching from the terraces, and not all goals are preventable.

Not everything that happens to us is down to randomness. Many diseases are preventable if we make healthy lifestyle choices, and conceding goals is often down to bad defending. But realising that much of what happens to us is random can sometimes help us come to terms with the challenges life throws at us. Not everything in life is predictable.

## Explained by Randomness

It is the unpredictability of a football match from one minute to the next that produces the Poisson distribution after 90 minutes. We know the average number of goals scored in a match, but their timing is unpredictable. As a result, some scorelines become much more likely than others. The paradox here is that scores are explained by randomness. The fact that goals are very random in time makes the pattern in the results predictable. It's a difficult idea to get your head round, but it is true. Often the very fact that something is extremely random helps us to explain it and to predict how often it will occur. Randomness allows us to make all sorts of predictions about the future.

Mathematicians use this trick all the time. At the start of a new football season, or in the run-up to the World Cup or to the Oscars, newspapers will often run stories about a 'genius' mathematician who has predicted the probability of particular teams or films winning. These predictions often seem reasonable, and they are sometimes correct. But where do they come from?

I'll let you in on a secret. These geniuses are usually doing something very simple with the Poisson distribution and a bit of background information about the competing teams and films. One trick for modelling football results is to calculate a scoring rate and a conceding rate for each team and then simulate matches between them. For example, during the 2012/13 Premier League season Arsenal scored an average of 2.47 goals when playing at home and 1.32 goals when playing away. They conceded an average of 1.21 goals at home and 0.74 when away. By collecting such statistics for each team and then simulating matches between each pair, we can produce predictions for the coming season. An example of such a prediction is given in Table 1.1, where I have used data from the 2012/13 season and a model to predict the top four for the 2013/14 season.[7]

This prediction isn't so far away from what actually happened. In the real world, Manchester City were champions, two points ahead of Liverpool, and Chelsea came third. But this is just one of many possible simulated top fours that come out when I click 'run' on my computer. Each time I run the simulation, the teams meet each other both home and away, the scores are picked randomly, with an average based on their scoring and conceding rates, and I compile a league table based on the results. Each run gives a different result, some very different. Table 1.2 is another example.

**Table 1.1** The top four clubs after the first simulation of the 2013/14 season, based on the scoring rates of clubs during 2012/13.

| *Team* | *P* | *W* | *D* | *L* | *F* | *A* | **Pts** |
|---|---|---|---|---|---|---|---|
| Manchester City | 38 | 22 | 7 | 9 | 71 | 42 | **73** |
| Liverpool | 38 | 22 | 5 | 11 | 64 | 43 | **71** |
| Chelsea | 38 | 21 | 5 | 12 | 74 | 51 | **68** |
| Manchester United | 38 | 19 | 7 | 12 | 61 | 45 | **64** |

**Table 1.2** The top four clubs after the second simulation of the 2013/14 season, based on their scoring rates in 2012/13.

| *Team* | *P* | *W* | *D* | *L* | *F* | *A* | **Pts** |
|---|---|---|---|---|---|---|---|
| Liverpool | 38 | 23 | 7 | 8 | 68 | 37 | **76** |
| Chelsea | 38 | 22 | 8 | 8 | 75 | 52 | **74** |
| Manchester United | 38 | 22 | 5 | 11 | 72 | 43 | **71** |
| Manchester City | 38 | 19 | 8 | 11 | 64 | 42 | **65** |

As a Liverpool fan, I like this one a lot more. It represents an alternative reality in which Steven Gerrard didn't trip over in the crucial game against Chelsea, and Liverpool went on to win their first league title in nearly 25 years. Gerrard would perhaps have taken his positive energy on to win England the World Cup and have been knighted Sir Stevie G. There are lots of possible simulated alternative realities, so I may as well choose the one I like best.

Unfortunately, the objective scientist in me feels he has to report the full results of all the simulations. It takes a couple of minutes on my laptop to run the Premier League 10,000 times, and each time I get something different. As interesting as each of these alternative realities may be, individually they are unimportant. What is important is to summarise what happens over all 10,000. How often did different teams win the league? When we do this, we see that Liverpool won in only 11.5% of the simulation runs. Manchester United, who had won the title the season before, won 26.2% of the runs. Chelsea won 19.2%, Arsenal 17.6%, Manchester City 12.8% and Tottenham Hotspur 6.0%.

With hindsight, we can see that these predictions were a bit out. Manchester United changed manager and had a terrible season. Manchester City and Liverpool dominated, both teams scoring more than 100 goals. But this isn't the point. I'm certainly not going to claim that I have already created the best model of football. We are just at the start of our story, and I wouldn't want to give all my modelling tricks away at once.

The important point is that, while it is not completely right, this model based on randomness isn't totally wrong, either. The teams predicted as likely league winners are those that usually do well, and the league tables above do look like potential outcomes for a season, or at least not too different from what we might expect. And we got this without any real thinking. We just simulated goals as going in at random, with each team having a different scoring rate, and out came a reasonable-looking final top four. This is almost the exact opposite of Paul Gascoigne's picture of unpredictable football. Football is *very* predictable. Every weekend in the Premier League season, over 400 players spend 90 minutes running around and kicking a ball, and at the end of the season a big club from London or Manchester wins.

Prediction based on randomness is a large part of how mathematics is used in society today. When you are waiting in a telephone queue, an analyst has already looked at the rate at which calls come into the call centre and worked out how long people are willing to wait on the line. When the bank lends money to a small business or to a new homeowner, it has already worked out the probability of bankruptcy, and applied the Poisson distribution to work out how many bankruptcies it will have to deal with in the coming years.

Prediction isn't about saying exactly which club will win the league, exactly how long you will wait in a telephone queue or which company will go bankrupt. It is about using the frequency of past events to calculate probabilities of events in the future. All these predictions arise from a mathematical model based originally on German soldiers being kicked by horses. If you'd like a verbal analogy, you could say that waiting for Liverpool to score is like waiting for the number 19 bus to arrive on a Bank Holiday Monday: you see nothing for ages and then two or three come along at the same time. Through the model, I have made this analogy useful. Mathematics allows us to pin down the features that bus arrivals, football matches, bankruptcies, cancer cases and telephone calls all have in common. It then allows us to predict how often they will happen.

## The Real Story

Even when goals go in at random, mathematics can find a way of making predictions. But Gascoigne does have a point. The real stories in football are not about the randomness, they are about rising above the randomness. They are about the setbacks and the comebacks. When Alex Ferguson retired in 2012, and David Moyes led Manchester United to their worst season in over 20 years, this couldn't be explained away by a run of bad luck. When Germany destroyed Brazil with five goals in 18 minutes in their 2014 World Cup semi-final, this wasn't just part of a random sequence of goals. Brazil collapsed under pressure, and Germany capitalised.

The success of Fergie or the German football team can't be understood in terms of randomness: to do that we need to think about their inner workings. The irony is that events that aren't random are the ones that are more difficult to understand, more difficult to predict, but also much more interesting.

In my research work, it is a lack of randomness that produces the biggest challenges. My biologist colleague comes back to me a few weeks later and says, 'When there's no predator around, the fish are spread out at random, but when they see a predator they form a tightly packed, rotating mill.' Now, there is a real puzzle. Does one fish initiate the mill? How fast does the mill rotate, and do certain fish prefer certain positions? Why are mills the best formation for evading a predator? The questions become interesting when the random model fails.

As I go deeper into modelling in the chapters to come, the problems I will look at become less random. Player movements are highly synchronised, their network of passes is structured, the ball moves according to the laws of physics and the managers think strategically about tactics. The models we will go on to look at are very different, but the basic approach I take will always be the same. I make observations, which give me a set of assumptions. I turn those assumptions into equations and investigate them using computer simulation and mathematical solutions. I then compare the model properties with data from the real world.

The challenge for an applied mathematician is to choose the right model for the question of interest. If we are just interested in predicting numbers of goals over a season, then randomness is often good enough. But if we want to understand formations, movement and skill, then we need to understand structure. Personally, I am not satisfied with random explanations – I want to find out what is really going on. To do this I need to get closer to the players and watch carefully what they are doing. And that's exactly what we will do next.

CHAPTER TWO

# How Slime Moulds Built Barcelona

My dad's theory about football is straightforward. Football is about taking the chances when they come and not making mistakes. As he sees it, the ball bounces around between the players, backwards and forwards; sometimes near one goal, sometimes near the other. Now and again an opportunity arises. A forward is well positioned and a midfielder in the same team has the ball. A pass cuts through the defence, who are caught napping. The striker manages to keep control and sticks it past the keeper. Then it's back to kick-off and the whole process starts all over again.

Admittedly, my dad's theory has come from watching the third-tier Scottish side Dunfermline Athletic. During the hours spent sitting in East End Park watching the Pars play, he has seen them go through countless line-ups and managers, and move up and down the divisions, but he has discerned little difference in their style. In his opinion, football comes down to an occasional piece of skill by the attacking team or poor communication by the defenders. The rest of it is messy.

Gary Lineker, Alan Hansen and the other professional TV pundits would probably disagree with my dad's analysis, even if they were forced to watch a Dunfermline v Cowdenbeath local derby. But when I visit my parents in Scotland, and we watch *Match of the Day* on a Saturday evening, I can sympathise with what my dad is saying. The TV analysis focuses on the acts of 'brilliance' and 'genius' by the forwards, or the 'diabolical' and 'shocking' defending on the part of the defenders. The whole discussion revolves around the goals and the near misses, and one or two players are identified as the heroes or villains. The tactics are mentioned briefly, in the form of the line-up shown at the start of the match, but these are quickly forgotten, and the focus switches to the individuals.

If there is one thing I've learnt from John Sumpter, it's intellectual honesty. He says things out loud which many people think but may be too embarrassed to admit. In some ways, he is right. It is difficult to make out what is happening on a football pitch, even when you go to the games week in, week out. We have already seen that randomness plays an important role in goals, but my dad takes it a step further. His argument is that the surprising and unexpected nature of football can be explained by a lack of genuine structure to the game. For him, it is glimpses of determination and individual skill, or lack of concentration and carelessness, that decide a match. The tactics are there, but the manager's most important job is to inspire the players to produce the right performance at the right moment. Could Gary Lineker, who spent a good deal of his playing career waiting for the ball to pop up in front of him so he could punish sloppy defending, sometimes think in the same terms as my dad? Just how important is a team's structure and formation – isn't it individual skill that decides a football match?

One way to understand structure is to zoom out. When I study schools of fish, flocks of birds or herds of mammals, I don't start by focusing on one individual animal. I take a wide-angle view and examine the group as a whole, looking from a distance at the twists and turns of starling murmurations, milling balls of mackerel or the fast-moving antelope fleeing from a lion. From farther away we can see what the group is up to as a whole. This wider perspective is harder to get from watching football on the TV. The cameras follow the ball around and focus on the star players. The bigger picture is lost, and the individual details are amplified.

If I am going to persuade my dad that there is a genuine structure to football, I will have to start with a bird's-eye view of the game.

## The Demise of One–Two–Seven

The widest view we can take of football is of formations. Formations are indicated by, for example, 4–4–2, 3–5–2 and 3–4–3, where the numerals represent the numbers of defenders, midfielders and attacking players when the teams line up. More complicated systems, such as the possession-intensive 4–2–3–1 or the 4–1–2–1–2 with its 'diamond', reflect how teams would like to play in midfield. These formations give a rough overall idea of the intended strategy and the roles of the players.

Formations are the first evidence that structure is important in football. Some set-ups work better than others. At the first ever international football match between England and Scotland, in 1872, both teams adopted top-heavy formations: England with 1–2–7 and Scotland with 2–2–6. Despite the emphasis on attack, the teams cancelled each other out and the game ended goalless.

Football has changed a lot since then. The changes are partly in the rules. At the time of that first international, attackers had to have at least three defenders between themselves and the goal to be considered onside. This explains why England used seven forwards: they formed an offside trap all the way up the pitch. But formations have also changed during periods when the rules have remained the same. Figure 2.1 shows the line-ups adopted by four of the greatest teams of all time: the Hungarian national side of the 1950s, the Inter Milan team of the 1960s, Liverpool in the late 1970s and the Barcelona of 2010/11.

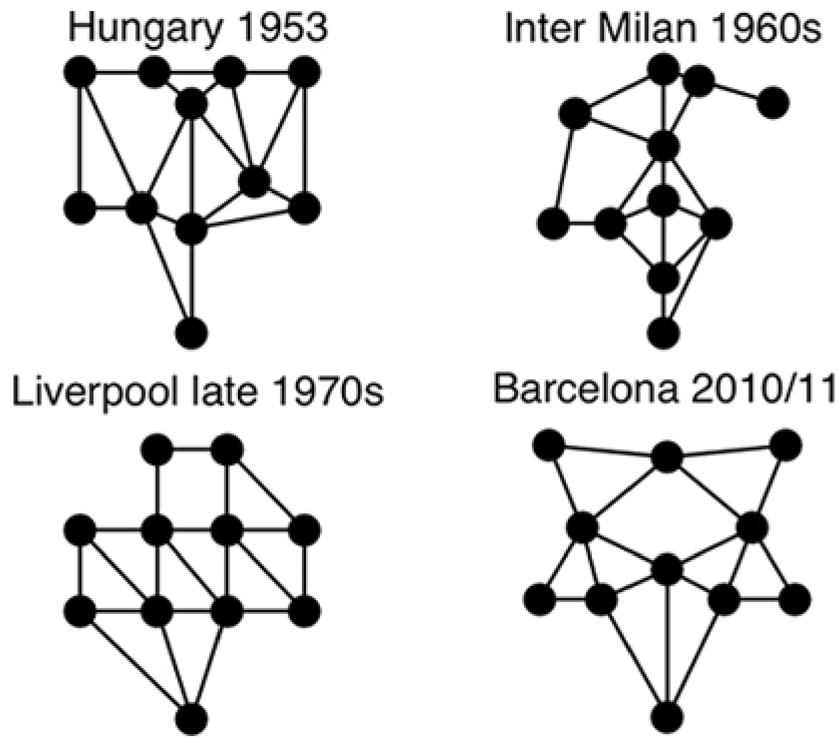



**Figure 2.1** Four formation networks from footballing history. In the Hungary 1953 formation, Hidegkuti is the central player directly behind the front four. For Barcelona 2010/11, Lionel Messi played as the centremost of the three forwards, Iniesta and Xavi were the left and right of midfield respectively, and Busquets played in the centre just in front of the back four.

These formations are presented a little differently from the ones we see on TV in the build-up to a match. First of all, I have taken away the names of the players so as to draw attention to the overall structure. Secondly, and more importantly, I have added links. These links are calculated based on a technique called the minimum spanning tree. I have calculated both the shortest and the second-shortest network that connects all the players, and drawn a link between players if they are included in either of these networks.[1] By connecting the players in this way, we can get a rough idea of how the team plan to move the ball around.

Looking at team networks allows us to see how tactics have evolved over the past 60 years. The first formation is of the great Hungary side of the 1950s, adopted from when they visited London in 1953 to play a friendly against England.[2] The player in the centre of the Hungary network, slightly behind the front four, is Nándor Hidegkuti. His position, with multiple links to other players, and his freedom to move allowed him to connect up players in a way that England were simply unable to cope with. The result was a 6–3 humiliation for the hosts. Hungary's final goal, a volley that earned Hidegkuti his hat-trick, was the culmination of a sequence of six keepy-uppies played over the heads of the bewildered England players.

The Inter Milan formation of the 1960s is often referred to as 'the net', and the network gives us some idea why. The midfield and the defence are a tangle of connections, making it difficult for an attacking side to pass through. Keeping a tight defence allowed Inter to hit opponents on the counter-attack. The Liverpool team of the 1970s and '80s filled the pitch with right-angled triangles, enabling them to execute their pass-and-move style. It was a simple but effective structure, where the players became interchangeable parts of a system. It may have been effective, winning both in Europe and in England, but the rigidity of the Liverpool network makes it not particularly pretty. Contrast this with the Barcelona team of 2010/11. Here, Xavi and Andrés Iniesta are links in a sequence of wide-angled triangles, with Lionel Messi lying at the tip of a diamond.

There are triangles in all the formations, but the ones in the Barcelona formation are particularly appealing to the mathematical eye. If you take any player in the team and rotate them through 360° you can see that they have nearby passing options in all directions. These options are equally spread out. So, for example, the central defensive midfielder, a role filled primarily by Sergio Busquets or Javier Mascherano in 2010/11, has five options, each consisting of the side of a triangle. He can pass directly back, as well as diagonally forwards or backwards to either side. Each player acts as a junction, where the ball can come in from one angle and can be quickly moved out in another direction. This makes it straightforward to do what Barcelona do best: control the ball and move it rapidly around the pitch.

## Train-track Triangles

Barcelona may have built the best triangles in football, but triangles were solving problems long before football came along. Consider the following problem. You are the mayor of a city consisting of a number of suburbs, and you want to build a train track connecting them all together. But you are short of cash and want to use the least amount of track possible. How do you connect all the suburbs with the least possible amount of track?

Figure 2.2 shows three plausible solutions for four suburbs. Have a look at these and think about which one uses the least track in total. If we use a bit of high-school trigonometry we can work out which is shortest. The one on the left uses three units. Each side is one unit long and we need three of them to connect everything up.

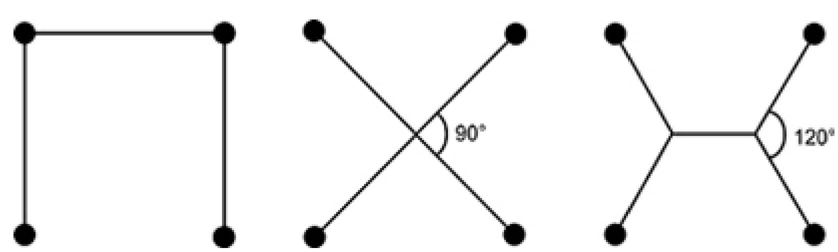



**Figure 2.2** Three potential solutions for linking up four suburbs (circles) with the least possible train track (solid lines).

The middle solution adds a junction in the centre, dividing the area into four equally sized triangles. The length of each of the two crossing lines can be calculated using Pythagoras's rule, and is $\sqrt{2}$. This gives a total length of $\sqrt{2}+\sqrt{2}=2.82$ units. This solution is like Hungary's positioning of Hidegkuti between midfield and the other forwards, or how Barcelona use Messi. Adding extra points gives triangles that reduce the total length of the connecting links.

If one extra connecting point is good, then using two is even better. The structure on the right of Figure 2.2 has a total length of $1+\sqrt{3}=2.73$ units,[3] less than for either of the other solutions. Triangles are again involved. Three branches come out from the connecting points at 120° angles. As is often the case in mathematics, the nicest looking, best balanced shape provides the best solution.[4]

Solving the problem of efficiently connecting four points on a square wasn't straightforward (and I'm not sure how many city mayors could do it). But that's only for starters. If you want to give yourself a real challenge, then try it for the five points on the corners of a pentagon. The answer again involves triangles, but the question is how to arrange them. Once you have done five, try the six corner points of a hexagon. It turns out to be an entirely new type of solution that works in this last case, but it still involves triangles. (For the answers, see note 4 on page 277.)

Let's make the suburb-connection problem really difficult. Let's think about how to solve the problem if we don't know where all the suburbs are, or even how many need to be connected. This is the problem faced by a slime mould called *Physarum polycephalum*. Slime moulds have no brain and consist of only a single cell. Their 'body' is a network of interconnected tubes that pump nutrients backwards and forwards. Slime moulds can be found on the forest floor or on trees. They usually cover an area smaller than a coin, but they can spread out quickly, shrinking when conditions are bad or expanding when there is a lot of food about.

It is when slimes are looking for food that they need to solve the suburb-connection problem. Inspired by this idea, my Japanese colleague Toshi Nakagaki decided to test whether slime moulds could build the Tokyo light rail and underground network. He and his colleagues laid out the slime's food as a scale model of the Greater Tokyo suburban area. They put oat flakes in a Petri dish: one large one in the middle to represent the city centre, and smaller oat flakes in positions corresponding to Shibuya, Yokohama, the airport at Chiba and other nearby areas. To solve the problem of connecting the oats, the slime mould would have to solve the same problem that Japanese town planners had solved when they designed the Tokyo transport system. Could the slime mould build efficient connections between its food resources?

The experiments worked perfectly.[5] The slime mould had no difficulty at all, building a network of triangles connecting up the oat flakes. Toshi compared the slime's solution to the real Tokyo transport network and found that, while they weren't exactly the same, they did have a similar structure. The slime mould's solution was just as efficient as the city planners', and a similar number of connections was used to link all the suburban oat flakes together. A comparison between slime and humans is shown in Figure 2.3.

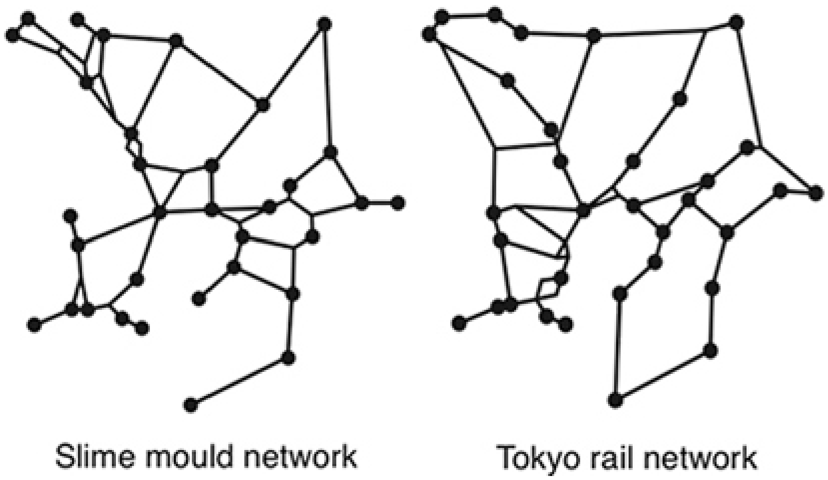



**Figure 2.3** Comparison of the network built by a slime mould to connect oat flakes (circles) laid out in the positions of Tokyo suburbs (left) and the real rail network (right). Reproduced with permission from The American Association for the Advancement of Science.

Triangular junctions are the central feature of the slime mould's tubular networks. Certain oat flakes become junctions that connect other points, so that the total length of tubes remains small. Notice that, like football formations, the angles at these junctions are large, and the network spreads evenly in all directions. Slime moulds don't build the absolutely smallest network that connects all the oat flakes: they also include a few loops, providing different ways of travelling between the same points. Toshi and his colleagues showed that these loops are very useful if the network is disturbed or broken. If one link in the network is cut, then the slime mould still remains connected and can shuttle resources using this alternative route. This is similar to when there is a breakdown on one part of the line on the underground. If the system is well designed, the whole network doesn't have to be shut down because of the failure of one line.

## Tiki-taka Tessellation

Networks of slime and railway services are physically very different from football formations. Football teams don't mark out their passing links with slimy tubes or steel tracks – they pass the ball to one another. But there are several recurring similarities. The first is the idea of covering the world with triangles. The slime moulds cover a small area of the forest floor, the Barcelona team covers the pitch with potential passes, and a good railway service covers the country with train lines. Another important similarity is that there are large angles between the different options available at connecting junctions. If we rotate through 360° round central points in the slime and rail networks, we find that there are equally spaced options in all directions, just as we saw for Barcelona.

There is a mathematical connection between wide-triangle networks and efficient use of space. The triangulation networks for famous football teams that I built can also be used to see how the team breaks the pitch down into zones.[6] For Barcelona, the resulting zones are as shown in Figure 2.4. On the left is the player network, together with the zones it creates (dotted lines). On the right the network is removed, leaving just the zones, and I've added which player played in which zone in 2010/11.

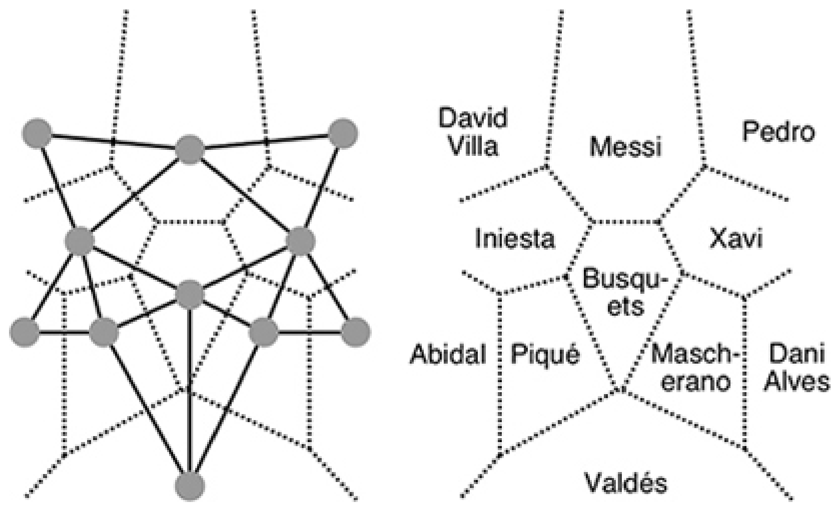



**Figure 2.4** The Barcelona 2010/11 network and zones. The formation network (solid lines) together with the zones (dotted lines) for each player (left); typical position for each player during the season (right).

The fact that the patchwork of zones in Barcelona's 4–3–3 has a similar symmetrical beauty to their passing network is not a coincidence – it is a mathematical necessity. If a team builds a network of wide-angled triangles then it also divides the area of play into well-spaced zones. Likewise, if each player occupies a clearly defined space, then a wide-angle network of triangles is created.[7] This point is crucial: it tells us that if we can solve one problem, that gives us the solution to the other one. If a team covers space well, the players will find that they have lots of good passing opportunities. If they move to receive a pass, they will find that they have created space. Players don't need to calculate all the angles to their team-mates, they just need to make sure they have enough space to receive the ball and pass it on.

Symmetry is the key to the style of play, often referred to as tiki-taka, adopted by Barcelona. Tiki-taka football is all about passing the ball rapidly between the players, with the aim of creating an imbalance in the opposition defence. To view tiki-taka mathematically we need to understand a bit more about how the zones are defined. We say that an opposition player is in Iniesta's zone if Iniesta is the closest Barcelona player to him. Each dashed line in Figure 2.4 marks the boundary between two zones.

When standing on any of these lines, an opposition player is equally close to two Barcelona players.

Imagine that I am standing on the boundary between Iniesta's zone and Messi's zone, so that I am equally close to them both. Standing on this line is probably the worst place to be on a football pitch, especially if Messi has the ball. If I move to close him down, the ball will quickly pass to Iniesta. If I back off to Iniesta, Messi is free to move forward. The lines in the diagram show opposition players where they *shouldn't* be when defending against Barcelona. Avoid the lines at all costs. Standing on them is like being caught in a footballing no-man's land.

## Flexi-zones

Train networks have very little flexibility. Once you have decided to build a line between York and London, or across Siberia, you are stuck with it. Slime moulds are a bit more adaptable. When food runs short, or part of the network is exposed to danger, the links contract and new links form elsewhere. But these changes still take several minutes or hours to complete.

Football networks have to be very flexible. Paths forward become blocked; new opportunities open up in the blink of an eye. Players who insist on waiting for their nearest team-mate to be unmarked before passing won't have the ball for very long. The structure of the team has to respond rapidly to changing conditions. Although a team can start with a particular formation, it has to be able to adapt its shape to the conditions, and this adaptation has to happen quickly. If your opponents spot an opportunity before you do, then quite soon you'll be chasing the ball instead of passing it.

All the movement and counter-movement on the pitch is what makes it so difficult to find a pattern. I think that explains where my dad's theory about football comes from: things change too quickly for him to keep track. Top footballers have spent tens of thousands of hours in training honing their reactions. They seem to respond almost intuitively to the game, moving immediately to find better attacking and defending positions. For many of us watching, it is hard to see where they are going or what they are doing, but for them it's second nature.

To get an idea of the structure, my dad just needs to get out on YouTube and watch some of Barca's greatest hits from 2008 to 2012. These videos usually feature Lionel Messi running past confused defenders, and he certainly catches the eye. But Messi is not the thing to concentrate on. Instead, download the video, play it in slow motion, and watch how his team-mates move around him. Typically, he will pass the ball directly to Iniesta or Xavi and then run forward, and a second later the ball is back in front of him. In the build-up to their best goals, Barcelona make four or five of these direct passes.

Now pause the video at exactly the point when Messi passes the ball, and look at how his team-mates are positioned.

At the top in Figure 2.5 is an example from a Champions League match between Barcelona and Greek champions Panathinaikos in 2010. Messi has the ball and is moving towards the goal, and two Panathinaikos players are moving to tackle him. The passing triangles show Messi's options. Xavi is directly in front of Messi, with Iniesta to his left. We can already see that their positioning is good because both can receive a direct pass. But by looking at the players' zones, at the bottom of Figure 2.5, we start to see just how good these positions are.

Both the defenders between Xavi and Messi are standing on their 'no-man's line'. They are running towards Messi, but it's too late – it's easy now for Messi to pass forward between them. Xavi returns the pass, and two seconds later Messi has the ball back again. In itself, this 'wall pass' between the players is school-playground football. But this passing is made possible by the way in which Xavi, Iniesta and Messi have broken the outside of the Panathinaikos penalty area into zones. Their opponents are left standing near the boundaries of these zones, unable to decide whether they should tackle or try to mark the receiver.

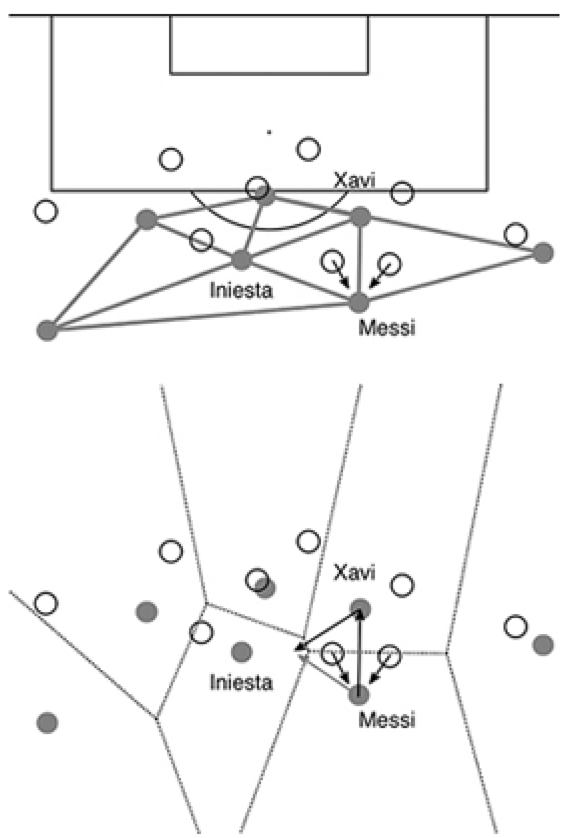



**Figure 2.5** Barcelona's positions outside the Panathinaikos penalty area five seconds before Messi's goal. Barcelona attacking bottom to top. Barcelona players' positions are marked with grey circles, Panathinaikos players' positions are open circles. Arrows indicate that the two defenders are running towards Messi. The passing triangles are given by the direct links between the players (top panel); the player zones are given by the dotted lines (bottom panel).

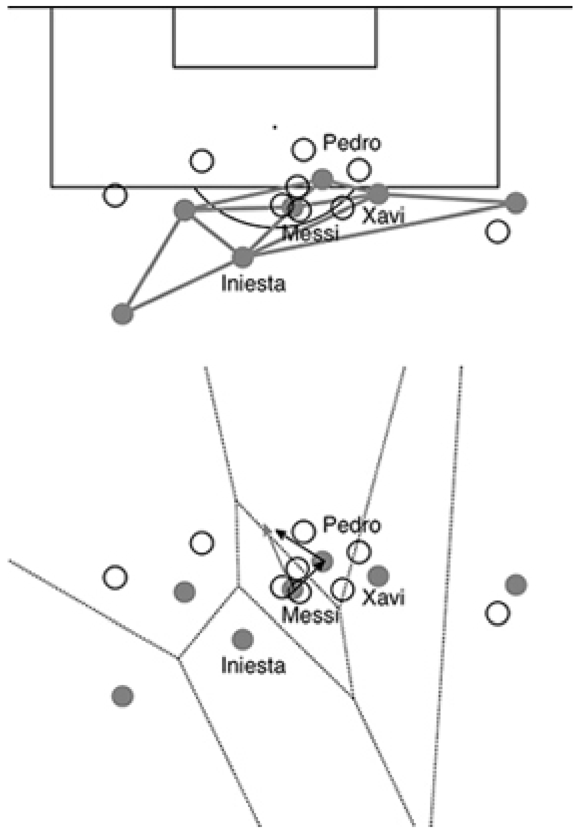



**Figure 2.6** Barcelona's positions outside the Panathinaikos penalty area three seconds before Messi's goal. See Figure 2.5 for explanation of the symbols.

Everything happens in less than two seconds. When Messi gets the ball back again, there is a new arrangement of players, shown in Figure 2.6. Two defenders throw themselves at Messi, but it's already too late. Pedro, who was marked by a defender on the corner of the box when Messi passed to Xavi, steps up to create a new zone. There are only a few metres between the players, but Pedro has moved away from his marker and is now standing perfectly positioned at the centre point of all four remaining defenders. One of the defenders is on the no-man's line between Messi and Pedro, one is halfway between Pedro and Xavi, and a third is stuck on the corner between all three. Pedro has created the maximum possible zone in the smallest possible space. Messi passes the ball out to Pedro, who taps it back, and Messi is one-on-one with the goalkeeper. The entire sequence of passing takes just four seconds, and one second later the ball is in the back of the net. Barcelona 3, Panathinaikos 1. Barcelona went on to score five.

Gary Lineker may have called this a brilliant run, while Alan Hansen may have picked out sloppy defending by Panathinaikos, but what is really going on is basic geometry. The Greeks invented the mathematical study of shape and position; these Barcelona footballers put it into practice. They have mastered the art of creating space on the edge of the penalty area. Most goals come not from a single act of genius nor a moment of inattention by the defence. They are the outcome of carefully planning how a team should play football. Whatever my dad may think, when football is played properly, goals are the outcome of the structures the players build as a team. When we slow the game down and look at the patterns, we can understand why certain teams do so well.

## Rules of Motion

Barcelona certainly taught the Greeks a lesson in geometry in 2010, but are the players doing the maths? Did Pep Guardiola sit down with the team before the match and go through a few triangulation and tessellation algorithms with them? As much of a genius as Guardiola is, I doubt he told Xavi, Iniesta, Pedro and Messi to create a Delaunay triangulation such that the opposition players would lie on the edges of their dual Voronoi diagram. All four of these players graduated from Barcelona's football academy, La Masia. The academy is world-famous for its footballing education, but doesn't include undergraduate studies in computational geometry. Nevertheless, these are the shapes and structures that the team creates. Barcelona use advanced geometry.

You don't need a complete mathematical grasp of geometry in order to use it. Schools of fish also 'use' a wide range of geometrical shapes. When they are travelling long distances, mullet adopt an oblong formation, with individuals packed more densely at the front. When under attack, sardines form rotating balls that expand away from the bills of hungry sailfish or other predators. But the creation of effective and beautiful collective patterns does not imply that the fish have understood the maths behind the shapes. It is unlikely that they have even understood which formation they are creating at any particular time. A fish in the middle of a rotating school just sees a lot of fish swimming forward. It can't tell how big a school it's in, and it may not even be aware that it's going round and round in circles.

The fish is just following a few of its neighbours and going with the flow. Instead of saying, 'Let's go round in circles' or 'Let's make an oblong shape', the fish follow a simple set of swimming rules. Fish tend to respond to the movements of the small number of other fish nearest them, primarily those in their field of view. They adjust their speed to stay together in the group, slowing down to avoid collisions and speeding up to avoid being left behind. When one fish suddenly changes direction, those nearby follow it. When researchers study these interactions, we often find that they are simpler than we first thought.[8] Simple solutions of speed changes and reacting to positions work better than complex solutions that involve calculating the positions of all the neighbours.

The simple rules adopted by fish provide us with a starting point for thinking about organisation in football. Our human brains are not necessarily any better than fish brains when it comes to following how other objects move around us. We can only track a handful of objects in detail, and the faster the objects move the fewer we can keep a handle on.[9] So while football players are able to react quickly to changes on the pitch, they cannot plan the exact arrangement of the team. When holding an offside trap, defenders have to keep their eye on one another and the opposition forwards. But it is impossible for them to know the positions of all 22 players and the ball.

Instead, players have to adopt simple rules for interacting with their team-mates. They have to know when to accelerate and when to slow down, as well as how to use space and respond to their team-mates. In the same way that natural selection has shaped the way fish interact, the thousands of hours professional players have spent on the training field shape how they move on the pitch. The former Ajax and Barcelona coach Rinus Michels used to emphasise that just practising 'wall' passing and shooting drills does not develop the sort of skill Messi showed when passing to Xavi and Pedro. Instead, coaches should plan exercises that teach players 'to recognise, in a flash, when a situation asks for playing a ["wall"] combination'.[10] These exercises involve developing a feeling for the game, so that the next step comes naturally.

As Messi made his run towards the Panathinaikos goal, Xavi, Iniesta and Pedro didn't start to calculate tessellations and triangulations of the defence. They probably didn't reflect at all about what they were doing. They employed the simple rule that they should move into space and pass the ball directly to feet. Now, in post-match analysis, I can admire the mathematical regularities of the passing network they created, but this emerged from the style of play they adopted. Just as a flash expansion of a school of fish away from a predator emerges from movements of the individual fish, so a goal emerges from a simple set of movements carried out by the players.

The goal Messi scored, and many others like it, came from a set of rules laid down many years earlier. When Barcelona copied the Ajax academy and set up La Masia, they put in place a system already tried and tested not just in Amsterdam, but also by millions of years of evolution. Slime moulds have mastered triangles, and fish have mastered changes in speed and the use of space. Barcelona wanted to train players who could master all these skills. La Masia didn't need to teach them advanced geometry, it just needed to make sure their youngsters had the right rules of motion. Those rules were established on the training field: learning to pass and move, to twist and turn. When Messi found himself outside the box with nine Panathinaikos players between him and the goal, he didn't need to think. He just carried out what was, to him, the simplest and most natural movement in the world.

CHAPTER THREE

# Check My Flow

When my son and daughter were about six or seven, and started playing football, the first thing I noticed was The Clump. Most of the players, goalkeepers included, would charge towards the ball as it bounced around at random into feet, heads and hands. A few solitary players departed The Clump, maybe to pat a passing dog, pick flowers or simply lie face down on the ground, but the majority joined the chase. Occasionally, the kid lying face down would be the best-positioned player on the pitch as the ball inexplicably flew out of The Clump and onto his or her head. But on the whole, the most successful players were those who were most determined to get the ball and could push the others out of the way most efficiently.

Some adults seem to believe that The Clump is an unavoidable state of children's football, and that it isn't until the children get older that they can understand positions and tactics. Worse still, some parents see The Clump as a method of sorting the wheat from the chaff. The boys and girls who are willing to get stuck in at an early age are the ones who are made for football, while those running around at the touchlines pretending to be an aeroplane should try another sport, such as badminton or table tennis. All too often, those who find themselves inside The Clump continue with football and those on the outside spin off into another activity.

As Barcelona's La Masia has proved, The Clump has nothing to do with good football. Football is about structure, and creating structures doesn't require players to follow complicated rules. The question at all levels of football, from six-year-olds to professionals, is how to get these rules into the minds of the players: how to make them play in a way that creates opportunities and makes football a proper team sport. The answer is to get players thinking about movement and positioning.

Movement and positioning are central to my own research. The locusts, sticklebacks, chickens and meerkats I have studied in my professional life may not be quite as chaotic as kids' football, but the challenges in making mathematical models of them are similar – as are the challenges in understanding the professional game. I want to know which dynamic rules the individuals use. Cannibalistic locusts chasing their neighbours to take a bite out of them; sticklebacks following their neighbours to get away from a predator; chickens pecking at the ground near their neighbours to find food; meerkats calling to their companions before setting off for home – all are examples of these dynamics. Kids or professionals chasing or passing a ball is another.

So when I, together with a bunch of other eager dads, took on the job of managing my son Henry's team, I knew there would be interesting dynamics involved. And I wanted to know how they worked.

## Why the Piggy Always Wins

The simplest passing exercise is the footballing version of piggy-in-the-middle. Two 'attackers' stand on opposite sides of a square and pass the ball to each other, while one 'defender' in the middle tries to intercept the ball. In Figure 3.1, the attacker on the left has the ball, and the defender is trying to block a pass.

When the defender is in the middle, it is simple for the attackers: the player with the ball passes, and the receiver runs down the line and collects it. The defender is poorly positioned and it is easy to make the pass. Once the pass is made, the defender must decide where it is best to move to in order to prevent the next pass. She has two options: either she can close down the player with the ball or she can mark the receiver. It depends who she is nearest to. If she thinks she can get in front of the player with the ball before a pass is made, she should go towards the ball. Otherwise she should move in front of the receiver.

A good way of summarising how the defender should move is with a diagram called a flow field – see Figure 3.2. The arrows show how the defender should move, depending on her current position. If she is near the player with the ball she should try to block. Otherwise, she should run towards the receiver and try to intercept. The flow field gives us a single picture that explains what the defender should do in all possible situations. The attackers can move, of course, but this will just change the directions of the arrows. The flow field summarises the entire dynamics of piggy-in-the-middle.

This picture reveals why piggy-in-the-middle is not a good training exercise. The defender has two options, both of which allow her to take the ball. If she follows the arrows towards the player with the ball, she can block the pass as it is made. If she places herself in front of the receiver, she can intercept the pass before it arrives at its target. Both these positions, in front of the player with the ball or in front of the receiver, offer the defender a strong advantage. Whatever the players with the ball do, the defender will eventually get closer to one of them and prevent passing.

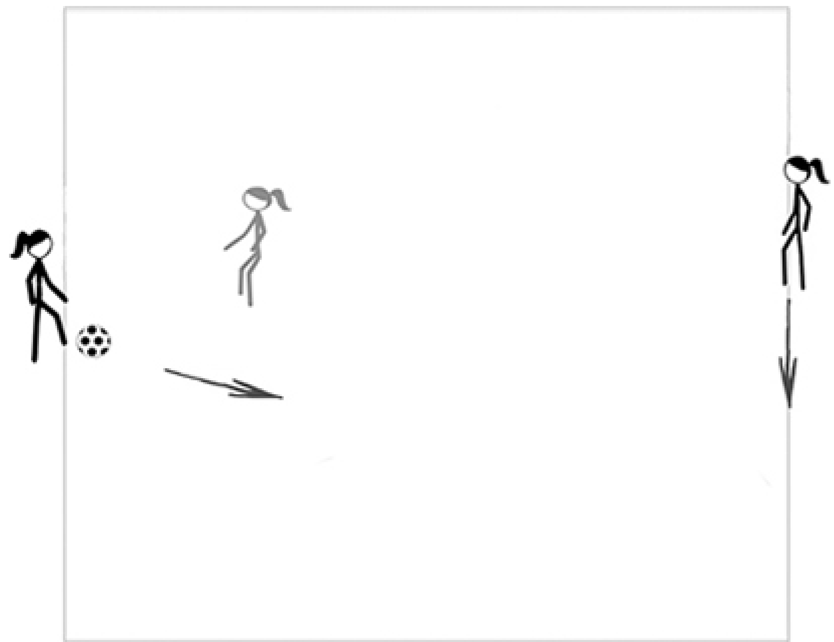


**Figure 3.1** Piggy-in-the-middle. The two attacking players on the edges of the square attempt to pass the ball to each other. The defending player in the middle tries to intercept the pass.

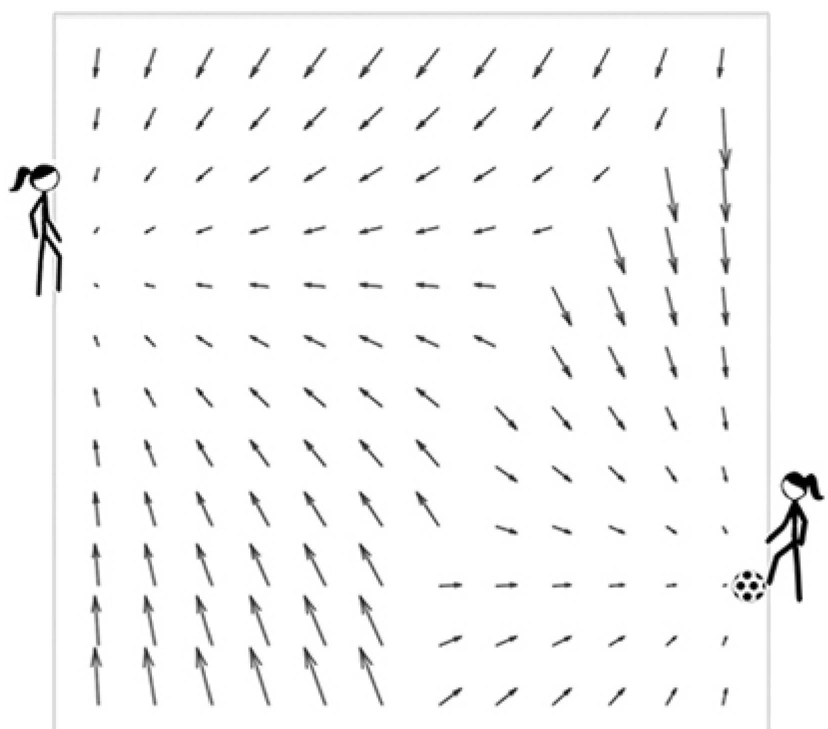


**Figure 3.2** How the defender in piggy-in-the-middle should move dependent on her current position. The arrows show, for each position in the square, the direction of movement of the defender.

Flow-field theories like this are fine, but they need to be tested. So I set up the piggy-in-the-middle exercise with Henry and two of his friends, Frank and Edvin, who by now were 10 years old. They had learned a lot during the four years since the team was set up, and were keen to try out my new training exercise. We put down cones to mark the four corners of a square, and asked Henry and Edvin to stand on opposite sides and pass to each other.[1] Frank was the piggy defender. We placed sports GPS trackers on the three boys' shoulders to track their movements, at five measurements per second. Unfortunately, the Ekeby Astroturf where we train isn't yet equipped with multiple cameras and ball-tracking technology, like those used in Champions League matches, but these positional measurements were sufficient to work out how the boys moved around.

The results confirmed my predictions. Henry had the ball, so Frank ran up to Edvin and stood just in front of him. Edvin tried to move along his line but Frank followed. Henry dribbled backwards and forwards, but couldn't find a way to make a pass. Then he stopped, looked at me and started shouting that this was a rubbish training exercise. When Henry did finally try to make a pass, it was easy for Frank to take the ball. We repeated the exercise a few more times, and Henry's complaints got worse and worse. Eventually, after one attempted pass, the ball bounced off Frank's foot and into Edvin's face. I had one kid shouting at me, one lying on the ground crying and clutching his head, and Frank was left standing in the middle feeling guilty about the whole thing. No, piggy-in-the-middle is not a good training exercise.

The problem with piggy-in-the-middle isn't that it's played by children; professional players would have had the same frustrations as Henry and his friends. Training exercises involving straight-line passing don't create good flow. Instead, we need to find exercises that create opportunities and movement.

To give the boys a chance to train properly, we got their friend Elias to join in. He stood at the bottom of the square. The only option now for the defender is to run towards the ball and try to block before a pass is made, because he can't mark both free players. The attacking players should pass the ball as soon as possible after they receive it, preventing the defender from closing them down. The track from Frank's GPS for this four-player set-up is shown in Figure 3.3, together with an illustration of the passes made.

The wiggly line from the GPS shows how Frank ran during a 63-second sequence. The point labelled 'Frank' is his position when he finally intercepted the ball, and the line shows his path up to that point. The line is like a vapour trail, so that the intensity shifts from light grey, through darker grey, to black as we move forward in time. At the start of the sequence, where the line is faintest, Frank challenges Henry, who passes to Elias. The pass is shown as the arrow on the left of Figure 3.3. As Frank runs toward Elias, the ball goes back to Henry. Frank follows, but before he gets there Henry has passed out to Edvin. He and Henry are then able to pass the ball a few times between them, while Frank runs backwards and forwards. Frank defends very well in this situation, returning to the middle when uncertain so as to maximise his chances of interception. Eventually this pays off when Henry and Edvin make one pass too many in the same direction, and Frank gets the ball.

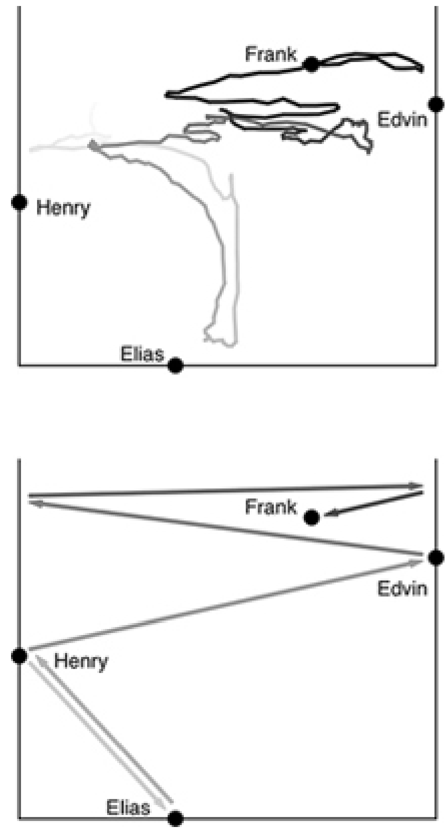



**Figure 3.3** Training exercise with 10-year-olds, showing Frank's position (top panel) and the passes made by Henry, Elias and Edvin (arrows in bottom panel). Darker shading indicates more recent events in time.

Professional football players use some variation of this exercise at almost every training session. They typically work on all four sides of the square, sometimes using smaller technical squares with one defender inside, other times with larger passing squares and two defenders. But the crucial step in making this exercise effective is the creation of dynamic triangles. For only two players and one chaser, as in traditional piggy-in-the-middle, the chaser just needs to mark one of the players and they then control the game. Introducing a third or fourth attacker adds an extra dimension. The passing square helps players to learn not only how to pass accurately under pressure, but also how to carefully track and create new opportunities.

These training-square exercises are simple compared with the rapid activity of a professional football match. However, my flow-field analysis of piggy-in-the-middle, Frank's vapour trail and the ball-passing arrows are a starting point for a whole range of methods of studying movement in match situations. These methods are just as relevant to the Champions League as they are to children's football. The key is identifying the players' flow: we need to find out where each individual's arrow points and then look at what happens when many arrow-following individuals interact. The principle is the same for two teams of 11 professional players, a crowd of supporters leaving the ground after a 0–3 home defeat, or even a V-shape formation of migrating birds. The task for a modeller is to find out where the arrows point and what happens as the members of the group follow the flow.

## The Right Social Convention

You may not notice it, but flow fields govern many of your everyday movements. For example, you are walking down a narrow corridor and a stranger is coming in the opposite direction. You have to let them pass, but do you move left or right? This is something that happens to us every day, but it can still feel a bit awkward. There is a convention in the UK, enforced by signs on the London Underground telling commuters to keep left. But it isn't unusual to find yourself dodging backwards and forwards trying to second-guess what the other person will do. When I lived in Oxford, I sometimes found myself involuntarily riding my bike into groups of foreign tourists, as I failed to work out which way they would move. Now that I live in Sweden, I am the hapless foreigner who doesn't know if he should be on the left or the right.

It might appear that strong 'keep to the left' and 'keep to the right' conventions are the key to avoiding collisions. Perhaps it is these conventions that are to blame for locals and tourists bumping into one another. However, if we look closely at the dynamics, we find that things aren't that simple. This is exactly what Mehdi Moussaïd did as a PhD student in Toulouse, France. He set up a narrow corridor and assembled a bunch of undergraduate students to look at how they walked past one another.[2] In his first experiment, Mehdi asked one undergraduate student to stand in the middle of the corridor. Then he asked another student to walk past the stationary student.[3]

By filming the students from above, Mehdi was able to construct a flow field, just as I did for my piggy-in-the-middle game, but now using observations of adult pedestrians. Figure 3.4 shows the flow field for how the walking students moved to avoid collisions with a stationary student. Each arrow is the average reaction of students to the presence of a stationary person in front of them. The most obvious trend is that the moving students try to walk around the stationary students. All the arrows point away from him or her, which in itself is not too surprising. Once the walking student is about two metres away from the stationary student, there is a strong tendency for the moving student to veer off to the side. After all, we don't like walking straight into people.

What is surprising is that there was only a very weak preference for passing on the left or on the right. If you look very closely at the arrows in Figure 3.4, you can see that the ones pointing to the right of the stationary student are only slightly more prevalent than those pointing to the left. For these French students there is a weak bias towards passing on the right-hand side, but so weak that it is only just statistically significant.[4]

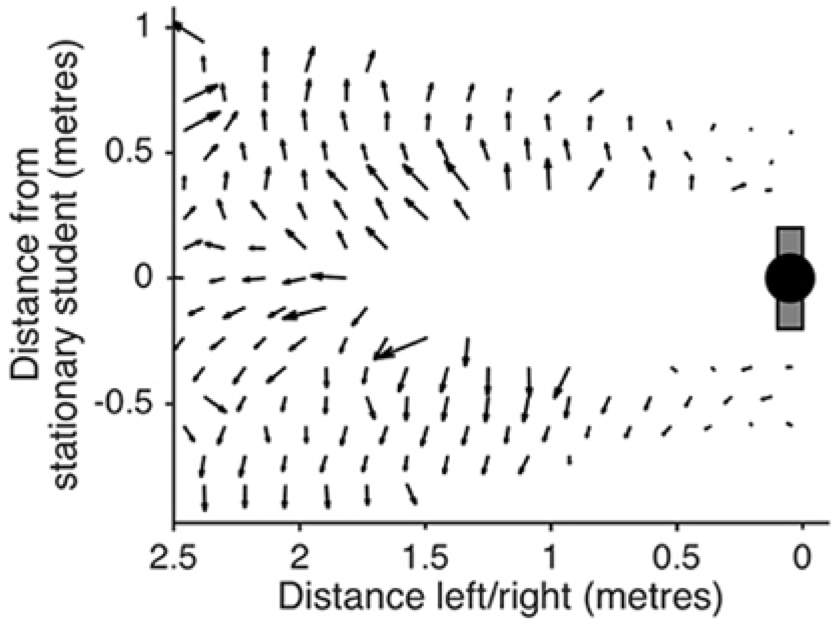



**Figure 3.4** The effect a stationary student (positioned on the right of the figure) has on a student walking toward them in a corridor. Each arrow is the average reaction of the moving student to the presence of the stationary student. An upward-pointing arrow indicates a tendency to pass the stationary student on the left, while a downward-pointing arrow indicates a tendency to pass on the right. Adapted from original figure by Mehdi Moussaïd.[3]

The power of mathematical models is that they allow us to extrapolate from what happens in one situation to predict the outcome of other situations. Mehdi used the outcome of the first experiment to predict what would happen for two people moving towards each other. He created a simulation in which two moving students follow the dynamics shown above, interacting as they come closer together. The simulated students approach each other in the corridor, simultaneously adjust their positions and successfully manoeuvre past each other.

Mehdi also noticed something else in his simulations. The simulated students passed to the right of each other in 77% of his simulations and to the left in only 23%. This was a much larger bias towards passing on the right than in his experiment with a stationary student. Mehdi could now test the model. He did an experiment where two students walked towards each other and passed in a corridor. The results of the experiment were essentially the same as those from the model: 80% of passing events were to the right, and 20% to the left. Moving French students showed a clear right-hand bias.

Mehdi's model and experiments challenge the idea that we have a strong ingrained national preference as to whether we pass people to the left or to the right. Individual French students had a weak preference for passing to the right. The stronger convention of right or left passing shows itself when people interact. As individuals move towards each other, they constantly update their direction and the small bias is reinforced. It is then that the right-hand bias appears.

Social conventions like these appear very strong when we look at a culture as a whole, but when we look at individuals they are difficult to detect. Mehdi's results explain why we need to work so hard to avoid bumping into tourists. The problem is not that the tourists stubbornly stick to their own conventions, even when they are abroad, but that the preferences are so weak that neither native nor tourist knows which way to go. Conventions only appear through repeated interactions.

## The Unbeatable Back

In football, an attacker's flow field would look very much like the one found by Mehdi for moving pedestrians. The attacker wants to get past the defender and will choose the shortest possible route. The conventions are very different for defence. Defenders don't try to avoid bumping into the opposition. Their first aim is to get into the path of oncoming forwards, and definitely not to politely let them go past. A defender's flow field points towards the player with the ball, not away from them. In one-on-one situations, if the attacker moves to his left, then the defender moves to his right, and *vice versa*. The attacker then has to fool the defender that he is going one way and take the other.

It is in this cat-and-mouse game between the defender and attacker that much of the individual skill comes into football. Dummying, stepping over the ball, switching between the inside and outside of the foot, and dropping the shoulder are all tricks that forwards can do to confuse defenders. Each dribbling master has their own speciality: the Cruyff turn, the Maradona double-swivel, the Messi shoulder-drop, the Ronaldo step-over – all designed to befuddle the defender.

Fooling a defender is harder than it looks on TV. In true one-on-one situations, where the defender is between the goal and the attacker, the odds remain in the defender's favour. While a defensive mistake is extremely costly, the onus is on the attacker to make the first move. The attacker has to cover more ground in order to round the defender while keeping the ball. For the defender the important thing is not to be fooled by the fancy footwork, and to watch the ball instead.

The Bayern Munich defender Holger Badstuber describes one-on-one defending as 'the great art'.[5] According to him, it is important to put an opponent under pressure immediately, to get very close and not give them any space. At the same time, a defender shouldn't go in too early, because this gives an experienced attacker the chance to take advantage. Instead, the defender should 'show' the attacker a route away from goal. The best defenders narrow down the options, and when the attacker is forced to make a move they recover the ball.

Selina Pan and her colleagues at the University of California at Berkeley developed a mathematical model of a defensive strategy similar to that described by Badstuber. The setting in their paper, entitled 'Pursuit, evasion and defense in the plane',[6] was slightly different from in football, but the principles are the same. Selina created simulated defenders and forwards and programmed their interactions. The defenders try to catch a forward before the forward crosses a target line. In her model, there is no ball and the defenders have to get sufficiently close to the forward to stop them, which is more like rugby or American football than 'soccer'. Another difference from a pure one-on-one is that she included two defenders: one defender between the goal and the attacker, and another who is pursuing the attacker. This is tougher for the attacker, but it is also more realistic. In a match situation, forwards have only a fixed amount of time to get past the last defender before the other defenders get back behind the ball.

The defensive algorithm proposed by Selina and her colleagues is all about narrowing down the space available to the attacker. Just as the secret to Barcelona's attack is creating spatial zones, the secret of good defence is to decrease the size of these zones. In Selina's algorithm, the defender in front of the goal first moves directly towards the attacker. Once the simulated defender is as close to the target line as the attacker, the defender moves instead towards the endpoint of the target line, blocking the attacker's advance. The algorithm follows Badstuber's advice of showing the opposing forward which way to go: the defender needs to get as close as possible while still preventing progress. The second defender's algorithm is straightforward: chase the defender and reduce the space they can move into.

To illustrate how their algorithm works, Selina and her colleagues wrote a computer game where the human player controls the forward, and the computerised defenders follow their zone-minimising algorithm. I adapted their game for a footballing situation, so that the aim is to take the attacking player past the two computerised defenders to the edge of the opponent's box. Figure 3.5 shows four of my many attempts to get the forward into the box.

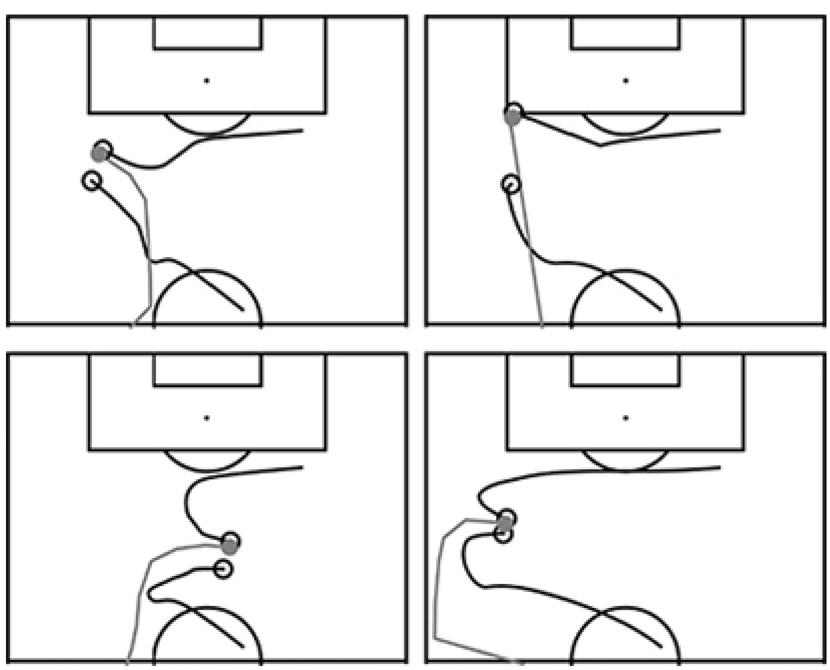


**Figure 3.5** My attempts to get past the defenders using Selina Pan's zone-minimising algorithm. My attacking player (solid circle) attempts to move past the defenders (open circles). The lines show the trails of how we moved during the game and the circles are the endpoints.

Winning this game proved impossible. While I could sidestep one of the simulated defenders, the defender nearest the goal always got in my way. When I tried to go down the middle, that defender pushed me out to the left. When I tried to get to the left-hand corner, the defender got there first. When I tried to dummy either through the middle or round the side, then both defenders narrowed me down. The defenders always won.

It wasn't just me. Beating the computerised defenders is impossible in theory as well as practice. Selina and her colleagues went on to give a mathematical proof that their zone-minimising algorithm always wins. Unless the attacker is closer to a point on the line he is trying to get to (in this case the edge of the box), then the defenders always catch the attacker. This explains why forwards such as Cristiano Ronaldo, Neymar and Luis Suárez are so much more revered than defenders like Badstuber. While defenders may well be great artists, when forwards go past them they are achieving the impossible.

The game of attack and defend is also one of nature's greatest arts. In creating her model, Selina was inspired by research on group hunting by lionesses. She cited work by Philip 'Flip' Stander, who runs a conservation project for lions in semi-arid and desert areas of Namibia. In the dry season prey animals are far and few between, and the lions can't afford to let their dinner escape.

Flip watched and recorded how the lionesses (it is the females that do much of the hunting) stalked springboks, zebras, wildebeest and even giraffes.[7] The pride works together, each of the lionesses adopting her own preferred role as they get close to their prey. Some are 'wingers', approaching from the side, while others are 'centres', approaching slowly or waiting in front of the prey. The wingers initiate an attack, while the centres wait and ambush once the prey tries to make its escape. Flip noticed that when the majority of the pride, which typically consists of around five to seven lionesses, hunt together, they capture prey more than twice as often as when only half of them hunt. Hunting in the desert is a team sport.

A lioness's position within her team is consistent from one hunt to the next. In one of the hunting groups Flip studied, from the Okondeka pride, he identified three centres, one left-winger, one right-winger and two lionesses that could occupy either wing. The lions then co-ordinate their positions to keep their prey between them. For example, when Okondeka's left-winger was missing for a particular hunt, then one of the more flexible wingers would take up the position on the left. With the hunters moving in on the prey from all sides, the prey's escape options are rapidly narrowed down.

When the chase starts, the lionesses don't need to communicate or plan. As I found when I played Selina's computer game, the zone-minimising algorithm produces the same narrowing-down of options as during a lion hunt: I always ended up sandwiched between the two computerised defenders. These defenders did not actively communicate with each other – they each used their position relative to my forward to decide how to move. Similarly, the lions each take up a good starting position, rapidly reducing the space available to the prey before going in for the kill.

As in group hunting by predators, the use of complementary roles is key to football defence. Just like lions, defenders hunt in pairs or packs. The first defender puts pressure on the attacker, attempting to force a mistake. A back-up defender gets into a position where they can intercept the ball if the attacker tries to pass. A team playing the 'pressing' style of football starts to narrow down their opponents as soon as an opposition player has the ball, preventing them from gaining space.

In 2013, Bayern Munich delivered a footballing demonstration of the principle found in lions by Flip and proved mathematically by Selina and her colleagues. Bayern adopted a pressing style when they took on Barcelona over the two legs of their Champions League semi-final. Instead of allowing Barcelona to build up, the Bayern forwards closed down the Barça defence at every opportunity. In midfield, Iniesta was denied any space by the physically intimidating and larger Bayern players.

Badstuber missed those semi-final games through injury, but I am sure that he enjoyed seeing how his 'great art' of defence defeated Barcelona's space-creating passing game. There is more than one way to play mathematically elegant football. While the Barcelona team of 2011 efficiently maximised the available space, the Bayern team of 2013 effectively minimised it. Bayern won the semi-final 7–0 on aggregate.

## Picturing Pirlo

Good defenders study in advance whether their opponents' forwards are stronger on their left foot or their right. But movement analysis can reveal more detailed player profiles. Using the tracking data of player positions that is now collected by multiple cameras at all top-flight matches, it will soon be possible to create dribbling flow fields for opposition strikers. Using the methods that Mehdi Moussaïd applied to pedestrians, scouts will be able to assess whether particular players have a tendency to try to go past a defender on the left or the right, and how that choice depends on distance from the defender. Instead of asking scouts to spend hours watching TV footage, managers could press a button on their phone at half-time and get a full graphical picture of how each opposition player moves in different situations.

More work needs to be done before graphical analysis can completely replace scouts, but it is already possible to use flow fields to characterise individual players. The Italian midfielder Andrea Pirlo is famous for combining laid-back movement with an ability to pass the ball with pinpoint accuracy. He seems to meander around the midfield at walking pace, but still manages to be perfectly placed to deliver an incisive and game-winning pass forward.

One of Pirlo's greatest performances was during Italy's semi-final against Germany in Euro 2012. He controlled the ball in midfield, making short passes near the centre circle, before launching longer passes up the field. It was with one such pass that he set up Italy's first goal. Figure 3.6 illustrates the sequence of passes made by Italy in the 30 seconds leading up to this crucial strike.

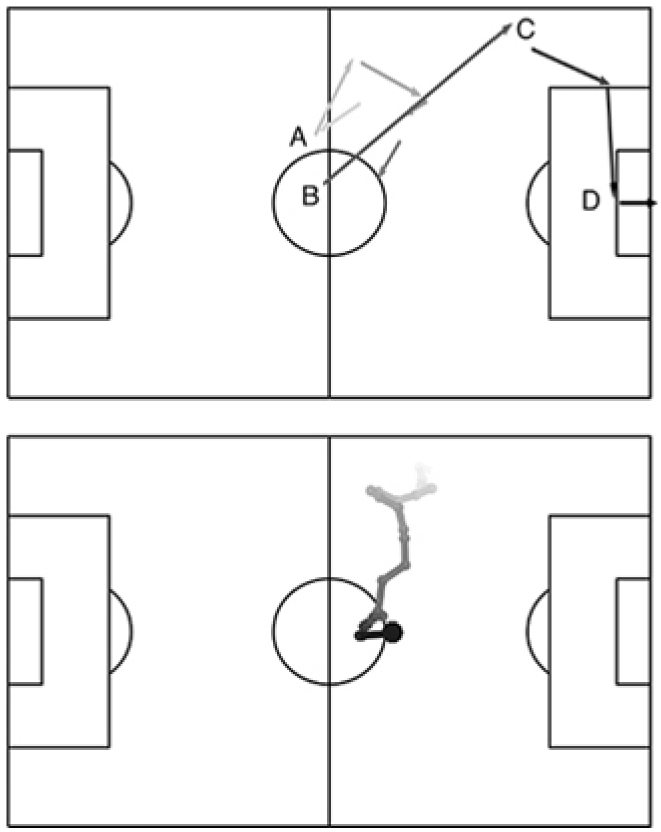



**Figure 3.6** Passes leading up to Italy's first goal against Germany in Euro 2012, showing passes made by Italy (arrows in top panel) and Mesut Özil's movement while chasing the ball (bottom panel). Darker shading indicates more recent events in time. Letters indicate events: (A) Pirlo's first pass; (B) Pirlo's second pass; (C) Chiellini receives the ball; (D) Balotelli scores.

The passing sequence starts from a throw-in, and the ball comes quickly to Pirlo. At this point, there's no clear opportunity for Pirlo to make an attacking pass. Mesut Özil, who has brushed himself off after having a claim for a foul denied, is near to Giorgio Chiellini, closing off any possibility of a pass down the left. So Pirlo rotates the ball round his team-mates, and after four passes it comes back to him. While the ball rapidly switches between the Italian players, Özil presses, trying to recover it. But for Italy, this is training-ground football. They are close together, playing the ball as they would inside a coaching square. Özil is made to run round and round in circles, just like Frank in my GPS experiment.

Unlike Frank, however, Özil doesn't manage to get the ball back. Instead, when Pirlo receives the ball the second time, he takes a few soft steps back behind the halfway line, looks up, and sees that there is now space on the left where Özil was earlier. He sends a long ball directly onto the chest of Chiellini, who is standing on the touchline. Chiellini passes the ball into space in front of Antonio Cassano. Cassano turns a defender and crosses the ball to a waiting Mario Balotelli. Super Mario heads the ball past the keeper. 1–0.

It's all very well admiring one goal, but my aim is to give a complete picture of what Pirlo is doing, just like the picture Mehdi Moussaïd created of French pedestrians. To capture the Pirlo style, I mapped out his passes from the entire match against Germany at Euro 2012. Pirlo made 66 passes in the game, 61 of which were successful. Figure 3.7 is a heat map and flow field of his passes during the game. The shading indicates the place on the pitch where Pirlo made passes from. We can see that Pirlo made a lot of passes from near or just behind the halfway line, usually from the middle of the pitch. The arrows indicate the general direction of Pirlo's passes made from various positions. These arrows are not the actual passes themselves, which were sometimes long passes from one side of the pitch to the opposite wing, and sometimes short passes to nearby team-mates. The arrows are instead a statistical fit that give a general picture of the direction in which Pirlo tends to move the ball, and from where.

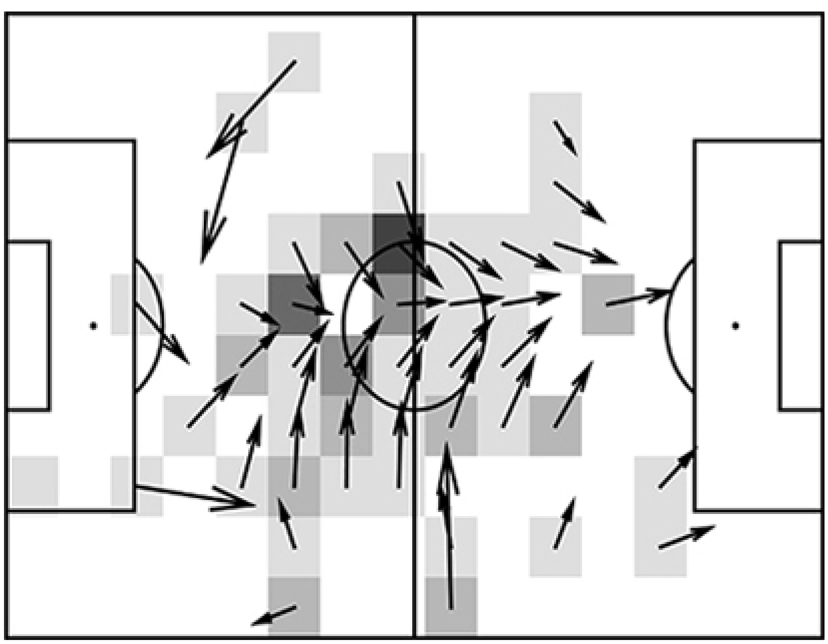


**Figure 3.7** Pirlo's passing during Italy's semi-final against Germany in Euro 2012. The intensity of the shading at each particular position on the pitch indicates how often Pirlo made a pass from this position.The arrows are a statistical fit to all 66 passes he made during the match. The average direction of passes is indicated by the direction of the arrow. The average length of the arrow is proportional to the average length of a pass.

The statistical fit I have used is created using a toolbox my research group developed.[8] The methods are similar to those used in weather forecasts, where meteorologists try to predict the movement of weather fronts. In weather forecasting, the meteorologists have previously observed fronts moving under different conditions across different parts of the country. Their forecast is then a statistical prediction of the most likely future development for each area. Similarly, when we look at the passes made by a player, we see that they are made from various positions on the pitch and in different directions. But if we assume that nearby areas of the pitch are likely to generate similar types of pass, then we can smooth out the differences. This smoothing reveals the general pattern for Pirlo, enabling the complicated passing data to be summarised in a compact but meaningful picture of the overall flow.

From the passing arrows in Figure 3.7, we can see that Pirlo tends to switch the side of play when he passes the ball. The lengths of the arrows are proportional to the typical length of a pass, but shorter than the actual passes made. Pirlo's passes typically switch the attack to the opposite flank. If he is on the left side of the pitch, he passes over to the right, and *vice versa*. In this match against Germany, Pirlo had a slight bias for passes to the left. This is exactly the area from which he created the goal.

Toni Kroos was given the job of controlling Pirlo. A few days before, in their game against England, Pirlo had played in almost exactly the same style as he did against Germany, successfully completing an incredible 115 passes. I don't know what homework Kroos did before the match, nor do I know what picture he had in his mind about how Pirlo would play, but Kroos failed to keep him under control. Italy won 2–1, and Germany were sent home from Euro 2012 to think things over before the World Cup.

In the future, when Kroos and other players are given the task of controlling a key opposition player, the manager may well send them a few flow-field maps of their quarry. These pictures won't tell them everything they need to know, but they will capture important aspects of their opposition's playing personality. Studying these images is like watching a weather forecast: they shouldn't completely determine how you plan your day, but they do ensure that you'll be prepared for the worst.

While Germany lost in the end, it was a close match. They had 14 corners and 20 shots, but only got a single goal – through a penalty by Özil in the dying minutes. Germany's own midfield star, Bastian Schweinsteiger, played in the same central-midfield position as Pirlo and made a similar number of passes: 60 completed passes out of 74 attempts. But Schweinsteiger's flow field looks very different from Pirlo's – see Figure 3.8. Schweinsteiger's passes came from all over the pitch. Sometimes he was forward on the right, crossing in towards goal; at other times he was back on the left laying the ball out to the wing to build up an attack. While Pirlo is the calm in the eye of a storm, Schweinsteiger is a whirlwind of energy engulfing the entire field.

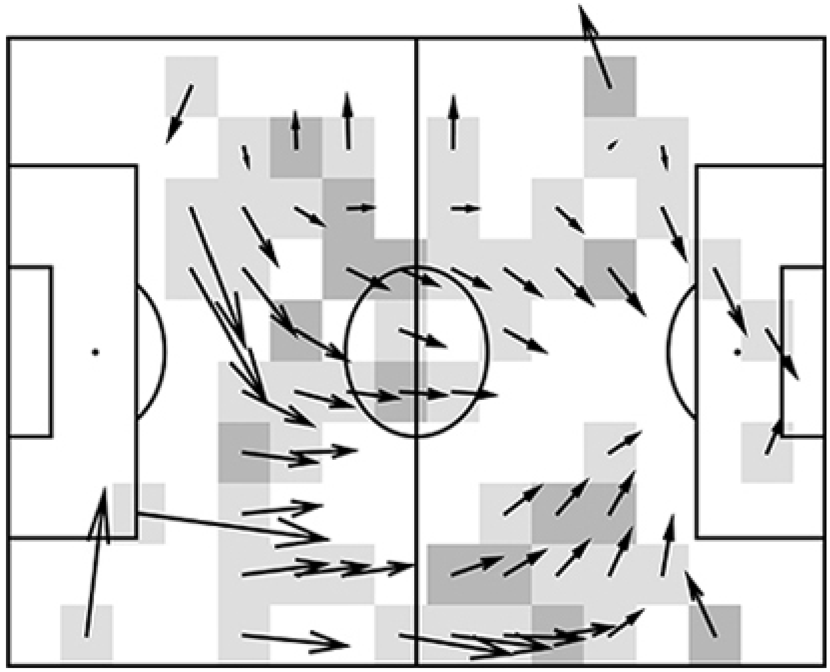


**Figure 3.8** Schweinsteiger's passing during Germany's semi-final against Italy in Euro 2012. See Figure 3.7 for details of how this plot was created.

The Schweinsteiger tornado forecast from 2012 arrived in full force in 2014. At the World Cup in Brazil that year, he was everywhere, making penetrating runs, rescuing corners, dribbling past defenders, completing neat one-twos, shooting from outside the box, heading from inside it and crossing the ball from one side of the pitch to the other. In the final, he ran 15km and made 94 successful passes from all over the pitch. He intercepted Messi at several crucial moments and was fouled seven times by an increasingly desperate Argentina. His phenomenal movement paid off, and after 120 minutes Germany were 1–0 winners and World Cup champions. The tornado had finally triumphed.

## CHAPTER FOUR

# Statistical Brilliance

Every year, someone is best – the player who scores the most goals or the athlete who runs the fastest. But now and again there comes along a person who thoroughly outclasses everyone else. In recent years we have seen Lionel Messi and Cristiano Ronaldo destroy crumbling defences and smash goalscoring records. We have seen Usain Bolt dominate the 100 metres and 200 metres, and with strength and ease. Other athletes, such as the Williams sisters in tennis, have dominated their sport over such long periods that they seem invincible.

Similar questions come up when we think about many of the problems facing society today. How often do we expect the financial ups and downs of worldwide economic crises to occur? How do we plan for increases in extreme weather events, such as storms and heatwaves? For global finance and climate change, these are big and important questions – perhaps even more important than the relative merits of Ronaldo and Messi. To answer them, we need the statistics of extremes.

## The Guessing Game

There are billions of people in the world, and only a very small number of them rise to the top. FIFA's official statistics estimate that around 265 million people actively play football.[1] Ronaldo and Messi are just two of these millions of players – extreme 'events' that occur in only 1 out of every 132 and a half million people.

It is not particularly fair to compare Messi and Ronaldo with the rest of us, tripping over our own feet on a Sunday morning in the park. Messi has been described, usually together with Diego Maradona and Pelé, as the best footballer of all time. The question we need to ask is how today's players compare with those we have seen in the past. How often do we expect an event like Messi to come along? Once every ten years? Once every generation? Or even once a century?

To measure extremes we need observations. We need to know the number of goals scored by the top scorer in a particular league in a season, or the hottest day of the year in London. Looking at these observations over previous decades can give us some idea about what will happen next year. If no one has scored more than 40 goals in previous seasons, then it isn't particularly likely that anyone will do so this year either. If it hasn't been warmer than 36°C on a summer's day in London, we shouldn't expect it to be much hotter next year.

In Spain, the Pichichi Trophy is presented each year to the player scoring the most goals in La Liga. The trophy has been awarded since 1929, but in 1986 La Liga established the current format of 20 teams, providing 38 fixtures per season per team.[2] Figure 4.1 shows a histogram of the number of goals scored by the winners of the Trophy between the 1986/87 and 2009/10 seasons. This histogram is similar to the ones for goals per match in Chapter 1, except now it is for total goals scored over the whole season by the most successful striker.

As of summer 2010, no player in La Liga had ever scored 39 or more goals in a single season. Nor did it seem particularly likely that anyone would break the record in the forthcoming season. Messi had managed 32 goals in the 2009/10 season, and Ronaldo was in second place with 26. If one of them were to beat this record it would take something extra-special. They would have to beat a record that had stood for 23 years.

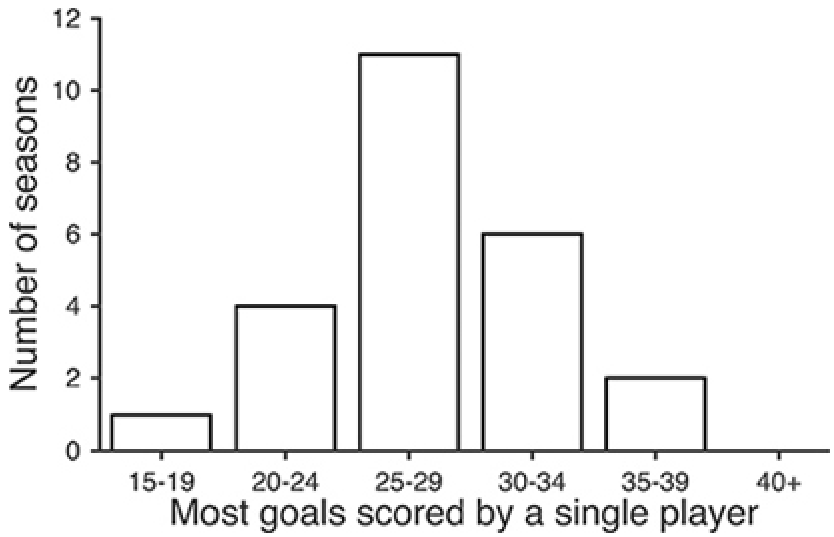



**Figure 4.1** Histogram of number of goals scored by the winners of the Pichichi Trophy (for most goals scored in La Liga, Spain) between the 1986/87 and 2009/10 seasons.

How do we assign a percentage probability to one of them beating the record during the 2010/11 season? One way is to think of it as a guessing game. Look at the histogram in Figure 4.1. Can you guess in which previous season the most goals were scored? Maybe you fancy the older Brazilian Ronaldo (but in which year and for which team?) or Raúl, or even Ruud van Nistelrooy? If you're on the ball, you'll remember the great Hugo Sánchez, who dominated the Pichichi Trophy during the late 1980s. In the 1989/90 season he scored 38 goals in 35 appearances for Real Madrid, more than one goal per game. But the chances are that most of us would have got it wrong. Without a good knowledge of Spanish football, most people's chance of guessing the right year would be around 1 in 23, or 4.35%. All we can do is pick a random year.

But now let's include the 2010/11 season in our guessing game. Imagine that it's the summer of 2010 and you are asked, 'By the end of this football season, which do you think will be the highest-scoring season of all time?' If you don't know the answer, then you should just treat 2010/11 the same as those earlier years. The chance of it being the season with the most goals by a single player is the same as for all the previous 23 years: 1 in 24, or 4.17%.

The guessing game provides a good rule of thumb for determining how likely it is that an extreme event occurs. Just because something hasn't happened before does not imply that it will never happen. Instead, a good estimate of the probability of an event occurring is 1 divided by the number times it hasn't occurred, adding one for the season we're currently interested in. In our case, that's 1/(23+1) = 1/24. In 2010, it looked like something extra-special would be needed to beat Hugo Sánchez's record.[3]

## Once in Every Lifetime

So, did Messi score more than 38 goals and beat the Pichichi Trophy record during Barcelona's wonder season of 2010/11? If you haven't checked Wikipedia already, I can tell you that the answer is 'No'. Messi scored 31 goals, one short of the season before. Ronaldo, on the other hand, upped his performance in La Liga, scoring a record-beating 41 goals and taking home the Trophy.[4] In hindsight, it is always difficult to say whether 23/1 were good odds or not, but it was certainly an achievement to surpass all previous goalscoring records. Ronaldo's feat might have seemed like a-once-in-a-generation achievement.

What happened next pretty much completely defied our guessing-game model. In the 2011/12 season Ronaldo scored 46 goals, comfortably beating all previous records. But Messi weighed in with 50 goals in 37 appearances. Not one but two players surpassed all previous records. In 2012/13 the goals continued to flow, with Messi on 46 and Ronaldo on 34. There was a small dip in 2013/14, with Messi out of the running and Ronaldo on 31, but in 2014/15 Ronaldo was back, banging in 48 during the season.

Messi's total of 50 league goals in one season is truly exceptional, but exactly *how* exceptional? The maths we have done so far allows us to say only how likely it is for a record to be broken, not by how much. How often do we expect 50 goals to be scored in a season of La Liga? We can answer this using slightly more advanced mathematics than my guessing game, employing a mathematical model called the extreme-value distribution.

The extreme-value distribution is a mathematical model of all types of extremes, be it hottest or rainiest days, wind speeds or goals in Spain. In order for it to apply, two assumptions must hold. The first assumption is that the number of goals scored in one season shouldn't affect the number of goals scored in the next. This is reasonable enough: as we saw in Chapter 1, the time of one goal has little or no effect on the time of the next. The second assumption is that there should be no year-on-year trend in goalscoring. This is less likely to hold since the balance between attack and defence can change over the seasons. We'll come back to this second assumption later, but for now we'll assume that it holds and see what the extreme-value theory tells us.

Figure 4.2 shows the histogram for the top goalscorers from the 1986/87 season to the 2013/14 season. The solid line is the curve for the extreme-value distribution.[5] Using the extreme-value model, I can now work out just how exceptional Messi's 50 goals were. Overall there is a reasonable enough match between the actual data observed so far to reassure us that the extreme-value theory could be useful. But if you look carefully at the bottom-right-hand corner of Figure 4.2, you'll see that Messi's 50 goals are above the curve derived from theory. The area under the curve corresponding to 50 or more goals is the small shaded area. Here the theory curve is quite a bit lower than the data histogram. In fact, the area under the 50+ part of the curve is only 1.36%, or 1/73, of its total area.[6] So the model tells us that we should expect a performance like Messi's to come along once every 73 years. The average life expectancy in Argentina is 75 years. Put in those terms, Messi really is a once-in-a-lifetime event.

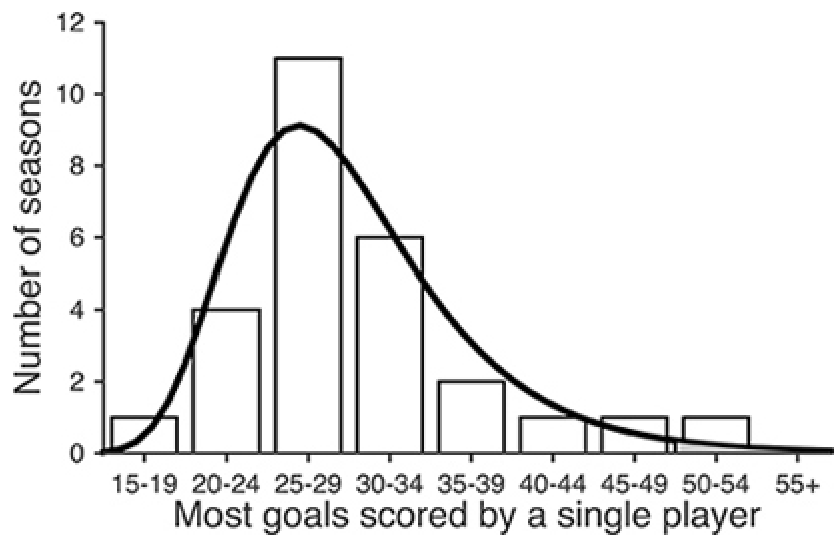



**Figure 4.2** Histogram of number of goals scored by the winners of the Pichichi Trophy between 1986/87 and 2013/14 (boxes) compared to the extreme-value distribution (solid line).

## New Extremes?

In January 2015, Messi was awarded FIFA World Player of the Year for the fifth time, moving ahead of Ronaldo who has won it on three occasions. Messi was not, however, the first player to win the award for the fifth time. I had the pleasure of watching the Brazilian five-times winner in a Champions League final in 2007, in a match played at a football ground less than a kilometre from where I lived at the time. The venue was Gammliavallen in Umeå, Sweden, and the player was the striker Marta. She is a truly exceptional player. When playing for the Brazilian national team she plays as a

Ronaldo-style centre-forward. In Umeå she played more like Messi: dribbling past defenders, creating chances for her team-mates and occasionally unleashing an unstoppable left-footed shot.

Marta was joint top scorer in the Swedish women's league in both 2004 and 2005, with 21 and 22 goals respectively. She came in second place in 2006 and 2007, with 20 and 26 goals. These are relatively small totals for seasonal scoring in Swedish women's football. The all-time goalscoring record in Swedish women's football was set in 2002 by another Umeå star, Hanna Ljungberg. She scored 39 times. Figure 4.3 shows the distribution of goals scored since 1982. As in Spanish men's football, the extreme-value distribution accurately captures the goals scored by the most successful player. The solid line of the theoretical distribution again lies close to the real goalscoring numbers. When we calculate the probability of the total of 39 being equalled or beaten, it turns out to be 3.16%. This makes Ljungberg, who grew up a short forest walk away from Gammliavallen, a once-in-a-generation goalscorer in Sweden.

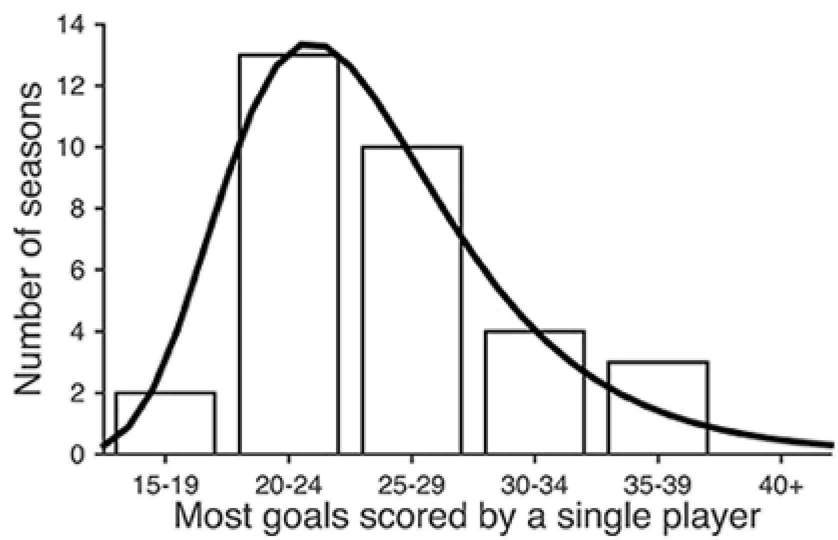



**Figure 4.3** Histogram of number of goals scored by the winner of the Swedish Damallsvenskan Skyttedrottning, the player scoring the most goals during the season (boxes), compared to the extreme-value distribution (solid line), between 1982 and 2014.

Like Messi, Marta should be judged on more than the sum of her goals. The thing that makes her really special is her ability to contribute to the team. In the 2005 season, four of the five top goalscorers played for Umeå, including both Marta and Ljungberg. They scored 79 goals between them in 22 matches. This total can be compared to the combined total of 81 goals scored in 38 matches by Messi, Suárez and Neymar in La Liga 2014/15. In the 2000s, Umeå were the Barcelona of the north.

Swedish women's football has Ljungberg, Spanish men's football has Ronaldo. Women's football has Marta, men's football has Messi. The challenges are different in different countries, and in the men's and the women's game. But we can see the same statistical regularities across different leagues. Extreme-value theory allows us to predict how often we will see players beat previous records and, for big events, to measure just how special they really are.

## The 10,000-year Storm

A field outside football where we really need to understand extremes is making policy decisions about the future. Political decisions should be based on long-term thinking about outcomes, rather than on short-term responses to events. And sometimes, politicians actually do think long-term. Dutch law states that 'the most important parts of the coastal defence system [should] be able to withstand a water level that on average is reached only once in 10,000 years'.[7] A demand like this from politicians poses a considerable challenge to scientists, who have access to data for the past 150 years and are being asked what might happen in the next 10,000.

Extreme-value theory allows us to answer such questions. The largest storm surges each year in the North Sea over the past 150 years fit the same type of extreme-value curve we used for goalscorers.[8] By looking at the area under the curve we can predict the frequency of extreme events. For example, the North Sea flood of 1953 saw surges of nearly 4 metres, corresponding to an event that occurs only once every 455 years. This doesn't mean that the North Sea flood won't return again until the year 2408. Instead, it suggests that the chance of there being a flood next year of the size of the 1953 event or bigger is 1 in 455.

The current sea defences in the Netherlands, and the Thames barrier downstream of central London, were designed with the 1953 flood in mind. By asking the reverse question, what the maximum yearly surge would be with a probability of 1 in 10,000 years or less, scientists can address the requirements imposed by Dutch law. This is exactly the calculation that the Royal Netherlands Metrological Institute have carried out. They have shown that the maximum expected surge is likely to be 5 metres, 25% larger than that seen in 1953. We can't be sure that the mathematical meteorologists are correct, but until the flood comes their estimate is the best that we have.

## Bolt from the Blue

Up to now I have assumed that the world doesn't change very much. We observe the largest flood surge, the hottest day or the top goalscorer each year, but we assume there is no year-on-year trend. In the models I've been discussing so far I've assumed that scoring goals in La Liga is equally tough every year, and that the climate is not steadily changing. These are big assumptions, and for the climate there is good evidence that this assumption is plain wrong. Scientists now agree that the climate is changing and that we need to be able to make predictions that account for these changes. Can we still predict the future when the world is changing?

To get a better idea of how we study change, let's look at a sport in which there is no doubt about who is best. Before Usain Bolt, the world record for the 100 metres decreased steadily, from 10.6 seconds at the beginning of the twentieth century to Asafa Powell's time of 9.74 seconds in 2007. This progression is shown in Figure 4.4. Ben Johnson's time of 9.79 seconds in the Seoul Olympics 1988 would have bucked this trend, but it was struck from the record books when he was found to have been using performance-enhancing drugs. As well as Johnson's, a few of the other times plotted in Figure 4.4 have been retrospectively rescinded as a consequence of failed drug tests. But drug-free times by Maurice Greene and Asafa Powell remain as part of the general trend.

Up to 2007, there was a clear trend of gradual improvement in the 100 metres world record. This average year-on-year improvement is shown as the dotted line in Figure 4.4: a steady average decrease of 74 milliseconds per decade over the previous 100 years. By 2007 the world's best sprinters were still getting faster, but only very gradually.

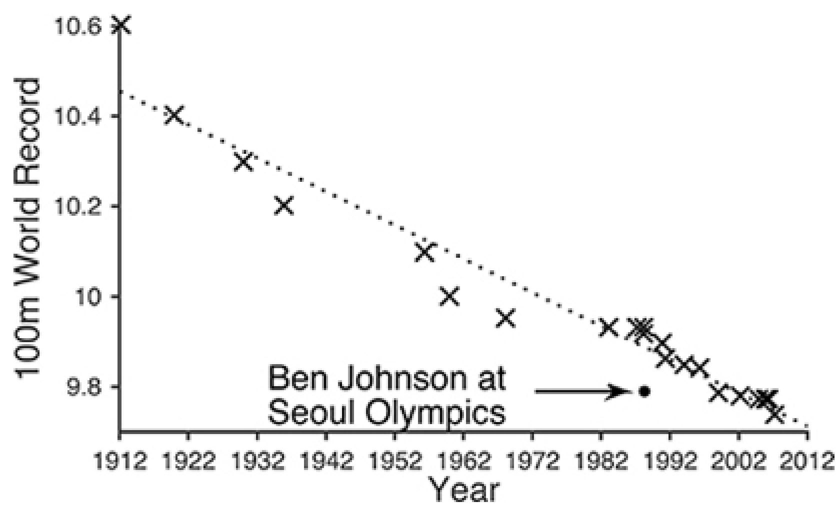



**Figure 4.4** World record times for men's 100m up to 2009. Crosses are world record times. The dotted line is the fitted trend line. Ben Johnson's time (circle) of 9.79 seconds at the Seoul Olympics was later disqualified.

In the lead-up to the Beijing Olympics of 2008 I don't think I or anyone else could have predicted the dramatic arrival of Usain Bolt. In Figure 4.5 I have extended the straight line for predicted future record times to 2032. Bolt's pre-Olympic time of 9.72 seconds in May 2008 was excellent but not unexpected, and it lies slightly below the predicted line. His 9.69 seconds in the Olympic finals that same year was both brilliant and surprising. We can see just how surprising by following the dotted line until it passes the circle at 9.69 seconds. This happens at some point in 2015 or 2016. We should have expected to have to wait until Rio de Janeiro to see someone achieve this time. Bolt was about eight years ahead of his time at that point.

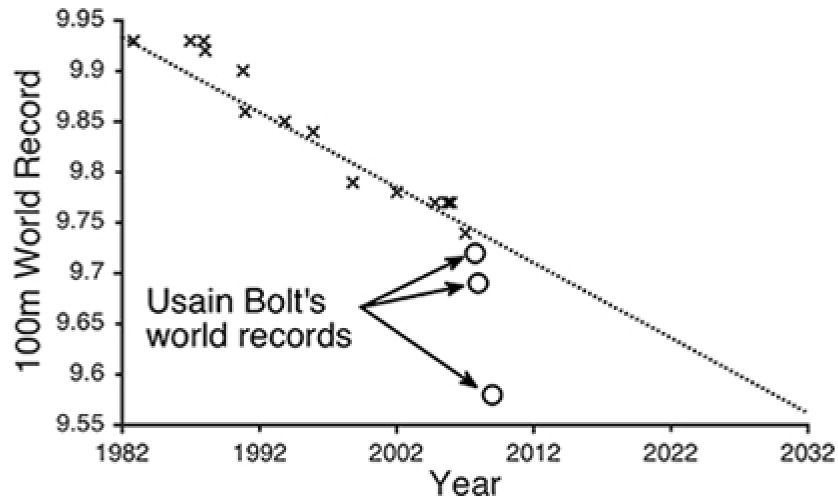



**Figure 4.5** World record times for the men's 100m from 1982 to 2015. Crosses are world record times. The dots are the fitted trend line.

It is when we come to the World Championships in Berlin in 2009 that we see just how far ahead of his time Bolt actually is. According to our prediction we would not have expected to see a time of 9.58 seconds until around 2030, more than 20 years after Bolt set his record. Bolt is not just brilliant, but unexpectedly brilliant.

Drawing straight lines through data in order to predict the future is fraught with danger. If we were to stretch that line way into the future, we could predict that in the year 3318 the world record time would stand at an amazing zero seconds! After that, our descendants will develop the capacity to travel back in time and run 100 metres in negative time. You don't have to be Einstein to see that this is nonsense. There must be some limit to how fast a human being can run, even Usain Bolt.

If anything, a straight-line fit to the data up to 2008 would have been seen at the time as overly optimistic. Increased participation in athletics around the world, combined with better training techniques and sports facilities, have allowed athletes to get faster and faster. But usually we would expect to see a diminishing return, with the rate of improvement slowing down over time. So Bolt not only outperformed a model, he outperformed an unrealistically positive model. He shattered not only the world record, but also our ability to make reliable predictions about the 100 metres. Looking at the times he recorded, it is difficult to know where we should start in predicting what will happen next.

## Game-changers

I can imagine one of two possible futures for post-Bolt sprinting. One possibility is that Bolt truly is a one-off. His combination of attitude, shape, size and demeanour make him unique, and we may have to wait a long time until we see another Bolt. A second possibility is that he has changed the rules of his sport. Bolt is taller than other sprinters, takes fewer steps and seems to be more relaxed in his preparation. By emulating his approach, other youngsters with a similar build may be able to adopt the same style. We haven't seen any of them coming up yet, but a 10-year-old watching the Beijing Olympics will be 22 in 2020. Imagine seeing, at the Olympics in Tokyo, a row of Bolt-a-likes sprinting for gold.

A similar question can be asked about Messi and Ronaldo. It is no coincidence that they both smashed goalscoring records in the same season. They undoubtedly push each other during the season, and they could even spearhead a new resurgence in attacking football, similar to that seen in the 1960s. Figure 4.6 shows the goals per game for the highest-scoring forwards since the beginning of the Pichichi Trophy in 1929. Some La Liga players of the past, such as Telmo Zarra in the late 1940s and early '50s, were regularly banging in more than one goal per match. They were playing in a different era, of course, with a smaller number of matches and very different formations. But this is precisely the point we have to consider now. Maybe Ronaldo and Messi are taking us into a new era of football?

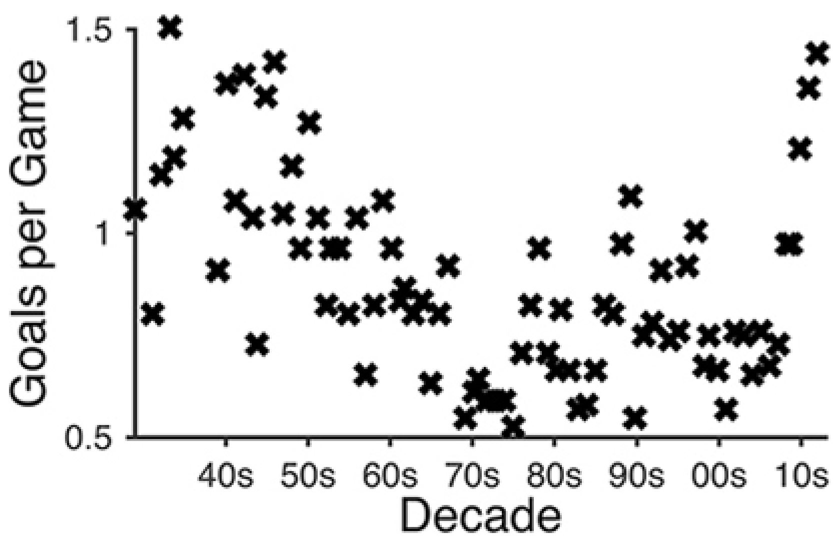



**Figure 4.6** Average number of goals scored per game by the winners of the Pichichi Trophy.

It is not just sport where the rules of the game change. Whenever there is a typhoon, a drought, a heatwave or a flood, we wonder whether we are seeing the results of climate change. The Intergovernmental Panel on Climate Change is cautious about attributing any single extreme event to emissions of greenhouse gases or other human activities.[9] However, in recent decades there has been a steady increase in the length of warm spells, as well as in the number of extremely hot days. There has also been an increase in the number of very rainy days, especially in North America and Europe.[10]

These sporting and environmental game-changers show us a fundamental limitation of statistics such as extreme-value distributions. I referred to this limitation earlier, when I wrote that I was assuming that there is no year-on-year trend in goalscoring. But it seems that there is such a trend: in the climate, in goals and in sprinting. These trends don't invalidate statistical models. Instead, we should see statistical models as the best we can do given what we currently know about the past. It's just that, unfortunately, we can't always know exactly what the future has in store.

## Objective Rankings

Football is a team sport, so it isn't fair to judge players solely on the number of goals they score. Cristiano Ronaldo shoots more than pretty much any other player in professional football. And if you shoot a lot, then the goals are likely to follow. The number of shots he makes is not only a measure of his own skill, but also that of the team around him. I have already shown how Messi depends on his team-mates' passing. Maybe it is defenders and midfielders such as Marcelo, James Rodríguez and Isco who deserve much of the credit for Ronaldo's latest tally in La Liga?

Midfielders and defenders cannot be judged purely in terms of goals. Luckily, there is a whole range of statistics that can be used to rate them: assists, number of shots on and off target, distance run, number of dribbles, interceptions, tackle ratios and clearances, to name but a few. These stats are available for each player after each match. The difficulty is turning these numbers into real measures of performance.

Reducing skill to a single measure is the aim of the Performance Index, presented each week on the Premier League's website. There you can find the Team of the Week, made up of those players who have performed best in the past week in each position. You can also look at the cumulative rankings and find out who is most consistent over the entire season. According to these statistics, Chelsea's playmaker Eden Hazard was the best player in the 2014/15 season. Other players agreed, with Hazard also chosen as the Players' Player of the Year.

While the players decide who to vote for as their Player of the Year on the basis of having played against them over the season, the Premier League's Performance Index is a purely statistical model. Its sponsors, EA Sports, claim that it is an objective index designed to 'settle disputes by using a comprehensive range of statistics'.[11] Often the players and the index agree. Usually most of the players nominated for the Player of the Year award also feature in the top 10 places of the Performance Index. However, although it works well, the Performance Index is not quite as 'objective' as its sponsors claim it to be.

Every index, no matter how objective we would like it to be, has to be designed by humans. The task of designing the Premier League Performance Index was given to two Manchester statisticians, Ian McHale and Philip Scarf. They took on a serious challenge. First, there is the difficulty of comparing goalkeepers, defenders, midfielders and strikers. Even within these roles there are sub-categories, such as winger or defensive midfielder, and different tasks assigned to players, such as marking or taking corners. Each player has a different role in the team, and their success is judged by different criteria. The second difficulty comes in teamwork. A defender may do very little during a match because the team manages to focus all its effort on attack, while a goalkeeper of a weaker side may make lots of good saves – but only because he has to.

Ian and Phil started by looking in detail at what players contributed in a game. They started by building a statistical model of how different contributions created goals. In each match they counted how often players performed certain actions, including passing, tackling, crossing, dribbling, blocking and clearing, and also how often they were shown a yellow or a red card. They then used a statistical fit to see how well the number of these actions predicted their own or the opposing team's number of shots on goal. What the statistical fit gives is a measure of the effect, positive or negative, on creating chances. They found that the more a team passes the ball, the more chances it creates, with successful crosses proving one of the most likely routes to goal. By fitting their model to Premier League data from 2003 to 2006, they estimated that each successful cross was worth about 10 standard passes in terms of shot creation. Getting sent off was equivalent to missing 41 interceptions.

Their statistical model allows the effect of the build-up to an attack that comes from the defence and midfield to be quantified. Ian and Phil then added in the probability of successful shots and saves, which allowed them to take into account the influence of strikers and goalkeepers on goal creation and suppression. The method is just about as 'objective' as it can get, in the sense that it is based entirely on the numbers of various actions performed on the pitch. It takes the actions of each player, relates them to goal chances and finds out which players contribute most.

Once Ian and Phil had created the model from historical data, they tested it on the 2008/09 season to determine the best players. They looked at how many of the various types of action each player performed during the season, and calculated the index for each of them. This was Cristiano Ronaldo's last season at Manchester United, where he was top goalscorer; Rio Ferdinand, Nemanja Vidić and Ryan Giggs all made key contributions to the side. Liverpool made a serious title challenge, with Jamie Carragher, Steven Gerrard and Dirk Kuyt as regulars. Gareth Barry and Gabby Agbonlahor helped Aston Villa qualify for Europe.

So who do you think was the best player according to the model? Well, it was Fulham goalkeeper Mark Schwarzer. He scored 7.29 on Ian and Phil's original Match Outcome

Index, coming in ahead of Barry in second place with 7.06 and Portsmouth's Sol Campbell in third with 6.86. Ronaldo didn't even make it into the top 20, nor did any player from Manchester United. In fact, the list featured only one striker, Chelsea's Nicolas Anelka, and one midfielder, Gareth Barry. All the others were goalkeepers and defenders.

No disrespect to Schwarzer. He had an excellent season, and was voted Fulham's Player of the Year. But a 37-year-old Aussie goalie topping the rankings was not exactly what the Premier League sponsors were looking for. While performances were important, it made sense to them that a team's final league position should also be part of a ranking.

In their scientific paper about the system, Ian and Phil stood by their match contribution model.[12] When they looked into why strikers and attacking midfielders weren't higher in the ranking, they found it was because they were squandering chances in front of goal – Frank Lampard in particular was found to be 'highly variable' with his shooting. Defenders were rated as more valuable because it was their interceptions and blocks that were preventing goals from going in. Defenders stop a lot more potential goals than attackers score, making their match contribution more significant.

Ian and Phil presented their scientific findings, but they were also pragmatic. After talking to the Premier League, they made a revised index, adding points for winning matches and total number of goals as well as success rate, assists and clean sheets. These amendments tilted the index back in favour of attackers, and Anelka, who was highly rated in both the old index and the new one, topped the list. This combined index is still used by the Premier League.[13] It is currently dominated by attacking midfielders and forwards, and they do tend to be the players we love to watch. But we should remember that, in purely statistical terms of performing actions that create chances for their own team or prevent chances for the opposition, these are not necessarily the best players. There are unsung goalkeeping and defensive heroes, working for Crystal Palace, West Bromwhich Albion and Stoke City, who never make it into the Premier League top 10, but still top Ian and Phil's list. These players are the true statistical stars of the Premier League.

## FC Analytics

It is impossible for managers to watch every match and evaluate every player around the world. Ideally, they would like to have access to some numbers so that they can find rising stars early on. The data is available. Teams can browse through massive online databases of players' vital statistics. A good starting point is the database inside the computer game *Football Manager*. Stats on real-life players are collected by players of the computer game, who watch lower-league and youth-team games and provide rankings. Recently, this database has been bought by sports performance company Prozone, which now supplies it to clubs.

The question is, how best to use all this data? In baseball, the use of statistics is well established. American baseball has a long history of using batting and pitching averages, run times and fielding indices to assess player performance. The success of using these stats in trading players is well documented in the book (and film) *Moneyball*, for example, and the stats approach is now widely adopted throughout the sport.[14]

In football, the team complicates the statistics to a much greater degree than in baseball. A defender might head the ball-recovery stats because of a specific role assigned to him by the manager, or because his team-mates keep losing the ball, or even because he is poorly positioned to make an initial interception. A forward might make lots of shots, but it could have been better for the team if he passed more instead. The challenge is to separate the individual from the team. Defenders playing in clubs in the bottom half of the league encounter more difficult situations than their counterparts in the top half. And strikers for the top teams are presented with more chances than those in teams that are struggling.

Individual players can't be properly rated without taking into account the team they play for. In their book *The Numbers Game*, Chris Anderson and David Sally explain the potential failure or success of a football team in terms of its weakest link.[15] They use an analogy between the essential role of each of the components of a spacecraft and the essential role each player plays in the team. Their advice to football clubs is to concentrate less on the small improvements to be gained by buying, say, a striker with a slightly higher conversion rate, and more on finding players who are a better fit with the team's structure, and so strengthening the squad where it is weakest.

Analytics companies are aware of the limitations of player statistics and are working hard to improve things. Dan Altman's company North Yard Analytics uses detailed passing data from performance analysis company Opta to assess the overall impact a player has on the field. Dan rates players in terms of how far they advance the ball in different areas of the pitch, and how they contribute to shots. For example, he broke down Everton defender Leighton Baines's performance during the 2013/14 season and showed that his match contribution was strongly focused in the opponents' half.[16] Compared with the average left-back in the Premier League, Baines was contributing substantially more when the ball was farther up the pitch, but substantially less when the ball was in his own half. Any team interested in signing Baines would need to have a strong left-sided centre-back who could cover while Baines created chances farther forward. Building a team is about fitting all the pieces together.

Statistical scouting is probably less important at the biggest, richest clubs, which already have an army of scouts looking for new talent and young players competing to get into their academies. But it is possible for smaller clubs to get an edge by keeping track of the numbers. In 2014, the Danish club FC Midtjylland was bought by Matthew Benham, owner of betting site Matchbook and sports modelling service Smartodds. Benham specialises in using mathematical models to predict results, and has created a vast performance database of players and teams. In betting, his statistical methods have earned him a fortune. At Midtjylland, as well as English club Brentford, which he also owns, Benham aims to integrate a mathematical modelling approach into the running of football clubs.

Midtjylland's sports director, Claus Steinlein, immediately saw the benefits of being owned by a mathematical modeller. 'Before we had one scout – and he spent half his time coaching,' Steinlein told *Guardian* journalist Sean Ingle. 'Now we have a team in London crunching the numbers and suggesting suitable targets.'[17] Midtjylland use the database to find players who potentially fit into the squad before they watch the player and speak to him face-to-face. Statistics is a starting point for scouting, not a replacement for it.

The scientific approach to football has been widely applied at Midtjylland, not just to scouting. They have a kicking coach who analyses in detail the spin the players put on the ball when they kick it. Using a thorough analysis of set pieces, they greatly improved their scoring success rate from free kicks and corners. At half-time the manager is sent a breakdown of the chances that his team and the opposition team have created, and he can adjust his tactics accordingly. In summer 2015, this scientific approach paid off: Midtjylland won the Danish league for the first time in their history.

Midtjylland perhaps represents the future of football, where science is properly integrated into the club's approach. But this analytical approach won't be based around naive statistical measures of player abilities. It will be focused on building up the team. Nor will analytics deny that players now and again achieve things that are truly unexpected and exceptional – no team wants to miss out on the next Messi because they were looking at a computer screen full of numbers. Integrating mathematics and science into football involves a balancing act that we are only just beginning to get right.

CHAPTER FIVE

# Zlatan Ibrah Rocket Science

'Zlatan Ibrahimović! I want to go and give you a man hug!'

These were the words of commentator Stan Collymore after he watched the gigantic Swede rotate his body vertically through 180°, meeting the ball in a bicycle kick and lobbing it over Joe Hart's head from more than 25 metres. Seconds before, Collymore was calmly describing the game as a 'worthwhile exercise' as England and Sweden looked forward to the World Cup qualifiers. Suddenly he exclaimed, 'Oh my God, an insane goal! I've just seen the most insane goal I've ever seen on a football pitch!' Collymore's voice then almost cracked as he declared his desire to hug Zlatan.

It was a spectacular goal.[1] The bicycle kick is one of the most celebrated ways of scoring in football because it involves such a high degree of co-ordination, anticipation, accuracy and timing. The player has to turn upside down at high speed, follow the ball and meet it perfectly. For the 'standard' bicycle kick executed inside the box, like Wayne Rooney's amazing goal against Manchester City, the player needs to make contact with the ball in the middle of the foot. This ensures that the ball will travel downward towards the goal. Zlatan was well outside the penalty area, so he had to use the tip of his boot to arc the ball over keeper Joe Hart and the defence. His 'bicycle lob' was a whole new type of shot.

## Newton Explains

The execution may be difficult, but the physics of Zlatan's bicycle lob is relatively straightforward. The main force involved is gravity, and the path of the ball can be worked out with the aid of Newton's equations of motion. Assuming no air resistance, the ball will follow a trajectory similar to that shown in Figure 5.1. Like all such trajectories, its shape is a parabola. I'm assuming that the velocity of the ball towards the goal is constant. The downward velocity of the ball increases over time as gravity causes the ball to accelerate towards the ground.[2] Gravity provides a constant acceleration, so the ball's initial upward velocity is positive, but decreases by an equal amount at every point in time. It is zero when the ball is at its maximum height and then becomes negative as the ball falls. As a result, the ball's trajectory is symmetric around the maximum, and it follows the same path down as it does up.

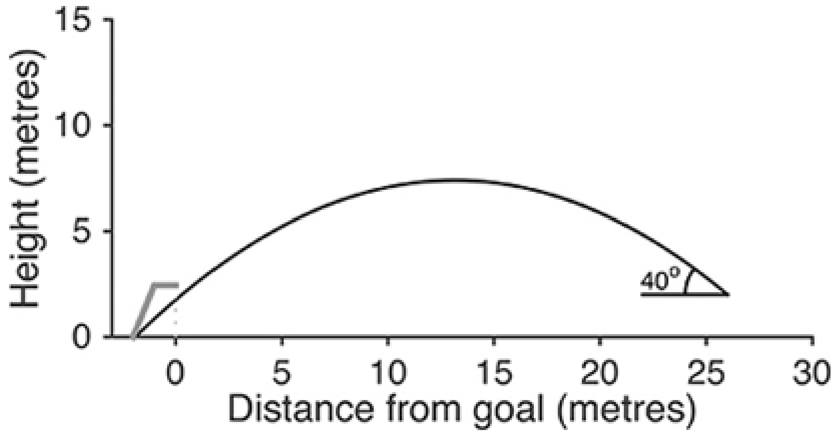



**Figure 5.1** The trajectory of Zlatan Ibrahimovic′'s bicycle lob, according to Newtonian physics.

Following the path of this parabola, the whole thing looks pretty simple. It is just a case of launching the ball at the correct angle and speed, then gravity will take over and the ball will land in the back of the net. But the problem is that the relationship between the angle of launch and where the ball ends up is far from straightforward. Figure 5.2 is a plot of six different shots made at different launch angles, two of which end up in the back of the net and four of which miss.

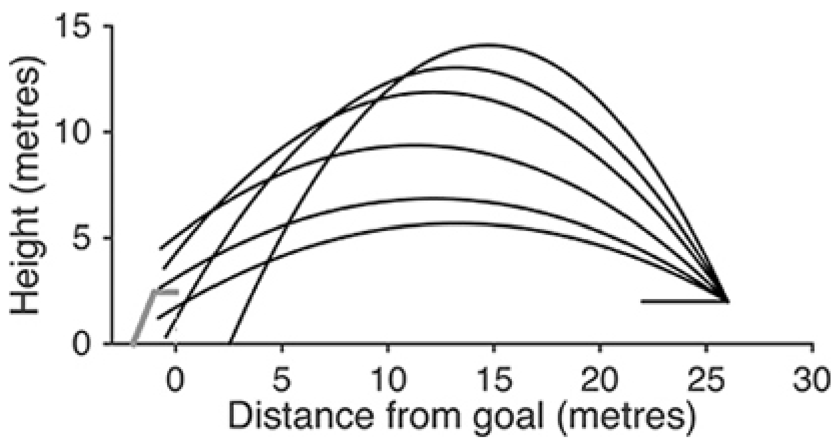



**Figure 5.2** How the trajectory of Zlatan's bicycle lob shot is determined by launch angle.

All these shots had the same initial speed, 17 metres per second, but the outcomes are very different. Whether the ball lands short, goes in or flies over depends on a complex relationship between speed and angle. We can work out this relationship by solving the equations of motion. Showing how the ball moves with time is a task you might be asked to do in secondary-school maths.[2] To solve for whether the ball goes in the goal, we need to rearrange the terms to find a condition for the ball to go under the bar, but not hit the ground before it goes in. This isn't difficult, but it does require a few mathematical steps.[3]

Figure 5.3 shows the combinations of angles and speeds that result in the ball reaching the goal. Think about how Zlatan might choose a point on this plot. For example, he could kick the ball at 25 metres per second at 40°, in which case it will fly over the bar. Or he might kick it at 15 metres per second at 30°, but then it will bounce just in front of the goal. If, as I showed in Figure 5.1, he hits it at 16 metres per second at 40°, then it will go in. The goal-bound combinations occupy just a narrow sliver through the space of all possible angles and speeds. If the ball is hit too hard it will go over the bar; if it leaves the foot at too large or too small an angle it will bounce and be cleared by the defence. The weight on the ball has to be just right. The sliver is formed in a way that makes predicting what will happen without mathematics very difficult. For example, a ball struck at an angle of 19° at 20 metres per second will go in, but just a bit harder than that and it will go over the bar. However, if the ball is hit at the same speed but at 65°, it will go up very high, come down and drop into the net. This is a difficult shot to make, since even a small increase in angle or decrease in speed will cause the ball to bounce before reaching the goal.

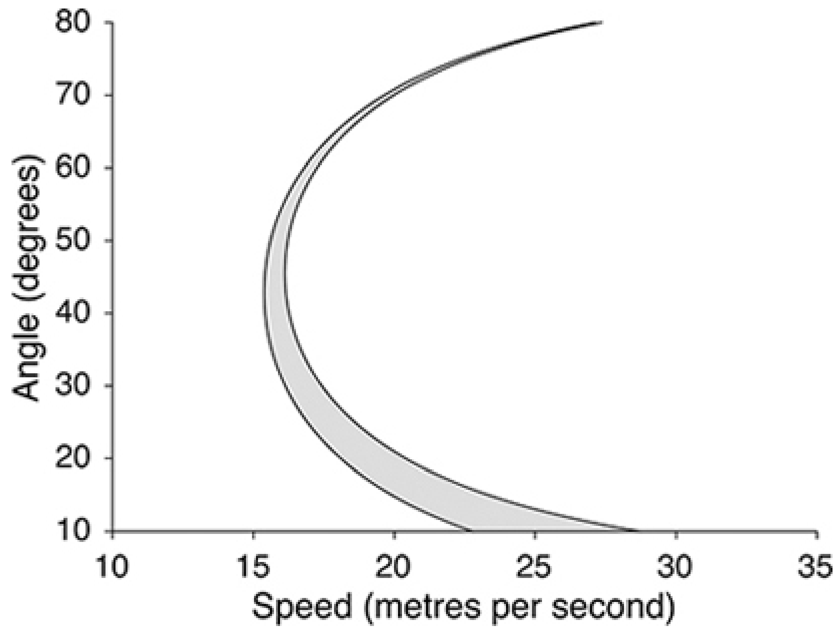



**Figure 5.3** How initial speed and launch angle determine whether or not the ball goes in the goal. The grey area indicates angle and speed combinations under which Zlatan scores. Combinations to the right of the grey area send the ball over the bar, while combinations to the left cause the ball to bounce before reaching the goal.

So how did Zlatan get it right? Well, there was some degree of luck in the goal. Hart found himself outside his box and didn't head the ball away far enough, so Zlatan was in the right place at the right time. But when he got the chance, he weighted the ball perfectly. I estimate, from my repeated viewing of the footage (with commentary on), that the ball left Zlatan's foot at somewhere between 16 metres per second at an angle of 40°, as shown in Figure 5.1. By choosing this slower speed, Zlatan left more leeway for error in the angle. Any angle between 30° and 50° would have put the ball in the net. If he had hit the ball harder, for example at 20 metres per second, the error margin would have been much smaller. Even when upside down, Zlatan minimised the probability of making a mistake.

## Sent it into Space

The bicycle lob has relatively simple aerodynamics. The main force is gravity, and the equations are the same as those you learned in school physics classes. Zlatan applies an initial upward force, and gravity provides downward acceleration. However, it's not quite as simple as that. In the calculations above I made things easier by ignoring the other forces at play here. As the ball flies through the air the resulting drag slows it down, and Zlatan also applies spin so that the ball rotates as it drops into the net. There is a lot to think about when modelling the motion of a ball in flight.

Luckily, a bunch of rocket scientists are on the case. NASA runs a whole research programme dedicated to ball aerodynamics. They have even created an online shot simulator, where you can enter the position, direction, forces and spin on a ball and calculate whether or not it will hit the target.[4] I don't quite have the resources available to NASA, but I did add drag due to air resistance to my own Zlatan bicycle lob simulator. One simulated goal is shown in Figure 5.4.

Air resistance is significant: it causes the ball to fall at an angle that is steeper than the launch angle. Figure 5.4 shows the ball being launched at 27°, but when it reached the goal its angle to the ground was closer to 80°. The fact that the ball is falling more steeply means that the slither of goal-creating speeds and angles shrinks. A slight overhit will go over the bar, and an underhit will bounce in front of the goal. When Zlatan hit that ball he put backspin on it to counteract air resistance, and make the path more like the gravitational parabola. As the ball spun from his foot, he must have known immediately that he had done something special. The ball curved exactly as he wanted it to, and fell perfectly into the England goal.

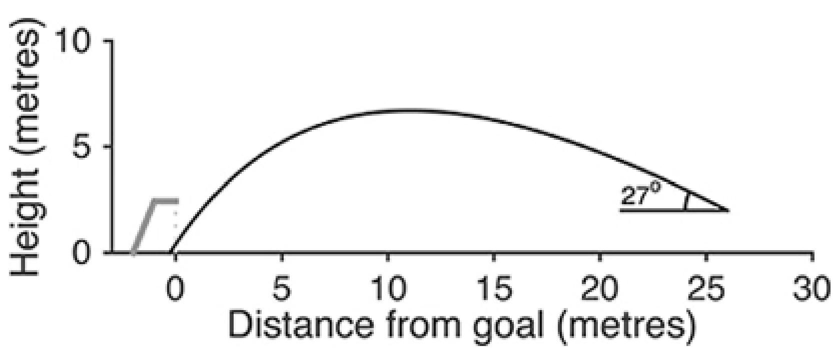



**Figure 5.4** Aerodynamics of Zlatan Ibrahimovic´'s bicycle lob, including drag.

The trajectory of every ball kicked in a football match is determined by details such as spin and air resistance.[5] At high speeds, the material from which the ball is made and its physical characteristics are also important. Manufacturers try to eliminate surface differences, though for all balls the outer covering consists of panels stitched together. Variations in the configuration of these panels and the stitching between them mean that different balls generate slightly different turbulence patterns around them when they are in flight. Different balls kicked in exactly the same way may have slightly different trajectories.

In the build-up to the 2014 World Cup in Brazil, NASA engineers decided to test various footballs. At the South African World Cup there were complaints that the Jabulani ball used in matches exhibited ‘supernatural’ movements. These claims are difficult to verify, and could be down to sour grapes from teams knocked out of the competition, so NASA decided to put the Jabulani in a wind tunnel and test it. The Jabulani was a much smoother ball than ones used previously, with only eight panels as opposed to the traditional 32. The stitching connecting the panels produces surface irregularities, so fewer panels make for a smoother ball.

It turns out, however, that a smooth ball doesn’t necessarily have a smoother flight. When kicked hard, the Jabulani would ‘knuckle’ – bobble around in the air. All balls knuckle when kicked at low speed with no spin, because the path of the ball isn’t stable. Since low-speed shots are easier for a goalkeeper to follow, knuckling isn’t a problem in this case. However, the Jabulani also knuckled when the ball was kicked at the typically very fast shooting speed delivered by top players. It was this effect that made it hard for keepers to deal with.

NASA went on to test the new Brazuca, to be used at the 2016 World Cup in Brazil. Although the Brazuca has even fewer panels, only six, the seams on the ball are longer and the panels are covered with little dimples. This surface roughness means that the Brazuca knuckles only at low speeds, and thus has a more reliable flight pattern when kicked hard.

Researchers in Japan went even further than the NASA scientists.[6] They built a kicking robot to repeatedly kick the Brazuca, the Jabulani and several other balls in exactly the same way. In each trial with the same type of ball, they placed the ball with different initial orientations so that the robot foot would meet different parts of the panelling. When kicked at 30 metres per second, a very hard shot in professional football, the flight of the Jabulani depended strongly on which part of the panelling was met by the robot foot. The researchers set up a target 25 metres away, had the robot kick directly forward with no spin and looked at where the ball hit the target. Depending on the orientation, the point of impact on the target could vary by up to two metres for the Jabulani – that’s a pretty big variation if you are trying to hit a goal 2.44 metres high. The Brazuca was much more reliable, and changing the initial orientation made only a small difference to the point of impact. However, this reliability held also for the more traditional 32-panel football, which the researchers also tested. As far as a robot footballer is concerned, a good old FIFA standard ball is just as reliable as the modern six-panel Brazuca.

## Luck, Structure and Magic

I have a reason for ending the 'On the Pitch' part of this book with a man hug for Zlatan. Although his bicycle-lob goal can be looked at mathematically, I have not attempted to completely reduce Zlatan's brilliance to equations. His wonder-kick and Collymore's reaction to it remind us that it's futile to try to reduce the whole of the beautiful game to mathematics and science.

In the five chapters we have spent on the pitch, I have looked at various aspects of football using a combination of randomness and structure. Randomness takes us a long way in explaining goals and measuring excellence; structure allows us to control space in both attack and defence, and measure the passing dynamics of the midfield. Away from football, similar approaches work for fish and lions; weather forecasts; climate change and 100-metre runners; sending balls in the air and rockets into space; as well as horse kicks, accidents and cancer. These are just a few of thousands and thousands of examples. Across biology, sociology and meteorology, mathematical analogies allow us to see clearly how randomness and structure arise. Mathematics and science are powerful, but can they explain everything?

I watched Zlatan's goal on TV at home with my family. My Swedish wife jumped in the air, screaming with joy, and performed a kung fu display. My daughter watched her mum in delight, grinning and holding her hands over her ears to cut down the noise. And my son collapsed in a new round of tears, cursing Zlatan, and sobbing that it was the substitution of his beloved Steven Gerrard that had led to the goal. This scene – in my living room, on the pitch, in Friends Arena where even the England fans applauded, and repeated across Sweden – reflects a passion for a game that is neither random nor structured. It was, quite simply, magical.

Zlatan's goal should remain partly unexplained – as should Rooney's bicycle kick against Man City, Giggs's run against Arsenal and Beckham's halfway-line goal against Wimbledon. We may never draw a final conclusion about Maradona's (second) goal against England in the 1986 World Cup or Pelé's 1961 *gol de placa*, where he ran the whole length of the pitch to score for Santos at the Maracana. The statistical regularities in Messi's and Ronaldo's goalscoring do not take away from the amazing variety of ways they have found to find the net. The 20-minute capitulation of Brazil to Germany in their 2014 World Cup semi-final; the two minutes in which Edin Džeko and Sergio Agüero brought the title to the blue side of Manchester for the first time in 44 years; the half-time sound of 'You'll Never Walk Alone', followed by Liverpool's three-goal comeback in Istanbul – these may all be partly explained by logic and reason, but they will always retain an element of legend.

Maths and science give us an edge. We can use scientific tools to reveal patterns and to tame randomness. Each time we apply a mathematical model, we get a clearer idea of how the world works. But mathematicians and scientists should recognise their limits: there will always be things in football, and in the rest of life, that we can’t fully explain. This shouldn’t worry us. It is something to be celebrated. The action on a football pitch will always remain a unique combination of luck, structure and magic. It is all three of these together that makes football what it is.

# PART II
# In the Dugout

CHAPTER SIX

# Three Points for the Bird-brained Manager

When I was a boy, I was not a fan of Jimmy Hill. I grew up in Dunfermline in Scotland, and was pretty much the only English kid at my school. Hill was, it was explained to me repeatedly, the embodiment of everything that was wrong with the English. He was pompous, overly sure of himself, and most of all 'up his own arse'. Every week he offered advice on *Match of the Day* that seemed designed to annoy and irritate. In 1982 he made his biggest *faux pas* when he described David Narey's opening World Cup wonder-goal against Brazil as a 'toe poke'. The Scottish moment of glory was destroyed by this arrogant Englishman, and Brazil went on to win 4–1. I had to listen to stories about what an idiot Hill was for the next 10 years, usually in a way that implied I was personally responsible.

Hill was the perfect example of a 'know-all'. His on-screen analyses never pandered to the passions or the feelings of the fans, but focused on what he saw as the facts of the matter. He was logical and rational and, on TV at least, he appeared to have an infallible faith in his own reasoning. My own upbringing left me in two minds about Hill. I loved mathematics, so I believed in the power of rationality and carefully ordered thought. But there is more to life than logic. There is the feeling a nine-year-old boy gets when he sees the small, insignificant country he lives in take the lead against the greatest footballing nation of all time. There must be something worthwhile in that?

So it is with mixed feelings that I am now going to champion Jimmy Hill. Not for his on-screen persona, but for his work behind the scenes in bringing strategic changes to football. In the late 1970s he pushed the idea of giving three points instead of two for a win in league matches. His proposal was adopted in England in 1981, and over the next decade the change to the three-point system spread to Turkey, Greece, the Scandinavian countries, Italy and Ireland. In 1994, Scotland finally followed suit, and in 1995 FIFA adopted three-point system as standard.

If even Scotland has heeded Hill's advice, then it is certainly worth looking more closely at the changes he advocated. We need to look strategically at how the number of points for a win changes a team's incentives to attack and defend.

## Expanding Point Pies

At first glance, two points for a win seems to make sense. There are two points up for grabs: if you win you get both of them, and if you draw you share them. The match is a fixed-size half-time meat pie. If you win, you take the whole pie, and if there is a draw then it is split down the middle. A three-point system doesn't obey school-playground mathematics: it implies that the pie gets bigger if someone wins, or shrinks when there is a draw. It doesn't make sense from what we know about meat pies.

However, while a two-point system may make sense, it fails to provide incentives to win. Imagine that you are playing a match against an equally skilful opponent. You can decide either to attack and go for a win, or to defend and play for a draw. But you know that attacking can leave you exposed and increase the chance of conceding goals as well as scoring them. For example, imagine a very attacking strategy that gives a 50% chance of winning and a 50% chance of losing, but no chance of a draw. With this strategy, in half your matches you will get two points and in the other half you will get nothing. The expected number of points is

$$(2 \times 0.5) + (0 \times 0.5) = 1$$

This is no better than if you use an extreme defensive strategy that guarantees a draw: $1 \times 1 = 1$. Since the draw is more certain than the gamble of going all out for a win, there is an added psychological factor against attacking football. It is better to be safe than sorry.

Jimmy Hill understood incentives, and these look very different under a three-point system. If an attacking strategy has a 50% chance of winning, then the average number of points expected per game is

$$(3 \times 0.5) + (0 \times 0.5) = 1.5$$

A defensive strategy that always gives a draw guarantees only 1 point, so it is better for you to play attacking football. It is also better for your opponents to attack, irrespective of what you choose to do. One team will always lose, but when two equally skilled teams play each other, on average they will get more points if they both go for the win.

These calculations assume that both teams are equally matched. This assumption doesn't reflect the reality of the Premier League, where the top teams have billionaire owners and can move in and buy up all the best players. A more realistic example is to imagine yourself as the manager of a mid-table team fighting to remain in the Premier League. Next Saturday you are travelling away to face Arsène Wenger's Arsenal. Your scout tells you that if you play attacking football you'll have a 32% chance of winning – just under 1 in 3. He also predicts that Arsenal have a 48% chance of winning, just less than evens. A draw is unlikely, with a probability of 20%. These percentages are meant as a hypothetical example – or, in my language, a mathematical model. In reality, scouts don't provide such numerically accurate information. But bear with me, and we'll see later that the actual numbers aren't important. What is important is how the relative strengths of your own team and Arsenal determine your and Wenger's strategy.

The scout's estimate is based on the assumption that both Arsenal and you play attacking football. Let's imagine that by defending you can halve Arsenal's chance of winning, but in doing so you also halve your own chances of winning. The same applies to Arsenal: if Wenger decides to defend, then he halves both Arsenal's and your team's chance of a win. We can summarise the scout's predictions in Table 6.1, which gives the probability of winning, drawing and losing for each of the strategies that you and Wenger could adopt.

**Table 6.1** The probabilities of win (W), draw (D) and lose (L) in the Attack/Defend game.

| | Arsenal attack | Arsenal defend |
|---|---|---|
| You attack | W32%, D20%, L48% | W16%, D60%, L24% |
| You defend | W16%, D60%, L24% | W8%, D80%, L12% |

How should you think strategically in this situation? To give yourself the best chance of winning, you should play attacking football. On the other hand, if you defend you have a 60% chance of taking home a point, and if Arsenal defend too then a draw is almost guaranteed. It doesn't matter how long you think about this, these percentages alone are not in themselves enough for you to decide what to do. To work out the best strategy you need to know how many points you will get for a win.

Let's look first at two points for a win. We'll start with the situation where both teams attack and work out how many points you can expect to get. You get two points for winning, with a probability of 0.32 (*i.e.* 32%), and one point for drawing, with probability of 0.20 (*i.e.* 20%). This gives an expected outcome of

$$(0.32 \times 2) + (0.20 \times 1) = 0.84$$

This is the number given in the 'You attack/Arsenal attack' entry in Table 6.2. Working out the expected points for all the different strategy combinations, depending on what you and Wenger do, gives the rest of the numbers in the table.

**Table 6.2** Expected points in the Attack/Defend game under the two-points-for-a-win system.

| | Arsenal attack | Arsenal defend |
|---|---|---|
| You attack | $(0.32 \times 2) + (0.20 \times 1) = 0.84$ | $(0.16 \times 2) + (0.60 \times 1) = 0.92$ |
| You defend | $(0.16 \times 2) + (0.60 \times 1) = 0.92$ | $(0.08 \times 2) + (0.80 \times 1) = 0.96$ |

With the two-point system it's always better for your team to defend. If Arsenal attack, then defending increases your average expected points from 0.84 to 0.92. If Wenger also chooses to defend, then you get the best possible outcome: 0.96 points. But there is little chance of a tactical master like Wenger making this beginner's mistake. Since you are both playing for a share of the same two points, he has the opposite incentives. It's much better for Arsenal to go on the attack, and there is only one rational tactical outcome: you defend for your lives and Arsenal bombard your penalty area.

With the three-point system, the situation changes. Table 6.3 gives the expected points for all the different strategy combinations, this time multiplying by three points for a win. If you attack, then you can now expect 1.16 points, which is more than the 1.08 you get from defending. There is no strategic change for Wenger. Arsenal are the stronger side, but you should still try to go for goals. While you will lose more matches against Arsenal in the long run, you will also win more, and get more points overall. The three points for a win make the expected points pie larger, and the result is more attacking football.

## Breaking up the Pecking Order

Just as the results of football matches depend on skill, the outcomes of animal interactions depend on size and strength. One of my favourite animal contests is that between the shore crabs native to northern Europe. These crabs have different-sized claws which they use for fighting, over both food and mates. Larger crabs tend to have larger claws, and smaller crabs have smaller ones. Biologist Isabel Smallegange, working on the beautiful island of Texel north of Amsterdam, set up arenas containing pairs of crabs of various sizes and a small number of mussels for them to fight over.[1] She found that small crabs avoided larger ones, letting them have the first pick of any food. The larger crab ate the best mussels, while the small crabs had to spend more time looking for food elsewhere. Mussels are like a fixed-size pie, and the result is an aggressive 'attacking' strategy by the stronger crab and a 'defensive' strategy of backing down by the smaller crab.

**Table 6.3** Expected points in the Attack/Defend game under the three-points-for-a-win system.

| | Arsenal attack | Arsenal defend |
|---|---|---|
| You attack | $(0.32 \times 3) + (0.20 \times 1)$ $= 1.160$ | $(0.16 \times 3) + (0.60 \times 1)$ $= 01.08$ |
| You defend | $(0.16 \times 3) + (0.60 \times 1)$ $= 01.08$ | $(0.08 \times 3) + (0.80 \times 1)$ $= 1.04$ |

Contests are most interesting when the two crabs are roughly the same size. In the build-up to such an encounter, the pairs spend a lot of time both eyeing each other up and cautiously avoiding each other. Once started, the contest lasts longer than those between opponents of very different sizes. These equally sized adversaries spend a lot of time pushing each other to assess who is the best fighter, whereas with two differently sized crabs one of them quickly backs down. This is exactly what we would expect from my Attack/Defend model for playing against Arsenal. When the probability of winning is roughly equal for both sides, there is an incentive to fight, especially if a crab is unsure how strong its opponent is.

Animal competitions are not simply one-off encounters with winners and losers. In fact, the social lives and hierarchies of many animals, including humans, are very similar to the structure of football leagues. Hierarchies are something that my friend and colleague, biologist Dora Biro, knows all about. She is a lecturer at the University of Oxford's Department of Zoology, which is full of Lords and Sirs. With these dignitaries as colleagues, and celebrities such as Richard Dawkins popping in to listen to seminars, it isn't too surprising that Dora has developed an interest in social interactions.

As fascinating as her peers might be, the hierarchies that interest Dora the most are those found in the lofts of homing pigeons. She and her colleagues have set up an automated system for tracking and measuring these pigeons, both when they are feeding inside the loft and when they are flying about outside.[2] The researchers measure how pigeons queue for food, and how they approach and avoid one another. By using a computer to track the birds' movements, they can automatically identify the way in which the pigeons interact. The algorithm can identify which pigeons are dominant. When one pigeon pushes forward and another moves away to the side, then the pushing pigeon is boss.

There is also a hierarchy in the sky. By attaching GPS to the birds and watching how they turn in a group, Dora and her colleagues can work out which bird is deciding where the other birds should fly. The birds that lead in the sky are not the same as those that dominate the food on the ground, but in both cases there is a strict hierarchy. If bird A leads bird B when flying, and bird B leads C, then bird A will lead when it is flying with bird C. The same rule is found to apply to all the birds in the loft when it comes to deciding which ones will back down in an interaction. In mathematics, we call this relationship transitive, and it turns out that transitivity is about 97% effective in predicting which pigeon will dominate. This is almost as strict as the lunchtime pecking order of Oxford professors. The higher your seniority, the better the seat you get in the staff restaurant for lunch.

For both pigeons and professors, transitivity makes life simple. As a young researcher at Oxford, I didn't try to steal any of the best seats in the restaurant from the departmental dignitaries. For pigeons, too, it saves a lot of trouble if you know your place in the pecking order. If a pigeon sees another, stronger pigeon lose a fight, then it won't go and start a scrap with the new champion. Transitive hierarchies are widespread in nature. They are seen in everything from ants, through birds and rats, up to chimpanzees and, of course, humans.

These hierarchies are a consequence of the finite resources that are being competed for. In pigeon lofts, as in much of life, the aggressive strategy is not always the best one – unless, that is, you are certain you will always win. Then you should make sure you are top pigeon.

Hierarchies might reduce fighting in the pigeon loft, but they do not create an exciting football league. If Newcastle United are beaten at home by Crystal Palace one weekend, and Palace lose in mid-week at home to Aston Villa, then Newcastle may be reluctant to play for a win when they visit Villa Park the following weekend. If all the teams start to think this way, defensive football will dominate. There may be some variation due to home advantage, but if this line of thinking is followed, weaker teams will defend and stronger teams will attack.

We can now see the powerful logic behind Jimmy Hill's three-point system. The two-point system provides a fixed-size pie for the teams to fight over. If you are weak, it is always better to compromise. As a result, a boring, transitive hierarchy will form where the stronger team tries to push its advantage and the weaker team defends what it has. The three-point system provides more incentive for the weaker side, and the result – in theory – is a more attacking form of football.

But was the theory borne out? Did 'three points for a win' break the dominance hierarchy in English football? Theory is one thing, but football is another. If the managers were acting rationally, and their players were listening, we could expect to see a decrease in the number of draws when the change from two to three points was introduced to English football in the 1981/82 season. There were 118 draws in 1980/81, compared with 121 in 1981/82: an increase of three, and not exactly strong support for the theory. Maybe Jimmy was wrong after all?

But two seasons are not enough to base a proper statistical comparison on. Figure 6.1 is a plot of the number of draws for the six Division One seasons before the change and the six seasons afterwards. Now we see that in 1980/81 there was an exceptionally low number of draws compared with the seasons before the change. The five seasons with the most draws were all before the change to the three-point system; the four seasons with the lowest number of draws came afterwards. This is enough data for a statistical test to be applied, and gives good support to the conclusion that the policy did encourage attacking football.[3] If we plotted goals per game over the same seasons, we would find a slight increase after the introduction of three points for a win. We have to give him credit: Jimmy Hill was right. He broke up football's pecking order.

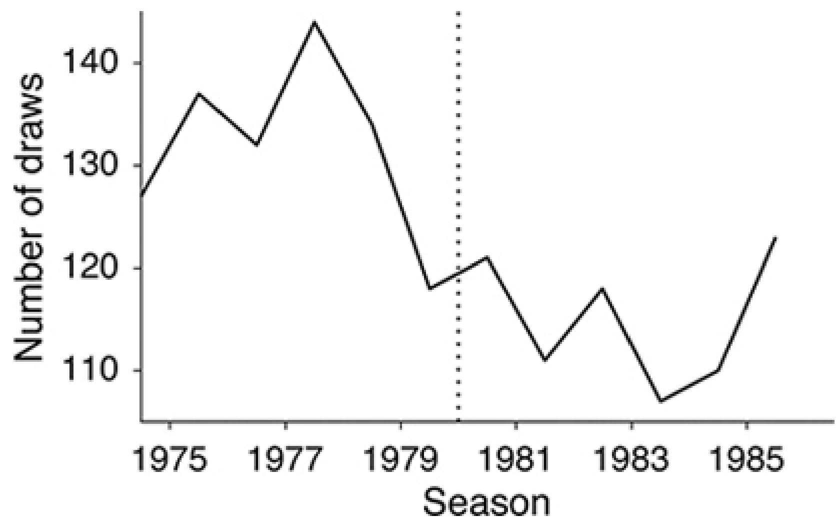



**Figure 6.1** Number of draws in the English First Division before and after the change to three points. The vertical dotted line indicates the year (1980) when the change from two to three points was implemented.

## Given Half a Chance

In my hypothetical meeting of minds between your lowly team and Wenger's Arsenal, I made quite a few assumptions. The probabilities of winning, losing or drawing were very specific, set at 32%, 48% and 20% respectively. In reality, we would expect these probabilities to change depending on the opposition, whether you are playing at home or away, the formations adopted, whether your star striker is injured, and so on.

This variety of possible outcomes isn't a problem for the mathematics of strategy. In fact, the beauty of mathematics lies in its power to make generalisations. We mathematicians don't usually work with specific probabilities, such as the likelihood of winning and losing: we try to tackle problems in a more general way. To do this we replace numbers with symbols that can take a range of values, and we prove results about the relationships between the symbols. For example, I assumed above that if you switched from attack to defence, your probability of winning would halve, from 32% to 16%, and that Arsenal's chance of winning would be halved, from 48% to 24%. To make the model general, we can assume that defending reduces the win or lose probabilities by a proportion we can call $p$. The symbol $p$ can be thought of as representing the effectiveness of the defence: $p = 1$ is a useless defence, whereas $p = 0$ is an iron curtain. In my earlier example I set $p = 0.5$, midway between these two extremes.

I can also replace the win, lose and draw percentages with symbols: $w$, $l$ and $d$. This allows me to set up a very general representation of a football match, where any outcome is possible. The question now is how the probability of different results, along with the effectiveness of a team's defence, affects the managers' tactical decisions.

Setting up the problem in terms of $w$, $l$, $d$ and $p$, and working through the algebra, I find that under two points for a win the weaker team should always defend.[4] If your opponent is more likely to win than you are (*i.e.* when $l > w$), then it is always better to defend. Again, we see a strict hierarchy where stronger teams attack and weaker teams defend.

The result is different with a three-point system. Here you should attack, provided the opposition's probability of losing is not more than twice your probability of winning (*i.e.* when $l < 2w$). On the other hand, if your opponents are twice as likely to win as you are, then you should defend. Under three points for a win, there is a wide range of situations where weaker sides should play attacking football.

In terms of mathematical sophistication, moving to symbols is a step up from numbers. And by making this step, I can now give a very general piece of strategic advice to managers whose teams are playing for three points: *Attack provided the opposition are not more than twice as likely to win than you are.* This advice applies irrespective of whether you are facing Arsenal or Accrington Stanley. You should assess your own team's strength, and that of the opposition. As long as your opponents' chance of winning is not twice as large as your own, then go for a result and never play for a draw. Given half a chance you should go for the result.

This strategic advice still leaves a lot of work for the manager. He must be sure that he knows the relative strengths of his own team and the opposition. One measure of this can be found in the bookies' odds. Another can be found by looking at the league table and comparing the number of wins. In the 2013/14 Premier League season, mid-table teams such as Crystal Palace, Newcastle United and Stoke City won about half as many matches as Liverpool, Chelsea and Manchester City up at the top. So mid-table teams shouldn't resort to defensive tactics against the top teams, but instead go into these matches looking for a result. On the other hand, those teams fighting a relegation battle should attack mid-table opposition – but when Wenger or Jurgen Klopp arrive with their superstars, they should concentrate on shutting down the penalty area and hope for a draw.

## Intuitive Design

We know from the statistics that more attacking football was played after the points change in 1981. But did the managers actually do the maths? Did they go through the calculation I made and work out that they should attack whenever they have at least half a chance of winning? It doesn't seem particularly likely. Football managers are, of course, talented individuals. They have to be able to get on with players, fans and club chairmen, and to have an intuition and understanding of the game, as well as organisational skills. Some logical thinking about tactics is also useful, but manipulating mathematical symbols is not something we expect football managers to have in their skill set. Yet they seem to have responded to changes in incentives in exactly the way that mathematics predicts. So how did managers solve the Attack/Defend problem?

To answer this question we need to go back to the birds, and also look at even smaller biological entities, such as cancer cells. It is not just pigeons that play the Attack/Defend game. A species of sparrow called the Dark-eyed Junco observes the same home-advantage rule we see in football.[5] When a group of birds is resident in an aviary and another group is added, there is less outright fighting than when both groups are placed in a completely new aviary. The 'home' birds threaten, and the 'away' birds back down. Because the birds usually compete over fixed-size resources, this is analogous to two-points-for-a-win football, and the birds adjust their strategies to reflect this. Obviously, the birds don't use league tables and mathematics. Instead, they follow patterns of behaviour that allow them to respond appropriately to different situations.

It is natural selection that gives the birds their intuition. Imagine a bird that is always timid, regardless of whether it is in its own territory or elsewhere. When it meets an aggressive bird, it is in trouble. The timid bird always backs down and gives up its territory. Pretty soon it will have nowhere to go, and no food, and it will die before reproducing. Birds that are always aggressive are also at an evolutionary disadvantage. Constantly getting into scraps, even when they are likely to lose, they will eventually meet an opponent bigger and stronger than they are. Natural selection does its job, and the aggressive birds die out. The best rule of thumb, which has evolved in Dark-eyed Juncos, is to respect home advantage. These birds behave more aggressively defending their territory and less aggressively away from home turf. The birds never explicitly work out why this strategy works, it is simply that the individuals adopting this strategy are the only ones left in the population.

An extreme example of this type of 'intuitive design' is seen in the strategy of cancer cells. Cancer does not 'plan' to take over the body, but the cells within your body mutate during your lifetime. Researchers at Johns Hopkins University have proposed a model for tumour cells that is quite like my Attack/Defend model.[6] The researchers looked at two types of tumour cell, one oxygen-rich and the other oxygen-starved. Both these cell types can produce energy using glucose, a sugar, but the oxygen-rich cells can also produce energy by combining oxygen from the blood with a different sugar, lactose. The competition between the cells comes when they both try to use glucose to make energy. If they both use this same source, they halve their energy production, and are forced to share the available glucose.

It is at this point that the oxygen-rich cells change strategy. Instead of sharing glucose, they switch to lactose-based production. Using lactose is not as efficient as when the cells had all the glucose to themselves, but it is better than sharing it with the oxygen-starved cells. Instead of competing for resources, the cells start to exploit oxygen from the blood. This is all good for the oxygen-starved cells, but not so good for the person whose body is playing host to this cellular competition. The cells start to work together, and the tumour they have created grows faster. No part of tumour growth requires intelligent planning. Cells that adopt the best strategy spread more quickly, and evolution leads them to increase in number.

If evolution can help cancer cells, then it can also help football managers. We can see this by simulating football management using a version of the computer game *Football Manager*. Unlike modern computer games, simulation models should simply capture the essence of management rather than all the realistic details. So my management simulation is more similar to the one I used to play on my Dragon 32 home computer in 1982 than the latest version on a dedicated games PC. In my simulation, instead of having lots of complicated graphics and information, for each season each team is assigned a 'strength' score between 1 and 100. When two teams meet in a simulated match, their probabilities of winning, losing and drawing are determined by their relative strength scores. The question is which strategy each team should adopt to survive the season. To determine this, we assign each team one of four rules that reflect the manager's intuition about the team's style of play. These intuitive rules are:

'Attack': This team plays attacking football against all opposition.

'Defend': This team plays defensive football against all opposition.

'Stronger': This team plays defensive football if the opposition is more likely to win than they are, but attacking football if they themselves are more likely to win. In other words, the team attacks only when they are the stronger side.

'Twice': This team plays attacking football provided the opposition is not more than twice as likely to win as they are; otherwise, they play defensively. The team defends only when their opponents are twice as strong.

In the first simulated season, five of the 20 clubs adopt each rule and play the Attack/Defend game twice against each of the other teams, with win, lose and draw probabilities determined by the teams' strengths. Table 6.4 shows final league positions over 20 seasons produced by one such computer simulation.

**Table 6.4** Example league positions for four different rules (Attack, Defend, Stronger, Twice) in a computer simulation of the Attack/ Defend model league. The top six are in bold, the bottom six in italics.

| Season | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Position** | 1 | 2 | 3 | 4 | … | 10 | 20 |
| **1st** | **Stronger** | **Twice** | **Twice** | **Twice** | … | **Attack** | **Attack** |
| **2nd** | **Stronger** | **Attack** | **Twice** | **Attack** | … | **Twice** | **Attack** |
| **3rd** | **Twice** | **Stronger** | **Twice** | **Twice** | … | **Twice** | **Twice** |
| **4th** | **Stronger** | **Attack** | **Attack** | **Stronger** | … | **Twice** | **Twice** |
| **5th** | **Attack** | **Twice** | **Stronger** | **Twice** | … | **Twice** | **Twice** |
| **6th** | **Twice** | **Twice** | **Twice** | **Stronger** | … | **Twice** | **Attack** |
| **7th** | Defend | Stronger | Stronger | Twice | … | Twice | Attack |
| **8th** | Attack | Stronger | Attack | Attack | … | Twice | Attack |
| **9th** | Stronger | Twice | Attack | Attack | … | Twice | Twice |
| **10th** | Attack | Stronger | Twice | Twice | … | Twice | Twice |
| **11th** | Attack | Stronger | Stronger | Twice | … | Attack | Twice |
| **12th** | Defend | Stronger | Attack | Attack | … | Attack | Twice |
| **13th** | Defend | Defend | Stronger | Stronger | … | Twice | Twice |
| **14th** | Twice | Attack | Twice | Twice | … | Attack | Twice |
| **15th** | *Stronger* | *Twice* | *Stronger* | *Stronger* | … | *Twice* | *Twice* |
| **16th** | *Defend* | *Attack* | *Stronger* | *Twice* | … | *Attack* | *Attack* |
| **17th** | *Attack* | *Stronger* | *Stronger* | *Twice* | … | *Twice* | *Twice* |
| **18th** | *Defend* | *Attack* | *Twice* | *Attack* | … | *Twice* | *Twice* |
| **19th** | *Twice* | *Defend* | *Attack* | *Stronger* | … | *Twice* | *Twice* |
| **20th** | *Twice* | *Defend* | *Defend* | *Twice* | … | *Attack* | *Twice* |

The first season is won by a team adopting ‘Stronger’, with another ‘Stronger’ team taking second place. ‘Twice’ and ‘Attack’ also do reasonably well, but most of the ‘Defend’ teams end up in the bottom half of the table. After this first season I apply natural selection to the 20 teams. I remove all the teams in the bottom six places (as indicated in italics in the table) and replace them with clubs assigned the rules that had been adopted by the teams that finished in the top six (indicated in bold). These changes could correspond to unsuccessful teams being relegated, or they could reflect a change in manager of the failing teams, with the new managers copying the rules of the successful teams. As a result, in the next season, there are two additional ‘Stronger’ clubs, and two fewer that ‘Defend’.

We continue the same process season after season. At the end of season three, ‘Defend’ is completely eliminated. It simply doesn’t pay off to play for a draw in every game. By season four, it isn’t going quite so well for ‘Stronger’: although initially the number of ‘Stronger’ teams rose to seven, now they are down to five again, and by season 10 there are no ‘Stronger’ teams left in the league; ‘Twice’ and ‘Attack’ dominate. Season 20 is a good year for ‘Attack’, but by this stage ‘Twice’ is the most common strategy. Figure 6.2 shows how the numbers of each strategy in the league evolve over 50 seasons.

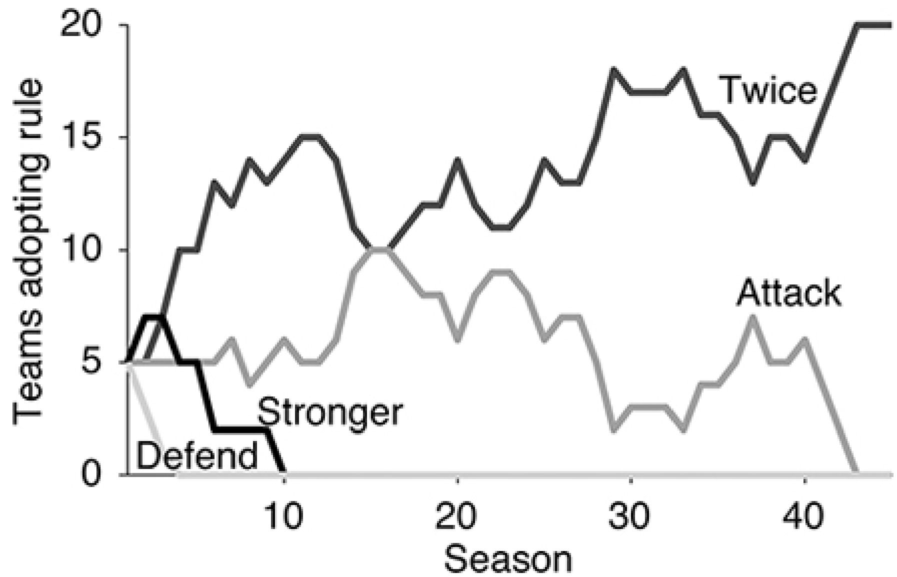



**Figure 6.2** Change in number of teams adopting the four different rules in computer simulation of Attack/Defend model league.

‘Defend’ disappears after four seasons, ‘Stronger’ holds out for 10, and ‘Attack’ lasts 43 seasons. Attack is a good strategy provided you are lucky enough to have a strong team, but it fails when your team is weak. Because the simulation changes the strength value of the team each season, ‘Attack’ is eventually beaten by ‘Twice’, which works well irrespective of the strength of the team.

Earlier in this chapter, I proved mathematically that ‘Twice’ is the best strategy. The important thing to notice here is that the ‘Twice’ rule evolves without the simulated managers of the simulated clubs having to do the maths. Managers may be no more logical than a flock of birds, but because successful managers keep their jobs and others copy their winning strategies, the ‘Twice’ rule takes over.

It is the intuitive rules adopted by the managers that evolve, rather than the managers themselves. Managers learn through a process of training, from the success of others and from their own mistakes. Over time, only the best ‘rules of thumb’ for football survive, and the managers who learn them become successful. The adoption of a strategy arises from a constant pressure to perform. These days, many Premier League managers last only one or two seasons at the same club. To survive, they do not have to be fantastic mathematicians, they just have to learn from experience. Over time, natural selection works its magic, and their intuition improves.

## It’s football – enjoy!

Jimmy Hill understood strategy when he proposed the change to three points for a win. He identified a clear practical problem: English football was getting boring, and teams were defending too much. His suggested change was a strategic tool to solve this problem, and it worked. The teams that finished the 2014/15 Premier League season in the top half of the table, but outside the Champion’s League positions, played some of the most exciting football to watch. Southampton, Swansea City and Stoke City all played with a positive style against stronger opposition.

Southampton, managed by Ronald Koeman, are a good example. They have one of the best defensive records in the league, but they can also turn defence quickly into attack. In the 2014/15 season, Southampton won away at Manchester United, but lost at home. They beat Arsenal 2-0 at home, but lost the return visit. Against Manchester City the gamble didn’t pay off, and they lost 5–0 on aggregate. But over these six matches, against wealthier and stronger opposition, they took six points. This is the maximum number of points they would have got if they had instead tried to hold out for a boring draw in each match.

The incentive to win, even at the risk of losing, makes for better football. Supporters of Premier League sides like Southampton want to see their team giving it their all against the big clubs. They want their players to take the game to the opposition. In part, it is a matter of honour for the fans of mid-table teams that their players are not scared to play football against the big boys. But when Koeman said ‘Why do we have to be afraid? It’s football – enjoy!’, as he did before Southampton’s away game against league-leaders Chelsea in 2015, he wasn’t just trying to boost morale. He was also thinking logically. No team worthy of being in the Premier League should be afraid of attacking, and those that don’t won’t be there for very long. Defensive football is not only ugly, it is also irrational.

CHAPTER SEVEN

# The Tactical Map

Can mathematics really make managers more successful? If a top-flight manager were to read this book, I can imagine that they would have one or two questions for me. In the last chapter I reduced football management to deciding between defence and attack. While this simplification is useful for thinking about strategy and incentives, it's not the whole story. There's more to management than rallying your team with a call to 'push up' or by encouraging them to 'fall back'. It's one thing to explain why Arsenal generally play attacking football, but quite another for me to start giving Arsène Wenger advice on how he should run his team. Should Wenger, Carlo Ancelotti or Pep Guardiola really start taking tactical tips from a mathematician?

To put the question even more directly: can I tell these managers something they don't know? They are the experts, after all. They live and breathe the game, and they have watched countless matches and worked with a host of star players. They have the intuition and the understanding that only years of experience can bring. What can a mathematician, with an amateur interest in football, offer the professionals?

The challenge of finding out things that no one else knows is why I do research. That challenge becomes even more interesting when I'm working with the leading experts in a field. Can I show the experts how to take a new look at their specialist area? This is how I have tried to work, throughout my research career in biology and sociology. I don't have an extensive training in either of these fields. In fact, I never even studied them at school, but I do have a different way of seeing things. I have always pushed myself to tell biologists and sociologists things that they didn't know, and now and again I have succeeded.

But I know my limitations. I don't believe that mathematics can provide grand new theories of biology, sociology, economics or football. What it does provide is a way of seeing things more clearly, as we found in Chapter 2 and 3 in our study of movement, passing and space. By taking a mathematical view we uncovered the geometry behind Barcelona, saw how Bayern defenders narrow down space, and characterised differences between Andrea Pirlo and Bastian Schweinsteiger. These insights don't completely revolutionise how managers see the game, but they are important because they sharpen our thinking. They give us an edge.

One thing I have learned from my work in biology and sociology is that in order for the message I want to get across to be understood, it must be straightforward. It is the job of the mathematician to take a lot of complex data and boil it down to a simple but powerful idea. In football, the message needs to be even more direct. In the dressing room at half-time, with the team behind by two goals and unable to keep control of the ball, the manager has to be able to convey a plan quickly. I try to imagine what I would do in that situation. The manager is lost for ideas, and in a final act of desperation he turns to me, the mathematician, and asks, ‘Can you see the problem?’

In this situation, I have to have ready a quick but clear explanation of what’s going on, so that he can act. There’s no point in me jabbering on about Voronoi diagrams, passing networks or vector-flow fields. The manager wants to see exactly what the problem is and how to fix it. To do this I need to show him something he can quickly understand and that sums up why his team is losing. I need to find the tactical map.

## Route-one England

One of the best ways for a mathematician to communicate results is through pictures. Figure 7.1 reveals some of the most important features of the Euro 2012 quarter-final between England and Italy. It shows the passing networks for both teams over the 120 minutes played, including extra time. Each point represents a player. A link between pairs of players indicates that they made 13 or more successful passes to each other during the match. Thicker lines indicate a higher number of passes.

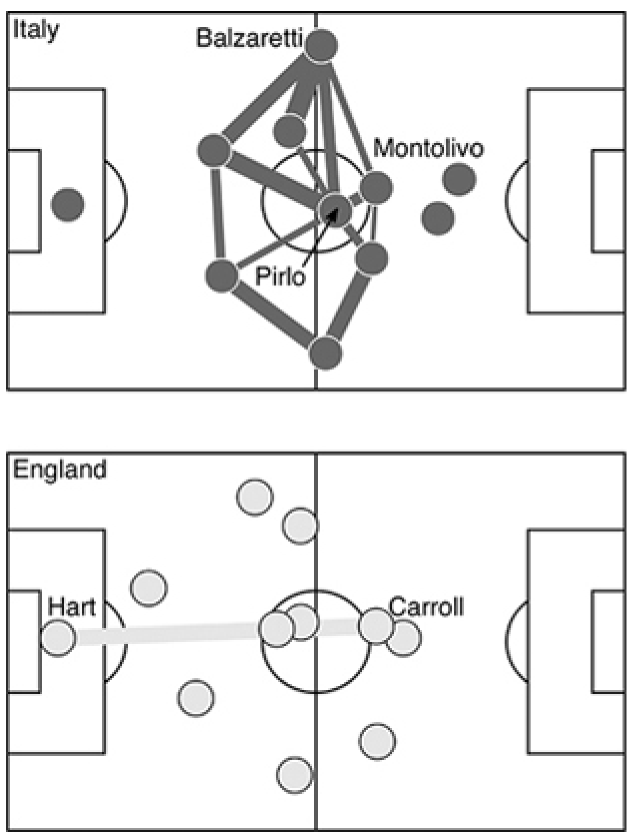



**Figure 7.1** Passing networks for Italy (top) and England (bottom) in the Euro 2012 quarter-finals, showing links between pairs of players that completed 13 or more passes between them.Thicker lines indicate a higher number of

passes. Both teams are attacking left to right.

It's immediately clear that Italy made a lot more successful passes than England. The focus was Andrea Pirlo. In Chapter 3 I showed how in 2012 Italy were always trying to get the ball to Pirlo, and this game was no exception. He dispatched the ball successfully 115 times, orchestrating most of the Italian attacks. England tried to get the ball to Wayne Rooney, playing longer balls towards him down the middle, or by crossing to him from the wings. As the match progressed, England's approach became even more direct. On 60 minutes, Andy Carroll came on. At 1.93m (6'4) he was the tallest outfield player, and controlled the airspace in the centre. Once he had recovered the ball after each of the many Italian attacks, Joe Hart kicked it directly up to Carroll (see bottom panel in Figure 7.1).

It's quite impressive that Carroll managed to win the ball so many times from Hart's kicks. Despite being on the pitch only half as long as most of the other players, he was one end of England's most successful passing partnership. But Hart and Carroll's success reveals failures elsewhere. England had no consistent passing network in midfield. Instead, the Italians dominated the match, with 68% of possession and 36 shots. Italy passed the ball forward and England tried to kick it over their heads. It was painful to watch as an England fan, with shot after shot flying towards the England goal and one desperate clearance after another.

So England struggled through 120 minutes and went out on penalties. It wouldn't have required a map of their passing network to predict that as a likely outcome. Many of England's major tournaments for the past 20 years have ended in exactly this way. However, while England's problems are plain to see, the insights into Italy are subtler. Look again at the connections in the network and notice that it looks like a wheel. Almost all the action is going in and out through Pirlo. All the other players who completed more than 13 passes with the same team-mate are also connected to him. There are six players connected to Pirlo, and four connected to Federico Balzaretti on the left wing, while the other Italian players have at most only two or three connections. Italy's network is too centralised.

How do I know that Italy were too centralised? Answering this question requires proper research, and mathematical sociologist Thomas Grund has already done the difficult part. Thomas looked at the network structure of Premier League teams during the 2006/07 and 2007/08 seasons, gathering a total of 76 match network observations for each team (apart from those teams relegated or promoted in 2007, which each had 38 networks).[1] These networks comprised a total of 283,259 passes made over the two seasons – enough data for him to draw some interesting general conclusions.

Thomas focused on two aspects of the networks. The first was a measure of passing rate: how many successful passes are made per minute of possession. To calculate this, he simply added up the number of passes made in each match and divided by the time the team had the ball. The second measure was of network centrality. This is more difficult to measure than passing rate, since there is more than one way to measure how centralised a network is.

Thomas settled on the following approach. He added up all the passes each player received and made, and compared these totals with the number of passes made by the player involved in the most passes. For example, in the Italy network above, Pirlo received the ball 78 times from the other seven players connected in the network. The second-most-passed player, Riccardo Montolivo, received it 39 times. So Montolivo received the ball on $78 - 39 = 39$ fewer occasions than Pirlo. The average of these differences over all players divided by the total number of passes made gives a number between 0 and 1. If all the passes had been made to Pirlo, this number would be 1, while if all the players had received the ball exactly the same number of times then it would be 0.

Armed with these two measures, passing rate and network centrality, Thomas looked to see how well they predicted match outcome in the two Premier League seasons. He found that those teams that passed more when in possession of the ball scored more goals. A team that passed on average five times a minute scored around 20% more goals per season than a team that passed three times a minute.[2] This difference is quite small, but it is significant, and it could make a major difference in the final league positions. Thomas also looked at centrality. Teams that focused their passing on only a few players scored fewer goals than teams that passed the ball more evenly between all members of the team. Here the difference is harder to quantify exactly, but roughly speaking, being decentralised gives a team an 8% advantage in terms of scoring rate.

It's goals that count, of course. In terms of goals, a 0–0 draw leaves Italy's centralised midfield looking no better a strategy than England's hoofing the ball up to Carroll, and Thomas Grund's analysis could explain why. Earlier in the tournament, Spain played Croatia. Like Italy, Spain had most of the possession, with control of the ball for 71% of the match. They also had a similar pass rate, 9.65 passes per minute while in possession, compared with Italy's 9.15 times per minute against England. But the difference lies in Spain's passing network, which is shown in Figure 7.2.

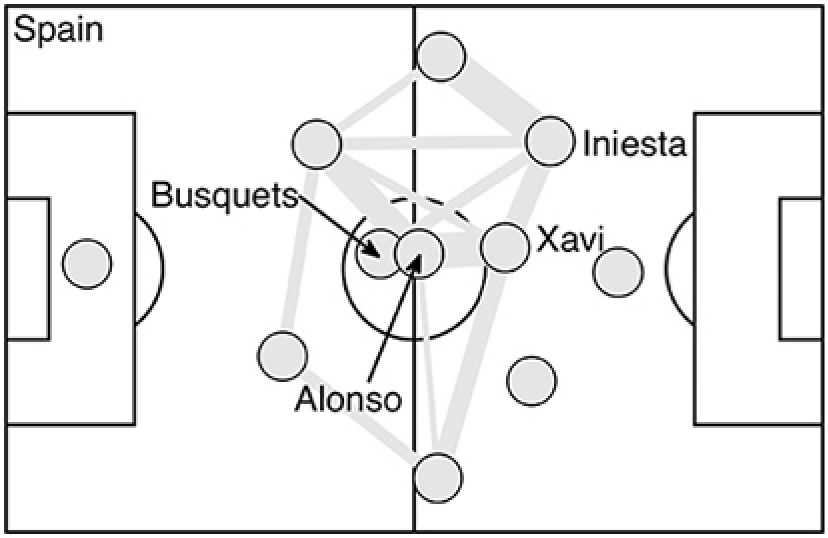



**Figure 7.2** Passing network for Spain playing against Croatia in Euro 2012, showing links between pairs of players that completed 11 or more passes between them.Thicker lines indicate a higher number of passes.

Instead of one key central midfielder, Spain have four: Busquets, Alonso, Iniesta and Xavi. The passes are distributed more or less equally. The calculated centrality measure for Spain is 14.6%, compared to 19.7% for Italy. This is a small difference, but from Thomas's study we would expect the team with the lower measure to be more successful – and they were. In their match against Spain, Croatia held out until the 88th minute, while England kept Pirlo and company at bay until the inevitable penalties. These may be small differences, but they are decisive: when Italy and Spain met in the final, Spain won 4–0. Decentralised passing triumphed.

## The United States of Abbymerica

Many teams have a focal player and a leader. For the US women's team over the past decade, this player has been forward Abby Wambach. By the time she retired in December 2015, she had scored 184 times for her country – that's more international goals than any other player in history. She has won the Associated Press Athlete of the Year and FIFA World Player of the Year awards. She has two Olympic gold medals, from 2004 and 2012. She was a key member of the team that came third in both the 2003 and 2007 Women's World Cups, and got a runners-up medal in 2011.

Wambach led the US from the front. She was known for intense team talks in which she screamed a string of expletives before telling her team-mates that they must all calm down.[3] She was a player who took responsibility for the whole team and led by example. In 2011, it was her 122nd-minute diving header that took the US to penalties against Brazil in the semi-finals. In the final against Japan, she again scored a header in extra time to put the US in the lead (but Japan equalised and won on penalties). In summer 2015, Wambach was 35 and had publicly stated that this would be her last

World Cup. She had never won the trophy, and Canada was her last chance to complete her set of honours.

Wambach's leading role can be seen through the passing network from the US's first group-stage match in 2015 against Australia. Figure 7.3 shows the major passing combinations made by the US during the game. Strikers usually receive fewer passes than midfielders, simply because it's harder to get the ball through the opposition defence and into the final third. But in Figure 7.3 we see that Wambach is well connected to her team over the whole pitch.[4]

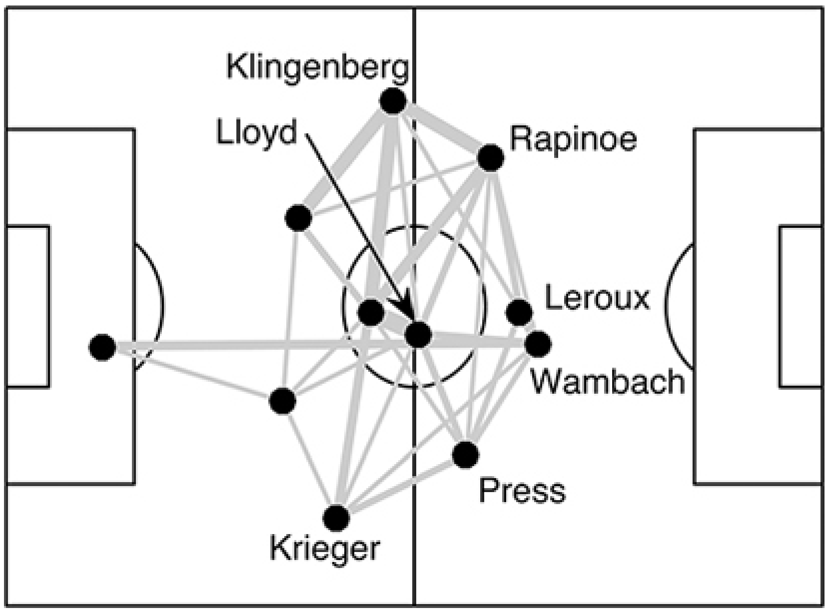



**Figure 7.3** Passing network for the US National team playing against Australia in the Women's World Cup 2015, showing links between players that completed five or more passes between them. Thicker lines indicate a higher number of passes. Original data analysis by Devin Pleuler.

The network shows Wambach's importance in the team's tactics, but may also reveal a problem for the team as a whole. Devin Pleuler, the soccer analyst who created the networks after the match, pointed out that Wambach received the ball directly from the goalkeeper, Hope Solo, just as many times as from the other big star of the team, midfielder Carli Lloyd. Likewise, the full-backs, Meghan Klingenberg and Ali Krieger, passed to Wambach the same number of times as wingers Megan Rapinoe and Christen Press. The connections going into Wambach reveal a long-ball strategy with the captain as the team's focal point. Unfortunately, Wambach struggled to make these balls count against Australia. The half-time score was 1–1, with the US taking an early lead via a lucky deflection and Australia equalising with a fine piece of rapid passing through the confused US defence. It wasn't Wambach's dream start to the World Cup.

Hidden behind the links to Wambach is another network structure. Both Rapinoe, on the left, and Press, on the right, connected well with the central midfielders and their respective full-backs. They provided width to the attack, and rather than relying solely on long balls up to a target, the US network was also decentralised. In the second half, it was these two factors that made the difference. Press scored first, perfectly positioning herself in the box for a cut-back pass from Sydney Leroux, and Rapinoe finished Australia off following a solo run down the left wing.

The US manager, Jill Ellis, had a difficult decision to make after this opening match. Wambach was captain and a leader, but the goals had come from elsewhere. In the end, Ellis went for decentralisation and Wambach was used primarily as a substitute. The network shown in Figure 7.4 presents how the US looked in their game against China in the quarter-finals without Wambach. Now the structure of the network was clearer. The team was led from the centre, by Carli Lloyd, who played and received the ball evenly in all directions. Lloyd also scored the only goal of that match.

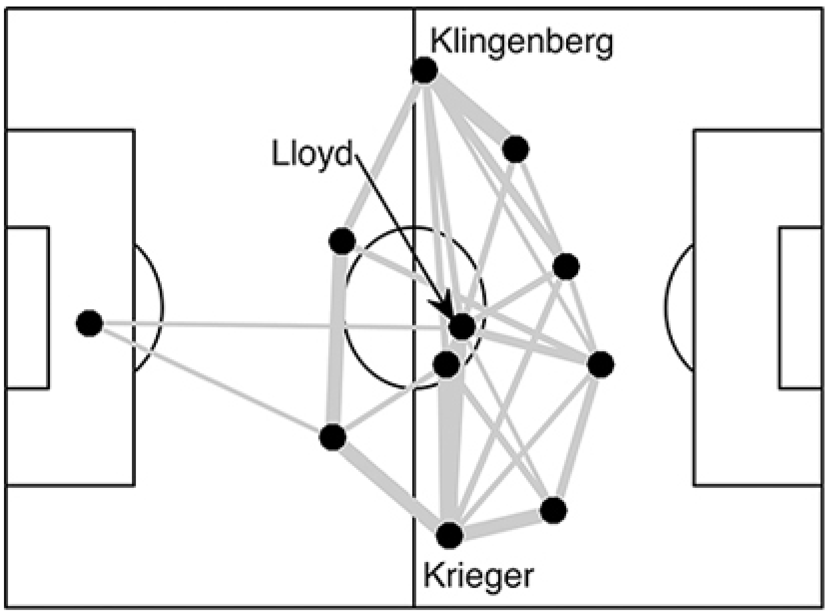



**Figure 7.4** Passing network for the US National team playing against China in the Women's World Cup 2015. For details see Figure 7.3. Original data analysis by Devin Pleuler.

Wambach still played an important role in the team. She scored the decisive goal against Nigeria that ensured that the US finished top of their group. Even on the bench, she pushed the team. In the huddle before the China match she was caught on camera instructing her team-mates, 'First 10 minutes, we get a fucking goal!' Her drive took the whole team to the final, again against Japan. This time there was no penalty shoot-out and no disappointment. Lloyd scored an incredible 13-minute hat-trick and the US won 5–2. Wambach came on for the last 15 minutes and celebrated as the US lifted their first World Cup for 16 years.

## Special Relationships

This view of passing the ball in terms of a network comes from mathematical sociology. When mathematicians approach the social world, they are looking to identify patterns that show how things work. The end result often comes in the form of a network. One of my favourite examples is the American high-school romantic and sexual network. To create this network, researchers went into a high school in the US Midwest and conducted in-home interviews to identify with whom students had had either a ‘special romantic relationship’ or a ‘non-romantic sexual relationship’ within the previous 18 months.[5] The network of these relationships is shown in Figure 7.5.

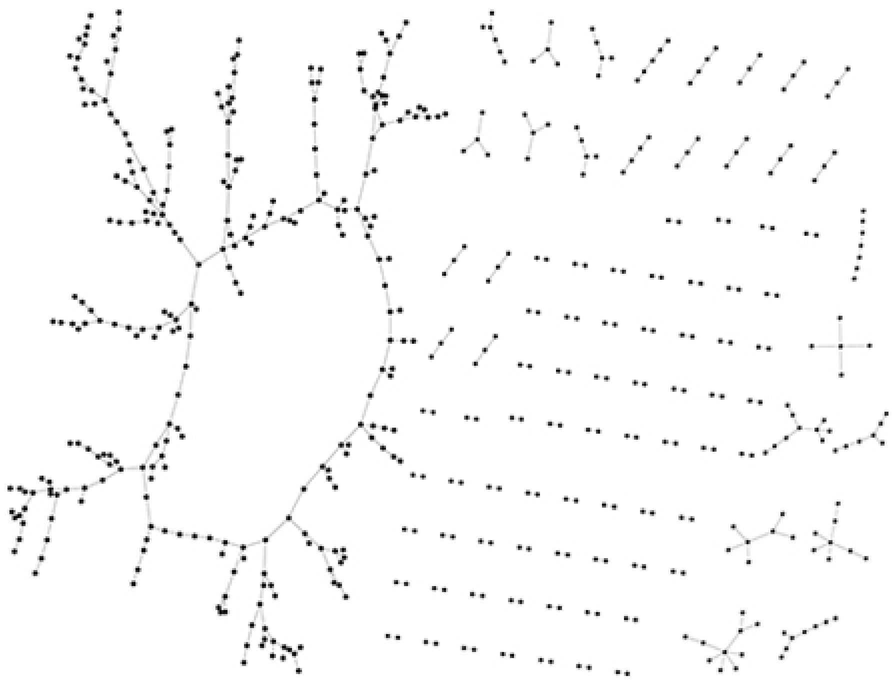


**Figure 7.5** Network of romantic and/or sexual relationships between students at a US midwest high school. Each student is a point, and a link between the students indicates a relationship. Data from the National Longitudinal Study of Adolescent Health (Add Health), a project designed by J. Richard Udry and Peter Bearman.

The entire romantic and sexual history of a year and a half in one US high school is encapsulated in this single picture. There are 63 isolated pairs, ten love triangles and a few small chains of various shapes and sizes. There is also one very large ring-like structure with branches coming out of it. This shape connects together about half the students – although they are blissfully unaware of it. All the students joined to the ring are connected by some form of romantic relationship.

One less obvious but interesting pattern is that loops are rare. If Alice has been together with Bob, and Bob has had either a romantic or a sexual relationship with Chloe, who has had a relationship with Doug, then Alice and Doug are very unlikely to get it on. At high school it just isn’t acceptable for two exes to become a couple. When I think about my friends, I can see that this is probably a feature of adult sexual networks as well. When one relationship hasn’t worked out, people don’t tend to look among their ex-partner’s friends and acquaintances to find a new romance.

Another property of this network is that it has a small number of central points. There are a few well-connected players, who dominate local groups, while most individuals have just one or two connections. These structural details are important in the understanding and control of sexually transmitted diseases. Disease wouldn't spread so quickly through this particular network, because it is disconnected and relatively sparse, but many adult sexual networks have stronger connections. Mathematical modellers use the structure of these networks to predict where a disease outbreak will occur and how large it is likely to be.

In the last 15 to 20 years, sociologists have mapped out many aspects of our social lives. Some of these maps are networks of friendships, some of work relationships and some of sexual encounters. Other maps are geographical, showing racial segregation in cities, rail and air connections, and urbanisation. These maps work best when they abstract away from unimportant detail and reveal the true structure of relationships. Maps should first give us a clear overview of the world we live in, and then allow us to zoom in more closely and look at the detail. Well-made maps allow us to explain what makes society the way it is, and do so in a way that single numbers, such as possession percentage in football or average number of girlfriends in high school, never can. Maps reveal structure.

## Fifty Shades of Bayern Munich

The maps we make of football depend on playing style. After Barcelona's success in 2011, more and more clubs adopted their approach. In the Premier League, Arsenal, Liverpool, Southampton and Swansea City all began to play a decentralised game with rapid passing. However, during his time as Chelsea manager, José Mourinho found a way of neutralising and defeating 'tippy-tappy' teams by crowding the box with defenders. And in Chapter 3 we saw that even Barcelona fell dramatically from their perch in 2013 when Bayern out-pressed them. The challenge for me is to map out the pros and cons of different tactics.

Much of this rich diversity of strategy can be seen in the Champions League semi-finals of 2015. Pep Guardiola's current club, Bayern Munich, took on his former club, Barcelona, while reigning champions Real Madrid faced a rejuvenated Juventus. Each of these clubs has its own unique way of playing the game. If I want to see how these footballing strategies have evolved, and what happens when they compete with one another, I need to dissect them. I need to find out how they are structured and how they perform. And for this I need data.

To classify each of these teams' style, I got in contact with Colm McMullan, who developed the online app Statszone for *FourFourTwo* magazine.[6] The app displays data from Opta Sports, who collect and present extensive 'on the ball' data from most major championships. The Opta data provides a complete list of who passed to whom, where the passes were made and at what time. It also has a full record of shots, tackles, blocks and so on. Nothing that happens to the ball is missed, and Statszone can be used to display it.

Over all the matches played by these four clubs in the Champions League, there are tens of thousands of passes to analyse. Bayern Munich under Guardiola made more passes per match than any other team in the 2014/15 Champions League. After the semi-finals they had made a total of 8,593 passes in 12 matches, of which 7,548 were successful. If I plotted all those passes out on the same diagram then the result would be a total mess. It would just be thousands of arrows pointing everywhere, and it would be impossible to discern any pattern in Bayern's play.

If I want to communicate something about a team directly, then my aim should be to produce maps that reveal the true underlying pattern. There's more than one way to build a map, so let's start by trying to get an overall view of where passes were made from and to by the various clubs. Figure 7.6 is a distribution map based on the Bayern data. The map divides the pitch up into 25 areas, with Bayern attacking from left to right. The lines originating from the centre of each area show the direction of the passes made from that area. The length of each line is proportional to the average length of a pass made in a particular direction. The shade of a line shows how often a particular pass was made: darker lines are directions in which passes were more frequent, while lighter lines indicate a lower frequency. For example, the Bayern goalkeeper Manuel Neuer usually starts play with a short pass out to the left-back. He passes less frequently directly forward towards the centre circle, but when he does, these passes are three times longer than his passes out towards the wings.

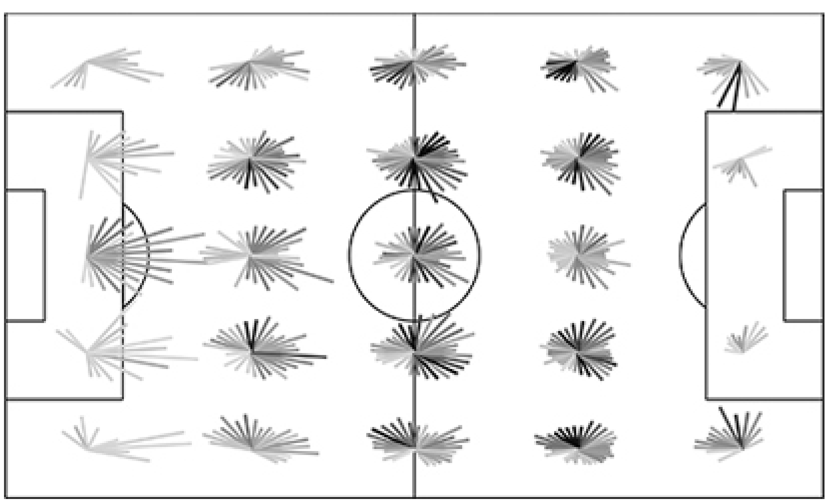


**Figure 7.6** Distribution of Bayern Munich's passes during Champions League season 2014/2015. The pitch is divided into 25 different areas. The lines emanating from the centre of each area indicate the direction of passes in that particular area. Length of the line is proportional to the average distance of a pass in that direction. The darker the shading of the line, the more often a pass was made. Bayern Munich attacking left to right. Data provided by Opta.

The distribution map in Figure 7.6 shows how Bayern control the midfield area in front of their opponent's goal. The dark lines emanating from the halfway line show that the ball is not played directly forwards, but passed out to the wing. From the wings, the ball is delivered to areas to the left and right in front of the penalty box, from where an attack is built up. Balls that go out towards the corner flag tend to be passed back in front of the goal area to restart an attack. This distribution map sums up Bayern's play over a whole season.

Juventus also got to the semi-finals of the Champions League, but did so with a very different style, as can be seen from Figure 7.7. Comparing the distribution maps of Juventus and Bayern, we see that Bayern play much farther forward and pass the ball around more than the Italian side. Juventus make a lot of passes from slightly to the left of the centre circle out to the wings. Many of these passes are the work of none other than Andrea Pirlo, who enjoyed his last season in European football playing for Juventus. Another team, another match, but the influence of Pirlo again stands out in their build-up.

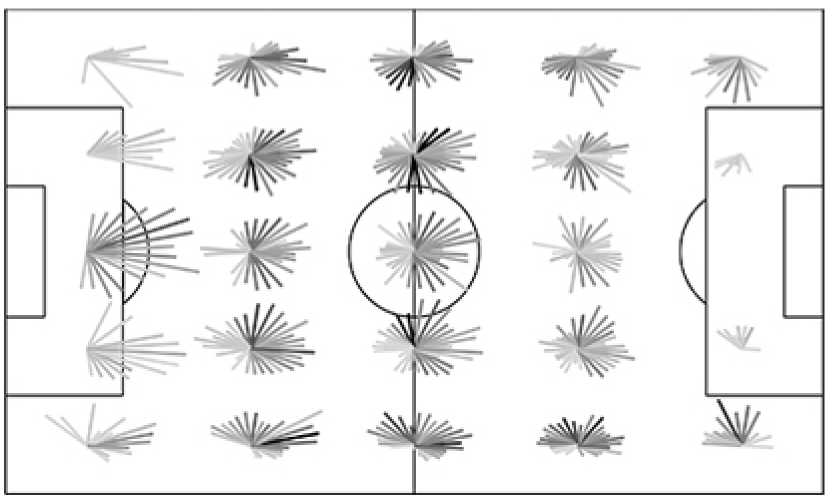


**Figure 7.7** Distribution of Juventus's passes during Champions League season 2014/2015. For details see Figure 7.6. Data provided by Opta.

Figure 7.8 compares the distribution maps of Bayern and Juventus when they are playing in their opponents' half. We see that Juventus make fewer passes in the opponents' half than the German side. The shading intensity is lower in front of the opposition's penalty area. Instead, many of their attacks originate from the right. Pirlo will feed balls over to that side of the pitch, then right-back Stephan Lichtsteiner and midfielder Claudio Marchisio link up, running down the wing and crossing into the box for Carlos Tevez and Álvaro Morata. The Italian club's intensity comes from the right-hand side of the pitch, Bayern's from the middle.

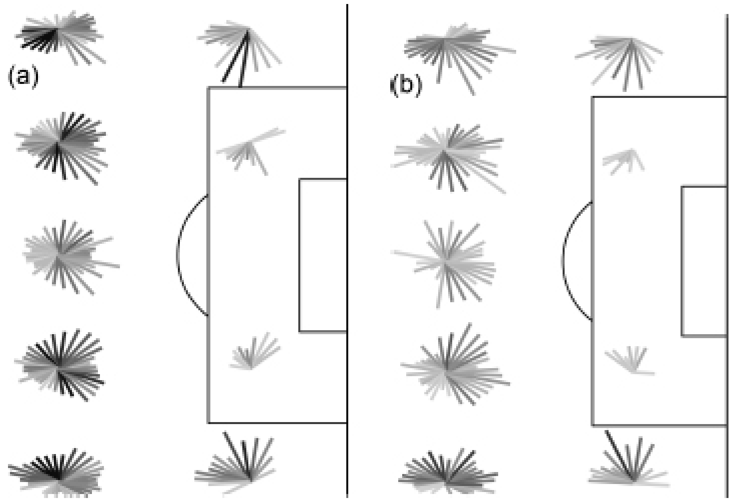



**Figure 7.8** Comparing Bayern Munich (a) and Juventus (b) passing distribution in the opposition half. See Figures 7.6 and 7.7 for details. Data provided by Opta.

## The Barcelona Jet Fighter

There is a lot of information contained in these distribution maps about each team's season, but they are by no means the whole story. In the semi-finals of the Champions League, Bayern met Barcelona. Seen in terms of a distribution map, Bayern and Barcelona don't look so different. Barcelona made almost the same number of passes as Bayern, just eight per match fewer than the German side, and the passes came from roughly the same positions on the pitch. This similarity in style is not too surprising. Guardiola invented the Barcelona style, and he had gone on to implement a new version of it at Bayern. In the build-up to the first semi-final, staged at the Nou Camp, the match was billed as a face-off between the managerial genius of Guardiola and the playing genius of Messi.

The score was 0–0 until 13 minutes from the end, when Messi scored two amazing individual goals in three minutes. For the second of these he dummied Jérôme Boateng, causing him to lose his balance and fall over, and then chipped the ball home over keeper Manuel Neuer's head. Neymar wrapped up the match with a breakaway goal in injury time. These goals were amazing, and the genius of Messi was widely celebrated, but they didn't come out of the blue: they were the consequence of a well-implemented passing structure.

Figure 7.9 shows a zonal-passing network for Barcelona. This type of network combines information about who passed to whom with information about where pairs of players passed to each other. I have created a zone for each player that shows the set of points from where he received and made a pass. For example, the arrow from Busquets' zone to Iniesta's shows that the average position of Busquets when he passed to Iniesta was slightly to the left of the centre spot, and the average position of Iniesta when he received the ball from Busquets was some way outside the centre circle. Passes from Busquets to Jordi Alba are made farther out on the left.

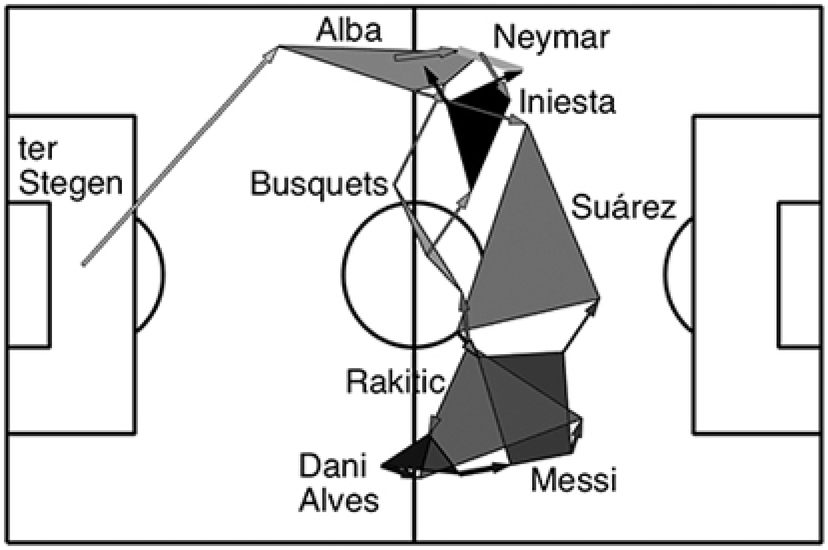



**Figure 7.9** Zonal-passing network for Barcelona playing against Bayern Munich in the first leg of their Champions League semi-final in 2015. The player at the origin of each arrow passed seven or more times to the player at the tip of the arrow. The origin of the arrow is the average position from which the first player made a pass, and the tip is the average position where the second player received the pass. The thickness of the arrows is proportional to the number of passes made. The shapes associated with the players contain all the arrows in and out to that player. Data provided by Opta.

In this zonal network I have included only those pairs of players who made at least seven passes to each other during a game. For example, Alba passes more often to Neymar than he does to Luis Suárez. Although there is a lot to take in, this figure gives a detailed picture of how Barcelona played over the 90 minutes of the match. They are well-balanced on the left and right, with Alba feeding Neymar and Dani Alves feeding Messi. Suárez then receives the ball from both sides. Notice in particular the links between Messi and Suárez. When Suárez has the ball near the centre circle, he passes it out to Messi's zone, receiving it back again as he approaches the box.

In this game, Barcelona surrendered control of the middle of the pitch. When I make the same type of zonal network for Bayern, shown in Figure 7.10, we can see how their players were focused on midfield. As in the corresponding network for Barcelona, Bayern are shown attacking from left to right – but they are stuck. Although Bayern passed more than Barcelona, the passes formed a circular motion. Xabi Alonso passed to Bastian Schweinsteiger, who passed to Thiago, who passed back to Alonso, who passed to Philipp Lahm, and so on. The arrows don't point in the direction of danger, as they do for Barcelona. Barcelona flew through the match like a jet fighter, with well-balanced wings and Suárez at the nose. Bayern built a roundabout for themselves, on which they drove round and round in circles.

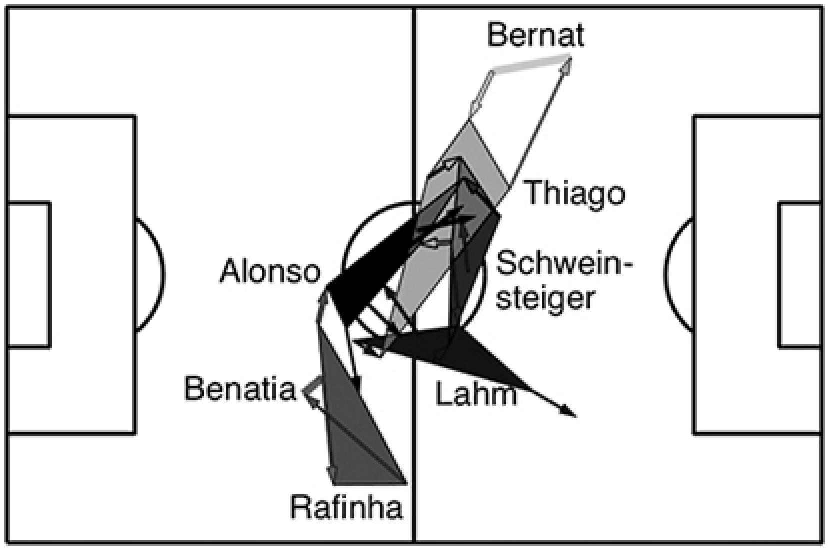



**Figure 7.10** Zonal-passing network for Bayern Munich playing against Barcelona in the first leg of their Champions League semi-final. Data provided by Opta.

## The Real Ronaldo

Before I start to get carried away again with Messi, Barcelona and their latest great Champions League side, I want to take a look at the other semi-final of the 2014/15 season. In Turin and Madrid a very different type of football was being played, but it was no less exciting or impressive. The Real Madrid side were built around a talented midfield releasing balls to three explosive forwards: Cristiano Ronaldo, Karim Benzema and Gareth Bale. Figure 7.11 is a map of their shots during their 12 Champions League games.

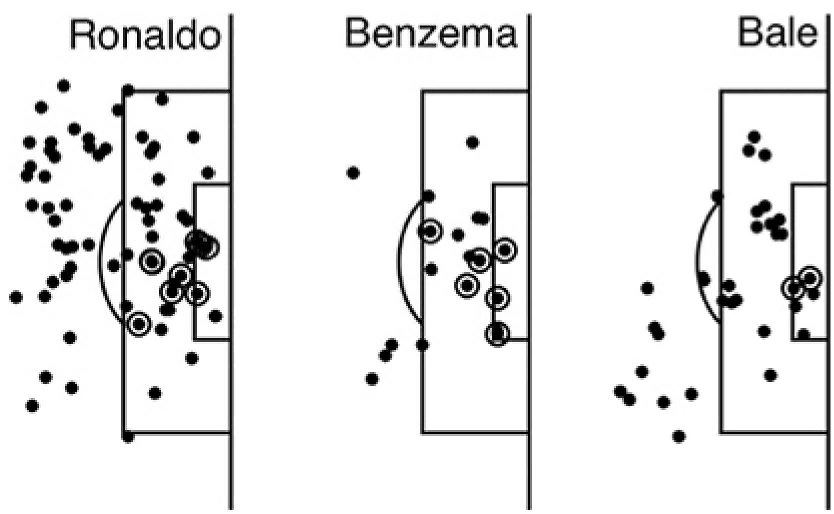



**Figure 7.11** Shots made by Cristiano Ronaldo, Karim Benzema and Gareth Bale during the Champions League season 2014/15. The dots represent the positions the players took a shot from, and the rings the positions from which they scored. Data provided by Opta.

Ronaldo gets a lot of shots in during a game, and he makes them from everywhere in and around the box. Benzema has fewer shots but enjoys a much higher conversion rate than Ronaldo, scoring with almost half the shots he makes from inside the box. In comparison, Ronaldo might look a bit wasteful: he took 35 shots from outside the box and didn't score a single time! Bale, who was criticised by some Real fans and the Spanish media during the campaign for being greedy in front of goal, was reasonably cautious. Bale has a few clusters of points he likes to shoot from: outside the box on the right, left of the penalty spot and from near the right-hand post. Unfortunately for him, he only managed to score from the last of these positions.

For any other striker, 35 misses out of 35 would look bad. But we should have learned by now to never underestimate Ronaldo. The reason he is a great all-round striker is that he creates danger all the time. Setting aside player differences for a minute, what Figure 7.11 shows us is that the most dangerous area for the ball to be is directly in front of the goal. All the goals that Ronaldo, Bale and Benzema scored came from this danger area, roughly 20m by 20m square. Any time the ball is in this area, the defence is in trouble.

Ronaldo excels at getting the ball into this danger area. We can see just how successful he is by studying Real's build-ups. In the period leading up to a shot, the ball typically moves around a lot as the attacking team tries to find a way through the defence. By marking every point at which the ball was played just before each of the Real Madrid shots during the Champions League season, we can get an overall picture of how they create successful attacks. Figure 7.12 is a risk map showing where the ball was played in the 15 seconds leading up to a shot from the 20m by 20m area in front of goal.

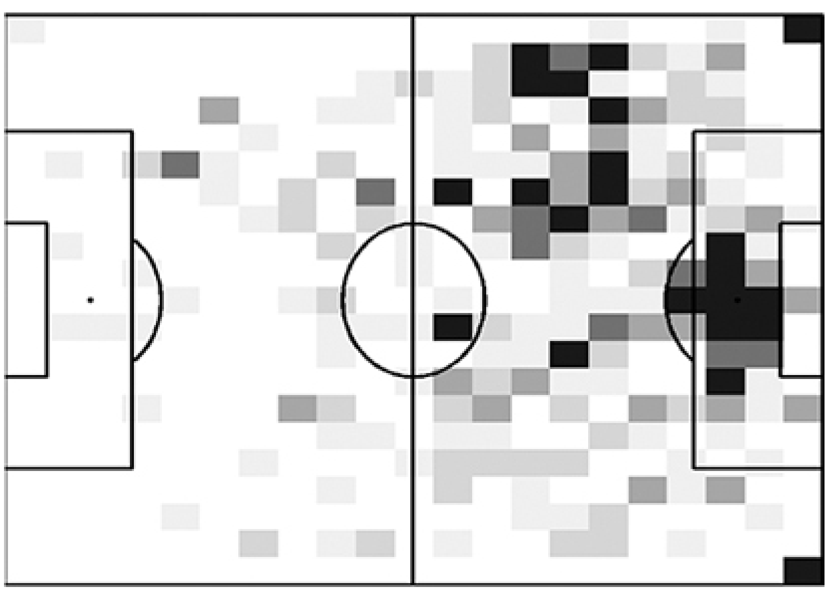


**Figure 7.12** Real Madrid danger-zones during the Champions League season 2014/15.The shading is proportionally darker in areas where the ball was located during the 15 seconds leading up to a shot from the 20m by 20m area in front of the opposition goal. Data provided by Opta.

The darker areas show places where there is a high risk of a Real Madrid shot from the danger zone coming within the next 15 seconds; the lighter areas show places where the risk is low. Corners are one clear risk-zone and, not surprisingly, if the ball is already in the box then the risk of a shot is high. But the most interesting risk zone is the hot area outside the box on Real Madrid's left. This area of the pitch is mainly occupied by Marcelo, who comes up on the left wing, and Ronaldo, who is more central. It is from here that dangerous chances are created.

There is always a risk of conceding a goal when Ronaldo is attacking. Either he runs into the box himself, or he passes to Benzema, Rodriguez or Bale, or he takes a shot from outside the box. The last of these strategies may not quite have worked out in the Champions League 2014/15, but the others did. Ronaldo scored seven goals in open play, plus three penalties, and made four assists. As a team, Real Madrid had a total of 224 shots, on and off target, during their Champions League campaign. That was 20 more than Bayern Munich, and nearly 50 more than Barcelona made before the final. Carlo Ancelotti had a simple but effective strategy for his side: get the ball, get it out to the left, get it back into the middle, and shoot.

## The Juventus Hull

Juventus were well aware of this danger from the left before they met Real Madrid in the semi-finals. Italian teams are famous for their well-organised defences and stinging counter-attacks, and this Juventus side was no exception. They had only 151 shots in the same number of matches that Real had 224, but they kept progressing through the group and knockout stages.

Unlike the Italian national team of 2012, this Juventus side was not built solely around Andrea Pirlo. Figure 7.13 shows the Juventus passing network for their first semi-final match against Real Madrid. The links in this network are weighted by the number of passes. This time I have not set a threshold for including passes, so all passes are shown here, even the single pass that Leonardo Bonucci made to Patrice Evra. The thicknesses of the lines show how common different passes were. While Pirlo is still central, he shares the role with Arturo Vidal, and the ball was often played out to Evra on the left or Lichtsteiner on the right.

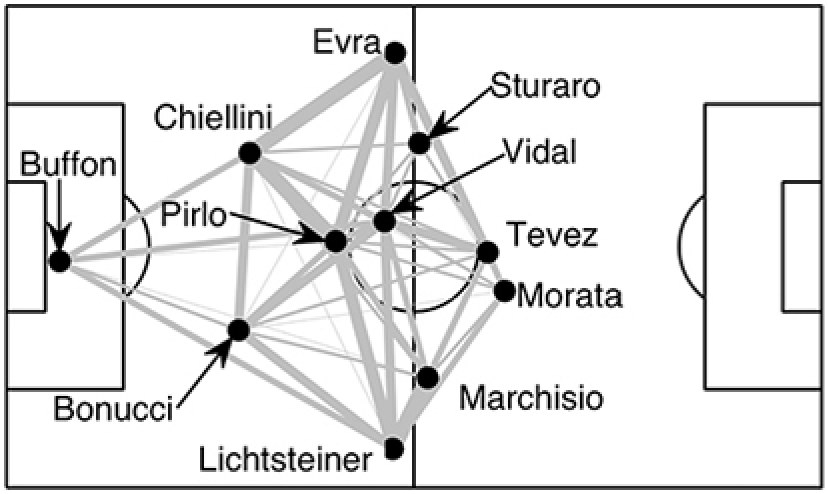



**Figure 7.13** Passing network for Juventus playing against Real Madrid in the first semi-final match between the clubs in 2014/15. Thicker lines indicate a higher number of passes between pairs of players. Data provided by Opta.

It was the Juventus defensive play that was most important against Real Madrid. Tevez and Morata are the only players whose average position is forward of the halfway line, with the others prioritising defence over attack. Each of the back four – Evra, Chiellini, Bonucci and Lichtsteiner – have key areas that they defend. On the left, Stefano Sturaro returns to help out in defence, while Marchisio comes back on the right.

To get a better feeling for how Juventus structured their defence in the first semi-final, I have created a defensive map for each player. To do this, I first looked at the points in the match where possession switched from Real Madrid to Juventus. The change of possession can come through a tackle, an interception, a goal-line save or any other way in which the ball can be recovered. Figure 7.14 shows how the 17 points at which Lichtsteiner intercepted the ball are spread out all over the right side of the pitch. The grey area containing these interception points is known mathematically as the convex hull. It is the smallest possible shape, without any kinks in it, that contains all the points.[7]

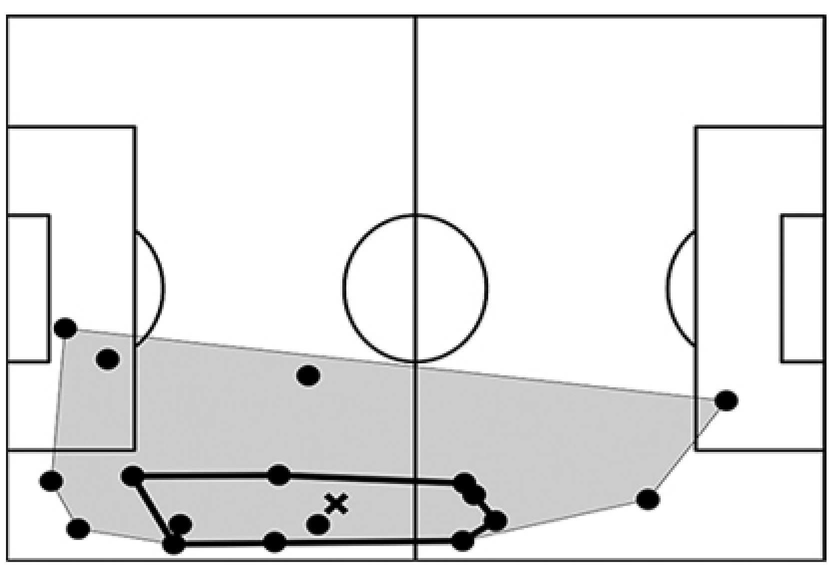


**Figure 7.14** Defensive actions of Stephan Lichtsteiner against Real Madrid. Circles show all the positions that Lichtsteiner intercepted the ball. The light shaded area contains all of these points. The solid line is the convex hull of the points 'nearest' to the average position of all Lichtsteiner's defensive actions. Data provided by Opta.

While Lichtsteiner made interceptions over a wide area, we want to get a better idea of where he typically operates. The average position of his interception points is marked as a cross. In order to find where Lichtsteiner's most important defensive contribution lies, I looked for all the points that are reasonably near his average interception position.[8] For these points, I then calculated a new convex hull. This is shown as the dark solid line. It is this smaller area that I call Lichtsteiner's defensive hull. By calculating these defensive hulls for all the players, we can see how they defended as a team.

The defensive hulls of Juventus in the first semi-final are shown in Figure 7.15. Probably the most important defensive hulls are those of Lichtsteiner and Marchisio, on the right, since they cover the opponents' left wing. Benzema was injured in this game, so Ronaldo moved into the middle, and Isco played on the left. But the focus of Real Madrid's attack remained down their left. Lichtsteiner and Marchisio regained possession 32 times between them, while Evra and Sturaro intercepted only 19 times. When Juventus went into a 2–1 lead after an hour of play, the left-sided midfielder Sturaro was replaced by the defender Andrea Barzagli, to create a defensive back five. Real Madrid came at Juventus from their left, and Juventus dealt with the pressure. The scoreline remained 2–1 until the end.

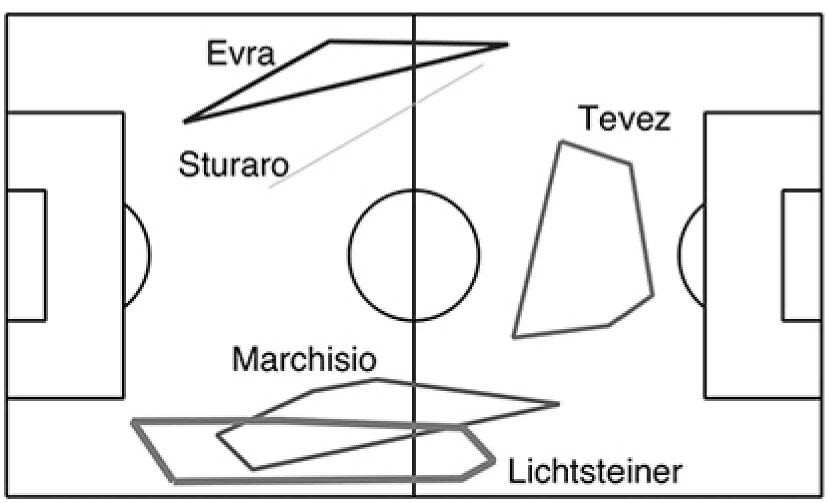



**Figure 7.15** Defensive convex hulls of five Juventus players against Real Madrid. See Figure 7.14 for details of how these hulls were constructed. Data provided by Opta.

The defensiveness of Juventus can be contrasted with the attacking football played by Real Madrid. Figure 7.16 shows the defensive hulls of five of the Real players during the match. Possession changes in Real's favour are made much further forward, regularly occurring in the Juventus half. This is typical of pressing high up and playing attacking football. Not everyone was pressing, however. Ronaldo gained possession only twice during the match, while Tevez (Figure 7.15) gained possession 18 times. It seems that the centre-forward at Juventus has to work hard defensively to create chances. It's not quite the same for the head of the Galácticos.

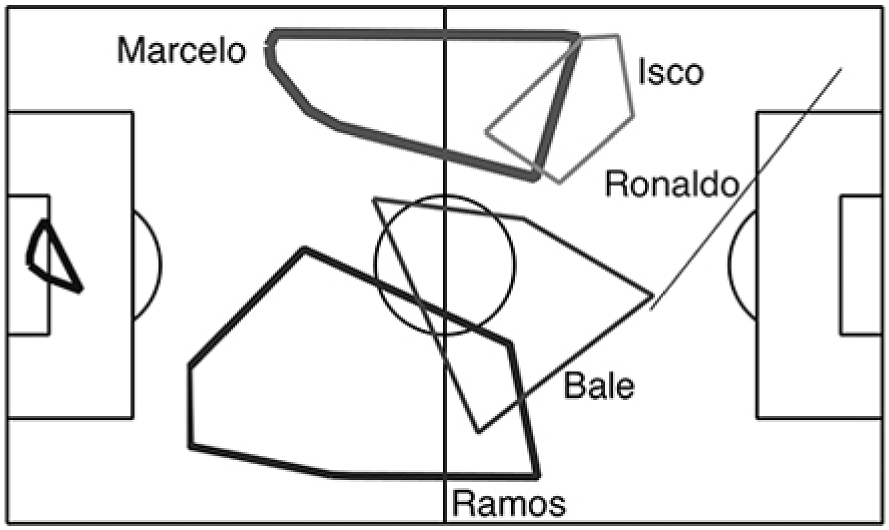



**Figure 7.16** Defensive convex hulls of five Real Madrid players against Juventus. Data provided by Opta.

Real Madrid's departure from the Champions League after a 1–1 draw in the return leg at the Bernabéu cost Carlo Ancelotti his job. Ronaldo scored a penalty to add to his goal in the first leg, and became, with Messi, the tournament's joint top scorer. His team had played explosive, attacking football. This style had been enough the year before for them to lift the Champions League trophy. But this time they fell before they could test themselves against Barcelona in the final.

## A Final Word

I performed my tactical analysis of the semis in the weeks leading up to the Champions League final. Most commentators were predicting a walkover for Barcelona, but I wasn't so sure. Plotting Juventus' defensive hull at the other end of the pitch from Barcelona's attacking jet fighter, it seemed to me that one could very well contain the other. I was looking forward to a good match.

Three minutes in, confidence in my defensive analysis of Juventus was slightly shaken. Barcelona scored a goal with a build-up that involved every player in the team apart from Suárez. And the only reason Suárez didn't finish off the move was that Ivan Rakitić ran onto the penalty spot and converted a pass from Iniesta. It looked as though Barcelona had opened up even more passing options, and that Juventus were completely unable to deal with their attack.

But the game settled down, and Juventus started to frustrate Barcelona. Ten minutes into the second half, Lichtsteiner made an interception on the front edge of his usual defensive hull. Marchisio received the ball and returned it with a back-heel to Lichtsteiner, who had now run to the edge of the Barcelona box. Lichtsteiner passed to Tevez, whose shot was saved, but Morata made no mistake on the rebound. It was 1–1. The goal was entirely consistent with what we had seen in the semi-final. Juventus were dangerous on the right, and it was from there that they got their goal. The game continued to be evenly balanced, but as it approached the end the Barcelona jet fighter proved impossible to stop. Suárez scored a second, and Neymar finished it off with a third in injury time.

No football match is predictable, but the tactical maps presented in this chapter can help teams prepare. I don't know the extent to which top clubs use visualisation tools like these. Managers each have their own method for studying the opposition and planning. So even if they employ analysts who go through the data, it isn't clear how the message contained in the data carries up to the manager or out to the players.

But my message is clear. To present information on anything, from a football match to a social network, analysts need to be able to present complex data in terms of accurate visual messages. By looking at the defensive hull of Juventus, Barcelona manager Luis Enrique would have seen how effective Lichtsteiner and Marchisio were in shutting down Real Madrid's attacks down the left and starting counter-attacks. By studying their zonal-passing network, Juventus boss Massimiliano Allegri would have seen how the forward-flying, counter-attacking Barcelona of 2015 differed from the intense passing of Pep Guardiola's side. These are the pictures that should be communicated to managers, and the tactical maps they should pay attention to. And maybe by doing so, even the most experienced coaches could find out a few things they didn't know about football.

CHAPTER EIGHT

# Total Cyber Dynamo

After a long week of squabbles and petty arguments, there is nothing better than spending a Saturday morning coaching a children's football team. Fewer temper tantrums, smaller egos and greater excitement make kids' football the perfect antidote to my working week in mathematics research.

Coaching kids also provides valuable insights into cooperation. The boys I coach have no problem with giving 100% for Upsala IF. But there are different ways of giving everything. There's charging after the ball, and there's looking up and passing. I stand at the edge of the pitch shouting at the kids to spread out. If only two or three of our boys could move into space, the game would change completely. Suddenly, one of our defenders controls the ball and sees that one of his team-mates is free. Unfortunately, before the other players get a chance to move into an effective formation, all six outfield players from the other team descend on the ball and manhandle my poor defender out of the way. 1–0 to the opposition.

Until the boys each know their part and play their role, it is difficult for them to see that the overall structure works. I know that if they play as a team, they are quite capable of opening up the opposition, but until they start playing in this way the opposition appears stronger. Frustration sets in and ball-chasing starts all over again.

This dilemma is shared by managers at all levels. Of course, professional players understand a bit more about positioning than 10-year-olds do, but the strategic problems for the managers of top clubs are similar. They need to worry about their own team much more than the opposition. If they can't get their team to play as a unit, their insights into the opposition will be more or less irrelevant. The same is true for all of us. At school, at work and at home, it is our interactions with those close to us that are most important. Getting on with our friends, our colleagues and our family is the key to a happy life.

Do football managers hold the answer to groups of people getting on together? I think they do. I believe that a good manager will create a structure such that it will pay for players to play for the team. This isn't an easy task for the manager or the players. For all of us there are always temptations to act selfishly, to let the team down. But I think football managers have, at least in part, solved this problem. They know the secret of good teamwork, and this secret is something we can all use in our everyday life. It may even be the secret to why cooperation has evolved in animals, and in humans. But in order to uncover this secret we need to think more about the underlying mathematics of teams. We need to think strategically about when it pays to cooperate, and when cooperation fails.

## Evolving Lazy Workers

In our everyday dealings with others, it is cooperation that usually provides us with the biggest headaches. Who should do the washing-up or vacuum the living room? Who should stay late at work and finish an important report? How do we get everyone to work together, instead of just focusing on making ourselves look good?

The mathematics we need to tackle these cooperative conundrums starts with the pros and cons. A ‘pros and cons’ list is something people draw up when faced with one of life’s big decisions: ‘Should I change my job?’, ‘Should I get married?’ or ‘Should I go on holiday with my family or go to the World Cup with my mates?’ The idea is to write down all the good and bad things associated with each possible choice. But, in a social world, lists don’t work. Lists don’t account for how others might adjust their behaviour in anticipation of your decision. For this, we need a pros and cons table.

In a pros and cons table we write down the good and bad outcomes for choices, but also consider what others might do. Let’s take an everyday social conundrum. It’s Tuesday evening, and you’ve promised to finish a report for work before Wednesday morning. The pro of writing the report is that your boss will be happy and you will feel like you have achieved something. The con is that you have tickets to watch your beloved Derby County against Nottingham Forest in the Championship. It’s life’s constant dilemma: job versus football.

Let’s add a colleague into the equation. Maybe he will work on the report during the evening. In this case, first thing in the morning you can paste in a few ideas of your own and hand the whole thing over to your boss as a collaboration. This is the perfect solution: your colleague does the bulk of the work, and you go and watch Forest lose. We’ll give this outcome three points for a win.

Wait a minute, though. What if your colleague is thinking the same way? He assumes you will write the report, and heads off to the pub with his friends. Manchester United are playing in the Champions League, and they’re showing it on the big screen. You both get into work on Wednesday morning and there’s nothing to show the boss. There’s a big argument, and you lose an important contract. We’ll give that zero points for a loss.

There are other possibilities. What if you write the report and your colleague goes to the pub? Well, you miss the football, but things are easier in the morning. And at least your colleague is grateful. It’s not exactly a draw, but it’s probably worth half a point. The last possibility is that you both work on the report. Now it is a masterpiece. Your boss is overjoyed and wants to make it up to you. Next week United are visiting Pride Park in the cup, and he buys you tickets so you can go together. It’s a better result than a draw, but nothing makes up for missing a home game against your local rivals. We'll give this outcome two points.

I summarise this dilemma in a 'pros and cons' table – Table 8.1. This table tells you what you should do if you know what the other person is going to do. If you think your colleague will go off to the pub, then the best thing to do is to cut your losses and write the report. But if you think he will write the report, then why waste your time? You may as well let him get on with it. You can also think probabilistically. If you think there's a 50% chance that your colleague will write the report, then if you watch the football you would expect to get 1.5 points, but only 1.25 points if you write the report. It's still better not to work on the report. What you choose to do depends on how you think people around you will behave.

**Table 8.1** Pros and cons table for the 'shirk or work' model. The entries are the pay-offs you obtain, given your own and your colleague's strategies.

| | Colleague watches football (shirk) | Colleague writes report (work) |
|---|---|---|
| You watch football (shirk) | 0 | 3 |
| You write report (work) | 0.5 | 2 |

Large companies or state-run organisations face the 'shirk or work' dilemma all the time. The company employees work together on different projects in pairs and small groups, and there are opportunities for individuals to either get actively involved or not really pull their weight. Since new groups and projects form all the time, it can be difficult to keep track of who is contributing and who isn't.

Let's assume that the company starts off with everyone highly motivated and keen to impress. Quite soon, one of their smarter employees notices that she can work a little bit less on a project. At first no one else notices, and with most projects everyone in the group is still working hard. After a while, however, another employee has the same idea and also starts piggybacking on the hard work of the others. This continues to spread as more and more people see the benefits of letting other people do their work for them.

We can use an evolutionary model to work out how the proportion of shirkers and workers changes with time.[1] We start with all individuals as workers, apart from one shirker. These individuals meet in pairs and get points according to Table 8.1. Most workers will get two points, because they meet another worker. But the worker who is paired with the shirker gets 0.5, and the shirker gets three. This gives shirking a higher chance of spreading in the population as others copy the strategy, and, slowly but surely, it does increase.

Shirking doesn't increase indefinitely, however. Once two-thirds of the company's employees are shirkers, then the average pay-off for a shirker is

$$\left(\frac{1}{3}\times 3\right)+\left(\frac{2}{3}\times 0\right)=1$$

because she has a 1/3 chance of meeting a worker and getting full points, but a 2/3 chance of meeting a shirker and getting zero points. Likewise, the average pay-off for a worker is

$$\left(\frac{1}{3}\times 2\right)+\left(\frac{2}{3}\times \frac{1}{2}\right)=1$$

It doesn't matter if you shirk or work, the average pay-off is the same. So shirking will never take over completely. In the extreme where everyone tries to avoid work, it becomes better for a smart individual to write the report and get half a point.

The situation for this company is pretty dire. Two out of every three employees avoid contributing properly. So the probability that both employees on a project shirk responsibility is 2/3 × 2/3 = 4/9. Only one project out of every nine (1/3 × 1/3 = 1/9) will involve two workers who take responsibility. Everyone follows their own interests, and ultimately no one is happy.

## Tribal Hunting

Seeing the world in terms of games such as 'shirk or work' is a big problem for cooperation. The models predict that if we follow our selfish interests, cooperative strategies will decrease in the population. Biologists think of these problems in terms of the selfish gene. If the worker strategy is a gene that encodes for cooperative behaviour, then it isn't going to do very well when it has to compete with shirker genes. When a shirker meets a worker, the shirker will get all the benefit and pass on her genes to the next generation. As a result, the model predicts that evolution does not favour selfless cooperation, even though this might be best for all individuals in the long run.

However, these theoretical predictions don't square with observations. In nature we see cooperation everywhere. I have already talked about how lions hunt in teams and fish coordinate their motion away from a predator. These are just a few examples. Birds call to warn of approaching danger; chimps groom each other; ants leave pheromone trails and honeybees dance to show their nest-mates the way to food; and meerkats take turns to watch out for dangerous predators. The list goes on and on. In some cases, such as in honeybees and ants, worker individuals never reproduce, and just work to allow the queen bee or queen ant to produce all the eggs. If evolutionary theory is a recipe for selfishness, then why are animals so nice to each other?

Thankfully, cooperation is just as common in humans as it is in animals. We do nice things for one another all the time. Workplaces function smoothly and people do what is expected of them. We also help one another with childcare, engage in the community and remember to go and visit our grandmothers. We even have fire crews, police and soldiers who risk their lives for us. There are problems now and again, but people are on the whole generous, helpful and often selfless.

One answer to why animals and humans cooperate is genetic relatedness. If you help your nieces and nephews, then they will do better in life. They share many of their genes with you, so by helping them you help your genes indirectly. This was the idea put forward by the biologist William Hamilton, summarised in this simple rule:

$$rb > c$$

What this says is that *r*, the probability that you share any particular gene with another individual, multiplied by *b*, the benefit you give to that person by helping them, should be greater than the cost, *c*, that you incur by helping. If you apply Hamilton's rule in your own life, you can use it to work out how much time you should spend kicking a football about with your niece or reading books to your mum's half-brother's grandson. Your niece shares ¼ of your genes, while your daughter shares ½ of your genes, so the rule suggests that you should invest twice as much time playing football with your own daughter as your niece. Your mum's half-brother's grandson shares $\frac{1}{32}$ of your genes, so 3.75 minutes should be long enough for that bedtime story.[2]

Before you start setting the timer, I must clarify how Hamilton's rule should be interpreted. It should not be taken as a guideline for modern living. Humans are nice to one another, often irrespective of whether they are family members or not. Today we live on a planet of seven billion people and contribute to communities far away from our original families. Instead of being a rule for everyday relations, Hamilton's rule is part of an explanation of how our cooperative behaviour evolved in the first place.

Our hunter-gatherer ancestors lived in small groups of relatives who competed with other groups for food. By helping members of their own group, and chasing away members of rival groups, they monopolised resources. One theory is that human cooperation prospered within these tribes of relatives.[3] Together they would perform dangerous hunts and care for one another's young. Helping could come at a cost: a group member could be trampled by an angry mammoth while trying protect their niece, but this genetic propensity for helping the group would be passed on as the tribe prospered. As agriculture and then industry rapidly took over from hunting and gathering, family groupings became less strong. However, in the twenty-first century the tendency to cooperate with those close to us remains encoded in our genes.

A similar evolution away from tribalism can be seen in football. When Celtic won the European Cup in 1967, all but one member of the team had been born within 10 miles of Celtic Park. The exception was Bobby Lennox, who was born in Saltcoats, 30 miles away. There is something special about a group of players who shared the same religion, working-class values and droll sense of humour conquering the whole of Europe. It wasn't that the best football players happened to be born in Glasgow in the 1940s, it was that their tribal team spirit carried them to victory. It is an achievement of which Celtic fans remain proud to this day.

Celtic's achievement will probably never be repeated. The Manchester United class of 1992 and the homegrown Barcelona squad of the 2000s provide some form of modern-day counterpart. However, while these groups of players started playing together when they were kids, they were brought to their clubs from all over the UK and Spain, and in some cases from abroad. Current big-spenders Manchester City and Paris Saint-Germain field hardly any players who have come up through their academy systems.

These changes don't make us any less passionate about our team. Football may have started with a group of local lads representing a town, but in the modern era clubs have players from all over the world in their squads. The Newcastle United squad of 2014/15 contained more players who were born in France than in England, and included nationals of Argentina, Holland, Wales, Switzerland, Jamaica, Ivory Coast, Senegal, Spain and Martinique. Just one of their players, Sammy Ameobi, was born in Newcastle. But in listening to the Toon Army chant 'Demba Ba, Demba Ba, Demba Demba Ba, NaNa NaNa NaNa NaNa Na Na Na' to the tune of KC and the Sunshine Band's 'Give It Up', you could sense the same passion as in the 'Shearer Shearer' of a decade earlier. That was until Demba Ba was sold to Chelsea, of course.

It is remarkable how quickly fans get behind their players. Humans can't help forming bonds with one another, irrespective of race and religion. Somewhere in our evolutionary history, when we lived in family groups, these bonds may have been between kin. But as our interactions have widened, so too have our social bonds. There is still racism in football – and this shouldn't be ignored, of course – but neither should we forget our amazing ability to form attachments with one another. It is an important part of what makes us human.

## Super-linear Dynamo

A tightly knit group of players taking on and beating a team of individual superstars is the essence of many of the most enduring footballing stories. It is what makes the Celtic 'Lisbon Lions' victory over Inter Milan in 1967 so special. Another equally remarkable story revolves around the three separate Dynamo Kiev sides of the 1970s, '80s and '90s that triumphed in Europe. On three different occasions, Valeriy Lobanovskyi took over as manager of Dynamo and led them proudly into European competition. They won the Cup Winners' Cup in 1975 and then again in '86. Then one last time, in 1999, after many of the best players had left Ukraine for better-paid football elsewhere, Lobanovskyi took Dynamo to the semi-finals of the Champions League. This time it wasn't to be, with Dynamo surrendering a 3–1 home lead to a visiting Bayern Munich, but Lobanovskyi's double comeback still encapsulates the tale of the committed underdogs working hard for one another and for their manager to achieve greatness.

There was more to Dynamo's success than gritty determination and local pride. There was also mathematics. The young Lobanovskyi was not only a brilliant footballer but also academically talented. He received a gold medal in mathematics when he graduated from secondary school, and went on to study engineering at the National Technical University in Kiev. There he learned about the new discipline of cybernetics, an area of science which was attracting increasing interest in the USSR at the time.

Cybernetics is the study of how parts of a system interact to create structures. In the 1960s and '70s it was a new way of thinking that went on to have a profound influence on both Western and Soviet science. The term is a bit dated now, and has been taken over by science fiction, but at the time it was a catch-all name for a new mathematical way of viewing complex systems. It is a way of thinking that has laid the groundwork for how researchers like me use mathematical modelling today. We don't call it cybernetics anymore, but the tools from this era were the forerunners to those I applied in the earlier chapters of this book, especially when I looked at the dynamics and structure of football.

When he was a manager, Lobanovskyi didn't have access to the type of automated analysis tools that I can now just open up on my laptop. But he was already thinking mathematically about football in the 1970s. Tactically, Lobanovskyi emphasised the importance of pressing: when the opponents had the ball, his whole team had to put pressure on them, irrespective of where they were on the pitch. Then, as soon as possession was gained, the whole team had to move to attack. In order to enforce his approach, he would set statistical targets for the players and then post their performance on a noticeboard after the match.

Even more groundbreaking than Lobanovskyi's use of statistics was the way he incorporated the cybernetics philosophy into his coaching. He was the first trainer to consider the football team as a mathematical system. Jonathan Wilson, author of *Inverting The Pyramid: The History of Football Tactics*, writes that Lobanovskyi saw football as 'two sub-systems of 11 elements, moving within a defined area and subject to a series of restrictions'. For Lobanovskyi, the remarkable property of football was that 'the efficiency of the sub-system was greater than the sum of the efficiencies of the elements that comprise it'.[4] He was first to realise that a team's performance can become more than the sum of its parts.

Figure 8.1 shows three plausible curves for how the performance of a team could increase with the sum of the players' efforts. In the curve in the middle the performance grows linearly, as a straight line, with effort. Here, the performance of the team is exactly proportional to the sum of the players' efforts. In the left curve performance grows sub-linearly, and is proportional to the square root of effort. Now the performance is *less* than the sum of the players' efforts. On the right, the performance grows super-linearly, and is proportional to effort squared: the team performance is now *more* than the sum of the individual efforts.[5]

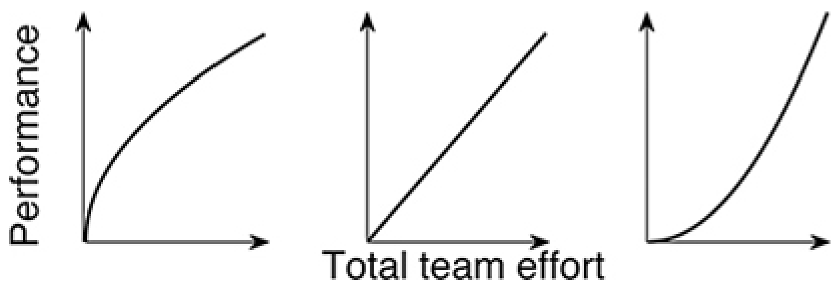



**Figure 8.1** Three team performance curves: sub-linear (left), linear (middle) and super-linear (right).

A team's performance curve depends on how they all play. A sub-linear curve implies that a team is less than the sum of its parts, that it has some form of redundancy. Imagine a forward who is already marked by a competent defender. If an additional defender starts to mark the same forward, it may make it slightly harder for that forward to score, but the improvement in the efficiency of the defence – and the team as a whole – will be marginal. There are diminishing returns to asking more and more defenders to mark a player.

Lobanovskyi argued that this type of redundancy should be avoided. Players should be able to trust one another to fulfil their individual roles and find a way of creating new space and movement. How the players synchronise their movements is important. If one player fails to function, the whole team falls apart. A football team can and should be super-linear. Modern coaches would agree with Lobanovskyi. Positioning, passing networks and the timing of runs all rely on the players combining their efforts. These factors make the team much more than the sum of its parts. If we plot the total effort of the team against its overall performance, we would expect it to grow super-linearly, as in the right-hand curve in Figure 8.1. If a professional team's performance is less than the sum of the players' efforts, then there is a problem.

## Rallying the Ants

Super-linear teams evolved millions of years before Lobanovskyi was born, and to find these teams you don't need to look much further than your back garden. Ant colonies consist of thousands of interchangeable parts. Each ant contributes to collecting food and rearing new ants, but if one ant disappears there are many others that can take its role. This makes ant colonies the ideal system in which to study teamwork. The question that I and my colleague, Dutch biologist Madeleine Beekman, set out to answer is how the number of ants in a colony affects its team performance. Do ants become more than the sum of their parts as the colony increases in size?

To answer this question, Madeleine got out her ant vacuum cleaner. She used it to suck up ants from one colony and deposit them in another, allowing her to create ant colonies ranging in size from a hundred or so to several thousand. Ants communicate by leaving a trail of pheromone, a chemical signal, when they find food. If one ant leaves a pheromone trail and another ant finds it before it has evaporated, then the second ant will also find the food, and leave more pheromone. The question is, what will happen first – will the pheromone evaporate, or will another ant reinforce the trail? The probability of these events depends both on the size of the colony and on how many ants are already on the trail. In small colonies it is less likely that another ant will find the trail, but in large colonies it is much more likely.

By turning this verbal description into a mathematical model, we were able to predict the team performance curve for the ants.[6] We have already seen three theoretical performance curves, in Figure 8.1, showing how performance can increase sub-linearly, linearly or super-linearly with effort. For the ants we predicted a fourth type of performance curve, shown in Figure 8.2. Now there is no longer a smooth increase in performance with effort: instead, there are two separate lines for performance with a jump between them. The bottom line in Figure 8.2 represents a performance flatline, while the upper line represents successful performance. The downward-pointing arrow for 300 ants shows that small colonies will always perform poorly in food collection, tending toward the flatline. The upward-pointing arrow for 1,100 ants shows that larger colonies will always perform well. Madeleine confirmed both these predictions in her experiments. Small colonies couldn't make a pheromone trail to food, while very large colonies could do so with ease.

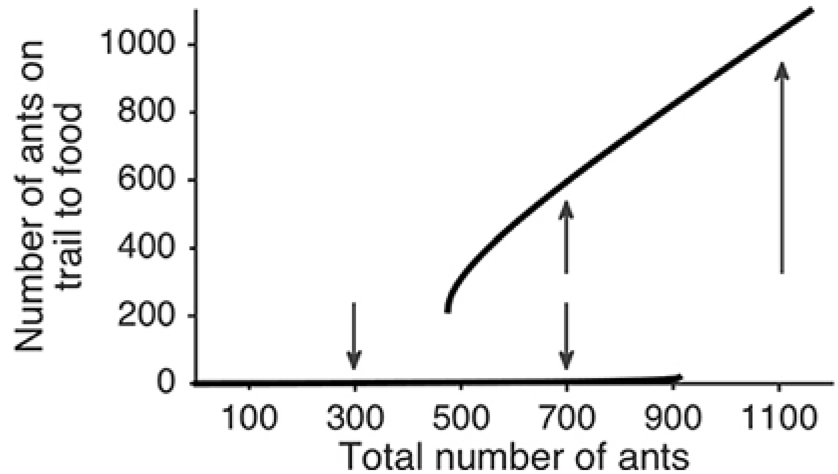



**Figure 8.2** Team performance curve for ants. The two solid lines shows how the number of ants finding food increases with the total number of ants. The lower line is the case when only small numbers of ants find food. The upper line is the case when there are enough ants to make a trail. When lots of ants find food initially they reach the upper line (indicated by upward pointing arrow); when few find it they drop to the lower line (downward pointing arrows).

While small colonies are always unsuccessful and large colonies are always successful, the performance of medium-sized colonies can have two different levels. The arrows in Figure 8.2 show that for 700 ants the performance level reached depends on how many ants form the trail initially. An ant colony that starts out performing poorly, with few ants on the trail, will remain on the lower line. But if the ants are initially performing well, with an established trail, then performance increases and the trail stabilises at the upper line. Madeleine was able to test this prediction in the ants, showing that the long-term performance of medium-sized colonies depends on how many ants initially formed the trail.

My model and Madeleine's experiments show that the same team with the same number of members putting in the same effort doesn't always have the same level of performance. This has important implications for football teams. Think about how quickly the tables can turn. One minute a team are powering ahead, putting pressure on their opponents; the next minute they are on the back foot as their opponents attack them. It's not unusual for the same team to be both successful and unsuccessful when they are using the same strategy.

Part of the explanation for these big variations in performance lies in Figure 8.2. I have redrawn this performance curve in Figure 8.3, marking different points that could arise during a game. Imagine that the team are in state A. They are putting in a reasonable amount of effort, but their performance is flatlining. Then a Steven Gerrard or Roy Keane character issues a rallying call. He pushes up their effort level, albeit temporarily, moving the team effort up to point B. As a result, the performance is lifted to point C, and the team are flying. But such an effort is unsustainable. Even the flapping arms of Gerrard or a barrage of four-letter words from Keane can't sustain such a big effort. But it doesn't matter – as the effort level decreases again, the performance stays on the higher branch. Even when the effort drops back to point D, exactly the same as the level the team started with, the performance remains much better than before.

These transitions are why a charismatic leader is worth so much to a team. He or she may only be able to lift the overall effort by a small amount, but that can be enough to achieve a top-level performance. Once the performance level is up, the effort can drop a bit but the performance level is maintained.

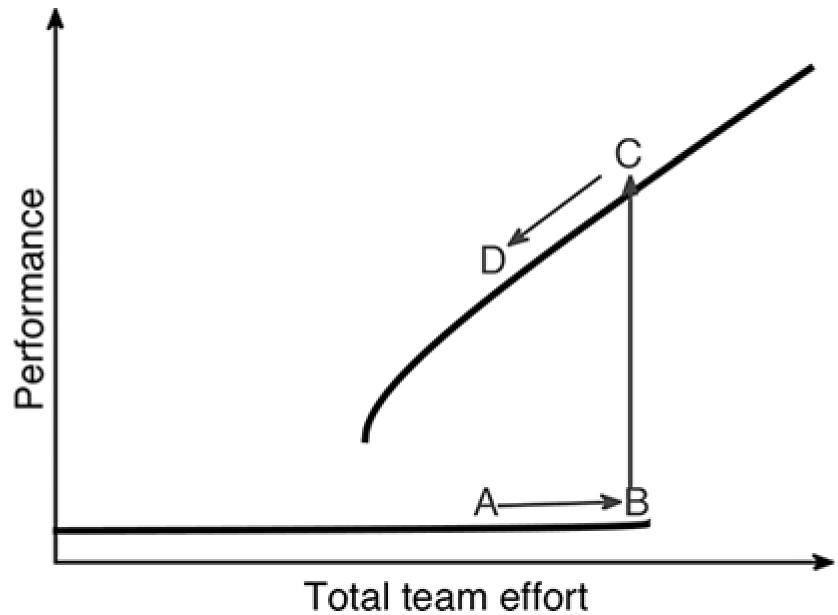



**Figure 8.3** Team performance curve in football. The team starts at point A. As effort increases it moves to point B, at which point performance jumps to point C. Effort can now decrease again, but performance remains relatively high at point D.

But we need to be careful here. The switch in performance can work in the opposite direction: complacency can creep into a team that is performing well. Imagine that a team is at point D in Figure 8.3, and their effort decreases a bit. There is now a risk that their performance will drop off at the other end of the curve. Once it is down, it becomes much more difficult to lift it again. The team needs to get all the way back up to A and then up to B, before a performance jump. This requires everyone in the team to be on the top of their game. And with each player having different motivational factors, this may not be easy to achieve.

## Star Commitment

I started this chapter with the question of cooperation. How do football teams, companies and animal groups prevent shirking? One answer to this question was through genetic relatedness, but Lobanovskyi's super-linearity idea gives us an alternative answer. Biologists and company directors can actually learn from football managers. Managers don't punish players for not getting on with their jobs – instead, they create a team structure that ensures that it always pays to cooperate.

To understand why super-linearity promotes cooperation, consider the problem managers face with getting players to contribute to the team plan. With lucrative transfer deals and agents eager to increase players' salaries by securing a move elsewhere, winning matches is not the only thing on a footballer's mind. The classic example is the disgruntled star, possibly looking for a transfer to a better-off club. I'm not going to name any names, but there are plenty of players who always have one eye on their next career move. In following the role assigned him by his manager, a star such as this may not shine so much as an individual. The question for our star is whether it is better to play as part of the team or to try to look as good possible as an individual. The challenge for the manager is to convince every player that it is worthwhile to follow the team plan.

Let's consider a team with a super-linear performance curve, one in which performance is proportional to effort squared, as shown in the right-hand plot in Figure 8.1. Each player's effort can be ranked on a scale from 0 to 1. If all the players contribute 100%, then the sum of their efforts is 11. For the super-linear performance curve, the team's effectiveness is 11 × 11 = 121. If one player contributes 0% and the others give 100%, then the sum becomes (10 × 10) + (0 × 1) = 100. If only one player contributes, then the effectiveness is just (1 × 1) + (0 × 10) = 1. If we assume that the players share equally in the glory of their performance, then their own pay-off is 121/11 = 11 when everyone puts in 100% effort. It is 100/11 = 9.09 if all but one player gives 100%, and 1/11 = 0.09 if only one player puts in any effort.

We can now look at how a disgruntled star might assess his pros and cons. The dilemma for him is not so much whether or not he will work hard on the pitch, but whether he should make an effort to help achieve the team plan. Instead of working for the team, he can try to make himself look good, by running at goal when he should have passed or by letting his defensive work slip. Let's assume that all the effort that he doesn't put into the team can instead be put into making himself look good. Table 8.2 summarises the pay-offs for a star player who puts in either 100% or 0% team effort.

Start by looking at what happens when the rest of the team gives 100%. If the star player puts in 100% effort, then he gets a big benefit from the team performance, 11 points. On the other hand, if he puts in 0% effort then he gets his personal glory (+1), but the team as a whole has lost out and now gets only 9.09. Our underperforming star gets a total of 10.09 points, less than the 11 he would have got if he had stuck to the manager's plan. If the rest of the team is committed to the game plan, then it is also in the star's own selfish interest to follow the plan.

**Table 8.2**. Pay-off table for the 'star commitment' model. The entries are the pay-offs the star player obtains, given his own and the rest of the team's efforts.

| | Team gives 100% effort | Team gives 0% effort |
|---|---|---|
| Star gives 100% effort | 121/11 = 11 | 1/11 = 0.09 |
| Star gives 0% effort | (100/11) + 1 = 10.09 | (0/11) + 1 = 1 |

All the players, even a showboating superstar, have an incentive to play for the team because the effectiveness of the team increases super-linearly. When the team is functioning properly, then if any player fails to play his part the whole team suffers, including the freeloader. The same argument holds if the star decides to contribute only 90%, or any other reduced level of commitment.[7] It never pays to give less than 100% if everyone else is pulling their weight.

From this analysis, it may seem that cooperation shouldn't be a problem. If players are working in an efficient team, and that team is more than the sum of its parts, then there are incentives for everyone to play their part. All the manager has to do is get his team working together, and everyone will be happy. Unfortunately for the manager, things are a little more complicated than that. If we look at the second column in Table 8.2, we see a very different situation. When the rest of the team isn't putting in the effort, what should the star do? If he gives 100%, then, once his small contribution is divided up between himself and his team-mates, he looks even worse than the others. If he gives 0%, the team still fails, but at least he looks like the best of a bad lot. Now the incentive is to let the rest of the team down.

This is the nightmare scenario for managers, as the players' confidence in him and in one another falls apart. The team isn't functioning, and there's no incentive for any player to commit to the failing plan. The same logic that applies to the star player applies to all members of the team. It is not in the interest of any of them to start giving 100% when no one else is committed. The puzzling thing here for the manager is that, apart from the players' attitudes, nothing has really changed. The team has the same potential to perform as when it was more than the sum of its parts, but it is no longer in the interest of any individual to put in that extra effort.

Teams that are more than the sum of their parts are disproportionally stronger when everyone is committed – but they are also disproportionally weaker when their parts start to fail. If a team strengthens rapidly as contributions are added, then it will also weaken rapidly as they are taken away. If contributions drop sufficiently, then suddenly it is in nobody's interest to invest in a failing team. It is easy for the manager to motivate players when it is in their own interest to do things his way, but motivating them when it is not is not easy.

The paradox is that by making the team more than the sum of its parts, the manager also makes it vulnerable to spectacular failure. In the model developed in this section, if the combined effort of the players falls below 50%, then it is no longer in any of the players' interest to contribute to the team. So whether a manager succeeds or fails depends on how well they function from the start. To avoid collapse, the manager needs to build up trust and make the players commit again. He needs to do this simultaneously for all the players, rebuilding faith within the team before the chairman and the board lose faith in him.

## Collective Individualism

Madeleine Beekman is not the only Dutch person to understand super-linearity and its possible paradoxes. For the past 50 years, football coaches in the Netherlands have been experts on this concept, and it lies at the basis of their long-term success. Louis van Gaal has said that management is all about instilling 'team discipline and self-discipline, individual responsibility and collective responsibility. Only then does the whole become more than the sum of the parts.'[8] When he said this, he was talking about the brand of 'total football' that developed at Ajax and in the Netherlands in the 1970s and '80s. There has always been room in total football for superstars such as Johan Cruyff and Arjen Robben, but for van Gaal the challenge was to get them integrated into the team.

In 1988, Lobanovskyi took a Soviet Union team, consisting largely of his own Dynamo players, to the European Championships. They played excellently and made it through to the final. There they faced a Dutch side managed by the father of total football, Rinus Michels. On the pitch, Michels' style demanded the same intense degree of teamwork and pressing as did Lobanovskyi's. However, while Lobanovskyi exercised tight control over his players, Michels accepted that he was working with stars 'who all happen to be millionaires'.[9] AC Milan strikers Marco van Basten and Ruud Gullit had to be motivated to play as part of the team. To create the right atmosphere, Michels would take his players on hour-long walks in the countryside, helping them bond and so that he could observe their personal chemistry. As in my 'star commitment' model, Michel created a team structure in which van Basten and Gullit would benefit more from contributing to the team plan than from simply playing their own game. The super-linear combination of these individually brilliant players was then the key to success.

The final of the Championships saw these two teams meet: Lobanovskyi's Soviet team, where individualism was suppressed so that collectivism would shine, versus Michels' Dutch individuals, brought together by a common goal. The teams had played each other in the group stage, and then it was the Soviets who triumphed, packing the midfield and forcing the Dutch to play long balls. But in the final it was the turn of the Dutch stars to shine. The first goal was a header by Ruud Gullit, but it was the second that was special. Marco van Basten's volley over the keeper from a narrow angle is still recognised as one of the greatest international goals of all time. But it all started with a team plan. With the game in the balance, the left-back Adri van Tiggelen pressed and recovered the ball as the USSR passed into the Dutch half. He immediately turned defence into attack and charged forward, laying the ball off to midfielder Arnold Mühren just outside the box. Afterwards Mühren admitted that he was tired, and the cross he made was not his best effort. But that's where superstar quality is needed, and superstar quality is what van Basten had. He hit the ball first time, and the match and the tournament were decided.

Cooperation can be simplified by thinking in terms of models. I started this chapter with the 'shirk or work' dilemma. I then moved on to families and ended with super-linear superstars. These three models can be used to explain many of the examples of cooperation we see in nature. When mathematical biologists argue about why so many species of animals cooperate, or economists try to explain why people are irrationally generous, they are usually trying to identify which of these three models is the most appropriate for explaining their observations. These questions continue to fascinate scientists, and generate a lot of debate and controversy in both the natural sciences and the social sciences.[10]

Successful managers are well aware of these models of cooperation and are skilled in using them. They probably don't formulate them in quite the way I have here, but they will know what I'm talking about. They have to get their players to form family-like bonds, they have to spot shirkers, and they have get their team to play as more than the sum of its parts. They also have to accept the ups and downs of cooperation, as morale collapses or returns for no apparent reason. Creating cooperation is not an easy job. I'm glad that I only have to implement it on the football pitch at weekends and can spend the rest of the week studying it in the peace and quiet of my office. But difficult as it may be, creating cooperation is one of the most important challenges we face.

CHAPTER NINE

# The World in Motion

Each field of academic research creates its own small world. Researchers study something in incredible detail, such as how fish respond to the position of the rest of the shoal or the shape of ant trails, and they get sucked in. They discuss whether position, speed or orientation is the most important factor in fish-turning responses. They argue about the chemical composition of the ants' pheromones.

I have to admit that I am sometimes guilty of getting lost in small academic microcosms. Just now it is football, but before that I was into fish movement and before that it was ant trails. There's something captivating about working intensely on a subject in minute detail. But this focus on details can lead us scientists to become too involved in our own world, talking to one another in our own language, and forgetting that there are other equally interesting worlds elsewhere.

So I was very excited to discover that football researchers had already found out about my fish and ant research, even before I began researching football. Shortly after I started writing this book, I went to Google Scholar to scout out what proper sports scientists were up to, to find out what their research world looked like. I got quite a surprise. One of the first articles I came across was a review entitled 'Sports teams as superorganisms' by Ricardo Duarte and colleagues in Lisbon.[1] They had picked up on the same analogy I used in the last chapter between teams and ants, claiming that 'team performance analysis could benefit from the adoption of biological models used to explain how repeated interactions between grouping individuals scale to emergent social collective behaviours'. And not only that – one of their primary sources was a review article I had written in 2005, 'The principles of collective animal behaviour'.[2] Sports scientists were already using my research to inspire their own thinking about teams.

I write scientific papers because I hope that my findings will be read by a wide range of people, not just those working in my own field. In 2005, my interest in football was limited to an occasional game of 'drop in' five-a-side at the university sports centre, and following Liverpool's progress in the Champions League on TV. There was certainly nothing in my review article about the game. So it was particularly pleasing to find out that a few years on, researchers working in Lisbon had found inspiration in my writing.

Ricardo Duarte and the Lisbon researchers are not alone in finding inspiration from biology. Natalia Balague, professor at the University of Barcelona, writes that training exercises for team sports should not ‘inform the athlete about a theoretically ideal motor output, but create tasks where skill can solve the constantly changing situations’. She quotes a leading Spanish coach as saying of his team, ‘When I see them moving like a flock of birds I know they are playing well.’[3] Paul Power, lead data scientist at sports analysis company Prozone, starts his presentations for football clubs with videos of fish schools. His talk then moves from the ocean to the pitch, emphasising the similar dynamics of players chasing the ball and sardines dodging a shark.

Many of the mathematical models I have developed in this book have drawn analogies with animal movement: the space creation of Barcelona and the schooling formation of fish, the hunting of lions and the narrowing down of space by Holger Badstuber, and the super-linear teams of ants. And it seems that sports researchers are starting to think the same way that I do. Models of animal behaviour are being adopted as models of team performance: we can use the way we study the co-ordination and movement of animals to improve the co-ordination and movement of footballers.

## Data Overflow

Figure 9.1 shows the positions and directions of 16 players at a particular moment during an eight-a-side training match for the FC Nürnberg reserves. The lighter team’s left-winger has the ball and is looking to make a forward pass. In the centre of the pitch, two dark-grey defenders are marking two light-grey attackers, while making sure they hold an offside line.

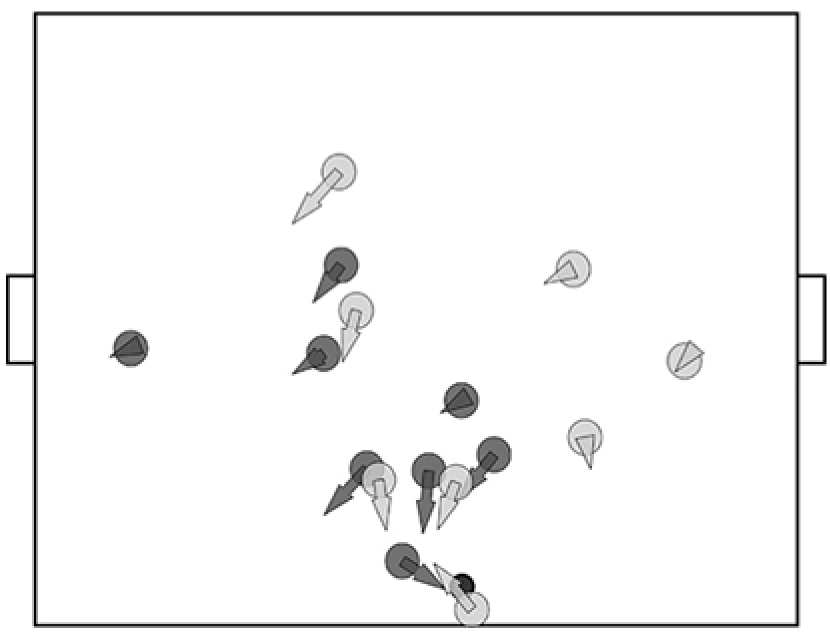


**Figure 9.1** Positions and directions of players during a training match between FC Nürnberg reserves. Circles are the positions of the opposing light and dark teams, while the arrows show the direction and speed the players are moving.The longer the arrow, the faster the player is moving. The ball is the small black circle.

This training match was special, because it was one of the first times that real-time football tracking data was made openly available.[4] All players had sensors placed on both of their boots, measuring the position of their feet. The ball also contained a sensor, as did the goalkeepers' gloves. The positions of the feet and hands were measured 200 times per second, and the ball's position 2,000 times per second. This provided 8,400 three-dimensional positional events per second. The result was a 6GB data set covering 60 minutes of play – and nearly 120 million numbers to crunch.

This is part of the information explosion that is facing football today. For a competitive match, tracking 22 players at 200 measurements per second, plus the ball at 2,000 measurements per second, in three-dimensional coordinates over 95 minutes, produces

$$\left[(200 \times 22 \times 2) + (200 \times 2 \times 4) + 2{,}000\right] \times 3 \times 60 \times 95$$
$$= 212{,}040{,}000 \text{ numbers}$$

And that's just for one match. A season of Premier League football would generate nearly 100 billion numbers.

A few years ago, the manager of a Premier League team would watch his own team's match on Saturday. On the Sunday he would put on a video of the next opponent's latest matches, as well as watching highlights from the other teams in the league. Now he has access to millions of data points for his own players and the opposition. Added to that, he has data on performance during training as well as fitness metrics of all his players. His job is to take these billions of numbers and reduce them to a few simple sentences that explain to his players the strategy for their next match.

It's no surprise, then, that football teams now employ experts to get the most out of the data they collect. Manchester City has a large team of analysts who collate team performance data after every match. This allows them to identify areas of the pitch where they are performing well and areas where they are failing. Liverpool's Head of Sports Analytics, Ian Graham, has a PhD in theoretical physics from Cambridge. He provides goalkeepers and forwards with a breakdown of how their positioning changes scoring probabilities. Bayern Munich have a massive database of all their players' movements during every match, and for many of their training sessions, too. After every match, Bayern's Head of Match Analysis, Michael Niemeyer, provides a high-tech data presentation in the 'auditorium' for the rest of the coaching staff. Bayern's players can exchange comments and ideas on the data and the match on a Facebook-like forum. Every top club has its own numbers expert.

The challenge facing these experts is how to turn millions of numbers into an informative picture. The picture has to sum up the essence of a match, much like my tactical maps from Chapter 7, but this time put together from all those millions of measurements. Alina Bialkowski is a member of a team of researchers at Disney Research in the US who are rising to this challenge. Quite why Disney is trying to understand 'soccer' isn't immediately obvious, but they have certainly made a good start on the problem.

During one of their first studies, Alina and her colleagues focused on using the data to identify subtleties in formations. In their scientific article they don't reveal which top-tier league they analyse, but it is a league in a country in which 4–4–2 is the dominant formation. Figure 9.2 shows four examples of team formations adopted. Each of the tiny symbols corresponds to the average positions of different players during each half of each match played during an entire season.[5] For example, for team A in the top left figure, the left-back is a cluster of triangles and the two forwards are the smudge of circles and diamonds at the front.

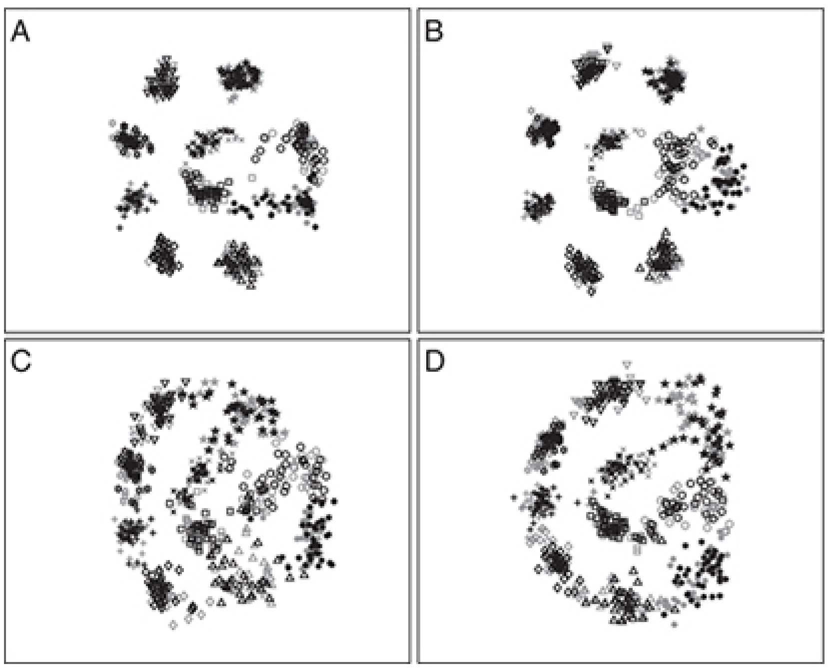



**Figure 9.2** Team formations studied over a season. Symbols indicate the average position of players during one half of every match during a season. Figure adapted from the original by Alina Bialkowski and co-workers.

Different teams play 4–4–2 in different ways. Some of these variations are well-established tactics. For example, team A has a striking pair playing parallel to each other, while team B has one deep-lying forward and one up front. But other differences between these formations are subtler. Team C is more flexible in attack than teams A and B, with a forward who switches between lying deep and playing in a pair. Team C also has a variety of different positionings in midfield, and is generally more spread out. Team D plays both 3–4–3 and 4–4–2 and moves around its attacking formations.

Identifying formations is just a first step. In a follow-up article, the Disney researchers looked at how likely different attacking situations were to result in a goal.[6] Every tenth of a second, different features of positioning were calculated, including how far defenders moved away from their usual positions, the speed of player movements, and how many defenders there were between the attacker with the ball and the goal. From these features the researchers were able to show that counter-attacks often provided the best opportunities for goalscoring. An attack by the opposition is the ideal time to start planning your own team's attack. The model is still work in progress, but ultimately the aim is to provide the probability of each different match situation resulting in a goal. From this, coaches can design their own strategies and identify areas where the opposition is particularly dangerous.

## Line-up

The next challenge in football analysis is to move from static descriptions of formations and positioning to dynamic analysis of player interactions. Let's return to the FC Nürnberg eight-a-side practice game. Figure 9.3 shows the players' positions and directions about one second after those shown in Figure 9.1, after the left-winger has passed the ball over to the right. Now nearly all the players are running in the same direction, although at different speeds. The players near the centre are moving the quickest, since they are furthest out of position.

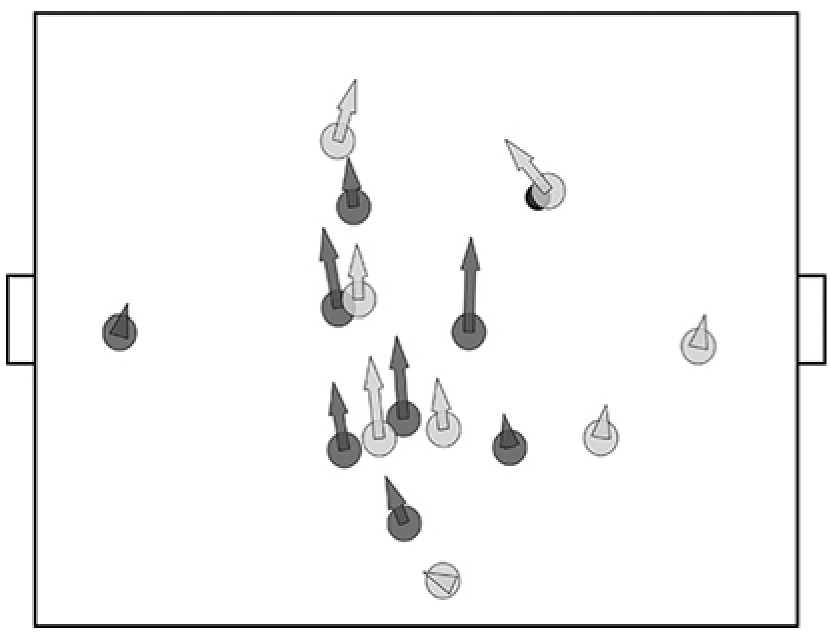


**Figure 9.3** Positions and directions of players during the FC Nürnberg reserves training match. Snapshot taken just over one second after Figure 9.1.

We can measure how co-ordinated the teams are by adding up the directions of all the players. If I take each of the direction arrows and stack them up end to end, then I get the picture for the two teams shown in Figure 9.4. Stacking up the players in this way shows the degree to which the team members are moving in the same direction. The players in both teams are aligned with one another. But the defending, darker team is slightly more co-ordinated than the attacking, lighter team. The whole defending team moves together in order to prevent any unnecessary spaces opening up. For the team with the ball, the situation is different. The attacking midfielders follow the direction of play, but the left-back and the left-winger are moving differently to create space to allow the direction of play to be changed again.

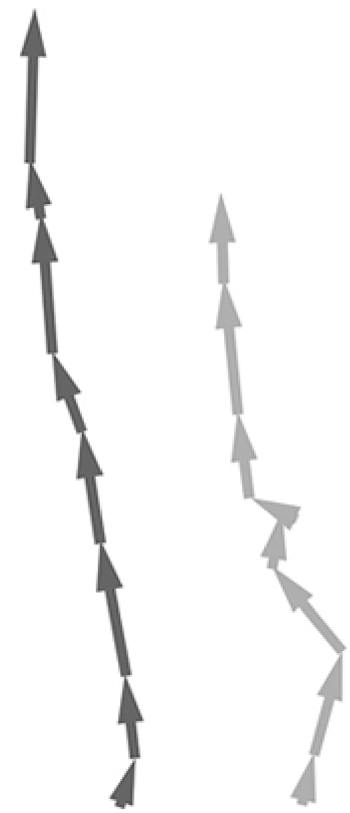


**Figure 9.4** Measuring the alignment of the players in both teams by adding up their direction arrows.

Co-ordinated movement is key to successful defence. This starts with the back four stepping up in unison, but is important for the whole team. By spacing themselves evenly across the pitch, the defending team creates a net that prevents further progress. Part of this co-ordination is achieved by following the ball, but it also relies on the players tracking one another so that the net remains tight.

At first sight, it may appear that keeping a tight defensive net requires players to track all their team-mates. However, back in 1995, Tamás Vicsek and his colleagues developed a model to show that group alignment can be achieved without a ball to follow, and even with very limited knowledge of the positions and directions of the rest of the team.[7] Imagine that you are running about on a 100m by 100m field (a bit less than twice the area of a football pitch) with 43 other people (exactly twice the number of people as in a standard game of football). You each have a speed of 12km/h – a fairly brisk jog. This field has the strange property that if you run off one side of it, then you come back on the other. So if you run off the top you reappear at the bottom, and if you run too far to the left then you re-emerge on the right. Of course, no such pitch exists in reality, but here you are imagining yourself inside a mathematical model. So slightly strange assumptions are OK for now.

According to Tamás's model, you should look at your nearest neighbours, let's say those within 10m of your current position, and start to move in the same direction as they are heading. Let's imagine that all other 43 players stuck in this field do the same, and see what happens. Figure 9.5 shows a simulation of the positions and direction of movement of all 44 of you, after 1 second, 20 seconds and 2 minutes of running around.

At the start, on the left in Figure 9.5, everyone is going in a random direction. The grey circle in the middle gives your location and shows your neighbours within a 10m radius. You start heading in the direction of your nearest neighbour. After 20 seconds, in the middle of the figure, small groups have formed. You are in one of the groups heading upwards to the left, but other groups are moving in different directions. After two minutes, as shown on the right of the figure, you have already travelled a couple of laps of the field. As you run around, you are still only following a few neighbours, but now the whole field is moving in roughly the same direction as you are.

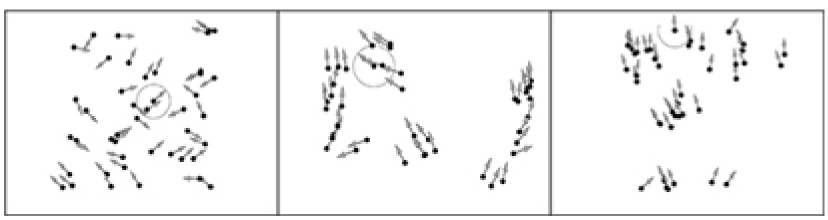


**Figure 9.5** Position and direction of 44 simulated people running around a field with teleporting edges. Simulation results after one second (left), 20 seconds (middle) and two minutes (right). The grey circles indicates your position and which other people you follow.

It takes less than two minutes for all 44 people to become aligned, and even after 20 seconds most of them are moving in the same direction. You and the others co-ordinate without a leader, without speaking or communicating, and without planning. All you need to do is run around in roughly the same direction as those nearby, and the group co-ordinates.

You can do this experiment yourself. Just find some people and an open area, and tell them to run roughly in the same direction to those nearby. Since you probably don't have a top-to-bottom teleporter, you will have to tell them to stay on the field. But the result will be similar: very soon you'll form a circular motion, either clockwise or anticlockwise, following one another around the open space.[8]

In Tamás's model, the teleportation from one side of the field to the other means there are no restrictions on the direction of motion. This recreates the feeling of being in the middle of a gigantic flock of starlings or a swarm of locusts. For example, a murmuration of starlings at dusk can consist of several thousand birds, but each of them responds only to a few nearby neighbours. Similarly, a flying swarm of locusts can cover tens or even hundreds of square kilometres, yet individual locusts respond primarily to those just a few centimetres away. Tamás's model helps explain how these flocks can form without an external signal telling the locusts which way to go. The locusts don't need to follow the wind or the sun – their local interactions alone are enough to get them all going in the same direction.

Tamás's model has been extensively tested on locusts, fish, birds and other animals.[9] The details of how species interact are different, but the principles are the same. Local interactions between nearby individuals allow the group as a whole to co-ordinate and align. This is good news for football players. If small-brained insects can move together in massive groups over vast distances over several months, then it shouldn't be difficult to get 11 players to move in unison up and down a football pitch for 90 minutes. Tamás's model implies that players don't need to track the positions and directions of all their team-mates and those of the opposition in order to co-ordinate. They just need to keep track of those nearby. Global co-ordination will follow automatically.

In a study of a Portuguese Primeira Liga team, Hugo Folgado and his colleagues found that synchronisation varies for different positions.[10] Defenders and central-midfield players show the greatest co-ordination. They move together, either marking the opposition when the ball is on the opposite side of the pitch, or pressing when the ball is on their side. The forwards have the lowest levels of co-ordination, trying to confuse the opposition players with unpredictable runs. This is similar to the pattern we saw in the FC Nürnberg eight-a-side practice game, with the defending team more aligned than the attacking team.

Synchronisation could provide a good overall measure of how hard a team is working as a unit. In the pre-season matches played by the Portuguese team, the players were more synchronised when playing against other Primeira Liga opposition and less synchronised when playing teams at a lower level. Similar measurements of an English Premier League team showed they were more synchronised when playing with a less congested schedule.[11] When there were only three days between matches, the players ran just as fast and as far as when they had six or more days between games, but there was less overall team synchronisation. It is plausible that too much football affects concentration more than physical condition. And when it comes to performing at the very top level, how a player interacts with others is just as important as physical fitness. Teamwork is about keeping track of those close to you.

## The Subtle Leader

Imagine that you are walking across a park with a colleague, on your way to lunch. You are deeply engaged in office gossip. As you walk back to work, you realise that your colleague is veering off to the left, towards the duck pond. You usually walk to the right, past the church tower. Neither of you wants to break off your conversation to discuss which way you should cross the park. Talking about your direction seems a bit trivial when you are about to get to the bottom of departmental politics. But as you walk forwards, you sense two opposing forces pulling you. One force pulls you towards the route you usually take, and the other keeps you together with your colleague.

When flying together, pigeons face the same type of social navigational problems as we do when walking together, but on a much larger scale. The first time a homing pigeon is released from a new place, it uses a combination of smell, magnetic cues and a compass based on the sun's position to choose its route. If it is released many times from the same place then it starts to memorise the route, identifying specific landmarks along the way. The pigeons often use the church towers and the railway line as familiar navigational aids. Pigeons are idiosyncratic. After being released five or six times they have established their own route home and tend to stick to it on future releases.

This isn't so different from how we navigate. When walking across the park with your colleague, you both have a different established route, the one you usually follow when you're alone. But once you are together, you prefer to stay together. Exactly the same phenomenon occurs for homing pigeons. When two pigeons are released together they are forced to choose between taking their own way home and staying with their partner. Pigeons can't discuss which way they are going to go, and have to rely on the movement of their neighbours to decide. Social forces pull them together.

I worked on modelling these social forces with Dora Biro, my colleague working in Oxford from Chapter 6, who has also studied pecking orders. Let's call our modelled pigeon Liverpool, which is the nickname given to one of Dora's pigeons after he got lost and made a long flight up north from Oxford. When in a pair, Liverpool has two forces acting on him, one towards the neighbouring pigeon and the other towards a landmark such as a church tower. When the neighbour is nearby and in front of or next to Liverpool, he tends to adopt the same direction as that neighbour. When the neighbour is farther away, or directly behind, the force of the neighbour weakens and Liverpool is attracted instead towards the church tower.

By building these forces into a model, Dora and I could predict what would happen when two pigeons follow similar forces, but in opposite directions. Let's assume that Liverpool's flying partner is attracted to a duck pond, and Liverpool to the church tower. In the model, when the tower and the pond are close together, the pair reach a compromise, flying down the middle. This is like you and your colleague walking straight across the park, halfway between pond and tower. But if the two landmarks are farther apart, one of the pigeons is forced to move farther away from its own landmark in order to stay with the other pigeon. As it does so, the attraction towards the landmark weakens and the attraction to the other pigeon strengthens. One of the pigeons becomes the leader and the other the follower.

This was exactly what Dora found in her experiments.[12] When there were small distances between the pigeons' landmarks they would compromise, but when the distances were large one pigeon would lead and the other would follow. It turned out that leadership had very little to do with navigational skill. Liverpool was not the most accurate of birds. After all, he got his name after he missed his target by 250km. Nevertheless, Liverpool is a leader. Whenever another bird was paired with Liverpool, it would fly closer to Liverpool's preferred route.

Today, Liverpool is almost 20 years old, has had a string of younger girlfriends and has fathered many baby pigeons. But not all leaders are the dominant birds in the loft – leadership is subtler than that. It turns out that the birds that become leaders are the ones that fly faster when alone.[13] When in a pair they tend to move just slightly ahead of their partner. As the faster bird turns, the slower bird follows. These faster birds may or may not fly in the right direction, but they assert their authority by moving slightly ahead.

You and your colleague also give off subtle signals as you walk across the park. We make small navigational decisions all day long without discussing them. Otherwise, life would be a constant round of 'Shall we go this way?' or 'Should I pass you on the left or the right?' Through these subtleties, some of us become leaders and others followers.

In football, everything happens so fast that in-depth discussions are impossible. Players who quickly understand the subtleties of their team-mates' movements are those who are best able to read the game. Some players instinctively exert their authority, and others follow. Mate Nagy, a statistical physicist working both with Dora and with Tamás Vicsek, created a method for detecting these leadership subtleties. He originally developed his method for pigeons, but then realised that it would work equally well on a football pitch.

Mate's idea was to look for small lags in players' changes of direction. Working on data from a leading European club, he first calculated the alignment measure we looked at on page 167, pairwise between all players. He then moved backwards and forwards in time to find the point where alignment between players was maximised.

To give you an idea of how Mate's method works, Figure 9.6 is a stylised example of two players turning. The top player lags behind the bottom player, adopting the same direction as the bottom player 0.3 seconds later. This is a stylised example, and during a real match it isn't usually as clear who turns first, but Mate developed a method for detecting subtle leadership lags. He could use this to pick out both leader pigeons and leader players.[14]

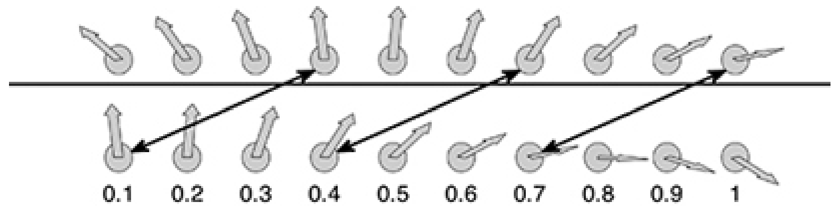



**Figure 9.6** Lags between direction changes in two players in seconds. The grey circles and arrows show the direction of two players at 10 time points within a single second. The black arrows mark the lags between times at which the players are facing in the same direction.The top player turns 0.3 seconds after the bottom player.

Working together with sports scientist Rui Marcelino, Mate used his method to identify a network of leaders and followers in a top-flight match. I can't reveal the club or the game, because clubs are sensitive about tactical information leaking out. But the team in question went 1–0 down in the first half at home against a big rival. It was time for someone to take a lead and, according to Mate and Rui's analysis, there was one very clear leader. This particular midfielder was followed by both the defenders and by one of the star forwards. He led moves in both halves, whenever his team had the ball.

While this midfielder led the attack, he didn't dominate possession or make the most passes. It wasn't simply a question of him using the ball to lead the play, but of the other players responding to his movement. What marks this player out is that he is the team captain. His team was behind and he took charge, and in the second half his leadership paid off. The forward who had been following his captain most closely equalised. A few minutes later, another player was brought down in the box and the team took the lead with the penalty. Final result, 2–1. The captain wasn't the player celebrated in the newspapers after the match, but, in terms of movement, the victory was very much a result of his subtle leadership.

## Pressing Power

As I mentioned earlier, big clubs are very cautious about sharing data describing the detailed movement of their players. This makes progress difficult for researchers like me, and analysis of collective team movement remains in its infancy. So far I have described a set of tools for looking at positioning and alignment, but there is so much more that could be done. Over the past ten years, several research groups, including Iain Couzin's group in Konstanz, Charlotte Hemelrijk in Groningen, Jens Krause in Berlin, Audrey Dussutour and Guy Théraulaz in Toulouse, Irene Giardina and Andrea Cavagna in Rome, as well as Dora's group in Oxford, Tamás's in Budapest and my own group in Uppsala, have been decoding the rules of motion of animals. This is a Europa League of research talent, all focused on animal movement. The results have been remarkable, and we now have a good grasp of how animal 'teams' move together. But these results have been possible only because we have collected large quantities of data, and shared this data with one another. It should be possible to revolutionise football in the same way as we have revolutionised animal movement, but to do this we need data.

Compared with the vast array of international researchers working on animal motion, the detailed analysis of player movement remains something of a one-man show. Paul Power is lead data scientist at Prozone, and has responsibility for game intelligence. Paul has a background as a coach, working at Sunderland, where he also completed a Masters in sports science. During his studies, he read Ricardo Duarte's 'Sports teams as superorganisms' article and was inspired. He wanted to use the superorganism ideas to study football, and landed himself a job at Prozone.

Working hards-on as a coach has meant that Paul wants mathematical analysis to lead to solid practical results. While I have emphasised the idea of creating a tactical map for communicating information about a match or a player, Paul aims to take analysis all the way to the training field. He wants to use match data to design training exercises. His approach goes back to the philosophy of Dutch manager Rinus Michels, who saw the manager's primary role as creating drills that encourage a particular style of play. The manager shouldn't have to tell the players what to do during a game if he has made them practise the most effective actions before the match starts. Paul's aim is to use data on player movements to find out what works on the pitch, and then use this information to decide what happens on the training field.

One of the most important tactical aspects of modern football is when and how much you should press your opponents. Pressing is when the team without the ball harass the team in possession as much as possible. At least one player should hound the opposition player with the ball, and the other players should look to block potential passes. It is a truly collective action, and one in which the whole is much more than the sum of its parts. As Barcelona coach Luis Enrique says, 'If one player does not execute the pressing – including your goalkeeper – you've got a big problem.'[15]

There are different forms of pressing. Enrique, and his predecessor at Barcelona, Pep Guardiola, both advocate pressing in the opponent's half to recover the ball as soon as possible. Jürgen Klopp's Borussia Dortmund also specialised in this form of 'counter-pressing'. The remarkable Bayern Munich side of 2012/13, the one that destroyed Barcelona in the Champions League semi-finals, pressed pretty much everywhere. This is an effective strategy if the team is extremely fit, but it can tire the pressing team as much as it does the opponents. During 2014/15, José Mourinho's Chelsea took a different approach, known as 'deep pressing', where they allowed other teams to come at them. As Liverpool and Arsenal have found out in recent seasons, they defend intensively in their own third. Each of these pressing styles has its pros and cons. The question is when to adopt a particular style and how to carry it out.

The key to pressing is narrowing down options. The diagram on the left in Figure 9.7 illustrates a situation in which the defender has nearly eliminated a passing option for the player with the ball. The player with the ball is looking to make a pass to her team-mate, below. The opposition player is attempting to block the pass. In this figure the player with the ball can pass in any direction, and the defender can intercept any ball coming within 1m of her. The other attacker can receive the ball if it comes within 1m of her, so it would require an extremely accurate pass to get the ball through. This situation is very like my piggy-in-the-middle game in Chapter 3, in which correct positioning by the defender always prevents passing. In the diagram on the right in Figure 9.7, the situation has changed: now there is a third attacker, who is not being pressed by the second defender, and a pass can be made.

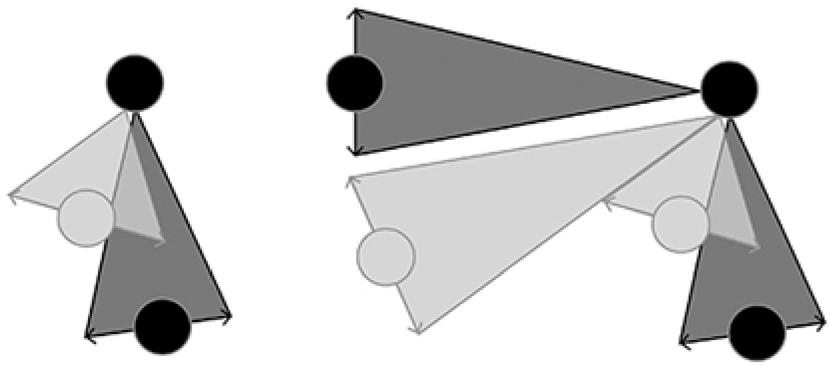


**Figure 9.7** Situations where a passing opportunity between two players is blocked (left) and where an additional player opens up a new passing opportunity (right). The black circle at the tip of the triangles is the player with the ball. The dark grey triangles indicate the range of potential passes to the team-mate (also represented by a black circle). The light grey triangles indicate the range of passes that would be blocked by an opposition player (represented as a light grey circle). Adapted from an original drawing by Paul Power at Prozone.

What Paul and his colleagues did was apply this model of viable passing options to detailed tracking data from matches. For every tenth of a second of the match they used the 1m interception/reception criterion as a way of classifying a potential pass as viable or not. They then built up a network to show which players were available for a pass from the player with the ball. Figure 9.8 shows such a network for Lyon during the first 15 minutes of a Ligue 1 match against Marseille. At this stage in the game, Lyon players have quite a few options. The thick arrow pointing from Yoann Gourcuff to Alexandre shows that they are both usually available for a pass. Nabil Fekir, the other forward, is available less often, but overall there is an array of forward-passing options to move the ball up from the back.

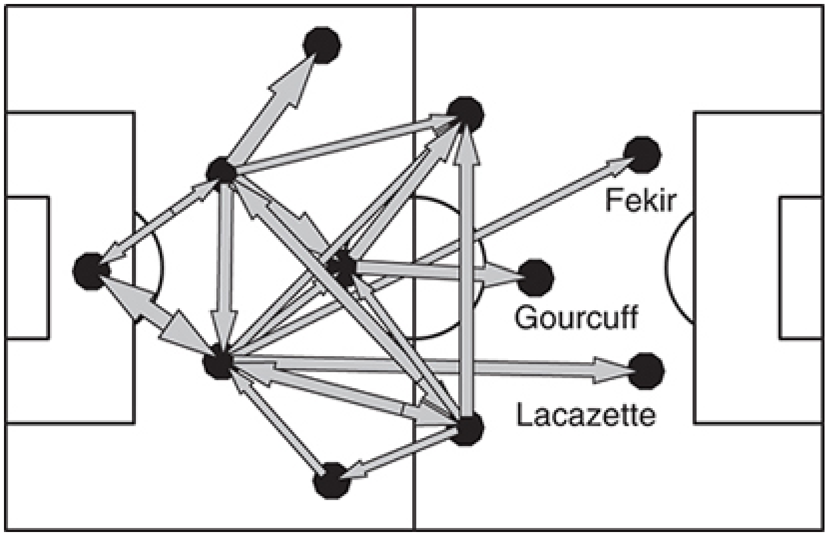



**Figure 9.8** Potential pass network for the first 15 minutes of Lyon's match against Marseille. The thickness of the arrow between two players indicates how often a player was available for a pass. Adapted from an original drawing by Paul Power at Prozone.

After this initial period, Marseille started to press effectively. In the second 15 minutes of the match the network had changed dramatically – see Figure 9.9. Now there are very few arrows in the network, and those that do exist point back towards the goal. Paul showed that this lack of options correlates with the degree of pressure exerted by Marseille. At the time, Marseille, under the management of Marcelo Bielsa, were one of the most effective pressing teams in Europe. Bielsa believes in high pressing levels everywhere on the pitch. In the 2014/15 season his team pressed 43% of on-the-ball actions by the opposition, compared with an average of about 25% by English Premier League teams.

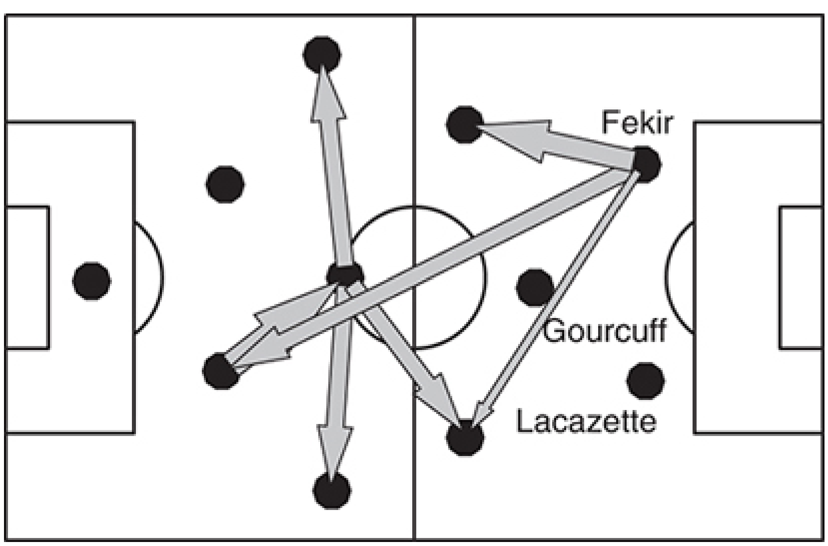



**Figure 9.9** Potential pass network for Lyon's match against Marseille during the second 15 minutes of the match.

But something went wrong for Marseille at the start of the second half. For a short period the pressure dropped. On 64 minutes, Lyon put together 10 passes to take the ball from their own box up to Gourcuff in front of Marseille's goal. Marseille were off-balance, and, presented with the chance, Gourcuff didn't miss. By plotting the passing options and pressure through time, Paul showed that each pass Lyon made resulted in a drop in pressure and an increase in passing options for their players.

Paul wanted to take his observations of single matches like the Lyon v Marseille game and find general principles for applying a press. So he and his colleagues took the entire Prozone data set – 260 million data points – for an English Premier League season and looked at a range of factors that are generally believed to lead to successful ball recovery. He studied pressing in two forms, the 'counter-press', which is applied directly after the ball has been lost during an attack in the opposition's half, and the 'deep press', which is applied when defending the final third. The principles for these two forms were very different. For counter-press, Paul found that a player first has to apply pressure within about 2.3 seconds of the opposition taking possession of the ball. Furthermore, a second player has to press the player with the ball within 5.5 seconds. Once this two-man press is applied, the opposition player who has just gained the ball starts to have doubts and is forced to turn back.

Pep Guardiola called this principle the six-second rule. Paul Power's data analysis backs it up and refines it. Guardiola reportedly timed his Barcelona players during training to see whether they could regain possession within six seconds. Paul's work suggests that coaches should set the stopwatch for two points in time. One player should be pressing before 2.5 seconds, and a second player should become involved by 5.5 seconds. After he completed his analysis, Paul summed it up as '260 million points of data, two coaching points. That's what it's all about.'[16] This simple idea of getting a two-man press within 5.5 seconds can be repeated over and over again in training exercises with forwards.

The conclusions about deep pressing were very different than those for counter-pressing. When the opposition has the ball and they are coming at a defence, then the single most important factor is reducing the speed with which the ball is moving towards the goal. Only one defender should approach the player with the ball – the others should block the remaining channels. If more than one defender moves to block the ball, then the viable-passing network opens up for the attacking team, and other routes to goal become available. Paul worked closely with one manager who, probably inspired by the Guardiola and Klopp styles, tried to get his defenders to play in flexible roles. It didn't work, and his team kept conceding unnecessary goals. Paul could show him why the traditional four-man defence, with well-defined zones, works. Deep defending is all about stability; counter-defending is all about dynamism.

## Putting It Together

I don't believe that players flock round their captain like pigeons, or that Barcelona play like a school of fish, or that the Bayern defence consists of a pride of lionesses. That would be daft. We can't train goldfish to pass a ball, and lions may well eat the opposition after tackling them. Nor do I believe that football players can learn new skills from watching a David Attenborough documentary. I don't think that you will become a better footballer from watching a flock of pigeons outside the local supermarket or staring at ants on your garden path. It isn't through birdwatching that we improve on the pitch.

The connection between biology and football lies instead in the way we do the maths. Imagery of flocks and swarms might inspire us, but it is mathematics and statistics that give us insight into performance. Models of one aspect of the world can be transferred and applied elsewhere. And now, as sports scientists become increasingly serious about understanding the organisation of football teams, we need to be especially clear on this point. The challenge is to take the techniques we have learned from collective animal behaviour and apply them to collective football.

The work by Paul, Mate, Alina and their colleagues is just a start. There are many more mathematical methods that can be transferred from animals to players. For example, recent studies of fish have taken visual networks a step further, using movement data to reconstruct how fish react to one another inside a school. These networks can then be used to identify which fish are key in initiating direction changes.[17] On the football pitch, similar methods could be used for evaluating passing options, identifying missed opportunities and finding which players initiate play in different situations.

Given the economic incentives, I was surprised to discover that the study of animal motion still has an edge on the study of footballing motion. But this will change as clubs start to release their high-resolution data on player movement. When statistics on corners, passes and shots first became widely available, academics, bloggers and amateurs alike quickly seized them and generated a whole range of new and valuable insights.[18] Soon after that, the clubs started to listen to them and adopt their methods. The same will eventually happen with the detailed analysis of player movement. We may well uncover new unthought-of formations where rapid passing can break down all forms of press, or we may end up proving that long balls played up to a lone striker is the only way to go. Whatever the findings, there is a bright future for collective soccer analytics, and it will be fun to see how it plays out.

PART III

# From the Crowd

CHAPTER TEN

# You'll Never Walk Alone

At the end of the 2014/15 season I went to Steven Gerrard's final home game for Liverpool. This was a special match, not so much because of the football, but because of the opportunity it gave the Anfield faithful to celebrate their captain and what he had given to the club. Gerrard was Liverpool's talisman for 15 years.

In a football stadium, the main way the fans get their message across to the players is through song. That day, 'You'll Never Walk Alone' rang out more loudly than usual, but it took second place behind the two most popular Gerrard chants: 'Stevie Gerrard Is Our Captain' and 'Impossible Forty Yards'. The crowd united in celebration of their home-grown star.

My favourite moment was when Gerrard, in an attempt to get Liverpool back in the game at 1–2 down, sent one of his trademark forty-yarders flying into the Kop stand, nowhere near the goal. The Kop responded with a 'What the fucking . . . What the fucking . . . What the fucking hell was that?' It was a spontaneous reaction, singing a song usually reserved for wayward shots by the opposition, to show a humorous appreciation of the frustrations their captain was feeling. The crowd's joke was subtle, but it was made almost in synchrony by thousands of fans. Gerrard, still annoyed with himself, slowly lifted his hands and ironically applauded the Kop.

## 1, 2, 4, 8, 16, . . .

Football songs work best when the all the supporters join in. In a few cases, such as 'You'll Never Walk Alone' and 'El Cant del Barça', the song is initiated at the start of a match by the tune being played over the PA. But in the vast majority of cases, football songs start spontaneously. This was certainly the case when the Kop responded to Gerrard's overly optimistic shot. One fan starts, his friend next to him joins in – that's 1 + 1 = 2. If each of them inspires one more fan, then we have 2 + 2 = 4, then 8, and then 16. The growth is slow at first, but at each point in time we double the number singing. In the 13th doubling, 4,096 inspire 4,096 and we have 8,192 people on song. This is exactly what it means to grow exponentially: we keep on multiplying the current population by itself, and quite soon that population is extremely large.

The singing fans multiply just like bacteria on a juicy piece of steak. Under suitable conditions, the bacterium *E. coli* can double its population every 20 minutes. If you leave your meat outside the fridge overnight, then one bacterium can become several million.[1] In her classic textbook, mathematical biologist Leah Edelstein-Keshet takes this bacterial growth to its (un-)natural conclusion.[2] Doubling every 20 minutes, it would take a $10^{-12}$ gram bacterium a little less than two days to grow to weigh as much as planet Earth. It's hardly surprising, then, that it takes just a few seconds for a song to fill a football stadium.

Although Leah's bacteria model is fun, it obviously isn't correct. In order for it to work, the whole globe would have to be one gigantic piece of meat. And it isn't. A better description of growth should account for limitations. When there were 20,000 West Ham fans singing 'I'm Forever Blowing Bubbles' in a 35,016-capacity Upton Park, there was no room for a further 20,000 to join in. West Ham's move to the Olympic Stadium increases their match-day capacity to around 60,000, but it won't take much longer to fill the ground with song. Doubling 40,000 gives 80,000, and chanting capacity is reached.

The answer to this modelling problem is the S-shaped curve, the most useful and universal mathematical model of growth. We need to make just two assumptions in order to get an S-shape. The first is the same as for exponential growth: each singing fan is copied by one other fan, who isn't already singing. We then note that the number of fans is limited, so that each fan can begin singing only once. Putting these two assumptions together results in the S-shaped growth curve shown in Figure 10.1.[3] Growth is slow at first, but exponential. The singing population doubles on every 'round'. Once half the crowd is singing, the growth slows again, and the number finally stabilises with everyone singing. It still takes just 16 rounds to get 10,000 people chanting.

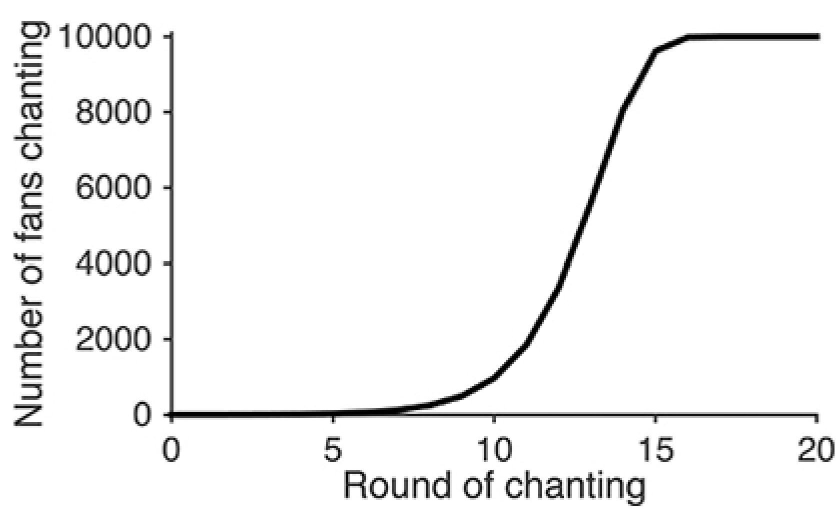



**Figure 10.1** The increase of chanting over time through social contagion.

## Clapping Contagions

The spread of ideas, the spread of disease and the spread of chanting – these are all forms of contagion. In chanting, the contagion is social: it is transmitted by people listening to and copying one another. Some infectious diseases, such as HIV, Ebola, influenza and SARS, are transmitted when people come into contact or close proximity to each other. Other diseases, such as cholera and some variants of hepatitis, spread via contaminated water and food. However they are spread, the increase in infections usually takes the form of an S-shaped curve. At first there are a small number of cases, but then exponential growth kicks in and the disease takes off. At this point, all we can hope for is damage control, trying to make sure that the number of infected individuals stabilises sooner rather than later.

The analogy between disease and chanting isn't especially easy to test. It's a bit difficult trying to do a controlled experiment at a football match. I could stand there humming the tune to 'Seven Nation Army' and see whether anyone joins in, but I've a feeling that no one would pay much attention to a lonely mathematician. A much better setting to measure applause is in the controlled environment of a classroom, and this is exactly what my colleagues Jens Krause and Jolyon Faria did.[4] They asked groups of first-year undergraduates to listen to a seminar given by a final-year student. Jens and Jolyon told the younger students to think about the presentation and take notes, but to remember to show their appreciation afterwards. The students knew they were being filmed, but they didn't know that the exact purpose of the study was to look at their clapping behaviour. If they had been told in advance that we were studying how they clap, they would probably have become self-conscious, and this would have affected how they behaved.

By noting every point in time at which each of the students' hands came together, we could see the spread of the clapping. These handclaps for one group of students are plotted in Figure 10.2. Each row is a person and each dot is a handclap. The students are ordered by when they started clapping. The second from the bottom row is the student who stopped clapping first, who also happened to be the one who started second. There is a variation between the students, some starting early, others later on. Some clapped vigorously and others more slowly. By plotting the data in this way we can visualise the whole sequence of applause of the whole group in one picture.

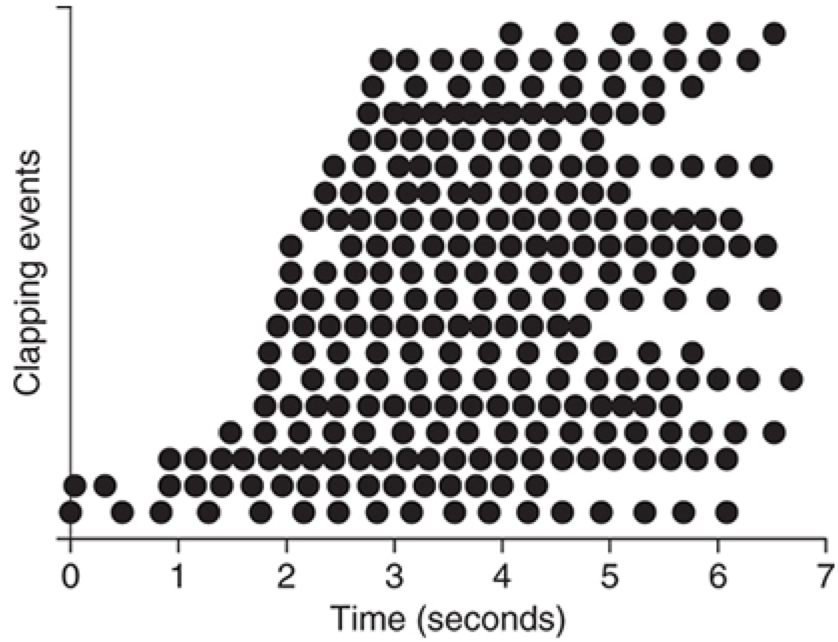



**Figure 10.2** Clapping events over time. Each dot represents a single student putting his or her hands together during a round of applause. Adapted from a figure created by Richard Mann.

Plotting the data like this also allows us to look at the factors that may cause a person to start clapping. Together with one of the researchers in my group, Richard Mann, we tested a whole range of possible reasons why someone would start clapping. Was it the time since the first clap anyone had made? Was it whether the person next to them was clapping or not? Was it related to how rapidly they clapped? Or was it just completely random? We found that the best model was based on the proportion of people already clapping: the more clappers there were, the faster others joined in. Clapping was a social contagion.

The growth curves for the handclaps over all the experimental trials reveal this social contagion. The black curve in Figure 10.3 is the distinctive S-shaped growth curve of the number of people who had started clapping. One or two individuals start things off slowly. The fastest growth is when about half the audience have started clapping, and the growth levels off when nearly everyone is applauding. The dashed curve in Figure 10.3 indicates the number of people who have stopped clapping, and we see the same S-shaped curve. We found that the moment at which an individual stopped their applause was 10 times more likely to be determined by the number of others who had stopped clapping than by the number of claps she or he had done. In other words, it is not how long you have been clapping, but whether those around you have finished, that causes you to stop. Just as football chants die out, leaving one or two diehards shouting, so too did the students' applause.

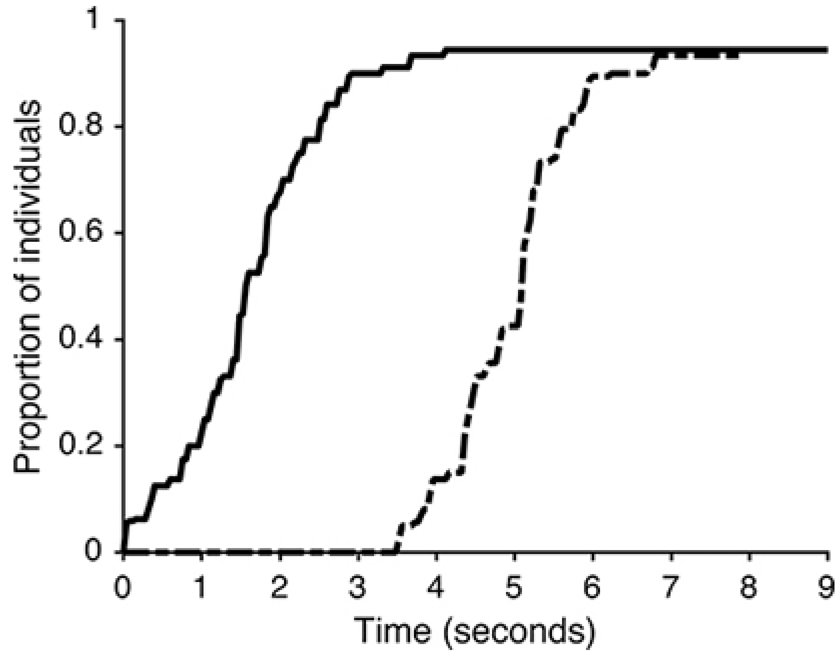



**Figure 10.3** Social contagion and social recovery in clapping. The solid line is the average number of students who have started clapping. The dashed line is the average number who have stopped clapping. Adapted from a figure created by Richard Mann.

The way clapping stops is a form of social recovery, a tendency to recover from or stop performing an activity because others have already stopped. Social recovery works in the opposite direction to social contagion, and is very different from the type of recovery from most diseases. Diseases run their course, and we get better. Apart from a mild psychological effect, we don't generally recover faster from the flu if our friends are feeling better. But clapping students do 'recover' faster from clapping if their friends have 'recovered'. The number of people clapping increases rapidly as people start to clap, and then decreases rapidly as the applause starts to die out. The reason it drops off so rapidly is all down to social recovery.

Once we had determined the factors involved in clapping, we could use our model to simulate applause. We produced our own simulated series of clapping events and came to a surprising conclusion. We ran 10,000 simulations, and found that on most occasions the simulated students would clap about 10 times each, but in some simulations there could be more than 20 or 25 claps per person. Some groups of simulated students were really showing their appreciation! What was going on? We always expect a variation in simulations, but these extended applauses were twice as long as average.

Long bouts of clapping can be explained by a failure to recover socially. If very few individuals have stopped clapping, then the chance of anyone else stopping too is small. Until a few people initiate the end of clapping, the group just continues. The clapping goes on, not because what they are applauding was particularly good, but because no one wants to be the first to stop. Now when I get a long round of applause after I've given a seminar, I don't let it go to my head. It doesn't mean that the audience is enthusiastic, just that they are not very coordinated.

Applause after a seminar is rather different from applause at a football match. Much as I would love to believe that students appreciate my solution to a differential equation in the same way that Liverpool fans appreciated Gerrard's free kicks, I know it can't quite be true. The initiation of clapping in the classroom is more a question of not wanting to stick out, rather than a wild desire to take part. Despite these differences, football fans and students both have an underlying desire to fit in and be part of a social group. At matches, bouts of singing may well continue, not because the fans are feeling particularly passionate, but because no one wants to let the others down by dropping out. At other times the singing stops abruptly, leaving one or two fans looking rather silly.

## Luis Suárez to Arsenal

We don't have to visit a football stadium or a classroom to become part of a social contagion. Psychologists have found it in smoking, binge eating and alcohol consumption. The more your friends smoke, eat and drink, the more you smoke, eat and drink. Sociologists have found S-shaped curves in medical innovations, donations to disaster funds, the spread of new music genres and the adoption of new crop types. It has even been suggested that the chance of finding a job or committing suicide is socially contagious. Academics from all fields have repeatedly reinvented the S-shape curve to describe these experimental findings. Everything from fashion and fads to culture and civilisation has been modelled as an 'S'.

The spread of news in the media is dominated by social contagion. We discovered this when we published our results on clapping. On the Monday morning after our paper came out, Richard Mann was called by a journalist who intended to write a short piece about our study. Richard answered questions and got on with his work, and everything went quiet for three days. Maybe the press weren't so interested in clapping? Then suddenly his phone started ringing like crazy. The BBC, National Public Radio in the US, *Scientific American*, the *Times*, the *Guardian* and *Slate* magazine, as well as French and German national media, all wanted to know about our clapping work. Social media filled up with links and discussion about the study.

Our own science had generated the social contagion we had written about, and – just as we had predicted – the media interest faded away as quickly as it started. On the Friday night that week, 130 hours after the first media contact, our research group had its summer party. When the BBC World Service rang that evening, none of us were available for comment. I rang them back on Saturday morning, but they weren't interested any more. Clapping was old news.

Contagion is even more important in online social media than for traditional news outlets. The internet is set up in a way that exaggerates contagion. Google's PageRank search algorithm works by counting the number of links to a page to give an estimate of how important the website is. So if other people have linked to a website, you are more likely to find it and link to it yourself. The most popular sites are the ones that are the most contagious. Facebook and Twitter have the same principle, with 'likes' and 'favourites' moving popular posts up your news feed. We lock into certain trends, not because they are the most interesting, but because they happened to take off first.

The clearest examples can be seen in the spread of rumours. Figure 10.4 shows the volume of Google searches for 'Luis Suárez Arsenal' during the summer of 2013.[5] Suárez was a Liverpool player at the end of the 2012/13 season, but there was speculation that he might make a move south. The search data has the same type of curve as we saw for participation in clapping. Interest in a pairing between Suárez and Arsenal starts in late June, and then grows rapidly in the first half of July. It reaches a peak in mid-July and tails off during August, much in the way that it started. This pattern is consistent with both social contagion and social recovery.

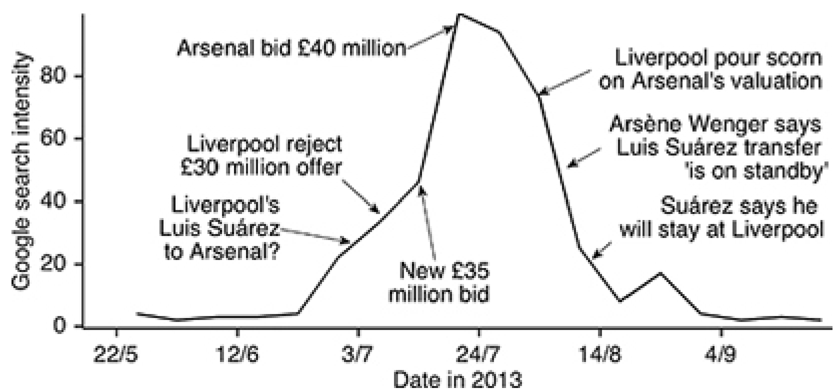



**Figure 10.4** Web searches as reported by Google Analytics and news events around potential move of Luis Suárez to Arsenal.

We should be careful here. Labelling transfer rumours as social contagion on the basis of Google searches alone is problematic, because there are also real external events involved. In some cases there are reasons to believe that a rumour is actually true, even when it is about a transfer. Completely separating reality from rumour is usually impossible, but there are several interesting observations to make about how people searched and how the newspapers reported the possible transfer. First of all, the searches started slightly before the link between Suárez and Arsenal was widely reported in the newspapers. When the *Guardian* published a Rumour Mill piece on 'Liverpool's Luis Suárez to Arsenal' on the 4th of July, fans were already searching for news on the subject. By the time the first bids were 'officially' placed, the number of searches had risen to half its eventual peak. The peak itself was reached when Arsenal lodged their maximum bid.

Like most social contagions, the interest in Suárez' potential transfer died off quickly. While Arsenal failed to place a bid higher than the £40 million plus £1 that supposedly activated a clause in Suárez's contract, the newspapers continued to run stories almost every day on a potential move. These stories continued until the 22nd of August, when Arsène Wenger admitted that there was now 'no chance' of signing the player. But despite media interest, the searching decreased. It may be that we were already being fed such vast quantities of Suárez/Arsenal gossip that we didn't need to search for it any more. But I think it is more likely that we had collectively lost interest in the whole story and stopped searching. Despite the newspapers continuing to write articles about Suárez through the rest of that summer, we were already in a 'recovery' phase of the rumour epidemic.

The type of social recovery we saw in audience applause makes perfect sense in the context of rumours. If I hear a rumour and pass it on to you, but you say that you've already heard it, then you are less likely to go and tell it to someone else. People don't want to lose face by spreading out-of-date gossip.

## Stand Up, Sit Down and Turn Around

Despite its popularity elsewhere in the world, Mexican waves have never really taken off at British club grounds. British football fans tend to think they are a bit naff. Standing up and waving because the person next to you has stood up has no part in the serious business of being a supporter. Appreciation of football is too tribal for waves; both sets of fans engaging in co-ordinated standing and sitting is an unnecessary distraction. True fans should sing about their own team, offer well-thought-out tactical advice to the players and manager, and put down the opposition in the clearest possible terms.

The mood of the crowd may be different, but the principles of the Mexican wave are very similar to those for chanting. The Hungarian physicist Illés Farkas was fascinated by the waves that had been seen in the crowds at the World Cup in Mexico in 1986. Together with Tamás Vicsek and Dirk Helbing, he created a model based on the same analogy between the spread of disease and social contagion that we found for applause.[6] Seated fans were classified as susceptible to catching the disease, standing fans as infected with the disease, and those who had recently sat down again as having recovered. Illés and his colleagues placed these simulated fans in a 64,000-seater simulated stadium with 80 rows and 800 columns, with most people sitting down but a small cluster of excited fans standing up in one particular part of the stadium.

They found that it took very little to set off a Mexican wave. When the infected fans stood up, the wave would first take the form of a blob, which would grow outwards to become a crescent and finally settle down into a straight line round the ground. The

Hungarian researchers found that the wave travels at about 22 seats per second, so it takes just over half a minute to travel all the way round the stadium. Waves move fast.

Mexican waves can be fun for the fans, but a matter of life and death for fish. Rapid propagation through the group is exactly what fish need if they are to survive attacks by a predator. By placing fish in a stadium-like ring and frightening a small number of them at the front of the group, my colleague Teddy Herbert-Read was able to watch how a fright-wave passes through the school. One example of just such a wave is shown in Figure 10.5. The four images show the sequence of turning. In the top one, the 'frightening rod' has not yet extended. In the second image, the rod has just popped out and fish at the front turn away. In the third image, just 0.7 seconds after the previous one, the turning wave has spread through two-thirds of the fish. And after 1.4 seconds, as the final image shows, all the fish have turned away and formed into a tightly packed group moving away from the rod.

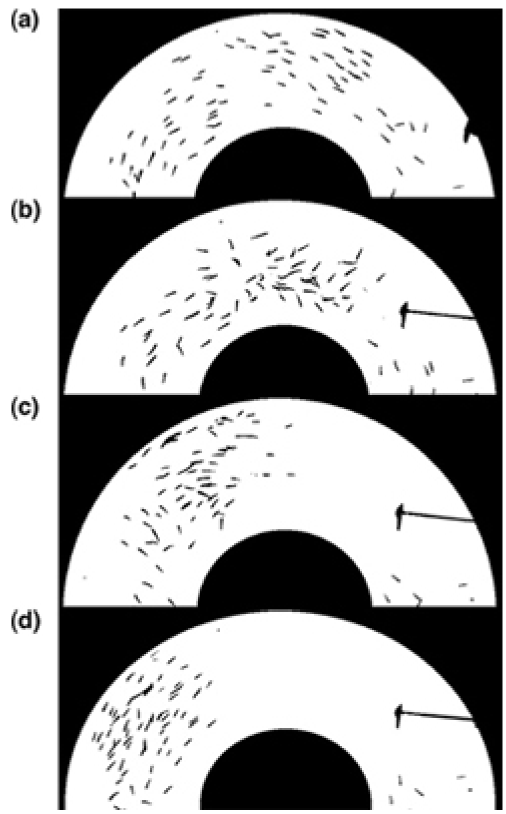



**Figure 10.5** Fish escape wave. Initially the fish are swimming clockwise round a circular tank (a). The frightening rod pops out and the front fish move away from it (b). After 0.7 seconds most of the fish have turned and are facing anti-clockwise (c). After 1.4 seconds nearly all of the fish are facing anti-clockwise and moving in a tightly packed school (d).

Using Teddy's data, and a variation of Tamás Vicsek's flocking model from the previous chapter, we created a model of how the wave propagated in fish.[7] This model shows that the fish typically form schools that allow for wave propagation. In model simulations, escape waves could pass through any size of group, from small lab groups of 100 fish up to tens of thousands of individuals. Good wave propagation is important for fish that live in big schools. If a predator attacks on one side of a big group, then within seconds fish on the other side will be moving away from it. Waves propagate information about attacks.

Unlike fish-school waves, human-crowd waves are not entirely local. Fish react only to their nearest neighbours, while when we take part in stadium waves we anticipate the wave's arrival. Illés and Tamás demonstrated this anticipation, both through their simulation and by doing an online survey of 'wavers'.[8] Most respondents to the survey said that they watched the wave go round the whole ground and anticipated its arrival. Teddy did his fish experiment in Sydney, and I visited him there during 2014. Given our interest in waves, we decided it was very important to experimentally confirm the non-local nature of Mexican waves for ourselves. So we went to the one-day international between England and Australia at the Sydney Cricket Ground.[9]

The rivalries are just as fierce, but the action isn't quite as intense in cricket as it is in football. And after a few hours (and a few beers), crowd waves are inevitable. As well as providing a 48,000-person research arena, the SCG contains a very useful experimental manipulation, called the Members' Stand. Spectators in this part of the ground refuse to participate in the wave. Roughly 47,000 people stand up, cheer and create a wave that moves rapidly round the ground until it reaches the Members' Stand – at which point the wave stops, only to restart on the other side of it. So this Mexican wave is non-local – in other words, it doesn't just pass between individuals sitting next to each other but can jump over the obstacle posed by the members. While the wave should have been passing through the Members' Stand, everyone else in the entire ground booed. The period of booing was equal to the time the wave would have taken to pass through the members. The fact that the wave continued with the same speed on the other side of the Members' Stand shows that the crowd were anticipating the wave's arrival long before it came to them. It takes more than a few cricket purists to stop the wave.

The failure of cricket connoisseurs to stop the wave should be a warning for those traditionalist British football fans who don't join in. As waves become more common at England international matches, it won't be long before we see them spreading through English club grounds. Very few people are immune to social contagion.

## The Mosh

By day, Jesse Silverberg works on the border between physics and biology, writing scientific papers on nanowires and plant roots. By night, he can be found jumping around at heavy-metal concerts. It was at one of these concerts that he realised there was a connection between his day and night lives. Just like plants and fibres, heavy metal fans are also soft biological matter. And they can also be modelled.

Deliberately bumping into other heavy-metal fans at a concert is known as moshing. The collective noun for moshers jumping about is a mosh pit, usually observed near the stage where the band is performing. A tangle of bodies moves around, people colliding, occasionally popping out of the pit, then making their way back in for more mosh. Moshers aren't the most complicated of creatures. People go to concerts to enjoy themselves, not to reflect about what they are doing and why. So in building a mosher model, Jesse and his colleagues at Cornell University made very simple assumptions about behaviour.[10] They assumed that all concert-goers fell into one of two categories: those who like to bounce around and those who don't. The active, bouncing moshers were subject to three different forces. The first force was a tendency to *follow* in the same direction as those around them, the second was a tendency to *mosh* around at random, and the third was the inevitable force caused by *crashing* into others. The passive bystanders were subject only to the last of these forces. When active moshers crashed into passive bystanders, they bounced off them. A mosh pit is generally a 'friendly' environment. Bystanders tolerate a bit of mosh, and if anyone falls over they are quickly picked up again.

With the rules established, Jesse and his colleagues set about creating a mosh simulator.[11] The moshers are initially positioned at random and move around according to the follow, mosh and crash rules. Quite quickly, one of three main mosh patterns was found to arise, as shown in Figure 10.6. The first is the classic mosh pit with random bouncing around in the middle. These pits form even if the simulated moshers are initially spread out through the crowd. Moshers don't need to decide on a place to make their pit, they just hop around and eventually a pit is formed with all the active moshers together. The second pattern is a circle pit, in which individuals move round and round in circles. These occur if moshers are a little less random and tend to move more in the direction of others. If moshers reduce their randomness still further, they no longer form a pit, but instead move in trains through the crowd.

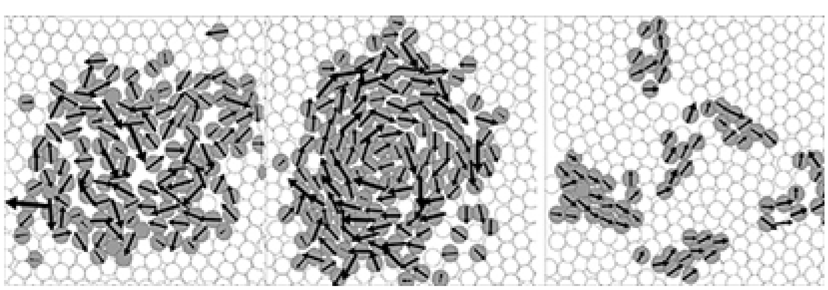


**Figure 10.6** Three patterns of movement from the mosh pit simulator: bouncing around in the classic pit pattern (left), circle pit (middle) and 'trains' (right). White circles are passive moshers and grey circles are active moshers. Arrows point in the direction of motion. Reprinted with permission from the American Physical Society.

Jesse and his colleagues found videos of moshing on YouTube that were sufficiently high-quality for them to identify both the classic pit and the circular pattern. They measured the rotations and the number of individuals involved, and they were consistent with the model. They didn't present any evidence of mosh trains in their article, but, thinking back to my own student days, these were relatively common. I can't say I was ever a metaller, but I do remember jumping around to music and suddenly finding myself in a group moving in a stream through the crowds. On the basis of my student memories, combined with Jesse's more careful study, I think we can conclude that all three patterns have been widely observed.

## People Don't Panic

So what is the point of making a model of a mosh pit? There are two reasons, and the first is simply to explain the moshing patterns themselves. The three assumptions in the model, of following, moshing and crashing, require no intelligence to be displayed by the heavy-metal fans. No one has to shout, 'Let's make a circle pit!' or 'Let's move through the crowd!' They don't even need to shout 'Come and mosh over here!' Everyone just bounces around, and the pattern is formed. The model tells us that no organisation at all is needed to make patterns. They really do emerge just from mindless moshing.[12]

The second reason for modelling mosh pits is that it helps us predict what could happen if we pack individuals together in other situations, not least on Saturday afternoons at the local football ground. Building a new stadium involves getting the planning right from the start, and computer modelling is used for all aspects of modern design. Simulations of aerodynamics ensure that there is less wind on the pitch, while fresh air can still circulate through the stands. Acoustics are modelled so you can hear the loudspeakers and the taunting of the visiting fans. Similarly, by using crowd simulations, designers can make sure that you get to your seat faster, that you have to queue for no more than five minutes for your half-time pie, and that you don't have to leave during stoppage-time in order to beat the traffic.

It is in simulating crowds that moshing-type models come in useful, especially when we think about evacuating a stadium. Since the disasters in the 1980s at Bradford and Hillsborough, all-seater stands have greatly improved the situation. Despite these improvements, safety remains an important issue. If 90,000 fans need to be quickly evacuated from Wembley Stadium, then there has to be a fast and steady flow of people out of the ground. If engineers can build models of how people move in escape situations, then they can design easy-to-evacuate stadiums.

Building good models means understanding how people will behave in an evacuation. It is a common misconception that crowd disasters are the result of panic or aggression, with individuals trampling on others in their rush to get out.[13] This is simply not the case. Even in the most terrible of crowd disasters, such as Hillsborough, individuals show great courage in helping one another. Death and injury are caused by crushes, where people become tightly packed and can no longer move freely.

So rather than focus on panic, we should concentrate on the difficulty of predicting the outcome of our actions in crowded situations. In the moshing study, circle pits and mosh trains emerged not because the moshers had planned them, but as an unexpected consequence of jumping around in a particular way. The way a crowd behaves is not easily predictable from the actions of the individuals involved.

In an escape situation the patterns that emerge are different from those in moshing, and experiments are needed to understand the details of pedestrian flow. In Chapter 3, we saw how Mehdi Moussaïd studied pairs of pedestrians passing each other. Mehdi and his colleagues then scaled up the experiments to crowding 60 people in a laboratory ring. He asked half of them to move clockwise and half anti-clockwise to see how they got past each other. Through these experiments, Mehdi established the 'rules of thumb' people use to navigate around one another when they are packed together. People tend to keep moving in the general direction of their target while making sure that they are always moving into free space. We don't push directly towards where we want to go, but we try to predict where others will go and move into the spaces they leave behind.[14]

For obvious ethical reasons, Mehdi can't do experiments in very crowded, dangerous situations. So he built a model of situations in which individuals follow a 'move into space' rule within a narrow street or corridor. For low and medium levels of crowding, Mehdi could compare the model with his data and checked that it captured the movement patterns. It did. He then turned up the crowd density in the simulation to a dangerously high level – and it was then that the waves appeared. When two individuals tried to move into the same place, one of them would stop and let the other go first. When this individual stopped it would trigger a wave of stopping that passed back through the group. The waves were very similar to those observed in schools of fish, but they were caused by stopping, not turning. One individual stopping near an exit would cause those behind them to stop, and a stop-wave would move back through the entire corridor. As the individuals stop, they produce a wave of pressure that passes back through the group, while those moving through the door produce a wave of pressure in a forward direction.

Stop/start waves like these were filmed in the hours before one of the worst crowd disasters in recent history, when at least 345 pilgrims died during a crush at the Hajj pilgrimage in Mecca in 2006.[15] It remains unclear whether these stop/start waves actually caused the subsequent crush or were an early warning that a disaster was imminent. But what Mehdi's model shows is that stop/start waves are less likely to be caused by people deliberately pushing, and more likely to be caused by some individuals stopping to let others into space.

The frightening aspect of stop/start waves is that once crowding reaches a high density, they are unavoidable. When I find myself inside a crowded area, on the way out of a football ground or at the train station after a match, I hear people shouting, ‘Stop pushing!’ or ‘Just wait!’ They get frustrated, because while on one side of the crowd people are passing through the turnstiles, on the other side people are stuck or moving backwards. I try to remind myself that other people aren’t necessarily pushing any more than I am: they are just stuck in a forward-moving wave, while I’m stuck in one that is going backwards. We are all just soft matter being moved around by the crowd.

To find the answer to crowd-safety problems, we need to look further back in time than when people are crammed together at a train station. We need to go right back to the design of our public areas and of big events. Safety is taken very seriously in modern stadium design, and part of that process is evacuation simulation. For example, PSV Eindhoven have run a full range of ‘what if’ simulations to test what would happen in different evacuation scenarios.[16] The models they used are less detailed than the ones described above, but they are a step in the right direction.

Analyses can also be run after the fact. After 21 people died and more than 500 were injured at the Love Parade in Duisburg in 2010, Dirk Helbing and Patrik Mukejri dissected the day’s events in the context of the different models available. They concluded that panicking or intentional pushing was unlikely to have been the cause, and consistent with Mehdi’s model, they found that ‘things can go terribly wrong in spite of no bad intentions from anyone’.[17] There remain many challenges in modelling crowds in dangerous situations, and models are an important part of the solution. Disasters such as Love Parade, Mecca or Hillsborough should never have happened, and hopefully we can use what we have learned from models to stop them happening again.[18]

CHAPTER ELEVEN

# Bet Against the Masses

When I teach statistics at the university, I like to start with an experiment. I take a jar filled with sweets with me and place it on the table at the front of the lecture theatre. I then ask the students to look at the jar and write down how many sweets they think are in it. The student who gets closest to the correct number gets to keep the sweets. They aren't allowed to talk to their friends, and they hand in their guesses to me without showing them to anyone else.

Once I have all the answers, I make a histogram of the results. Figure 11.1 is one of these histograms for one of my smaller classes, based on the guesses of 19 students in computational physics. The histogram shows the number guessing in each range. These are some of the most intelligent, hard-working students at our university, so it is interesting to see how they do when confronted with a practical problem. They had a wide range of guesses from 37 up to 300, with quite a few going for around 40 to 60, and a small number believing that I was generous enough to hand out several hundred sweets. Their average guess was 102, and the median was 90.[1]

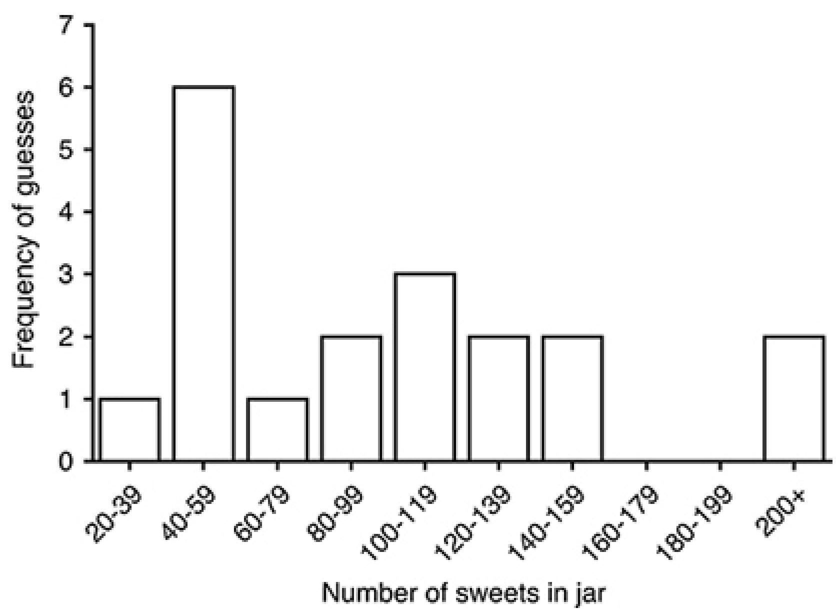



**Figure 11.1** Histogram of guesses by my students of the number of sweets in a jar.

So how many sweets were there? In my jar there were exactly 104 sweets, only two away from the average guess. No single student got it right, so the two students who guessed 100 shared the sweets between them. But in fairness, the prize should have been shared between them all, as the average group guess was closer than any individual guess. The group was collectively wiser than any one individual.

## The Wise Crowd

My classroom experiment is an example of the phenomenon known as the Wisdom of Crowds. The Wisdom of Crowds idea took off in 2005 when James Surowiecki published a book with that as the title.[2] His thesis was that, for many tasks, larger groups of non-specialists can be smarter than a smaller group of specialists. My guessing game with the sweets supports his thesis: together, the students were better than any individual in the group at counting sweets.[3]

Each year I use a different container and a different selection of sweets, but there is a general pattern in the results. Quite a few students underestimate the number of sweets, guessing about 40 or 50 fewer than are actually in the jar. But a small number of students drastically overestimate, guessing two or three times the correct answer. In doing so, these two types of guessers, the underestimators and the overestimators, cancel each other out, and the average guess is close to the true value.

Since Surowiecki's book was published, two research groups have conducted this experiment on a larger scale than I have. At an exhibition in Berlin, Jens Krause and his brother Stefan collected 2,057 guesses of the number of marbles in a large glass jar. The true answer was 562. The guesses ranged from 40 to 1,500,[4] but the average was only 8.4 marbles out, at 553.6. Again, the crowd was pretty close to the correct answer. Andrew King and his colleagues conducted a similar experiment with sweets in a jar during an open day at the Royal Veterinary College.[5] This time the average wasn't particularly close after 82 guesses: 1,396, compared with the true answer of 751. But the median was spot on, at exactly 751.[6]

If you like to place a bet now and again, these Wisdom of Crowds experiments can't be ignored. You might think you know more about the game than most of the other punters, and you may well be right, but that's not the point. In the experiments, most of the participants were not particularly good at guessing the numbers of marbles and sweets. In Andrew's experiments some of the participants guessed over 10,000, while others guessed there were fewer than 50. But as a group they got it right. It is this single entity, the wise crowd, that you have to beat in order to win at the bookies, not the performance of any of the individuals who make up the crowd.

## Beating the Crowd

To see why crowds are a problem for you as an individual punter, consider the over- and under-betting market for corners. Many bookmakers quote a spread, for example 10–11, for the number of corners in a match. Punters can then bet on whether the actual number will be over or under the spread. So if you think there will be more than 11 corners in a particular match, then you can place a £10 per point over-bet. If there are 16 corners, then you make a profit of £(16−11) × 10 = £50, but if there are only 8 corners then you pay £(11−8) × 10 = £30 to the bookmaker.

The same rule applies if you bet under: you pay out £(16−10) × 10 = £60 if there are 16 corners, and you win £(10−8) × 10 = £20 if there are 8. The trick for a bookmaker is to set a spread such that half their clients are betting under and half are betting over. If one person bets under and one bets over, then the bookmaker's profit is guaranteed to be £10, no matter how many corners are taken. Once the bookmaker has this half-and-half balance, then whatever the outcome they will make a profit.

I want you to imagine yourself in a world where you are a genius, surrounded by people who know nothing about football. The other punters bet pretty much at random, and the bookmakers have no idea how many corners there are likely to be during a football match. But you happen to know that the average number of corners is 10.5 in a 90-minute professional game. Sometimes it's a few more, sometimes a few less, but on the whole between 9 and 12 corners is typical.

Now imagine that our not-very-well-informed bookmaker decides to set a spread of 4–5 corners at the start of betting. Then the random punters arrive. Let's assume that all these punters have their own idea about corners. Some think it's going to be an attacking match, with maybe 20 corners; others think there will be fewer than the bookmaker has predicted – only two or three. There are even punters who think there won't be a single corner. Overall, let's assume that each punter is equally likely to predict any number of corners between 0 and 22. Thus 1/23 = 4.35% think there will be no corners, 4.35% think there will be 19 corners, 4.35% think there will be 11 corners, and so on. It shouldn't be difficult to turn a profit against this bookmaker and these punters, should it?

The problem is that there are lots of punters. Imagine that the first punter who comes along guesses that there will be eight corners. He looks at the spread of 4–5 and decides that he can make a profit and puts on a £10 over-bet. Given the likely number of corners, this is a very good deal. Then the next punter arrives. She is even more optimistic, guessing 14, and also bets over. Pretty soon the bookmakers find that they are taking a lot of over-bets. A few conservative individuals bet under, but most punters bet over. The bookmakers may be uninformed, but they are not completely stupid. They know they need a 50/50 balance of over- and under-bets. So after a few bets are placed, they increase their spread by one point to 5–6. But even this isn't high enough, and the over-bets continue to come in, forcing the bookmakers into a further upward adjustment, to a 7–8 spread.

To see what happens in this situation, I built a model. I assumed a simple rule for the bookmakers to apply in order to balance their books. The bookmakers set one counter that tracks the number of over-bets and another counter that tracks the number of under-bets. As the bets come in, the bookmakers increase the appropriate ‘over’ or ‘under’ counter. If, at any point, the ‘over’ counter exceeds the ‘under’ counter by three, the bookies increase their quoted spread by one point. They then reset both counters to zero and start counting again. Likewise, if they count three more under- than over-bets, they drop their spread by one point and again reset both counters. Figure 11.2 gives one example of how the bookmakers’ spread changes with the number of bets placed. It doesn’t take long for the bookmakers to get the spread right. After 30 bets the spread is 7–8, and by 70 bets it is 10–11. This correction doesn’t require the bookmakers to learn anything about the frequency of corners during a match – they are simply adjusting the spread to balance their books. Too many over-bets came in, so they moved the spread up. The whole process happens quickly, because bets come in from large numbers of punters, and the spread is adjusted automatically.

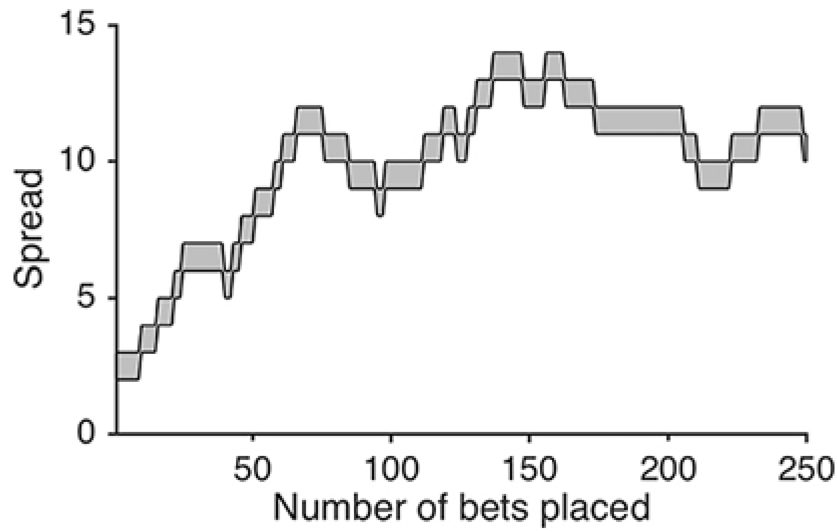



**Figure 11.2** Change of spread in my bookmakers’ model. The shaded area indicates the spread set by the simulated bookies.

Even if you understand the corners game, to make a profit you need to get in before the bookmaker has a chance to correct their mistake. While it’s worthwhile betting on a 7–8 spread, once the spread is at 10–11 you are out of the game. When you’re playing against a large number of these uninformed punters, the probability of you getting a bet in before they do is small. Another opportunity to make a profit is to pounce at one of the few periods when the spread drifts away from 10–11. For example, after 130 bets the spread is 12–13 for a short time, giving you a potential under-bet. Again, you have to be quick: five bets later the spread is back down to 11–12, and then 10–11. In the real world, making a profit is even more difficult than in my model. Bookmakers seldom make such glaring errors in their initial odds, and punters don’t have such wide-ranging opinions about corners. Even with the extreme levels of ignorance I've assumed here, it's difficult to beat the bookies.

How can an uninformed bookmaker and a bunch of people, some of whom believe that a whole match can be played without a single corner being conceded, stop a clever person like you from turning a gambling profit? The answer lies in the average guess of the crowd. While the punters were equally likely to choose any outcome between 0 and 22, their average guess was 11.[7] This is exactly the average number of corners which you knew all the time. Although fewer than 10% of the other punters guessed in the 10–11 range that you knew to be correct, the spread quickly moves towards the overall punters' average. The bookmakers use the collective wisdom of the crowd to set their spreads – and, smart as you may be, you can't get in quickly enough to turn your knowledge into money.

## Not So Smart

We mathematical modellers must always question the assumptions we make. My model of spread-betting shows that if the average guess of the crowd is close to the true answer, bookmakers will push the spread towards that average, and thus towards the true answer. But this is a very big 'if'. Why should we assume that the average guess of lots of uninformed people about the number of corners in a game is correct?

This question is a real challenge to the whole Wisdom of Crowds concept. So far I have presented limited experimental evidence: just three jar-counting experiments, each done in a different way. Economists have conducted more thorough experimental tests. One of the first of these was the Iowa Electronic Market, which is a non-profit betting market for Presidential and House of Representative elections in the US. A more recent example is the prediction website PredictIt, where you can bet on a UK exit from the EU, whether North Korea will test a nuclear weapon, and whether Hillary Clinton will become United States President.

There is some evidence that these markets work. The Iowa Electronic Market gave a very clear Obama win prediction in 2012. In the weeks leading up to this and other recent US elections, the Iowa market gave a more accurate prediction of the outcome than opinion polls.[8] However, both the bookies and the polls were completely out when it came to predicting the result of the 2015 UK general election. At the start of polling day, the odds for David Cameron and Ed Miliband becoming the next Prime Minister were roughly the same, but Cameron and the Conservatives won a majority that no one had predicted. Crowds can't always be counted on to predict the future.

In their experiment at the exhibition in Berlin, Jens and Stefan Krause posed an extra question to the public. They asked the same people who had taken part in their marbles-in-a-jar experiment 'to estimate how many times a coin needs to be tossed for the probability that the coin will show heads each time to be roughly as small as that of winning the German lottery'. To be honest, I find this question a bit confusing, even as a mathematician. But let's work through it. In the German lottery there are 49 balls, numbered from 1 to 49. To win, all six balls that pop out of the ball machine have to be on your lottery ticket. The probability of the first ball being on your ticket is 6/49, since you have 6 numbers and there are 49 balls. If the first ball is one of your numbers, the probability that the second ball is also correct is 5/48, since you have 5 numbers left and there are 48 balls left in the machine. The process continues, with a probability of 4/47 with the third ball, and so on down to 1/44 for the last ball. So the probability of winning is

$$\frac{6\times5\times4\times3\times2\times1}{49\times48\times47\times46\times45\times44}=\frac{1}{13,983,816}$$

Now let's turn to tossing the coin. The probability of getting a head on the first throw is ½. The probability of getting two heads in a row is ½ × ½ = ¼, and three heads it is ½ × ½ × ½ = ⅛. If we keep on multiplying by a half, the probability becomes very small very fast. If we do it 24 times we get

$$\frac{1}{2}\times\frac{1}{2}\times L\ \times\frac{1}{2}\times\frac{1}{2}=\frac{1}{2^{24}}=\frac{1}{16,777,216}$$

This is not exactly the same as the chance of winning the lottery, but in terms of these small probabilities it's as close as we can get. The answer to Jens and Stefan's problem is 24.

In Berlin, only 6% of the people who made a guess got the answer correct. 10% of them thought the answer was 1,000, over 15% went for the maximum allowed answer of 1,500, and 4.5% thought the answer was just six. The probability of getting six heads is 1/64, while the probability of getting 1,000 heads in a row is so small that it would take half a page to write out all the digits. Most people were not particularly good at answering probability questions, and this group performance wasn't particularly impressive. The average guess was 498, 20 times the correct answer.

Jens and Stefan's experiment shows us that for mathematical tasks, skilled individuals outperform the herd. If you pay attention at school and study hard, you can learn how to solve problems that other people can't. There are mathematical steps to answering probability problems that must be learned. This knowledge can't be compensated for by asking a lot of different people to guess the answer. Unless the people you ask know the proper procedure to be followed, then it is unlikely that their average answer will lie near the true value. Averaging uninformed guesses cannot solve problems that require specialist reasoning. The crowd can guess how many sweets there are in the jar, but it can't do maths.

## Around the Corner

The only way to find out if the crowd is wise is to ask it. During the Women's World Cup in June 2015, I decided to put one particular crowd to the test. I had joined members of Stockholm University's Department of Zoology for after-work drinks, and asked the 25 researchers the following question: 'In the Sweden v USA World Cup match tonight, how many corners will there be in total?' The balance between men and women in the group was pretty even, and they came from all over the world, but mostly from the US and Europe. I asked the question early in the evening, so they'd only had about one beer each. But their knowledge of football was somewhat limited. Most of them didn't even know that the Women's World Cup was on, and one Chinese researcher asked me 'What's a corner?' So these guys were certainly not experts.

They had to write down their answers straight away, without conferring and certainly without looking at their phones. From the histogram of their guesses, shown in Figure 11.3, you can see that as with the sweets and marbles guessing games, the crowd had a wide range of opinions, ranging from zero to 34! While the range of guesses was large, the average guess was 11.26, which is very close to the typical number of corners in a match. The median guess was 9. The median is the point at which the bookies should fix their spread because it guarantees a profit for them. And, when I checked the spread-betting on corners for Sweden v USA at the bookmakers, it was 9–10. Spot on. The Stockholm zoologists, who had no expert knowledge, made the same guess as the gamblers and bookmakers who were determining the spread.

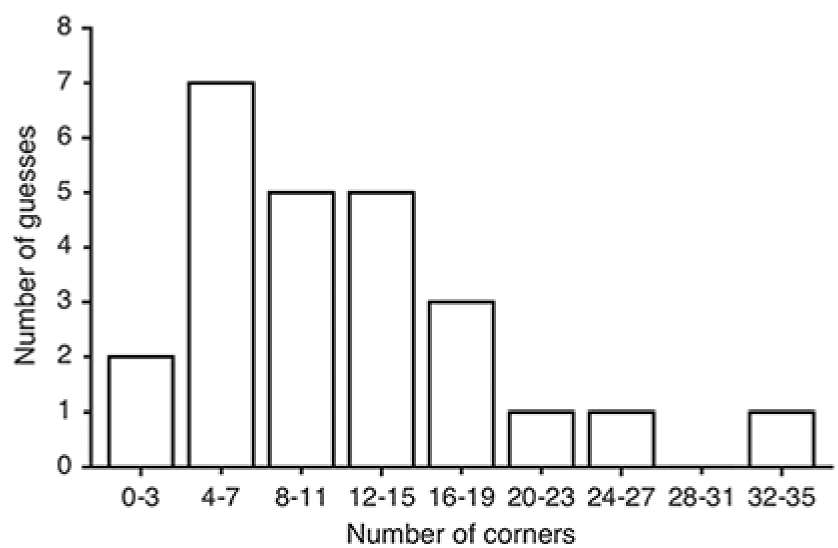



**Figure 11.3** Histogram of guesses of the number of corners in the Sweden v USA match in the Women's World Cup by members of the Stockholm University Zoology Department.

This is a one-off experiment, and a pub is an even less controlled setting than a statistics class, but the results are consistent with the earlier experiments with sweets and marbles. When a crowd is confronted with a guesstimation problem, the guesses will be wide-ranging, but the median or average will be a reasonable estimate.

While the estimate might be reasonable, there is of course no way of truly predicting how many corners there will be in a match. The USA and Swedish teams are unbribable, and no player is going to kick the ball out for a corner unless there's no other option open. What about some of the earlier matches in the Women's World Cup? There had been 12 corners in the game between Nigeria and Sweden a few nights before, but only three in the game between the USA and Australia. There's no particularly useful information to be gained there. While we can speculate, it is difficult to see how we can improve on an estimate of around 9 to 11. As it turned out, there were 18 corners in the Sweden v USA match. Some lucky gamblers will have taken home eight times their money on a corner over-bet, but it wasn't anyone from Stockholm's Department of Zoology. The closest guess in their group came from two people who chose 17. It was an unusually high number of corners.

What my corner experiment shows is that taking the average guess of the members of an uninformed group can yield estimates that are reasonable and reflect reality. But crowds are not able to see into the future. Using uninformed guesses to predict the exact number of corners in a match, or even if there will be more or fewer than average, just isn't possible. The Wisdom of Crowds is remarkable, but it is not magical.

## Premier Predictions

Before the 2014/15 season, journalist and ex-footballer Joe Prince-Wright wrote down his prediction of the final positions of the 20 teams in the Premier League. He then posted it on the NBC website for all to see.[9] This was a brave move, because it's almost impossible to correctly predict all the positions. But it *is* a true test of footballing knowledge. It's easy to be smart after the fact, but predicting the future can in time reveal whether or not a journalist knows his or her stuff.

Prince-Wright's top six were: (1) Chelsea, (2) Man City, (3) Arsenal, (4) Man United, (5) Spurs and (6) Liverpool. And these were exactly the top positions at the end of the season. There are thousands of football pundits writing for different newspapers around the world. Simon Gleave has collected and blogged about predictions for the last few seasons. In 2014/15 he found only 17 journalists who were brave enough to go public with a prediction for the final Premier League table.[10] Of these 17, Prince-Wright was the only one to successfully predict the top six. The most common errors were to place Liverpool above Spurs or to overrate Arsenal.

Further down the league table, however, Prince-Wright's predictions started to wobble. He thought Everton and Newcastle would finish in the top ten, but they both ended up in the bottom half. He had Leicester down for relegation, and it looked as though he would be right until they pulled off a miraculous late-season comeback to finish 14th. And he underrated Swansea and Southampton, who both had good seasons and finished in the top half.

To measure Prince-Wright's ability as a footballing prophet, I took his predicted positions and calculated the difference between his predictions and the actual outcome – see Table 11.1. There is no difference for any of the top six clubs, but there are differences (either positive or negative) further down the table. The average difference between prediction and outcome is 2.3, so Prince-Wright was typically out by just over two positions.[11] This prediction is reasonable, especially when we remember that even on the last day of the season teams can often move up or down a couple of places.

**Table 11.1** Comparison of Joe Prince-Wright's predictions with the outcome of the 2014/15 season.

| Position | Team | Prince-Wright's prediction | Difference |
|---|---|---|---|
| 1 | Chelsea | 1 | 0 |
| 2 | Manchester City | 2 | 0 |
| 3 | Arsenal | 3 | 0 |
| 4 | Manchester United | 4 | 0 |
| 5 | Tottenham Hotspur | 5 | 0 |
| 6 | Liverpool | 6 | 0 |
| 7 | Southampton | 10 | 3 |
| 8 | Swansea City | 13 | 5 |
| 9 | Stoke City | 8 | -1 |
| 10 | Crystal Palace | 12 | 2 |
| 11 | Everton | 7 | -4 |
| 12 | West Ham United | 14 | 2 |
| 13 | West Bromwich Albion | 18 | 5 |
| 14 | Leicester City | 19 | 5 |
| 15 | Newcastle United | 9 | -6 |
| 16 | Sunderland | 16 | 0 |
| 17 | Aston Villa | 15 | -2 |
| 18 | Hull City | 11 | -7 |
| 19 | Burnley | 20 | 1 |
| 20 | Queens Park Rangers | 17 | -3 |

So can Prince-Wright actually predict the Premier League? To answer this question we need to go all the way back to Chapter 1 and my alternative-reality simulations. There I used the goals scored home and away by each team in the 2012/13 season to predict the outcome of the 2013/14 season. In some of the simulations, my model accurately predicted the top of the table, while in others it produced slightly different positions. What was clear from these simulations was that the scoring rate in one season is a very good predictor of the season to come.

In fact, as a simple rule of thumb, the position in which a team finishes the season before is a very good indicator of its position in the coming season. Table 11.2 compares the 2013/14 and 2014/15 seasons.[12] Prince-Wright's prediction of the top six was certainly better than the prediction based on the results in 2013/14. But further down the table, the previous-season predictions start to look better than his. Stoke, West Ham and Southampton all finished in more or less the same position in the two seasons. Burnley and QPR went straight back down, in the same order as they'd come up the season before. Overall, the previous-season prediction was marginally better than Prince-Wright's, with an average difference between the seasons of only 2.2. If Prince-Wright or anyone else had simply presented the previous season's final league table as their prediction for the season ahead, then it would have been pretty good.

Of the 17 pundits identified by Simon Gleave who had predicted a final Premier League table for 2014/15, only one did better than a prediction based solely on 2013/14. Mark Langdon of the *Racing Post* got Liverpool and Spurs the wrong way round for fifth and sixth, but had the top four spot on. He also did better than Prince-Wright for clubs finishing lower down the table, correctly predicting success for Stoke, Swansea and Southampton. The average difference between his prediction and the outcome was only 1.8. But while Langdon was the best-performing pundit, and Prince-Wright was second best, all the others were quite a long way off. Figure 11.4 compares their predictions with the positions of the teams from the season before. Langdon was the only expert above the solid line, and the other experts made around three positional errors on average.

**Table 11.2** Comparison of positions during the 2013/14 season with the outcome of the 2014/15 season. When calculating the difference for teams that were promoted the season before, the positions were assigned in the order in which they came up. Leicester City, who won the Championship, were assigned position 18, Burnley position 19 and Queens Park Rangers position 20.

| 2014/15 position | Team | 2013/14 position | Difference |
|---|---|---|---|
| 1 | Chelsea | 3 | 2 |
| 2 | Manchester City | 1 | -1 |
| 3 | Arsenal | 4 | 1 |
| 4 | Manchester United | 7 | 3 |
| 5 | Tottenham Hotspur | 6 | 1 |
| 6 | Liverpool | 2 | -4 |
| 7 | Southampton | 8 | 1 |
| 8 | Swansea City | 12 | 4 |
| 9 | Stoke City | 9 | 0 |
| 10 | Crystal Palace | 11 | 1 |
| 11 | Everton | 5 | -6 |
| 12 | West Ham United | 13 | 1 |
| 13 | West Bromwich Albion | 17 | 4 |
| 14 | Leicester City | 18 | 4 |
| 15 | Newcastle United | 10 | -5 |
| 16 | Sunderland | 14 | -2 |
| 17 | Aston Villa | 15 | -2 |
| 18 | Hull City | 16 | -2 |
| 19 | Burnley | 19 | 0 |
| 20 | Queens Park Rangers | 20 | 0 |

Using the previous season as a benchmark for comparing pundits' predictions isn't completely fair, because the final table for 2014/15 happened to be very similar to the one for 2013/14. It is rare for there to be only small changes in position from one season to the next. For example, there are much larger differences between the final tables for 2012/13 and 2013/14. In 2013/14 Manchester United finished six places lower than the previous season, and Liverpool five places higher. That season also saw West Bromwich Albion, Fulham and Norwich drop seven to eight places, from mid-table safety into a relegation battle.

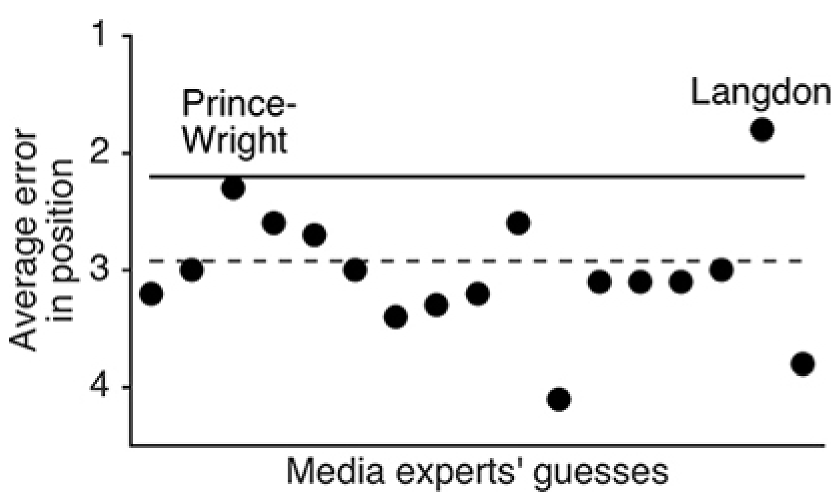


**Figure 11.4** Comparison of expert predictions of the 2014/15 season (black dots), the positions of the teams from the season before (solid line) and a benchmark based on the average change in position over five seasons (dashed line).

The dashed horizontal line in Figure 11.4 marks the average change in position, of 2.92 places, of Premier League clubs over the five seasons to 2014/15. It shows us how good we would expect predictions to be if the experts used guesswork based on the previous season. Only five of the 17 experts sit above the dashed line, and 12 of them are below it, so there is no evidence that the experts really know what's going to happen.

Things get worse if we look at how the experts did in predicting the 2013/14 season, as shown in Figure 11.5. Now, 13 out of 14 experts find themselves below the dotted line. None of them predicted the change of fortunes for Norwich, Fulham and West Brom, and few of them thought it would go as badly for Manchester United as it did. Most experts are simply unable to predict the outcome of a season.

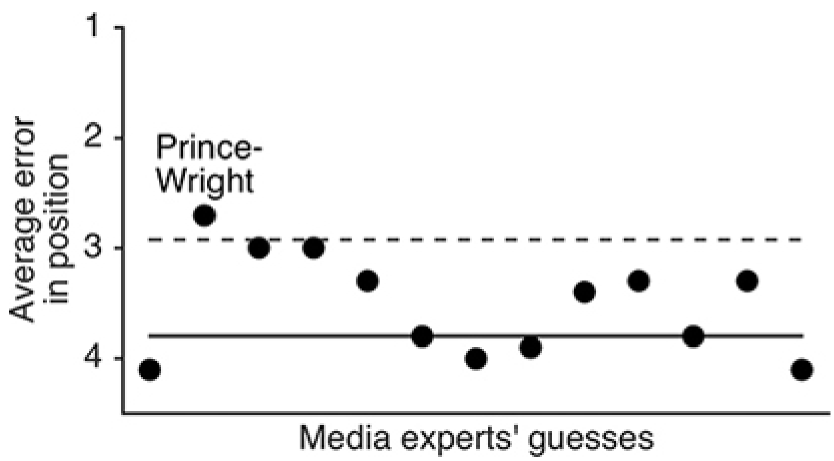



**Figure 11.5** Comparison of expert predictions of the 2013/14 season (black dots), the positions of the teams from the season before (solid line) and a benchmark based on the average change in position over five seasons (dashed line).

I don't have Langdon's predictions for the 2013/14 season, but Prince-Wright made the best predictions of those surveyed. On this basis, it may appear that he really can predict the future. However, before the start of the 2015/16 season, Prince-Wright made a new set of predictions. He predicted a second title in a row for Chelsea, with Manchester United coming a close second. Liverpool and Spurs would fail to qualify for the Champions League, and Leicester, Norwich and Watford would be relegated. Halfway through the season, when I completed this book, Chelsea were in serious trouble and Leicester were league leaders. So it looks as though Prince-Wright's powers have finally failed him. This book was published in May 2016, so you can now check his predictions yourself. But I think I can safely conclude that there is still no expert who can reliably, over a number of seasons, predict the outcome of the Premier League.

## Barcelona 4,539, Kilmarnock 1,093

Simon Gleave, who collected these media expert predictions, told me that 'a great deal of prediction is about being lucky rather than skilful'. Even with the possible exception of Joe Prince-Wright, it is difficult to argue with this conclusion. Simon is head of analysis at Infostrada Sports, a company specialising in sports data and intelligence. In one project,[13] Infostrada have co-developed the Euro Club Index, a tool for ranking team performance. The aim of this index is to give a statistical measure of team performance based on encounters between the teams.

Each club in the top division of each European league is assigned a number of points. At the start of the 2015/16 season, Barcelona had 4,539 points, Real Madrid 4,342, Bayern 3,953 and Juventus 3,712. The highest-ranked English club were Chelsea, in seventh place with 3,635 points, while newly promoted Bournemouth were in 165th place with 2,052 points. Scotland's highest-ranked team were Celtic, with 2,480 points and in 75th place. Kilmarnock, who had narrowly avoided relegation from the Scottish Premier League, were in 481st position with just 1,093 points.

When one team beats another, the winner takes points from the loser's index and adds them to its own. So when Juventus beat Real Madrid in the Champions League semi-final in May 2015, Juventus gained 28 points and Real Madrid lost 28. When Juventus went on to lose to Barcelona in the final, they transferred 18 points to Barcelona. Victory over a higher-ranked team gives more points than beating a lower-ranked team. So Juventus took more points from Real than they later lost to Barcelona. If the mighty Killie were to beat Barça in a competitive match, they would take hundreds of points from them, but if Barcelona won then Kilmarnock would only lose 1 or 2.

Like the Premier League Performance Index we looked at in Chapter 4 for players, I wouldn't call the Euro Club Index for teams an 'objective' measure. The exact rules for points are set up in advance by humans, and the index gives a statistical ranking that is purely results-based. Nevertheless, for the 2014/15 Premier League, the Euro Club Index performed just as well as the best human predictor. The average difference between the Euro Club Index prediction and the outcome was 1.8, exactly the same as Mark Langdon's.

Simon Gleave also collects data from other model predictions. Out of 28 predictions he collected for 2014/15, Euro Club Index was one of three that performed best. However, the fact that the overall positions in 2014/15 didn't change much from the season before was an advantage for ranking methods based on past performance. For the less predictable season of 2013/14, the Euro Club Index was the worst of all the models Simon looked at. The only human expert to be beaten by the index was former Liverpool goalkeeper David James, who was out by 11 places when he predicted that Everton would finish 16th.

So you can consult crowds, models or experts, but there is still no reliable evidence that any of these can make long-term predictions that beat last season's benchmark.

## Cup Final Rumour Day

Sometimes people are just a bit wrong, but at other times they can be very wrong indeed. Human history is filled with examples of us temporarily believing things that later turn out to be total nonsense. Lemmings engage in mass suicide, the Great Wall of China is the only human-made object visible from the Moon, Napoléon Bonaparte was short, Barack Obama is a Muslim – these are just some of the false rumours or ideas that many people still believe and tell one another today. Rumours about football transfers and celebrity pregnancies are even more common. As I showed in Chapter 10, these rumours spread rapidly as people share them with one another. Often little consideration is given as to whether they are true or not.

If we start passing on rumours without checking them ourselves, then the Wisdom of Crowds breaks down. When I did my sweets-in-the-jar and number-of-corners experiments, I was very careful not to let my subjects confer. All the guesses were independent and based on the participants' best judgement. I didn't want to let any of them who may have thought they knew best influence the others.

When Andrew King and his colleagues did their sweets-in-the-jar experiment at the Royal Veterinary College, they also looked at what happened when people knew what others had guessed. In a variation of the original counting experiment, they told later visitors the average of all previous guesses. These visitors had access to more information than in the original experiment: not only could they see the jar themselves, but they also knew what others had guessed.

But knowing more doesn't always help us to make better judgements. By chance, the first visitors tended to overestimate the number of sweets, mainly guessing over 1,300, when in fact there were 751. The next visitors looked at these large guesses and at the jar. They thought the guesses must be too large, so they guessed slightly lower. As more guesses were made, the average guess decreased slightly, but remained over 1,000. The visitors didn't fully trust their own judgement and overestimated the number of sweets.

A failure to trust our own judgement and go with the crowd is behind much of our decision-making. When we are unsure, we tend to follow the suggestions of others. Think of the following situation. It's FA Cup final day, Newcastle United versus Aston Villa, and there's a queue of friends at the bookmaker waiting to place bets. They have to decide whether to bet on Newcastle or Villa to win the Cup. The first punter has no information to go on, and is forced to decide for herself. We'll assume that Newcastle are the stronger team, and that she has a 70% chance of getting it right and backing them, but a 30% chance of choosing Villa.[14] The next punter, also with nothing to go on, faces the same choice, so again has a 70% chance of choosing Newcastle. But it occurs to the third punter that he can either decide for himself or save himself a lot of effort by asking the first two what they think. Assuming that they are honest, and that they agree with each other, there is only a 15.5% chance that both of the two who have already placed bets will get it wrong.[15] So for the third punter it's safer to go with his friends' judgement than to trust his own. If they disagree, then he should ignore them and make up his mind for himself. The logic that applies to the third punter also applies to all of those behind him in the queue. If they choose two others who have already made up their mind, ask their opinion and bet accordingly, then they have a better chance of getting it right than if they try to assess the evidence for themselves.

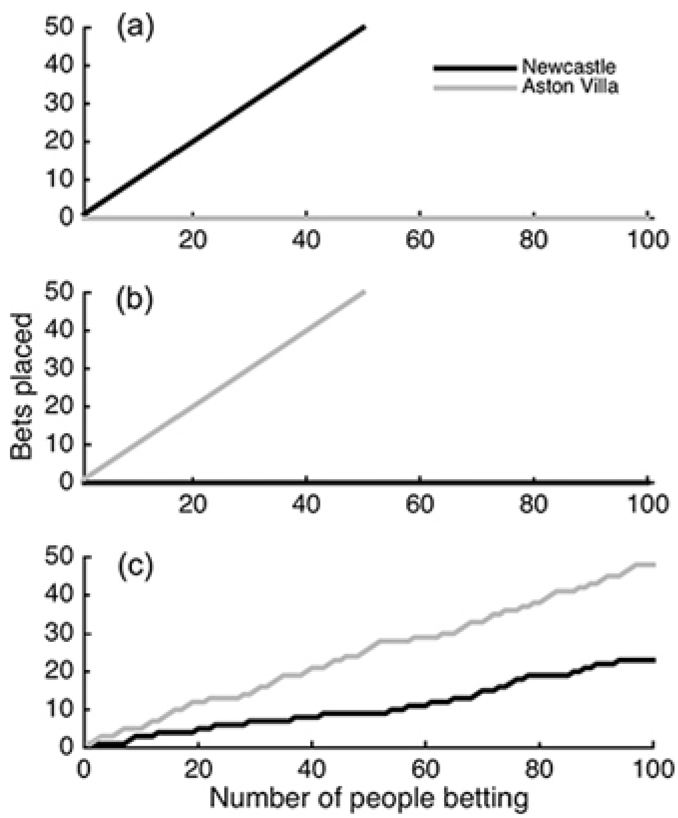



**Figure 11.6** Simulation of the model where individuals use the decisions of others to make up their own mind which team they think will win the cup. Three simulations are run for the same parameter values. In the first, both the first two punters choose Newcastle (a – occurs in 49.0% of simulations). In the second simulation both the first two punters choose Aston Villa (b – occurs in 9% of simulations). In the third simulation one punter chooses Villa, the other Newcastle (c – occurs in 35.5% of simulations).

So that's what they do, and Figure 11.6 shows three different simulated outcomes of this model. In the first simulation, the first two punters chose Newcastle, so when the third punter came along he concluded, from asking the others, that Newcastle were the more likely to win. The fourth, fifth and all the other punters all came to the same conclusion. In the second simulation, the first two punters by chance chose Villa. This happens in only 9% of the simulation runs, but the consequences are striking. Because the first two went for Villa, the third punter chose them too, and everyone else followed suit. Everyone was saying Villa, so there was no reason to believe anything else. In the third simulation, the first punter went for Newcastle and the second for Villa. The third had to make his mind up for himself, and went for Newcastle. From then on, when the punters further along in the queue asked those who had come before them for advice, sometimes they followed the two that fancied Newcastle, sometimes they went with the two who said Villa, and sometimes they made up their own mind when their two friends didn't agree. It wasn't until after quite a few punters had placed their bets that a consensus for Newcastle emerged.

Given what they have heard about the match, all of the punters in the model behave rationally. They realise that they can improve their chance of getting it right by following the advice of others. More often than not, their strategy works. In most cases the crowd goes for Newcastle, who are the better team on this occasion. But sometimes the crowd gets it wrong: if, by random chance, the first two individuals plump for Villa, everyone else just follows their advice. Even in simulations where the first two punters couldn't agree, most of the time Newcastle become favourites, but sometimes Villa take off. Who emerges as the favourite depends very strongly on what happens at the start.

The model reveals a paradox. On the one hand, we are collectively wise, and copying others provides us with information on which to base our decisions. But on the other hand, when we copy we lose that wisdom. We all want to know what everyone else thinks in order to get a better picture. But once we have found out the opinions of others, our independent judgement disappears.

At the start of this chapter, it seemed that individual punters were up against an impossibly wise crowd. Bookmakers just averaged opinions, and the odds reflected the collective wisdom. But as we have looked more closely at other models and experiments, we have seen subtleties in the Wisdom of Crowds concept. Jens and Stefan Krause's experiment shows that the crowd isn't good at skilled tasks involving mathematical reasoning. Simon Gleave's collection of Premier League predictions shows that even so-called experts probably don't really know much more than anyone else. And, paradoxically, we see that when individuals in the crowd start to talk to one another, the crowd loses its collective judgement.

Together, these observations suggest that there's an opportunity for someone who would like to try their hand at beating the bookies. It may just be possible for a mathematically competent individual who isn't easily swayed by other people's opinions to make a profit on football betting. I'm not entirely sure it can be done, but it would certainly be fun to find out.

CHAPTER TWELVE

# Putting My Money Where My Mouth Is

I was over the moon when Bloomsbury said they would pay me to write a book about football. I have written a book before, but it was an academic text, and hasn't really been a money earner. I worked full-time for a year to write *Collective Animal Behaviour*. Four years later, with various Swedish taxes deducted, I received £500 in royalties for my efforts.[1] That was about 31 pence per hour worked, paid out a long time after everything was finished. So when Bloomsbury offered not only to pay good money for my work but also to give me the cash up front, it was like winning the pools.

What better way to use an unexpected windfall than to invest it in something I know well? If Bloomsbury have faith in me, then I should repay that faith – by taking their advance and investing in what I have learned while writing this book. It's time I stopped theorising about the game and put my money where my mouth is. Let's predict some football results and win some money.

## Odd Probabilities

Before I start throwing my money around, it's important that I get the basics right. I need to understand a few facts about odds, probabilities and how bookmakers make their money. Modern gambling is far from straightforward. There is a wide range of online bookies, offering odds in different formats. Odds-comparison sites allow you to compare bookmakers, and there are even sites where you can trade your bets with other punters. Gambling is no longer simply a matter of going into the bookmakers and putting a fiver on your favourite team.

In the UK, odds are stated as fractions, such as 3/7. This means that for every £7 you stake, you will win £3 if your bet comes off. Even for a mathematician this is not entirely straightforward. When I place a bet, I don't think £7 is the obvious amount of money to put on City to win. I usually want to make bets of £1, £2, £5, £10 or £20. These are solid, everyday numbers, not fancy primes. Odds of 3/7 mean that if I go into a bookmakers and bet a fiver, then to calculate my potential profit I have to work out 3 × 5/7. Three prime numbers in one equation. OK, it isn't really that difficult – the answer is £2.14 – but it does take a second to think it through.

Things get even trickier when we compare odds. Which are better odds, 3/7 or 5/11? To see why, multiply the top and bottom of the fraction 3/7 by 11, giving (3 × 11)/(7 × 11) = 33/77. Then multiply the top and bottom of 5/11 by 7. (5 × 7)/(11 × 7) = 35/77.

Now we can compare the odds directly. 35/77 > 33/77, so 5/11 > 3/7. 5/11 are the better odds. But not all people, and I am included in this group, can do this type of mental arithmetic quickly. UK odds use numbers in a way we may well not be familiar with from other everyday experiences.

While UK odds are a bit awkward, in the US they are just plain crazy. They start off OK. If the UK odds are 2/1, then the US odds are +200: if you bet $100 and win, then your profit is $200. The two-orders-of-magnitude jump is perhaps to be expected. Vegas gamblers roll with hundreds of dollars, while UK punters nurse their £5 kitty. But for favourites, the odds suddenly switch from positive to negative. So, if the UK odds are 3/7, then the US odds are stated as −233. This doesn't signify a pay-out of $233 if the favourite wins, as a naive mathematician may think, but that in order to make a $100 profit, you'll need to have bet $233. As a result, all US odds are a number greater than or equal to 100, preceded by a + or a − to indicate whether the number is your profit or the amount you need to bet to win $100.

For most of this chapter, I'm going to use the European convention for odds, and UK currency, the simplest combination for a British mathematician. European odds are easy to understand. They tell me how much I'll get back if I win on a £1 bet. If the odds are 1.5, then after winning a £1 bet I'll have £1.50. If I lose, I'll not have my pound any more. Simple. For UK odds of 3/7, the European odds would be presented as 1 + 3/7 = 1.43. So if I bet £1, I will have £1.43 if my bet comes off. European odds are like interest rates: to calculate our potential profit, all we have to do is multiply by the odds.

## When Should I Place a Bet?

I've never really gambled on football. From watching friends who do, my overall impression is that mathematics isn't really on their mind when they bet. For example: I'm watching a Champions League match with my friend John. Atlético Madrid are playing Olympiacos, and Atlético are two goals down after 31 minutes. The European odds for an Atlético win go up to 7.00, and this number flashes up on the screen. 'Champions League matches can be pretty unpredictable,' says John. 'A tenner on Atlético would certainly liven up a Tuesday evening.' He gets out his phone and places the bet. An hour and a half later, and the final score is 3–2 to Olympiacos. John is £10 poorer, but we have both jumped up and down a lot more than we would have done if he hadn't placed his bet. It can be fun to gamble now and again, without thinking too much about the details.

This is totally irrational behaviour on the part of John. He has no idea how likely Atlético are to win, nor has he worked out how likely the bookmakers think an Atlético win is. And why £10? Why not just £1 or £100, 10 pence or 10% of his annual salary? He knows that with his £10 bet he'll be £60 richer if Atlético win, and £10 poorer if he loses, but what is his expected profit or loss on this bet? And how much is the bookmaker raking in from him and all the other amateur gamblers stuck in front of the TV on a Tuesday evening? None of these questions even pass through his head, but if he were serious about gambling, they are exactly the questions he should be asking.

If you take just one thing from this chapter, it should be this: before you look at the odds and place a bet, always calculate the probability of your prediction. Odds are not the same as probabilities. The odds tell you your potential profit, but a probability gives you an estimate of the likelihood of an event or outcome. Often when we talk about the future we use probabilities: 'I'm 99% certain that those two are an item' or 'There's a 30% chance of rain tomorrow.' And when you're deciding whether or not to bet on a win, you should also be thinking in terms of probabilities: 'There's a 50% chance that Chelsea will win the league again, and just a 1% chance for Leicester City.' Before you part with your cash, and before you even look at the odds, ask yourself the following question: 'What is the probability that my team will win?'

It's three weeks before the Community Shield between Chelsea and Arsenal in 2015. Chelsea were Premier League champions last season, but FA Cup winners Arsenal have a point to prove. I would rank the teams as reasonably equal, maybe giving Chelsea the edge. So my prediction is that Chelsea have a 55% probability of winning, and Arsenal 45%. But there's also the possibility of a draw after 90 minutes. Between 20% and 30% of matches end in a draw, depending on the competition. I don't know so much about the Community Shield, so I'll set my probability of a draw to 25%. The probability that the match will decided at the end of ordinary time is then 75%. So my estimate for the probability of a Chelsea victory after 90 minutes is $0.55 \times 0.75 = 41.25\%$, and $0.45 \times 0.75 = 33.75\%$ for Arsenal.

Only now do I allow myself to look at the odds. I visit a leading UK bookmaker's website and see that the odds are 13/10 for a Chelsea win, 12/5 for the draw and 21/10 for an Arsenal win. In European odds, these are Chelsea 2.3, draw 3.4 and Arsenal 3.1. So this bookmaker roughly agrees with me. They think Chelsea are more likely to win than Arsenal, and because the draw will pay out the most, they think that's the least likely outcome. But what I really need to know is whether I should place a bet at these odds – and on which outcome.

To answer this question, I need to calculate how much money I expect to have after the match if I place a £1 bet. For Chelsea, I estimate the probability of them winning to be 41.25%. What I now need to know is how much I expect to have, on average, if I place this bet. It is this *expected outcome* that is central to deciding which team to bet on. I want to know how much I expect to have after I place the bet, assuming that the probabilities I have assigned to each team winning are correct.

Let's calculate the expected outcome for a bet on Chelsea. If I bet £1 on Chelsea, then there is 41.25% chance that I'll have £2.30 after the match. The probability of this outcome multiplied by these odds is 0.4125 × 2.3 = 0.9487. Likewise, I estimate the probability of Chelsea not winning to be 58.75%, an outcome which would leave me with nothing. Thus my overall expected outcome for a £1 bet on Chelsea is

$$(0.4125 \times 2.3) + (0.5875 \times 0) = 0.9487$$

This is just under 95 pence, less than the £1 I started with. So Chelsea are not a good bet. Betting on a draw looks even worse: a similar calculation shows that I would expect my £1, on average, to become 85 pence. However, Arsenal are starting to look like a good option: I find that I would expect my £1 to grow to £1.046 – a small but respectable interest rate of 4.6%.[2]

What can be hard to get your head around is that by betting on Arsenal I am not predicting that they will win the match. I've already said that I think the chance of them winning after 90 minutes of play is just over 33%, while Chelsea have more than a 40% chance of winning. So if I bet on Arsenal I'll most likely lose, even based on my own assessment. This can be difficult to grasp. We'd all like to tell our friends about how we picked winners. But if you gamble properly, you should end up backing just as many winners as losers. The trick to gambling is not 'picking winners' but maximising your expected profit. This expected outcome is calculated, as above, by multiplying the probability of each result by the amount the result pays out.

There is a simple general mathematical rule for deciding whether to bet or not. If you estimate the probability of a team winning to be $p$, and the European odds are $o$, then you should place a bet whenever

$$p > 1/o$$

Making this calculation takes just a second on your phone's calculator, and you should always do it before placing a bet. Work out your own probability of an outcome. Now visit a bookmaker's website and check the odds. On the calculator, press '1', '÷', type in the odds and press '='. If the number that comes out is less than your probability, then place the bet. If not, then think again. This check can save you from some very silly mistakes.

For Arsenal, $p$ = 0.338 and $1/o$ = 0.323. $1/o$ is less than $p$, so the condition holds. For both the draw and a Chelsea win, the condition fails. So, back in July 2015, I put my £1 on Arsenal and entered the world of gambling. Three weeks later I was £2.10 richer.

## How Good Do I Have To Be?

A win was encouraging, but I have to remind myself that I was lucky. I thought Chelsea would win, but so did the bookmakers. It was only because the odds were better for Arsenal that I placed the bet. Now I have to think about exactly what I'm up against. The bookmakers are setting their odds based on the betting decisions of all the punters. If the Wisdom of Crowds theory holds for these gamblers, then we should expect the odds offered by the bookmakers to reflect the true probability of teams winning, losing and drawing. In the last chapter, I showed how quickly bookies can adjust their odds to reflect the predictions of the crowd. Not only that, but they have a lot more experience of predicting football results than I do. Even if the crowd are a bit out, the bookies will have set their initial odds based on years of experience. To win, I have to do better than both the bookmakers and the aggregated knowledge of tens of thousands of gamblers.

Let's imagine, then, that the bookmakers' odds are a perfect reflection of the probability of the various teams winning, drawing or losing. Let's also assume that the bookmakers have set odds to reflect the number of people betting on each outcome. So, for that Community Shield match, if the odds for Chelsea to win are 2.3, then the probability of them winning is simply 1/2.3 = 43.5%. Similarly, the probability of Arsenal winning is 1/3.1 = 32.2%, and the probability of a draw is 1/3.4 = 29 4%. But there is a problem. 43.5% 32.2% + 29.4% = 105.1! We can't have a total probability that is greater than 100%. It just doesn't add up.

The reason the probabilities don't total 100% is that the odds aren't fair. That extra 5.1% is the bookmaker's advantage. To get the real probabilities, we need to correct for the profit by dividing through by 105.1. So the bookmakers' true probability of a Chelsea win is 43.5/105.1 < 41 3%, the probability of an Arsenal win is 32.2/105.1 = 30 7%, and for a draw it is 29.4/105.1 = 28 0%. If we now add up these three adjusted probabilities, we do get 100%. For a perfectly wise crowd and perfectly efficient bookmakers, these are the probabilities of each outcome.

If the odds perfectly reflect reality, then it doesn't matter which outcome I bet on – my expected profit is always the same. If I bet £1 on Arsenal in the Community Shield, I expect to get back 0.307 × 3 1 = 0 95, = 95 pence. The expected profit is the same for Chelsea, 0.413 × 2 3 = 0 95, and again I expect to have 95 pence. And – you guessed it – if I bet on a draw, I expect to get back 95 pence. On average, the bookmaker will take about 5 pence from me per £1 bet.

Most online bookmakers have about a 5% to 6% advantage built into their odds. You can work out their advantage following the same steps as above, replacing my Community Shield odds with any of the odds they provide. Take, for example, the European win, draw and lose odds offered by any bookmaker, convert them into probabilities by taking one over the odds, and then add up these probabilities. The difference between this sum and 100% tells you the degree of 'unfairness' in the bookmaker's odds – which gives you an indication of how good you have to be to win.

It is possible for you to improve your odds. Different bookmakers will quote different odds for the same match. Each bookmaker's odds give them a built-in advantage, but some bookies offer better odds for the favourite while others will favour the underdog. For example, in the third weekend of the 2015/16 season, one leading bookmaker was offering 10.00 on Newcastle United to win away at Manchester United, while its big rival offered 9.00. So betting on Newcastle with the first bookie rather than the second would add 11% to your expected profit. But the rival was offering 1.33 for a Manchester win, while the first bookie had only 1.28, so a bet with the rival would give you a 4% boost.

Having an account with just one bookmaker is, to put it plainly, daft. Even if you use just four different bookmakers, you can find big differences in the odds offered. I signed up with four leading bookmakers and compared the odds offered using a comparison website. By taking the bookmakers with the best odds, their advantage for the bets I looked at dropped from 5% or 6% to around 1% or 2%. Sometimes the advantage was as low as 0.01%. Bookmakers are competing with one another, and they do this by offering lower odds on different outcomes. By shopping around it is possible to find better betting opportunities.

A 1% or 2% disadvantage may seem small, but it is on every bet. So if I place one £10 bet with a bookmaker's advantage of 2%, then my expected outcome is to have £9.80 after that bet. After the second bet I will expect to have £9.60. After ten bets I'll have £8.17, and after a year of making one bet a week, I'll have an expected capital of £3.50.[3] This is bad, but using just a single bookmaker is much worse. If you have one betting account with a single bookmaker who has a 6% advantage, then a year of gambling once a week with them will turn £10 into 40 pence. The other £9.60 is theirs to keep. It's little wonder that bookmakers try to attract business with 'free spins' and £100 cash start-ups. These disappear quickly once you start gambling.

## The Challenge

The 2015/16 Premier League season is about to start, and my aim is to win money by betting on matches in the first five weeks. My budget is £500, plus the £2.10 I've already won thanks to Arsenal's Community Shield victory. So that's £502.10 in total. I've set myself a few rules. First of all, I'm going to use mathematical betting strategies. When I watch football for pleasure, I obviously form opinions on which teams are playing well, but I'm going to set those opinions aside for the duration of my betting experiment. I set up models in advance, based on data from matches and the bookmakers' odds, and then use these models to make predictions. The models are all in place before the betting period begins, and although I will tune them slightly, I'll stick to them throughout.

The second rule is that I build all the models from scratch, using the type of mathematics and statistics found in undergraduate economics, physics and engineering courses. I have some help in understanding odds and match probabilities from Robin Jakobsson, who works for a professional odds-setting company, OddsCraft. But I programmed all the models myself in Matlab, which took about three or four weeks.

The third rule is that all the data used in the models are easily accessible online. There is an amazing amount of match data available, from shot statistics to team-ranking indexes, and there are a large number of odds-comparison sites which include historical records. I use only these sources as input to the models – not any of the private data collected by clubs and companies.

The final rule is mainly for myself: it is to have fun. I was incredibly lucky to get paid to write this book and, win or lose, I'm going to enjoy using the money this way.

## Strategy 1: Strong Favourites, Well-Matched Draws

How well do the bookmakers' odds predict the outcome of matches? As we saw in Chapter 11, if the crowd is wise then we should expect the odds offered on matches to accurately reflect the results. But we also saw that the crowd can make mistakes, both because it makes poor calculations and because as rumours spread it can get carried away. To test how well the betting crowd does and how well the bookmakers' odds reflect reality, I downloaded the closing odds for Premiership matches during the 2014/15 season[4] so I could compare them directly with the actual results.

Let's start by looking at matches where the closing European odds for a home win were close to 2.00, so that a winning bet on the home team would roughly double your money. When a bookmaker sets odds of 2.00, it implies that they are predicting a win about half of the time. There were 35 matches in the 2014/15 season of the English Premiership with home odds between 1.90 and 2.10. Liverpool v Everton was one example (final score 1–1); West Bromwich Albion v Burnley (4–0), Manchester City v Manchester United (1–0) and Sunderland v Hull City (1–3) were others. These matches typically involved a home team slightly higher in the table than the visitors, and the odds favoured a home win.

For odds of 2.0, the bookmakers' predicted percentage of wins is 50%. So out of 35 matches we would expect around 17.5 home wins. In reality there were 19 home wins – very close to the expectation, and certainly within the bounds of statistical uncertainty. The bookmakers' odds were a good prediction of home wins for these matches.

But the bookies' odds aren't always so accurate. Let's take matches where a home win has odds between 1.33 and 1.43. Such odds are typical for Champions League contenders playing lower-table opposition. In 2014/15, odds in this range were often set for Manchester United's home matches, including their wins over Sunderland, Crystal Palace and Newcastle – as well as a defeat to West Bromwich Albion. Odds between 1.33 and 1.43 imply that the bookmaker predicts wins in between 70% and 75% of cases.[5] In fact, 25 out of 28 of these matches resulted in a home win, giving a win percentage of 89%, so here the bookies' odds underestimated the favourites.[6] Not only are big favourites likely to win, they are even more likely to win than the odds predict.

To see if these results are part of an overall trend, we need to find the statistical relationship between the probability given by the odds and the actual probability of a win. This is shown in Figure 12.1 for all 380 matches in the 2014/15 season, grouping together matches with similar home-win odds. The odds between 1.9 and 2.1 fall very close to the bookmakers' predictions because, as I have shown, these odds were a good predictor of match outcome. For odds between 1.33 and 1.43, the points lie above the dashed line because the bookmakers underestimated the proportion of home wins.

Looking at Figure 12.1, it is difficult to see whether there is any overall trend in underestimating strong favourites. Some dots fall above the diagonal line, while others fall below it. But by using a statistical method called logistic regression, I can find the best fit between prediction and outcome.[7] This fit reveals a small bias in the bookmakers' odds: the proportion of home wins increases faster than the bookmakers' win probabilities.

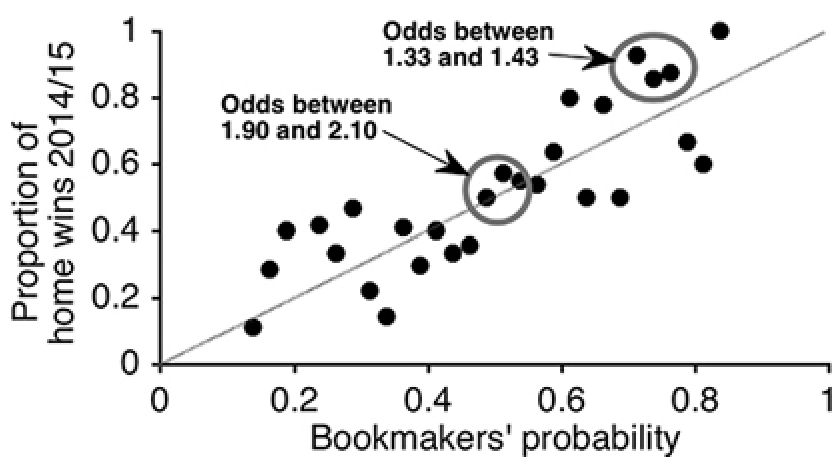



**Figure 12.1** Comparison of probabilities based on bookmakers' odds and match outcomes for Premier League season 2014/15. Outcomes (black circles) *above* the dotted line correspond to home wins occurring more often than predicted by the odds. Outcomes *below* the dotted line correspond to home wins occurring less often than predicted by the odds. Open circles indicate odds combinations that are discussed in more detail in the text.

Now, I have to be careful here. This trend isn't statistically significant for the 2014/15 Premier League season, but it is consistent with earlier academic research in economics and management. Several studies have looked at betting on football, American football, horse racing and other sports and found that gamblers prefer betting on long shots.[8] From a psychological point of view, this is easy to understand. Putting £5 on Chelsea to beat Burnley at odds of 1.30 in order to win £1.50 is much less fun than backing Burnley at 11.00 and potentially winning fifty quid. For most people, betting needs to create suspense, as well as produce the odd windfall. But for our mathematical model it's the bottom line that matters. Previous research, combined with the above analysis, supports the strategy of backing the favourite.

There is another pattern in the bookmakers' odds that may be exploitable by a gambling strategy. In encounters between two well-matched teams, draws are much more common than when one team is a strong favourite. Manchester City v Chelsea, Sunderland v Stoke, Liverpool v Arsenal and QPR v Swansea all ended in draws in 2014/15. In these matches the difference in the bookmakers' probability of each of the two sides winning was less than five percentage points. For example, in Liverpool's home game against Arsenal, the bookmakers gave Liverpool a 34.5% chance, and had Arsenal at 37.7%, giving a difference of percentage points. The bookmakers probability of a draw was then $1 - 0.345 - 0.377 = 27.8\%$. The game ended 2–2.

I looked at the 72 matches in the 2014/15 season where the difference in the win probabilities for the two teams was less than 10 percentage points. Of these matches, 25 resulted in a draw, giving a 34.7% probability of a draw. In contrast, the bookmakers' odds gave an average draw probability for these matches as 29.5%.[9] The odds are biased away from draws in favour of a win for one of the two teams. Again, there isn't a strong statistically significant difference between the odds and the results, but it is a trend that is worth considering. It could be that punters don't like betting on draws between equally matched teams, just as they don't fancy betting on favourites.

Although not statistically significant, the biases against equally matched draws and strong favourites are the clearest pattern I've found in last year's odds. They form the basis of a strategy I'll try – I'll call it my odds-bias strategy. For each match I first calculate the difference in the bookmakers' win probabilities for the two teams. If this difference is greater than 0.4, then I bet on the favourite. If the difference is less than 0.15, then I bet on a draw.[10]

## Strategy 2: Results Index

The Euro Club Index uses a system (presented in more detail in the previous chapter) of transferring points between clubs. The index is calculated purely on the basis of match outcomes. When a team wins, its index increases; when a team loses, its index decreases. This focus is the strength of the Euro Club Index. In the same way that gamblers tend to favour long shots and steer clear of betting on draws between well-matched teams, so we tend to see short-term patterns that aren't really there. Several studies of basketball and American football have identified a 'hot hand' effect, where gamblers favour the teams that have won their last few matches. The punters believe that a team is on a roll, and that it will continue into the next match. The strength of the 'hot hand' effect is debatable, and because it has been so widely discussed over the past 30 years there is probably a large number of gamblers who deliberately bet against it. But it is in our nature to overreact to short-term changes in form, and if the Euro Club Index really does reflect long-term patterns, then it should beat the odds.

The Euro Club odds do change, but slowly. Figure 12.2 shows the Euro Club rankings for the five highest-ranked teams in the first four weeks of the 2015/16 Premier League season. Chelsea are on their way down, having recorded three losses, a draw and only a single win. Even after this bad start to the season, the Euro Club Index still had Chelsea as the highest-ranked team in England in Week 3. That week they played away to West Brom, and the Euro Club odds gave Chelsea a 63% chance of winning, while the best odds from the bookmakers gave a win probability of 60%. Manchester City, who began the season with straight victories, were given a 46% chance of winning away to Everton by Euro Club, while the bookmakers gave them a 56% chance of a win. The betting public and the bookmakers were quicker to respond to the results in the first four weeks of the season than the Index.

Both Chelsea and Manchester City won their matches in Week 3, but when Chelsea were humiliated at home by Crystal Palace in Week 4, they swapped positions in the ranking. These observations alone are not enough for us to conclude anything about the relative merits of the Euro Club Index and the betting market. But we can see that Euro Club is more 'cautious' than the betting markets in how it evaluates changes in form. If the Euro Club Index is working properly, and the betting markets are overreacting to results, then following the Index should make a profit. I adopt the Euro Club Index as my second betting strategy.

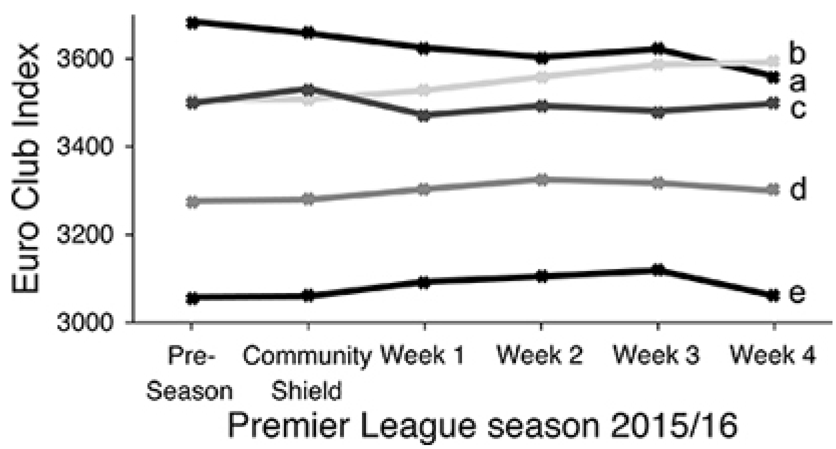



**Figure 12.2** Euro Club Index for the five highest-ranked clubs at the start of the 2015/16 season. Chelsea (a) dropped in the rankings, while Manchester City (b) rose. Arsenal (c), Manchester United (d) and Liverpool (e) underwent small changes.

## Strategy 3: Performance Indicators

Professional gamblers usually try to find indicators of team performance that other people have missed. We can all see how many goals a team has scored and how many matches it's won, but if a gambler can find an indicator that predicts future performance, and the rest of us have missed it, it gives them an edge. Gamblers tend to be rather secretive about what these indicators are and which ones predict match outcomes, so I'll try to find a few myself.

The performance indicator that we see most often on our TV screens is possession: the percentage of match time that each team has the ball. When our favourite team is 1–0 down but has more possession, it's tempting to think that they are on their way to equalising. Live betting sites show possession statistics and then offer long odds for a comeback. They want you to believe that more possession equates to better performance, and that you should back the team with the ball. Don't be fooled. Possession does not guarantee that a team is playing well. When two well-matched teams meet, the level of possession does not correlate with goalscoring rate.[11] If anything, statistics from recent Premier League seasons show that the team enjoying more possession is more likely to lose. This is because a team that falls behind wants to initiate attacks.[12] The opposition lets them have the ball, and waits for them to make a mistake. Being behind leads to more possession, and more possession does not necessarily lead to goals.

One of the other performance indicators provided by betting sites and shown on TV is number of shots, both on and off target. Shots are better predictors of outcome than possession, but they still tell only a small part of the match story. A speculative shot from way outside the penalty area is much less likely to go in than a well-controlled shot from in front of the goal, even if both shots are on target. To really know how well a team is creating goals we need to know where the shots were taken from.

It is straightforward to put together a model of where goals come from by breaking the pitch into a small number of zones and looking at the probability of a goal for shots from each of these zones. For example, we can define the goal area, the penalty area (minus the goal area) and outside the penalty area as three different zones. We can then count up how many efforts on goal were made from each of these zones, and how many of them went in. Taking a sample of shots and goals over previous seasons, the probabilities of a shot from these three zones going in were 32.2%, 12.4% and 3.4%, respectively.

This last statistic surprised me. Before I saw these figures, when I went to watch football I'd find myself standing up and shouting, 'Shoot!' as a striker neared the penalty area. Not any more. That 3.4% success rate is rather sobering: on average, it will take nearly 30 shots from outside the box to create a single goal. If the forward can force his way just a few more metres forward, his chance of scoring triples.

Finding the position of shots from previous matches is straightforward in the online version of *FourFourTwo*'s Statszone.[13] I just count the number of shots from different areas and weight them according to the goal probability. For example, in a home game against Arsenal in August 2015, Crystal Palace shot five times from outside the penalty area, five times from inside the penalty area but outside the goal area, and just once from inside the goal area. The number of goals you'd expect to be scored from these shots is

$$(0.034 \times 5) + (0.124 \times 5) + (0.322 \times 1) = 1.019$$

which is close to the one goal, a strike from outside the penalty area, that Crystal Palace scored in the match. Arsenal's expected number of goals calculated on this basis was also close to the tally of two that they scored. They had four shots from outside the penalty area, 13 from inside the penalty area but outside the goal area and just the one from inside the goal area, giving

$$(0.034 \times 4) + (0.124 \times 13) + (0.322 \times 1) = 2.066$$

expected goals. Arsenal are especially interesting, because it seems that Arsène Wenger has seen the shot-success statistics I've presented here. Instead of wasting chances, Arsenal nearly always try to get the ball into the box before shooting.

In their home match a week earlier, Arsenal lost 2–0 to West Ham. Arsenal's expected goals in that match had been

$$(0.034 \times 6) + (0.124 \times 15) + (0.322 \times 0) = 2.058$$

almost exactly the same as against Palace. West Ham's expected goals totalled only

$$(0.034 \times 2) + (0.124 \times 5) + (0.322 \times 0) = 0.686$$

It just happened that two of West Ham's seven shots went in. This match shows exactly why expected goals can be useful in gambling. Arsenal were unlucky against West Ham, but the next week against Crystal Palace their goals genuinely reflected the chances they had. West Ham had been lucky and went on to lose at home the next week to Leicester City. Expected goals give a much more useful performance indicator for future matches than actual goals scored.

Another important performance indicator is passing rate – the number of successful passes made per minute of possession. The more passes a team makes per minute, the more likely they are to score goals.[14] Moreover, teams that have made lots of successful passes in previous matches, whether they won those matches or not, are more likely to win their next match. It's not how much you have the ball that counts, but rather how you move it about.

The next step for me is to crunch numbers. To create a performance-indicator model, I took each team's expected goals and passing rates from previous matches and used them to predict scoring rates in subsequent matches. Using a technique called Poisson regression, I found that both expected goals and passing rates were better predictors of the rate at which teams score goals than previous results alone. This is exactly the requirement of a good performance-indicator model. The information most gamblers look at, previous results, is not as good at predicting outcomes as these performance indicators.

Table 12.1 gives passing rates and average expected goals, for and against, for all teams going into Week 4 of the 2015/16 Premier League season. The team's rankings are then calculated from these statistics, with higher passing rate and higher expected goals giving a higher ranking.[15]

There were several interesting matches in the week after these rankings were calculated. Two big-name teams, Chelsea and Manchester United, had games against lesser opposition with higher rankings. The big guys both lost. Chelsea, ranked 12th, were beaten at home by Crystal Palace, ranked 9th, and 5th-ranked Manchester United were beaten away by 3rd-ranked Swansea City. Placing bets according to this ranking in Week 4 of the season would have given profits of four times the investment. Palace were available at 11.00 to win on the Friday before the match, and Swansea at 3.60 Again, we can't read too much into one or two results, but there's enough here to suggest that performance indicators can be useful, and passing rate and expected goals make up my third strategy.

## Strategy 4: The Expert

With the three strategies I've outlined so far, I've tried to exploit biases in the way that gamblers assess football in order to beat the odds. I've identified statistical patterns – in betting odds, in long-term results and in performance indicators – that most people will probably have missed. But statistics don't tell us everything. Many things happen on the pitch that can't be captured by passing rate, expected goals and team rankings, and we expect footballing experts to pick up on these.

The expert I've decided to put my money on is Joe Prince-Wright, lead writer and editor for NBC sport's ProSoccerTalk. He lives and breathes the Premier League, sending news and reports to the US from London. On Twitter he comments on every important detail of every match, apparently not missing a single kick. He writes up to 20 articles a day for NBC, covering everything from the top tier to the lower reaches of the Football League, relating transfer rumours and injury news. As he works for a relatively small media outlet (in terms of football coverage), Prince-Wright has to check his own facts and make sure what he writes is correct. This could explain why, as we saw in the last chapter, he did so well in predicting the 2013/14 and 2014/15 Premier League seasons. When it comes to the Premier League, this man knows his stuff.

**Table 12.1** Ranking of teams going into Week 4 of the Premier League 2015/16, on the basis of passing rate and expected goals.

| *Ranking* | *League position (Week 3)* | *Team* | *Passing rate*[16] *(pass/min)* | *Expected goals for* | *Expected goals against* |
|---|---|---|---|---|---|
| 1 | 2 | Arsenal | 7.38 | 1.78 | 1.17 |
| 2 | 3 | Manchester City | 7.21 | 1.38 | 0.72 |
| 3 | 15 | Swansea City | 6.79 | 1.26 | 0.59 |
| 4 | 5 | Tottenham Hotspur | 6.55 | 1.31 | 1.14 |
| 5 | 4 | Manchester United | 7.40 | 1.11 | 0.77 |
| 6 | 7 | Liverpool | 6.52 | 1.27 | 1.09 |
| 7 | 16 | Norwich City | 5.70 | 1.42 | 0.72 |
| 8 | 12 | Southampton | 5.61 | 1.36 | 1.08 |
| 9 | 8 | Crystal Palace | 5.49 | 1.26 | 1.28 |
| 10 | 1 | Leicester City | 4.90 | 1.39 | 1.31 |
| 11 | 17 | Bournemouth | 5.99 | 1.12 | 0.97 |
| 12 | 14 | Chelsea | 6.45 | 0.99 | 1.16 |
| 13 | 9 | Watford | 5.49 | 1.17 | 1.07 |
| 14 | 10 | Stoke City | 6.19 | 0.95 | 1.29 |
| 15 | 13 | West Bromwich Albion | 5.91 | 0.97 | 1.22 |
| 16 | 11 | Everton | 6.39 | 0.85 | 1.19 |
| 17 | 6 | West Ham United | 5.59 | 0.96 | 1.45 |
| 18 | 18 | Newcastle United | 6.33 | 0.74 | 1.31 |
| 19 | 20 | Aston Villa | 6.13 | 0.69 | 1.12 |
| 20 | 19 | Sunderland | 5.33 | 0.75 | 1.43 |

And Prince-Wright isn't scared to forecast the future. As well as sticking his neck out with a full-season prediction, he makes his 'Premier League picks' each week. For each match in the week ahead he gives an exact scoreline, such as Aston Villa 0, Manchester United 2, or Swansea City 3, Newcastle United 0. He doesn't expect us to take these predictions literally, and qualifies each prediction with a 'Basically, free money', a 'So you're telling me there's a chance' or a 'Don't touch this' to show his level of confidence. But the scorelines do give an idea of who he thinks will win, and by how much.

To convert Prince-Wright's scorelines into the probabilistic predictions I need in order to compare to the bookmakers' odds, I create scoring rates. I think Prince-Wright sometimes exaggerates the number of goals a bit to make his point about the relative strengths of the two teams, so I adjust them to account for this. Let's take the prediction that Aston Villa will lose 0–2 to Manchester United. In reality, Villa do have some chance of scoring, so I set the expected goals to 0.5. Similarly, 2 goals for United is a high expected goals rate, so I mark their rate down to 1.5. In general, the equation for converting Prince-Wright goals $P$ into scoring rate $S$ is

$$S = P - \frac{P-1}{2}$$

This now gives me an expert-based strategy.

## Strategy 5: Ask My Wife

For help in solving life's problems, there is one person whose opinion I value over all others. From small difficulties, such as finding out where I left my keys, to large-scale project-planning, such as how we should raise our family, I trust in Lovisa Sumpter's judgement. Not only is she a fantastic wife, but she is also an associate professor of mathematics education and a qualified yoga instructor. So when it comes to working things out and taking a balanced approach to life, Lovisa knows what she is doing.

Lovisa has another talent. She is the only person I know who has won the football pools. Way before we met, when she was still a student, she correctly predicted the outcome of every one of the 13 matches in the Swedish football pools, Stryktipset. The chance of getting all the results right by picking at random is 1 in $3^{13}$ = 1/1,594,323. Unfortunately, it must been a slightly more predictable week than usual, and there were quite a few others who also got all 13. Lovisa shared the jackpot and received about £600. Not bad for a student, but not the world-cruise or house-buying jackpot that being the only winner would have provided. Lovisa has stopped following footballing form and playing the pools, but she remains proud of being one of the few people in Sweden to 'get 13 right'.

On this basis, I propose my fifth and final strategy: 'Ask my wife'. She will represent the typical punter.

## Putting It All Together

I now have five strategies, and there are three possible outcomes for every match. This is a problem. If the strategies don't agree, then I'll find myself betting on home win, away win and a draw for the same match. Given the bookmakers' built-in advantage, that's not a smart move.

The answer is to convert each strategy to outcome probabilities and then combine them to give an overall probability of win, lose and draw for each match. The odds-bias strategy already works on match probabilities, so these can be converted directly.[17] Likewise, the Euro Club Index already produces a probability for each outcome.[18] My performance-indicator and expert strategies both give scoring rates for both teams. For example, using Prince-Wright's picks I set Aston Villa's scoring rate to 0.5 goals per match and Manchester United's to 1.5. I then use these to run a Poisson simulation, just as I did in Chapter 1, with these scoring rates. For the Villa v United match this gives probabilities of 12.5% for a home win, 23.0% for a draw and 64.5% of an away win.[19] The performance-indicator strategy 3 also gives scoring rates that can be Poisson-simulated in the same way.[20]

My four betting strategies often make very different predictions. For Chelsea v West Bromwich Albion in Week 4 of the season, the odds-bias strategy suggests that Chelsea will win, while the performance-indicator strategy favours WBA. To reconcile these contradicting pieces of advice I take the average prediction of each of the four strategies and compare them with the bookmakers' odds by applying the $p>1/o$ test. I bet on a result only if it beats the odds. For Chelsea v West Brom, the probability of an away win from the performance indicators outweighs the odds-bias strategy, and my overall model favours backing West Brom.

Last but not least, there is 'ask my wife'. As Lovisa knows, my valuing her judgement is very different from my actually doing what she says – I'm far too stubborn for that. So she won't be allowed to have any input on my gambling. Instead, I'll withdraw £100 from our joint savings account, where I deposited the Bloomsbury cash, and give it to Lovisa to invest in her own gambling strategy in any way she wants to. Then I'll take out the remaining £402.10 and invest it in my other four strategies. And we'll see who does best.

CHAPTER THIRTEEN

# The Results Are In

If you invest yourself in a team, as a gambler, as a player, as a manager or as a fan, you put yourself into the hands of unpredictability. I'm used to drama from the trials and tribulations – and occasional triumphs – of Liverpool FC, and from watching my team of 10-year-olds in a penalty shoot-out for the district championships. But having a bet on every match over a weekend, and with my professional reputation at stake, means there's no respite. The goals are going in, the chances are being missed, and it's all being played out in front of me on TV or, if I'm out of the house, in live updates through the buzzing of my mobile phone. At times it looks like I'll win big; at other times it looks like I won't get a single correct prediction over the whole weekend. It's hard to remain rational when everything keeps changing.

## Lucky Luke and Calamity Jane

Imagine a Friday afternoon gambler – I'll call him Luke – who, starting with capital of £100, makes £5 bets on each of the 10 Premier League matches over five weekends. The left-hand graph in Figure 13.1 shows the progress of his gambling. It looks like he knows what he's doing, as after five weeks and 50 matches he has turned his £100 into nearly £250. Not only that, he also has a consistent record with reliable growth. Luke doesn't win every bet, but his capital steadily increases.

Now let's take another gambler, Jane, who has the same amount of money as Luke. She also places £5 bets. But things don't go as well for her – see the right-hand graph in Figure 13.1. She does OK for the first few matches, but from then on it's downhill. When Jane wins, she wins small, and when she starts to lose, she continues to lose, failing to win six or seven bets in a row. After 50 matches she has only about £20 left, one-fifth of her starting capital.

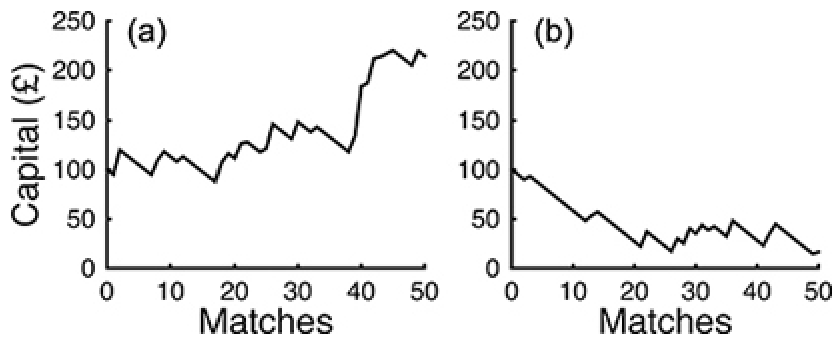


**Figure 13.1** Changes in capital for Luke (a) and Jane (b) in our gambling simulation.

Luke and Jane are computer simulations of two different models. In the first simulation, Luke places bets entirely at random. He picks home win, draw or away win with equal probability and then places his bet. The model assumes that the odds he is offered are a true reflection of the match-outcome probabilities, but the bookmakers have a small built-in advantage of 1.5%. But Luke is lucky: he continues to win even though he doesn't have a clue what he's doing. He is betting at random.

The Jane model is different. I've assumed she knows more than the bookies. Even accounting for the 1.5% advantage the bookmakers have, I've given Jane a 1% edge over them. But that edge isn't enough to escape randomness – calamity strikes, and she keeps losing. Jane is doing the right thing, but she can't get it to pay off before she loses her money.

Luke and Jane may be computer simulations, but they personify many real gambling stories. In the world of randomness, idiots can win and talented people can lose – and it can be difficult to separate the two. For simulated gamblers we have an advantage: we can run them many times and try to separate the luck from the skill. I simulated both Luke and Jane 10,000 times and recorded how much money they had left after 50 bets. Figure 13.2 shows the outcome. Each bar in the histograms indicates the proportion of simulations in which Luke or Jane had a particular amount of money. We can now see the difference between Luke and Jane. Around 8% of Lukes are bankrupt after 50 matches, while fewer than 5% of Janes run out of money. Overall, the Janes increase their capital and end up with more than £113 on average, while the Lukes finish with an average of £86.

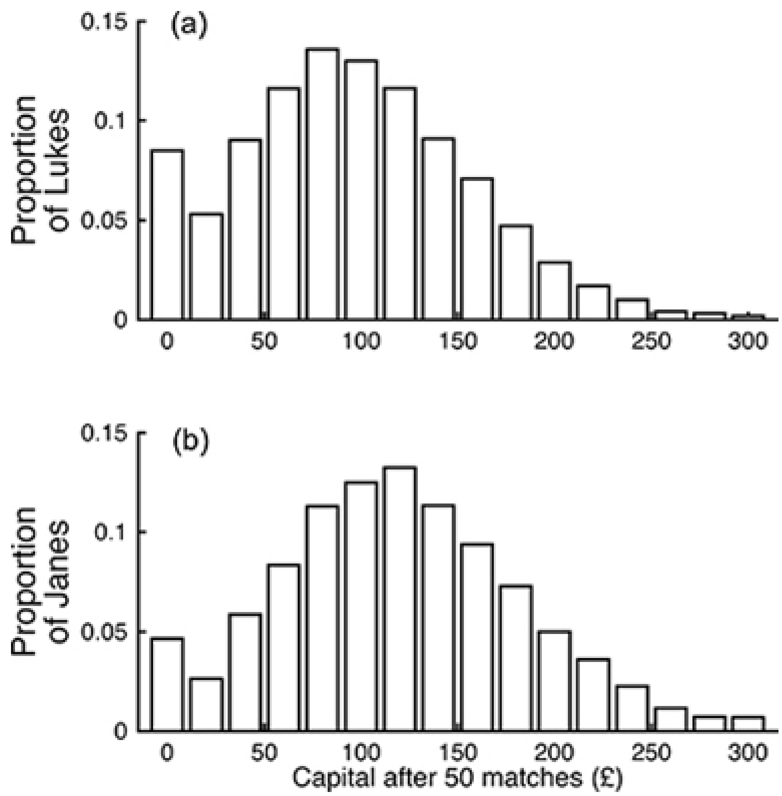



**Figure 13.2** Histogram of capital after 50 matches for Luke (a) and Jane (b) following 10,000 simulation runs.

In real life, of course, Luke and Jane don't get 10,000 chances to try out their betting strategies: they get one chance each. The simulations for Luke and Jane shown in Figure 13.1 may be extreme, but there is a reasonable chance that they would occur. The two distributions in Figure 13.2 are slightly different, but not that different. After 50 bets, Jane had more money than Luke in 62.6% of the simulations; in the other 37.4% it was Luke who had the most money in the end.

The problem for real-life Lukes and Janes is that no one really knows whether their strategy is best. All they can do is look at the results. After a profitable few weeks, maybe Luke starts boasting to his friends about his amazing new strategy. He imagines a pattern in his choices and visits tipping websites to tell others about his amazing new 'system'. Jane, though, becomes despondent. She thought she had a good solid analysis behind her strategy, but the results didn't go her way. She abandons it and tries something else. The world of gambling is full of Lukes and Janes, all fooled by short-term randomness.

## Betting Channels

It was with Luke and Jane in mind that I looked at how my four main strategy models had performed during the first four weeks of the 2015/16 Premier League season. Before I started placing bets, I looked at how a starting capital of £100 would have changed if I had bet according to each strategy each week – see Figure 13.3. The variation between the strategies is large. In the first week of the season, Prince-Wright got it right. My expert strategy based on his predictions suggested a big win away for Crystal Palace, a narrow win for Liverpool and a draw between Newcastle and Southampton, and all of these paid off. The performance-indicator strategy and the Euro Club Index both lost money. The odds-bias strategy, of backing strong favourites or even draws, made a few small winning bets. After a few weeks, however, Prince-Wright's luck had run out, but the performance-indicator strategy, which predicted Chelsea's decline, started to make a profit. The odds-bias strategy continued to make a steady return, while Euro Club Index suffered a steady erosion of its capital.

To construct Figure 13.3 I used a rule called the Kelly criterion to decide how big each bet should be. The Kelly criterion originated in a branch of mathematics called information theory. In the 1940s and '50s, engineers were trying to find ways of making the first digital communication channels more efficient. Two channels can be compared in terms of the probability of them transmitting reliable information. More accurate channels provide information faster than less accurate channels. John Kelly Jr drew a mathematical analogy between predicting sporting events and having a communication channel that transmits the result of the match before it has actually been played.[1] This channel may make errors, but if it's more accurate than the bookmakers' channel, then it's worth betting on. The more accurate your channel and the more money you have, the bigger the bets you should make.

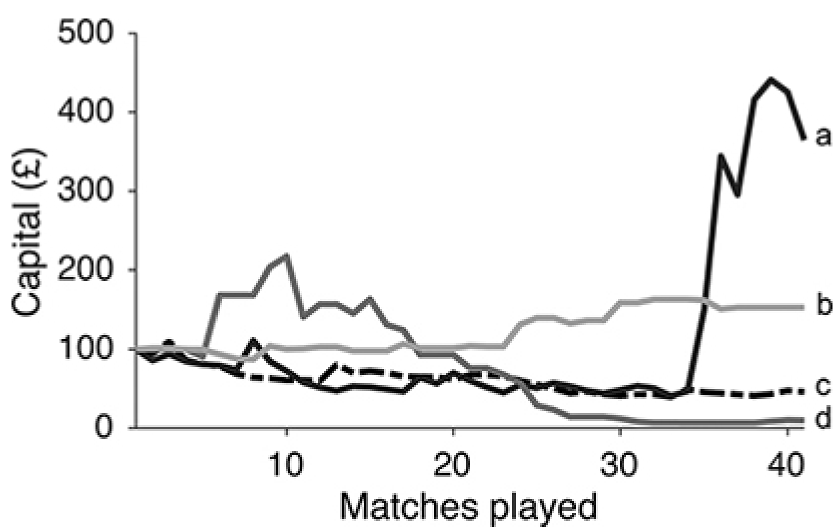



**Figure 13.3** How my four gambling strategies – performance indicator (a), odds bias (b), Euro Club Index (c) and expert (d) – would have performed for a starting capital of £100 over the first 40 matches of the 2015/16 Premier League season.

Based on this reasoning, Kelly came up with the following equation for the size of a bet:

$$\text{stake} = \frac{\text{expected winnings if model is correct}}{\text{(UK) odds offered}} \times \text{total capital}$$

These quantities are straightforward to work out. For example, in Week 1 of the Premiership, the expert strategy predicted that Crystal Palace would win 2–0 away at Norwich City. Using my Poisson model, this translated into a 64% probability that Palace would win. The UK odds for Palace to win away were around 2/1. Putting these numbers into Kelly's equation, we get

$$\text{stake} = \frac{0.64 \times 2 - 0.36}{2} \times \text{total capital} = 0.46 \times \text{total capital}$$

The Kelly criterion suggests putting a stake of 46% of the total available capital on Palace to win. This is a large bet that would have paid off. Palace won 3–1, and in Figure 13.3 we can see that the expert made a large profit on match 6.

The expert and performance-indicator strategies lead to wildly fluctuating results because they make predictions that are very different from the bookmakers' odds. When these strategies get it right, they win big, but when they go wrong they lose big, too. The Euro Club Index and the odds-bias strategies make predictions that are quite similar to the bookmakers' odds, with the former making a small but steady loss and the latter making a small but steady profit. The results so far are useful, but we should remember the story of Luke and Jane. There's no guarantee that any of these trends will continue once I start betting in the real world.

## Week 1: Bore Draws

The Friday afternoon before the first week of betting, I was nervous. Now it was all for real. I typed in the bookmakers' odds, last week's shot statistics, the Prince-Wright predictions and the Euro Match odds. To decide which bets to place, I weighted each model based on previous performance. I gave the expert strategy and the Euro Club Index, both of which had lost money in previous weeks, less influence in my combined model than the performance-indicator and odds-bias strategies. Everything was ready to go. I pressed 'run'.

My screen filled with a garble of text and numbers as it churned through the calculations. After half a minute the output table appeared: a list of matches, odds, bookmakers with the best odds and the stake size suggested by Kelly's criterion. I read them and was immediately disappointed: 8 out of 10 bets it suggested were on draws. Draws! I was looking forward to supporting different teams, cheering them on as they fought to victory and justified my strategies. But draws? How can I watch a football match and cheer on a draw?

I double-checked my code, but the result was right. It wasn't that the strategies predicted that draws were particularly likely this week, it was simply that odds for draws were particularly good. I had chosen to start my betting after a two-week break in the Premier League programme for international football, and the punters believed that the teams would return looking for results.

The punters got it right, and my strategies got it wrong. There was only one draw out of 10 matches. One of my predicted away wins, Spurs away to Sunderland, also paid off, but overall I lost £18.06 of the £48.75 I staked. Things would have been very different had Kelechi Iheanacho not come on as a late sub for Manchester City in their match against Crystal Place, and scored the only goal of the game in the 90th minute. But that's the nature of randomness.

Lovisa also made a small loss on her first week of betting, but she did win with the £10 bet she put on Manchester United to either win or draw against Liverpool. That bet was probably made just to wind me up, but in the end it didn't make much difference. Watching the Liverpool defence concede a penalty and then allow United to score with both the chances they got inside the box was much more depressing than the thought of Lovisa moving slightly ahead of me.

In this first week, the expert strategy performed just as badly as it had in the test period. Prince-Wright's tips were fun, but they weren't correct. But looked at in the context of previous research, this isn't surprising. Several studies have tested tipsters' ability to provide betting advice on individual football matches, and the tipsters haven't fared well. Experts are typically beaten by the bookmakers' odds.[2] In one study of the 2002 World Cup, experts' predictions were found to be no better than those made by people with no knowledge of the game.[3] So Prince-Wright has nothing to be ashamed of, but for the second week I dropped him from my overall model.

## Week 2: Indexical Decline

With the expert gone, the betting went better. The model backed six draws, three away wins and one home win. It told me to stake £27.88 over the ten matches, and I won £1.93. In footballing terms, this week was business as usual. Draws in the Midlands between Villa and West Brom, and Stoke and Leicester, combined with an away win for Watford at Newcastle proved sufficient for me to turn a small profit.

I now decided to take a closer look at the Euro Club Index. If I had used only this strategy from the start of the season, then after the 60 matches that had now been played my £100 would have become £33.21. My biggest concern about Euro Club was that its decline was so steady. Even when I used four different bookmakers and chose those offering the best odds, the bookies still had an advantage of around 1.9% per bet. Figure 13.4 compares the decline of the Euro Club Index with a benchmark decline based on the bookmakers' advantage with each bet. The trend in both cases is very similar. The Euro Club Index does a reasonable job of predicting football matches. The problem is that the bookmakers also do a decent job. Slowly but surely, the bookmakers' advantage was eating away the Euro Club's capital.

The Euro Club Index is a variation of a system known as Elo after its creator, physics professor Arpad Elo. It was originally developed for chess rankings, but it has since been used to rank everything from baseball and American football to Pokémon and *World of Warcraft*. It has been popularised and widely applied by Nate Silver and his team at FiveThirtyEight, who produce Elo rankings for the NFL.[4] Here, as with the Euro Club Index, winning teams gain points from the teams they beat, and their index improves over the season. Each week, FiveThirtyEight publish match predictions and project forward through the season to work out the chances of teams progressing to the play-offs and the Superbowl.

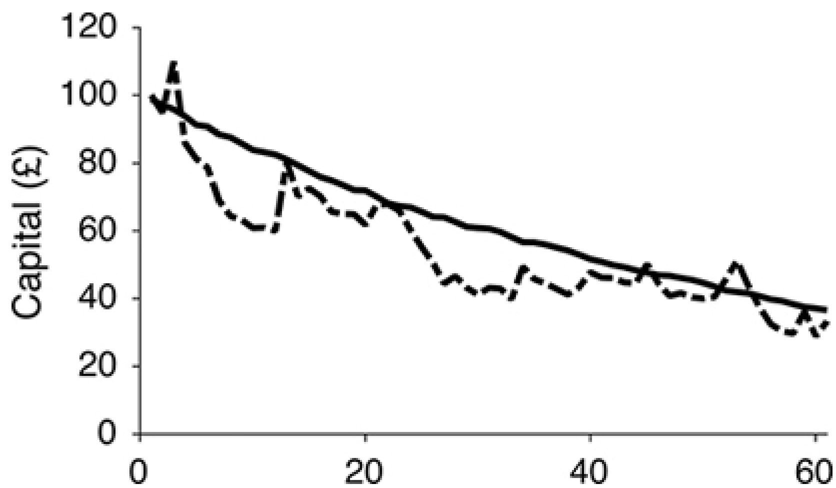



**Figure 13.4** How the Euro Club Index strategy (dashes) performed compared to a benchmark of random bets (solid) with £100 initial capital over the first 60 matches of the 2015/16 Premier League.

The Elo system is dependable. Grandmasters, quarterbacks and guild masters with high rankings are invariably those at the top of their game. But when it comes to betting, dependable isn't good enough. Elo lacks that all-important edge. Norwegian researchers Lars Magnus Hvattum and Halvard Arntzen implemented an Elo system and tested it over eight full seasons of football results across all four English divisions.[5] Testing their model across 16,288 matches was much more thorough than my evaluation of the Euro Club Index over 60 matches, but the results were the same. Their Elo system lost value at a rate that reflected the bookmakers' advantage. The Elo could track changes in team performance, but it couldn't make money.

There are variations of Elo that try to improve its predictive power. Nate Silver and his team have implemented something they call a Soccer Power Index for international football matches.[6] An important difference between this and the standard Elo is that the

Soccer Power Index incorporates information about team line-ups and the significance of each match when deciding how to exchange Elo points between teams.

To measure how well the Soccer Power Index performs, I downloaded FiveThirtyEight's predictions for the group stage of the 2014 World Cup and compared them with bookmakers' odds going into these matches. Betting on the 48 group-stage matches using the Soccer Power Index Kelly's criterion with a starting capital of $100 would have given over $400 in profit. Following FiveThirtyEight's predictions for the 2015 Women's World Cup would also have yielded a profit. Taking the $500 from the men's event and investing it in the women's would have produced a total profit of around $1,400 over the two tournaments. Not a bad result.

Despite this World Cup success, Nate advises punters not to bet on the basis of their weekly American Football Elo. Historical testing shows that it doesn't make money on the Vegas spread market.[7] Nate describes the NFL Elo as doing a 'pretty good job of accounting for the basic stuff'. It is reliable, but not exceptional. I have noticed that people tend to get a bit emotional about the World Cup, patriotically backing their own country despite all evidence to the contrary. There could be an opportunity here for dependable, unemotional indices to make money. When we return to the everyday business of gambling on domestic football matches, common sense kicks in and the bookmakers' odds are just as good a set of predictors as the indices.

I don't believe that index-based gambling can produce long-term profits in the Premier League. With so many Elo variations, including Glicko-2, TrueSkill and Power Ratings, I can't rule out the possibility that some variation of this theme can be used as a tool for making money. But for every gambling book or website claiming to have developed an indexing or ranking system that works, there's almost always a blog post showing that it doesn't.

The fundamental reason for my scepticism of betting using indices goes deeper than statistical tests. The fact is that many gamblers and bookmakers behave in a way that is very similar to how these indices are built. They look at previous results, try to work out a team's 'form', and predict the likelihood of results based on this form. This was what I could see Lovisa doing on Friday evenings. She studied the league tables, the run of results and the odds, and then tried to find a good bet. With tens of thousands of punters doing the same thing, the odds reflect a massive human-built index of performance.

I would love to see Nate Silver, Euro Club or someone else prove me wrong about the Premier League and produce an index that outperforms the odds. But until they do, I've dropped indices from my betting model. Euro Club can't compete with this wise crowd and the bookmakers who already fully understand the method. 'Dependable' isn't good enough when it's my own money that's at stake.

## Week 3: Unexpected Goals

It was Week 3 of my gambling and Week 7 of the Premier League season, and it was crazy. Spurs beat league leaders Manchester City 4–1. Arsenal beat Leicester 5–2 in a match where neither team seemed interested in defending. Late goals for Chelsea saw them recover from two down against Newcastle to draw 2–2, Everton made a three-goal comeback against West Brom to win 3–2, and West Ham equalised against Norwich in the 92nd minute for a 2–2 draw. These late changes in scoreline were all reversals on my betting positions, and they cost me. But thanks to a large stake on Spurs I broke even, earning a total of 33 pence on a £31.25 investment.

During this week there were 41 goals in 10 matches, compared with the usual average of around 2.7 per game. The chances created were about the same as usual, and the expected goals for this week were no different from previous weeks. It just happened that more shots went in.

This type of shot–goal randomness is a feature of all football. In their Women's World Cup semi-final against Japan in 2015, England controlled the game defensively, with Japan getting nowhere near the English goal. Expected-goals expert Michael Caley calculated that Japan had just 0.1 expected goals in open play. But Japan scored through a penalty, awarded for a tackle outside the penalty area, and then again with a last-minute freak own goal when Laura Bassett spectacularly lobbed her own keeper from the edge of the box. England's only goal also came from a penalty, so the game had no 'proper' goals. The 2–1 scoreline bore almost no relationship to the 0.1 v 0.7 expected goals created during the match.

Unfortunately for England, and all the other teams who have experienced a cruel departure from a knockout tournament, expected goals do not win matches. It can take time for some teams to transform scoring opportunities into goals, and other teams can get lucky with the small number of chances they have. Michael illustrated this point nicely with a study of Arsenal. In November 2014, Arsenal's results made them look bad. They lost 1–2 away at Swansea, 1–2 at home to Manchester United and managed just a 1–0 win away at West Brom. But their expected-goals map, shown in Figure 13.5, tells a very different story. Each square on the right represents one of Arsenal's chances during November, while those on the left are the chances created by the opposition. The bigger the square, the better the chance. Arsenal had more opportunities than the teams they played against, and Michael's expected goals were 5.3 in their favour. But the opposition converted their chances, leaving Arsenal with 3 goals for and 4 against.

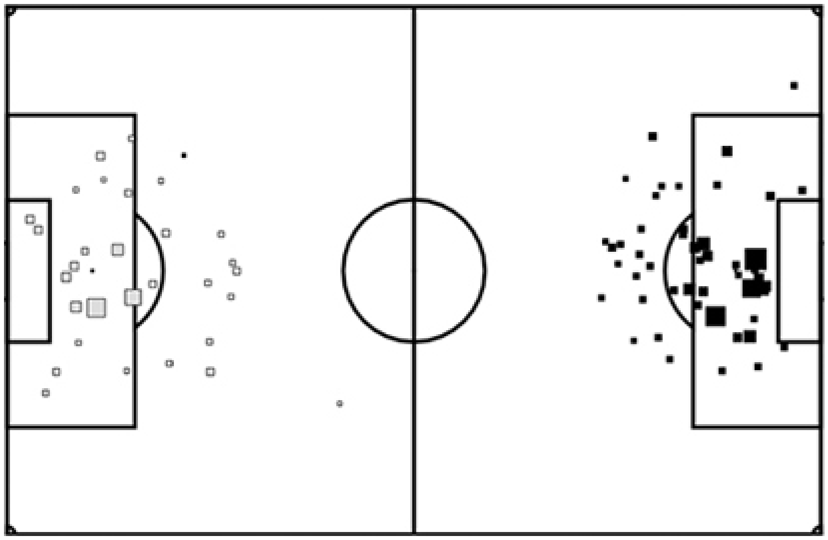


**Figure 13.5**. Expected goals for Arsenal in three Premier League matches during November 2014. The area of the square is the probability that a shot from that position would typically be a goal. Arsenal are attacking left to right, so their goal chances are shown on the right-hand side of the pitch as black squares; their opponents (Swansea, Manchester United and West Bromwich Albion) are on the left as grey squares. Calculations made and figure created by Michael Caley. Reproduced with permission. Data provided by Opta.

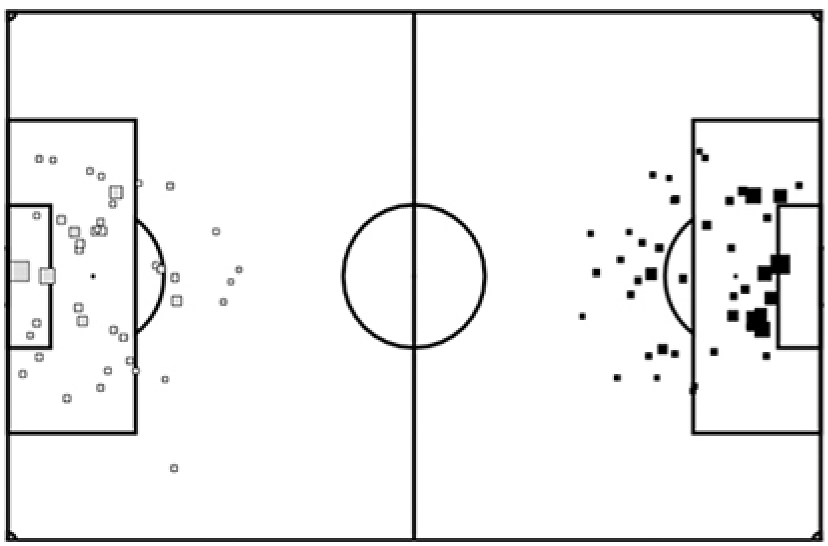


**Figure 13.6** Expected goals for Arsenal (black squares) in three matches during February and March 2015 against Leicester City, Crystal Palace and Everton (grey squares). For further details see Figure 13.5. Data provided by Opta.

From his expected-goals plots, Michael was confident that Arsenal would recover. He was right. Figure 13.6 is an expected-goals map for Arsenal's three straight wins against Leicester, Crystal Palace and Everton in February and March. The expected goals for these matches were fairly similar to those in November, but Arsenal were now converting their chances; they scored six and conceded two in the three matches. The expected goals started to reflect how the team were attacking and defending.

The question I was asking myself after the 41-goal weekend was when my performance-indicator strategy, based on expected goals, would come good. My model is simpler than Michael's. I use just three zones to define the quality of a chance, while

Michael uses a full set of Opta data. He includes the distance and angle from goal at the point of shooting, along with other details such as whether the shot was a header or a kick. These factors determine the size of the expected-goal squares in Figures 13.5 and 13.6. Michael, and many of the wide community of writers and bloggers working with expected goals, are not particularly interested in making money from gambling: they are using expected goals to understand more about football tactics. The Arsenal example is a perfect case in point.

After seeing the wide fluctuations in the outcome of bets based on expected-goal predictions, I agree with Michael. Expected goals, like the tactical maps I made in Chapter 7, are a way of summarising a match or a number of matches in a single picture. But I was becoming less and less convinced that it could be used to make money. From Week 4, I dropped the performance-indicator strategy from my combined model.

## Weeks 4 and 5: Satisfied with a Draw

Of my four mathematical betting strategies, there was only one winner – and it was the simplest. Over nine weeks, the odds-bias strategy of backing draws between well-matched teams and wins for strong favourites outperformed the other three. Figure 13.7 compares how the four models fared over 90 matches. If I had used only the Euro Club Index or the expert strategy, I would have lost most of my money: they would have turned my starting capital of £100 into £14 and £3 respectively. The performance-indicator strategy generated a lot more variation. On its own it would have made over £300 up to match 40, but then lost it all during the last 50 matches. Over the entire 90 matches it broke even. In contrast, the odds-bias strategy by itself would have produced steady returns over the 90 matches, eventually turning £100 into £240.

From Week 4 onwards, I bet only on the basis of the odds-bias strategy. In Week 4 I won £17.65 from evenly balanced matches between Everton and Liverpool, Bournemouth and Watford, and Swansea and Spurs that all ended in the predicted draws. In Week 5, I made a further £14.49. Wins for big favourites Chelsea, Manchester City and Arsenal, as well as a draw between Liverpool and Spurs, all paid off. The odds-bias strategy was making a reliable profit.

After 90 matches I could conclude with a reasonable degree of statistical certainty that the odds-bias model was profitable.[8] It consistently outperformed the bookmakers' odds. However, all betting advice comes with a word of warning. Even if the odds-bias strategy is making me a profit at time of writing, it can't be considered failsafe. If draws are undervalued on betting markets, then other gamblers can also find that out. If, for example, a respected mathematician writes a book claiming that backing draws between well-matched teams is a good strategy, then the odds can change. The bias will disappear, and the strategy won't work anymore. Likewise, it may just be that draws have been unusually common recently. As Jimmy Hill taught us in Chapter 6, under a three-point system, equally matched teams should risk everything to win. With a little bit of rational thinking by managers, draws may well decrease again.

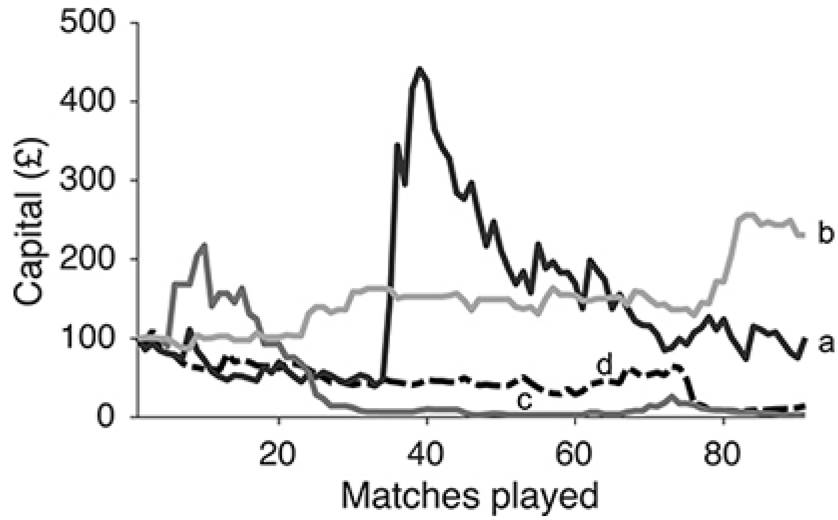



**Figure 13.7** How my four gambling strategies – performance indicator (a), odds bias (b), Euro Club Index (c) and expert (d) – performed for a starting capital of £100 over the first 90 matches of the Premier League 2015/16 season. I bet on the basis of a combination of these four models on matches 41–70, placing more money on the strategies that earned the most. From match 71 onwards I bet solely using the odds-bias model.

There are no universal mathematical truths about punter behaviour and bookmakers' odds. Academic studies of biases in betting markets have most often shown that favourites give the best odds, as I have found here. However, studies of other markets have found it better to back long shots. With the popularity of Asian handicaps and spread betting, it is possible that the market for draws is currently underexploited. But this could change. The only mathematical truth about gambling is that a thorough statistical study of the league, the country and the sport should be done in advance of placing bets.

I was feeling pretty pleased with myself at the end of the five weeks of betting. In spite of the bookmakers' advantage, I'd made a profit and I'd identified a statistically reliable strategy for future gambling. I hadn't been paying too much attention to Lovisa's betting over the last few weeks. I knew that she'd lost during the first few weeks, and I guessed that by now she'd be on a par with Euro Club Index or Prince-Wright. Without wanting to sound too condescending, I asked her if she could send me her list of bets so I could total them up and document her losses. She gave me a little smile, opened up the bookmaker's webpage and showed me the balance. Lovisa had won! She started with £100, and 18 bets later she now had £116.97. She had beaten the bookies.

Lovisa placed a smaller number of bets than I did, so statistically speaking she could still have been just lucky. But she'd done a lot better than I'd expected. It could well be the case that I am living with a person with a true talent for spotting winners. I will never know, because Lovisa was satisfied. She closed her account and took out her winnings. I didn't complain. It had been fun spending our Friday nights discussing current form, but it was also nice to go back to our usual routine. At the end of the next week we invested her winnings in a good bottle of wine and a film, and had a football-free Friday instead.

## Career Advice

I worked for about a month to develop my model. Then, in preparation for each weekend's betting, I spent a good part of my Fridays collecting data to feed into it. Building a betting model takes time and energy, and then there is the stress of waiting for the results to come in. I continued to bet on my model until the first week of December 2015. Of the £400 I initially set aside for gambling I had risked £388.76 and won a total of £108.30: a return of 27% over eight weeks. If I had continued to bet and make the same percentage profit, then I would have earned £325 by the end of the season. That's a potential 81% yearly return on my investment.

Before you mortgage your house to raise capital for a new career as a mathematical gambler, I should make it clear that there are no shortcuts to a fortune in gambling. And there are no guarantees. You need a large starting capital, a lot of patience, sound mathematical know-how and a cool head. If you do start to make a regular profit, you'll soon discover that bookmakers use algorithms to analyse the records they keep of their clients' activity and to detect consistent winners. If they think you're doing too well, they will thank you for your business, pay out your winnings and cancel your account. So you also need a large supply of credit cards with different names on them. And there aren't too many legal ways to arrange that.

My friend in the betting industry who helped me get going with my model, Robin Jakobsson, came into the business via this route. He started his working life as an academic, as a PhD student at a statistics department. Then he had the idea of applying his statistical approach to look at horse-racing odds, and he found consistent biases in backing favourites. He published an article identifying the biases and started to make money backing his predictions.[9] When the bookmakers finally closed his account, Robin realised he was on the wrong side of the business. He never finished his PhD, and instead started working for his former adversaries.

Robin's story isn't uncommon for talented mathematicians. Probably the highest-profile statistician and football-fan-turned-professional-gambler is Matthew Benham. He started his working life in the financial markets, but realised that he could apply what he was good at to football. In 2004 he set up the company Smartodds, which provides statistical information for gamblers. From there, he went on to set up Matchbook, a website for trading bets in real time. His fortune made, Benham invested his money in his boyhood club, Brentford, and Danish club FC Midtjylland. As I described in Chapter 4, he applied his analytical approach to his clubs. It worked. Brentford returned to the Championship for the first time in 20 years, and Midtjylland won the Danish Superliga.

So my advice for anyone with the requisite maths skills who would like to earn money working with football is this: join the game instead of trying to beat the odds. Find a job working for a bookmaker calculating win probabilities or creating novel betting markets. Or better still, take your mathematical skills to a football club, a player academy, a sports-analysis company or the media. Maths is a central part of football, and there are lots of opportunities for footballing stattos. Building tactical maps, calculating expected goals, devising fitness programmes, simulating the physics of shots, developing methods for tracking the movement of players on the pitch, preparing network-passing graphics for TV, helping the manager write out Attack/Defend game tables to outwit the opposition, and optimising the team's geometry – these are all potential jobs for Soccermatics nerds like us. Think about it: solving math problems, enjoying football and getting paid for it. And after doing exactly that for the last year, I can say with 110% certainty that there's no better way to earn a living.

# Part IV
# The Analysts

CHAPTER FOURTEEN

# Finding the Talent

Around the time that the first edition of *Soccermatics* came out, a revolution was taking place in football. Using match data made available online, amateur analysts were filling Twitter and blog posts with their own statistical and mathematical analyses of players and matches. Amateurs were starting to do what the professionals should have been doing years ago – analysing the game through passing networks, shot statistics and tactical maps.

Clubs are starting to take notice. In 2012, Arsenal bought performance analysis company StatDNA, and they are now using its services throughout their football operations.[1] During their league winning 2015–16 season, Leicester City's first-team analysts used Opta data to help inform post-match briefings given by the manager.[2] Data was also used by managers to argue their case: Manchester United manager Louis van Gaal waved around a Prozone printout at a press conference to prove his team weren't playing 'long ball' football.[3] And statistics is entering the boardrooms; the author of *The Numbers Game*, Chris Anderson, was appointed managing director of Coventry City. These clubs joined Manchester City, Bayern Munich and Liverpool in putting maths into their game.

The online data revolution certainly has its opponents, though. During the summer of 2015, Aston Villa sold or let go three of their best players: Christian Benteke, Fabian Delph and Tom Cleverley. They brought in 13 new players from across Europe, many of them on the basis of statistical recommendations from the recruitment department. The strategy failed catastrophically. They plunged to the bottom of the league en route to a hapless relegation, and manager Tim Sherwood was fired at the start of November.

Sherwood blamed Villa's use of numbers for his sacking, 'All that data analysis can be used for something, but it can't be used to pick your players', he told the *Telegraph*. 'Some of the data is not about goals, or assists, it's about "expected goals" when a player got himself in position to score, but didn't. What a load of nonsense.'[4] This was a damming assessment of a method widely applied both within clubs and by amateur analysts.[5]

Sherwood's resistance is futile, and the revolution cannot be stopped. I have witnessed it first-hand. After I finished writing *Soccermatics*, I was invited to write a column for the FourFourTwo website, using the tools I had developed to build tactical maps for analysing recent matches. I also started meeting with leading clubs, and talking to them about how they could better use data. Through Twitter I got to know a large number of bloggers and amateur analysts who were analysing the game using data. They were happy to share their experiences, and give feedback on different analysis techniques. And during the 2015–16 season, I saw how one by one the amateur enthusiasts, often younger than the players they were analysing, found themselves jobs at clubs and consultancies. The professional football world wanted data experts, and it found many of them on the internet.

## Start writing!

Omar Chaudhuri, head of Football Intelligence at consultancy 21st Club, gave me his advice for getting into the football business. ‘Start writing! My blog was effectively my CV, and without it I would not have even gotten a meeting in the first instance, let alone a job.’

Omar started blogging as a student in 2011, because he was frustrated by the lack of intelligent analysis of football in the media. His blog helped him get an internship working with Ian Graham, who later went on to become Liverpool’s director of research, and he then became the first data scientist at Prozone. A few years later he had moved to 21st Club, and has since worked with Arsenal, Spurs, Southampton, Crystal Palace and Everton, as well as clubs in Norway, the Netherlands and the United States. Omar’s journey from student dormitory to Premier League boardrooms took less than three years.

Omar believes that the best way to make a real difference to how a football club makes its decisions is through the boardroom. He told me that ‘the problem with working with managers, coaches and players is that, because tenures are so short, they are always focused on the next game.’ For the board, who have a long-term investment in a club, it is important that they plan on the basis of facts. Omar works as a kind of boardroom mythbuster. There are a lot of myths in football. His job is to find out which of these hold up to statistical and logical analysis.

Omar has proved a lot of myths to be wrong or misleading. ‘Teams hoping to get promoted from the Championship should employ players with experience of playing in that league’ is not true if you look at the numbers. ‘Managers should get their teams to play more aggressively’ is an unquantifiable and meaningless statement. ‘The wage bill in the Premier League team determines results’ is only true for the gap between the ‘big six’ and the rest of the league. Teams outside the big six can do well on smaller budgets. ‘The success of Spain and Barcelona means that clubs should look for shorter players’ is not established by research. Instead, there is a risk that teams get caught up in a rush to follow fads. Myth after myth fails when Omar starts to do his statistical checks.

Another common myth is that players are worth their transfer fee in terms of footballing ability. While the best players do cost more money, the price paid is often more complicated than a direct relation to ability. When Omar and I spoke, Paul Pogba had just joined Manchester United for a world record £89 million. Omar told me 'Pogba cost Manchester United a lot of money, but they also have the highest revenues in world football, have enormous fan expectations, and were in need of a central midfielder. Add all those things up and it probably does make sense for United to spend that amount. For other clubs it's not always so cut-and-dried.'

Talk to a few kids in your local park and it is easy to understand Omar's point. Every one of the young players I train have adapted their own variation of Pogba's Dab goal celebration and are desperate to 'be' him in the new edition of FIFA 17. Pogba is an excellent player, but like Gareth Bale, Cristiano Ronaldo and David Beckham before him, he is also a marketable product. Omar's job is to warn other clubs about the danger of getting carried away, because prices for the top players are increasing exponentially. A club's transfer policy should look to give long-term success, and not follow the headline trends.

## Row A; N'Golo Kanté

In the 2014–15 season, N'Golo Kanté topped the table of successful tackles per game for the whole of Europe. He was then playing for French Ligue 1 side Caen.[6] Sending a scout to check him out was a no-brainer, and lots of clubs did exactly that. It was Steve Walsh at Leicester City who was the most insistent with his manager about signing Kanté. Leicester bought him for a reported £5.6 million, won the Premier League, and sold him on to Chelsea for £32 million the next year. The tackle stat was simple, but it didn't lie.

One analyst I talked to described this as 'spreadsheet scouting'. The scouts have large spreadsheets where the columns include tackles made, interceptions, passes and dribbles. They sort the spreadsheets by the column they are most interested in. Row A shows the best player, Row B the second best, and so on. It is from there that they start their search. When Steve Walsh took over as director of football at Everton he signed Row B on his spreadsheet – Idrissa Gueye from Aston Villa. Like Kanté, Gueye also came from France to the Premiership in 2015, and during his first season ranked second only to Kanté in tackles and interceptions.

Teams don't rely solely on spreadsheets to buy players. All professional scouts agree that it is important to see a player in action before signing them. Footage of matches is valuable, but there is no substitute to sitting at the edge of the pitch and watching how a player reacts to the movement of play around him, and how he responds to teammates. A proper evaluation involves visits to several matches and, if possible, the opportunity to speak to the player and watch him train.

The real question for a club is about getting a good balance between initial statistical screening, the use of scouting networks, video analyses, watching matches live, and using data to double check decision-making.

Accounting for all of these inputs makes the job of picking players difficult. The difference in outcome for Aston Villa and Leicester City during the 2015–16 season is a perfect illustration of this point. Both teams followed the same principles of initiating a search using statistics. Both of them found undervalued players in France during 2014–15. Leicester signed Kanté and Riyad Mahrez and won the Premier League for the first time in their history. Villa signed Gueye and three other players from the French league, but were relegated with their lowest-ever points total.

The biggest difference between Leicester and Villa appears to lie in the mutual trust between the coaching staff and the analysts. While Leicester had found a way to integrate statistics into all aspects of their operations, the recruitment staff of Aston Villa and manager Tim Sherwood could not agree on how the stats should be used.

When I spoke to Rory Campbell, technical scout and analyst at West Ham, he emphasised the need for a club-wide analytics-driven strategy that combines all aspects of player assessment: from statistical analysis, through an understanding of personality and attitude, to seeing the players' strengths and weaknesses on the pitch. Rory's own background personifies this combination. As a schoolboy, he played for Arsenal until he was 16. He then went to Oxford University where he studied philosophy, politics and economics during the day and was a successful high-stakes poker player by night. After university, he took his coaching badges, working his way up from training his old secondary school team through jobs at Barnet and Cardiff.

These experiences have allowed Rory to understand how to combine football, analytics and psychology in decision-making. 'Whether its poker, economics, betting or football, it's about using all the information required to make correct decisions under pressure,' he told me.

Graduates from top-flight universities are becoming more common as club employees. Henry Newman studied philosophy and economics at the London School of Economics. He used spare time during his studies to learn everything he could about football, shadowing managers and taking all of his coaching qualifications. Since then Henry has worked in first-team and academy roles for Barnet, Charlton, Brentford and West Ham. He told me 'I can't say I have taken anything directly from my academic education and applied it in football, but it has shaped everything about how I approach the game.'

Rory and Henry have brought their analytical way of thinking to the clubs they have worked for. But it is their experience in all aspects of the game that allows them to apply this way of thinking.

Rory believes that many clubs still have a long way to go before they are properly exploiting data. 'The whole process of recruitment is the wrong way round at some clubs,' he told me. "Recruitment is often initiated by an agent and the clubs are then simply reacting to that. Of course agents are required, but clubs should work internally to identify players.'

Far too often, agents approach the club with names of players who could be interested in joining them, and then there is a discussion between data analysts, video analysts, trainers, the manager and the chairman about the qualities of the proposed player. At West Ham, Rory wants to be proactive instead. The club should decide on a style of play, and then set up their scouting system around that long-term strategy. To achieve this, Rory is developing algorithms that automatically identify candidate players who fit the club's long-term goals. By discussing the results of this analysis with scouts and agents, in footballing language, he expects a consensus to build up around players that is based both on statistics and conventional methods.

In Major League Soccer (MLS) in the United States and Canada, the process of integrating analytics into scouting has gone further than in Europe. MLS teams tend to be open to new ideas, especially if they come from other major American sports. Atlanta United are an interesting case study, because they have been built from scratch over the last five years, with the aim of starting playing in the MLS during the 2017 season.

Lucy Rushton, Head of Technical Recruitment at Atlanta, has used statistical analysis throughout the process of building up the team.[7] The coaching staff decides together what type of players they would like to recruit. Sometimes the process will start with Lucy performing a statistical search through the data for a player that matches requirements, and then a scout will be sent to look at the player. In other cases, a scout will have seen a player and Lucy will create a video montage of all of the actions performed by that player in a specific situation. At Atlanta, and many other MLS clubs, there is no division between stats and scouting.

## Mathball

When I talked to Premier League club analysts and scouts, they often mentioned the book and film *Moneyball*; the story of how Billy Beane used metrics to lead a baseball team, Oakland Athletics, to unexpected success. The Kanté story shows that a *Moneyball*-like approach might help find hidden talent in football, just as it did in baseball a decade earlier.

There are, however, limitations to statistics. Football consultant and founder of the Statsbomb website, Ted Knutson, is careful to qualify what stats can and can't do. He has developed 'player radars' that display various statistics in a format that makes them easier to take in. The number of successful dribbles, assists, number of times dispossessed, shots and goals are among the stats used to assess strikers. For assessing defenders the radars display tackles, long balls, blocks, aerial wins and interceptions. The radar gives an overall impression of the type of actions a player does most successfully.

Instead of hyping up his radar system as a solution to all scouting problems, Ted warns that 'the only thing [radars] represent is statistical output.'[8] These statistics change if players play in different leagues, for different teams or in different positions. We should also expect the statistics to change as players get older. Ted writes that 'Like any tool, [player radars] have strengths and weaknesses. In general, I have found it much easier to evaluate players with [statistical] information than without it.'

To illustrate his point, Ted looked in more depth at Kanté's 'Row A' interception statistic.[9] Leicester's style of play during 2015–16 meant they had less possession than other teams. Their plan was to let the opposition have the ball as long as they didn't come too near to their goal, then quickly recover it and counter-attack. So while Kanté topped the Premier League list of interceptions, when Ted adjusted the interception statistic to account for the time that Leicester were in possession of the ball, the number dropped. One of the reasons Kanté looked so exceptional was that he was playing in a team whose strategy was built on interceptions and counter-attack.

Football will never be *Moneyball*. Baseball is the sport of player statistics: singles, doubles and triples; runs batted in and home runs; run averages and strikeouts. Football is the sport of team patterns: triangles, strategy, passing networks, defensive synchronisation and a team becoming more than the sum of its parts. Simple stats such as number of passes made, tackles or pass success rate can't give a complete assessment of the effectiveness of a player.[10] A player who tackles a lot might do so because he is badly positioned in the first place, and a player with a high pass success rate might always choose the easy option.

Some side-of-the-pitch scouts argue that these problems are insurmountable. They don't describe the problems in exactly the same language I use above, and their criticism tends to be a bit more direct, but the sentiment is the same: stats don't give a proper picture of the game of football. I am more optimistic. As we start to get a grip of the mathematics behind football, we will be able to create models and build tactical maps that account for the interactions.

Currently, the model most widely adopted inside clubs is 'expected goals', the concept that Tim Sherwood despised so much. In Chapter 12, I developed a simple expected goals model based on three shooting zones. The idea was to assign each shot a probability of being a goal: shots from outside the box had a 3.4% chance of going in, shots inside the box were scored with probability 12.4% and shots in the six yard box had a 32.2% chance of success. The models used by clubs are more sophisticated than this; using the x and y co-ordinates of the shot, whether it originated from a cross or a through ball, if it was a header, volley or controlled shot and so on. Each club has their own way of measuring expected goals, but the basic methodology is the same: past shots from similar situations are used to give a statistical model of the probability that a certain shot results in a goal. A team's expected goals for a match is then the sum of all the chances it had during that match.[11]

Sherwood laughed at the idea of expected goals, because it gives credit to players and teams for shooting and not scoring. But here he misses the point. Expected goals are a measure of whether a team is generating good chances. It is obvious that, in the long term, teams that make more shots from better positions out-perform those shooting less often from further out. And statistics support this observation: teams with higher expected goals in the past are more likely to win their matches in the future. Expected goals is simply a way of measuring the quality of chances, and every manager who wants insight into how well their team is doing should be concerned about their team's expected goals.

Ironically, one of the first people to build expected goals for Premier League clubs now works for Aston Villa, the team that sacked Sherwood. Sam Green studied maths and physics at the University of Bristol, and it was while he was still a student that he started applying his knowledge of mathematics and statistical physics to the analysis of cricket and football. After a period working in the betting industry, the sports statistics company Opta recruited Sam.

Reading the analyses Sam performed during his time working at Opta, it is clear that he made good use of their extensive data sets. An article he wrote in 2013 about Manchester United's shot conversion was particularly revealing.[12] During Alex Fergusons last few seasons at the club, United had fewer shots than their title rivals, but they scored from more of the chances they created. Using 'expected goals', Sam showed that United scored more because they were shooting centrally, in positions that were more likely to result in a goal. However, he also suggested that even accounting for their better shooting position, their success was unsustainable.

Sam's prediction proved correct. The next season, under new manager David Moyes, United's shot conversion dropped dramatically and they finished 7th in the Premier League. To analysts looking at United underlying numbers, this change in fortune did not come a big surprise.

When I spoke to Sam, he emphasised that his current job as Head of Research at Villa is not just about calculating expected goals. He is building a system to put together all scouting information, both from scouts and from statistics. The scouts who watch the matches are still a central source of information for the club, and he wants to ensure that all their match reports are combined with statistical analyses to get a complete picture of potential signings. Again, the challenge recognised by those working in the business is integrating models like expected goals with traditional knowledge.

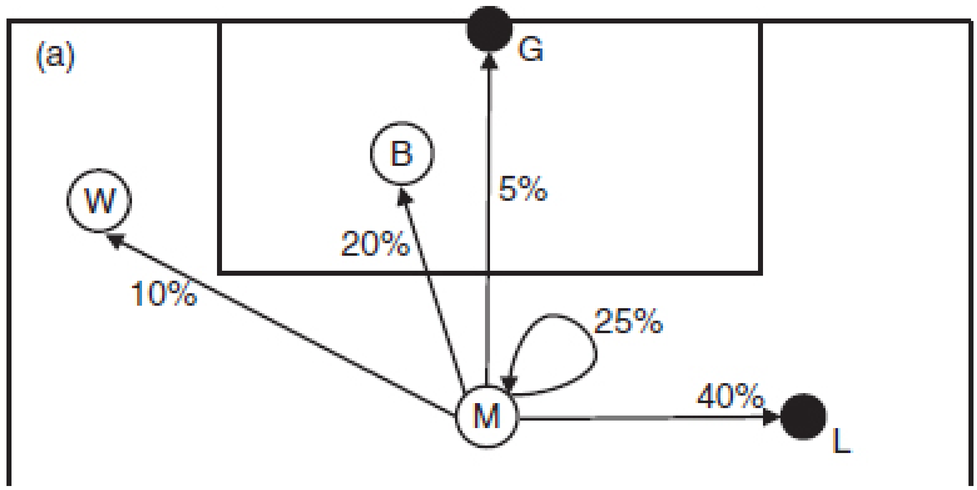
I asked Sam about the risks of working for a recently relegated club, especially when the ex-manager is doing interviews complaining about one of the methods he developed. “There are risks”, he told me, “but I have a brilliant job. I use interesting mathematics to work in football . . .”, he paused and smiled, “. . . and it does help that I am well remunerated.”[13]

## Goal Chains

In 2011, the start-up company StatDNA released their match data for the previous season’s Premier League to bloggers who were willing to try their hand at analysis. Amateur analysts were invited to enter a competition to see who could use this data to give the best understanding of the game. The winner was football blogger and software engineer, Sarah Rudd.

Sarah’s research paper and presentation has become something of a legend in the football analytics community.[14] It was the very early days of football analytics, and most club analysts hadn’t yet got as far as sorting their player spreadsheets, let alone talking about expected goals. So when Sarah took the stage at the New England Symposium on Statistics in Sports at Harvard to present her ‘framework for tactical analysis and individual offensive production assessment in soccer using Markov chains’, she was moving into completely new territory.

Sarah’s model divided the game in to a set of ‘states’, where each state describes where the attacking team has the ball and the arrangement of the defending team. A simplified example in Figure 14.1 (a) divides the pitch into states for Box, Wing, Midfield, Goal and Lost. The first three of these states describe where the attacking team has the ball. The ‘Goal’ state means that a goal has been scored, and ‘Lost’ means that the attacking team has lost the ball. The arrows leading out of the ‘Midfield’ state show the probability of reaching each of the other states: there is a 20% probability of getting the ball in to the ‘Box’, a 12% chance the ball is moved out to the ‘Wing’, and so on. These probabilities are also shown in the game transition table in Figure 14.1 (b); each entry in this table is the probability of the game moving from one state to another.[15]




(b)

| From/to | Midfield | Box | Wing | Goal | Lost |
|---|---|---|---|---|---|
| Midfield | 25% | 20% | 10% | 5% | 40% |
| Box | 10% | 25% | 20% | 15% | 30% |
| Wing | 10% | 10% | 25% | 5% | 50% |
| Goal | 0% | 0% | 0% | 100% | 0% |
| Lost | 0% | 0% | 0% | 0% | 100% |

**Figure 14.1** A Markov chain model of attack. There are three states for possession: M is having the ball in Midfield; W is having the ball on the Wing; and B is having the ball in the Box. Two additional states, G for Goal and L for Lost indicate the end of a possession. (a) States are represented as circles. Each of the percentage values indicates the probability that the next action is observed from when the ball is in midfield. (b) Probabilities of moving between states. Rows are the current state, columns are the next state, and the entries themselves are the probability of moving from one state to the next. The Goal and Lost states are special since they mark the end of the possession.

It is here that the mathematical idea of 'Markov chains' comes in. A Markov chain model assumes that the change between match states does not depend on what has happened earlier, only on the current state. So if the attacking team has the ball in midfield, the probability of it going into the box is the same, irrespective of whether the team has carried out a long sequence of passes, or if the ball has just ricocheted off a defender in the box. This assumption doesn't always hold in football, but it is a reasonable starting point.

Using the Markov chain assumption I can calculate the value of different match states. To do this I find the probability that the ball will eventually end up in the goal or be lost to the opposition. For the example in Figure 14.1 (a), I find that a ball into the box will eventually result in a goal in 25% of cases and the ball being lost in 75%. A ball in midfield ends up in the goal in 15% of attacks and will be lost during 85% of attacks, while a ball on the wing results in a goal in 12% attacks and is lost on 88% occasions.

These probabilities – Box 25%, Midfield 15% and Wing 12% –are thus the value of having the ball in each game state.

Sarah realised that these probabilities can be used to assign credit to players for an attacking move. For example, consider a central midfielder, a winger and a striker all involved in a goal. The winger cuts the ball back to midfield, the midfielder plays a through ball to the striker and the striker places it past the goalkeeper. How should we divide credit for the goal between these three players?

Usually, we give all of the credit to the striker for the goal, and the midfielder for the assist. But this is hardly fair on the winger who set up the move. The answer is to count how much each player improved the goal-scoring probability. When the winger passed to midfield, the probability of a goal increased from 12% to 15%, so the winger receives 15–12=3 points. When the midfielder made the through ball, the probability of scoring was increased, so the midfielder receives 25–15=10 points, and when the striker scores they receive 100–25=75 points for completing the move.

This measurement may still appear biased toward the striker, but if we count all successful passes by the winger, not just those that result in a goal, then they begin to accumulate quite a few points. Imagine, for example, that the winger completes 10 passes back to midfield and 5 passes into the box during the match. This gives the winger $10 \times 3 + 5 \times 13 = 95$ points for that match.

Sarah's method also allows us to see the difference between players who simply pass a lot and those that create chances. For example, imagine that the attack starts with the midfielder who passes the ball out to the wing, before a winger makes a successful cross in to the box. Now the midfielder receives negative points, 12–15=-3 points, because a pass to the wing typically reduces the probability of a goal. The winger receives 25–12=13 points for the pass in to the box. The method punishes players who make it harder for their own team to score goals by assigning them negative points.

Watching Wayne Rooney play in midfield for England during Euro 2016 brought this model to my mind. Rooney sent long cross after long cross out to Kyle Walker on the right wing. These balls flew with great precision, but they didn't greatly increase the probability of England scoring. Rooney completed 61 passes in the game against Iceland, and with each pass I became less and less convinced that England could win. Playing in a similar position for France against the same opposition, Blaise Matuidi made slightly more passes (77). But the important difference between Rooney and Matuidi was that the latter sent the ball forward to Dimitri Payet and Antoine Griezman. While England lost 1–2 against Iceland, France won 5–2.

Breaking down football into the three states of midfield, wing and box takes us some way to understanding player contributions, and Rooney might well learn something from this relatively simple model. But a full model requires a lot of data, and a lot of work. In her research presentation in 2011, Sarah Rudd broke the game into a total of 22 states, accounting for 11 different ball locations and two different levels of defensive pressure at each location. Now working for StatDNA and Arsenal, Sarah has access to detailed positional data of both defending and attacking players. This allows her and her colleagues to break the game down into hundreds or even thousands of different states, and they can analyse the outcome in each situation. As well as the position of the ball on the pitch, such a model could account for the total number of defensive players in front of the player with the ball, and whether or not an attack had arisen from a counter or from possession. A model like this would have 242 states and would still fail to account for all the different ways two teams can be organised on a pitch.[16]

It is here that the frontier of football analytics currently lies. The challenge is finding a relatively small number of fundamental states of play, and working out how likely a team is to score or concede a goal from each of these states. Different analysts have different solutions. One German television channel uses a concept called packing, which describes the number of players 'taken out' by a pass or dribble.[17] Other analysts I talked to discussed states in terms of getting the ball between and past the two lines of the opposition's defence. Marek Kwiatkowski, a former theoretical biologist turned football analyst at Brentford and other clubs, has called for an approach based on classifying different types of possession chains.[18] For example, counter-attacks, midfield passing and movements down the wing would each be one class of possession chain. The challenge is to automatically classify possessions and assess how well a team executes each chain type.

I was very keen to hear how Sarah Rudd's Markov chain model had developed after she was recruited to StatDNA and the company was bought by Arsenal. But, despite a lot of pleading on my part, Sarah politely declined an interview for this book. Arsenal use her approach both in assessing players and evaluating tactics, but they are understandably secretive about what they are doing. I don't know if Sarah and her colleagues have solved the problem of breaking down football into a reasonably sized set of important play states or chains. But I do know, from observing the attacking style of football Arsenal have played over the last 5 years, with short probing passes and continual movement to different states of play, that her model is having an influence on the pitch.

## Defending your Patch

Markov chains help solve the problem of evaluating attacking players, but they don't offer immediate insight into the performance of defenders. This was a problem that caught the eye of Thom Lawrence. Thom had been programming since he was a kid, working for a variety for companies and start-ups. But at the start of 2015 he felt disillusioned. ‘Coding had stopped making me happy,’ he told me. It wasn’t exactly that he decided there and then to become a football analyst, but he started a blog and analysed match data in his spare time; to do something that was ‘creatively nourishing’. He set up a blog called Deep XG, where XG stood for expected goals.[19]

Thom’s analysis started with finding players’ territories. It was through this question that I first met him on Twitter. He asked me whether I thought that techniques for finding animal territories could be used for identifying defensive ‘patches’. I showed him my work from Chapter 7 on defensive hulls, but it turned out Thom was already far ahead of me in thinking about the problem.

While my defensive-hull method looked at the areas players typically got hold of the ball, Thom had realised that turnovers of possession weren’t necessarily a good measure of a defender’s success. A player who has to tackle might do so because he is badly positioned in the first place. It is also plausible that a player who is regaining possession a lot is also *losing* possession a lot, with the ball bouncing backward and forward between him and the opposition.

Thom’s answer to measuring defence was to look at how far the opposition team advanced the ball within a defender’s zone or patch. The shapes of patches were defined in a way that was similar to my defensive hulls. Thom then measured defensive success. A player’s ‘patch score’ is proportional to

$$\frac{\text{Area of patch covered by defender} \times \text{Time opposition has possession within patch}}{\text{Distance ball progresses through patch}}$$

Players who allow the opposition to have the ball in their patch but don’t allow it to move forward are given a better score. As a result, Sergio Busquets ranked as the absolute best defensive player in Europe in the 2015–16 season. He occupies a massive space in midfield, where the opposition finds it very difficult to make progress.

When Thom published his patch method in a blog post in early March 2016, Liverpool’s Emre Can, Bayern Munich’s Joshua Kimmich and Tottenham Hotspur’s Eric Dier were all among the younger, high-ranking players, but top of the list in Europe was the relatively unknown 22-year-old Lyon defender, Samuel Umtiti. By May that year, the previously uncapped Umtiti had been named as part of the France squad for Euro 2016, and at the end of June 2016 he signed for Barcelona for €25 million. This was a spectacular validation of Thom’s method.

Despite these successes, Thom went to lengths to explain to me the limitations of his system. In the 2015–16 season, Leicester defied the model by allowing the opposition to advance down the wings, and only starting to defend near their box. This made their full-backs look bad statistically, while in reality they were successfully shutting teams out in the middle.

Thom is upfront about these and other limitations. Partly because of this honesty, clubs started to approach him for help, and he is now a full-time consultant. When I spoke to him he was working on one substantial project for a 'Europa League-level club'. Thom has also been involved in several smaller projects with other clubs and betting syndicates. It is possible, with a lot of hard work and a good idea, to make a living analysing football. And Thom is certainly having much more fun at work than he was a year ago.

Omar, Thom and Sarah all made rapid transitions from the blogosphere to boardrooms and consultancies. But it was clear when I talked to them that they already had what it took in terms of technical background. They all knew their football, and they all knew their maths. It takes a lot of hard work and study to get to the point where you can reliably advise clubs about the best signings, but once there, it is one of the best jobs in the world.

## The Worst Critics of Analytics

What do the old school scouts make of Markov chains, defensive patches and player radars? What do real footballing men think of confident economics graduates, nerdy computer programmers and brash Americans taking over their jobs and offering consultancy at inflated prices?

At the start of the 2016–17 season, *The Daily Telegraph* reported that the Fulham manager, Slavisa Jokanovic, was unhappy with the influence of American data analyst Craig Kline.[20] Fulham had brought in Kline to complement their traditional recruitment, and were using a 'ticks both boxes' policy: both the traditional scouting and the numbers had to be good in order for a deal to be made. Jokanovic was apparently furious with Kline when the latter refused to accept a player recommended by José Mourinho, because there weren't enough stats available on the player's performance.

This story, like Tim Sherwood's outburst in the same newspaper, typifies a new angle from which the media report on football. Analysts, like players and managers, are becoming the focus of attention. But it isn't clear from reading the news how serious these various conflicts really are. I spend a lot of my working day arguing with other researchers about how to interpret results, and we sometimes get frustrated with each other. That's part of any job involving analysis; the only difference in football is that the media are interested in every minute detail of what happens.

My overall impression, talking to both 'old-school' football scouts and to the new wave of data analysts, was not one of conflict. Omar Chaudhuri agreed with me. He told me that he had had one or two very frustrating discussions with scouts, 'but these were less than 10% of cases'. Most of the people working within football clubs were fascinated to hear about what maths might have to offer their game.

The hardest criticisms of analytics I heard during my research didn't come from the old-school footballing men – they came from the analysts themselves. Marek Kwiatkowski told me that, although stats-based scouting works, it 'is currently at the art-more-than-science stage.' Thom Lawrence was even harder on what he does – 'after a lot of hard work, I have no concrete evidence that proves my method is valid.' By this Thom means that it is difficult to validate any approach, not just his own. Since the work done within and for clubs is secret, the main forum for sharing ideas is via social media. Twitter allows people, including me, to share their match visualisations. But Thom feels that a lot more rigour is needed. 'You tweet something. People like it and you feel good about yourself for a bit. But there is no constructive critical feedback.'

When he said this, I understood exactly what he meant. I enjoy tweeting about football. But there is no peer review or replication of studies, as there is in my other scientific work. As a result, there is little or no verification that an approach works.

Henry Newman was also cautious about how far analytics has come. He told me, 'Just now its very cool to say you have used numbers to sign a player, rather than just having sent three people to have gone and watched him, but that's just the first step. The next step is to really prove that the numbers are working.'

To an outsider Thom and Henry's critiques might appear to cast doubt on the future of analytics. But being self-critical is important, whether you are a scout or a quantitative analyst. At West Ham, Rory Campbell told me a story about when he added a small 'but . . .' at the end of a glowing scouting report on N'Golo Kanté, while the player was still at Caen. West Ham passed up on Kanté and Leicester City snapped him up. Rory was still kicking himself, rewriting his report in his head.

Self-criticism is a characteristic shared by all the best analysts I have spoken to, both those that spend the majority of their time in front of a computer and the ones in the stands watching games. Never believe blindly in what you are doing. Always question and question some more. It is a basic principle of science, and it should be a basic principle of football analytics.

CHAPTER FIFTEEN

# Football's Intelligent Future

Over the past year I have talked to a lot of different people working inside football clubs, and I've heard a wide range of different frustrations. Many of the analysts and scouts I talked to didn't want to be named, both because they were often direct in their criticisms and because clubs are very sensitive to information leaking out. But they repeatedly came back to one important problem: the relationship between the club and the players.

Players train hard in the mornings. But outside of these training sessions and the games, they are left to do what they want with their extensive free time. After lunch the players go home. They sit in their luxury apartments and play Call of Duty and FIFA 17. Online gambling is also a big hobby. They invite friends over, play computer games, watch movies and hang out. The afternoon drinking sessions of the 1980s and 90s are gone, so the players aren't damaging their bodies, but intellectual development is not top of their priority list.

Very few clubs organise wider activities. Some have yoga sessions a couple of times a week. But unless these are made compulsory, the players often skip these, giving various lame excuses. Asking the players to study or broaden their minds is completely out of the question. It is difficult enough to get them to take their noise-cancelling headphones off during breaks.

Many clubs treat their players like spoilt children, scared to put too large a demand on their time. There is an ever-present threat of a player becoming dissatisfied with his club. Agents hover nearby, always ready to offer advice about a potential move. Everything possible is done to keep them happy, with the general rule being that they shouldn't be told to do anything other than to play football. The point made by those working inside clubs is not that football players are ignorant or stupid. It is rather that they are not being challenged mentally.

## Not Stupid

Torbjörn Vestberg and his colleagues have shown that football players are not stupid. In fact, players are very smart. Torbjörn visited clubs in the Swedish premier league and the third tier.[1] At each, he asked the manager to select two forwards, two midfielders and two defenders who were average for the level of football the team played at. The selected players completed a task involving connecting up a set of dots in as many ways as possible without repeating the same pattern. The task tested geometrical thinking, creativity and memory.

The players nailed the test. Both those playing in the top- and third-tier did about 50% better than the average person in the 'connecting the dots' test. The top-tier players were significantly better than those in the third-tier. Test performance was also correlated with performance on the pitch. In the season after Torbjörn conducted the test, the players who scored more goals and provided more assists were typically those who had done best in the task. The result applied both to men and to women. The better the football player, the better he or she was at connecting up the dots.

The term the researchers used for the type of intelligence required for connecting up the dots is 'design fluency'. Football players typically have less formal education than the average person, so the dot-connecting task is a fairer test of their intelligence than one involving written or verbal reasoning. This is an important distinction, because we shouldn't conclude from this study that football players are generally more intelligent than the person in the street. Players don't have above average verbal reasoning, for example. If they did we would probably have noticed it during TV interviews.

Instead, what we should take away from Torbjörn's study is that football players develop a capacity for abstract, creative thinking. This type of thinking is very similar to that used in many mathematical problems, in fields such as combinatorics and graph theory. The question for the team manager is how to utilize that type of intelligence. Clearly, many clubs and managers have not yet understood the intellectual potential of their players, but others have. And more and more of these managers are working in the Premier League.

## Geometrical Genius

As a player for Barcelona, Pep Guardiola was a defensive midfielder, known for his calm control and precise passing. This perspective from the back of midfield, a role sometimes referred to as the 'pivot', allowed him to see the geometry of the game – the passes made, the movement off the ball, and how players' positioning affected the outcome of a match. He took all of this experience to become one of the finest managers football has ever seen, first at Barcelona, then at Bayern Munich and, before the 2016–17 season, he joined Manchester City.

When we see managers, including Guardiola, jumping up and down on the sidelines and gesticulating to their players, we might get the impression that they are trying to communicate tactical details. But during his time as Bayern Munich manager, Pep Guardiola told their fan club, 'For me it's just impossible to sit. I'm nervous and have to do something . . . and I think they understand [what I am saying], but they don't.' Guardiola went on to joke that the fact his team doesn't listen 'is the reason we win.'

Like all good jokes there is an element of truth in it. The players don't need to listen to his shouting. A big part of Guardiola's coaching genius lies in recognizing that his players are also smart. He recruits intelligent and adaptable players. His aim is to then teach them about football, and let them work out how best to execute the plan on the field. Guardiola properly utilizes design fluency, which Torbjörn and his colleagues found in players. In his playing days, Guardiola could think tactically while playing the game. Now, as a coach, he expects his players to do the same.

In Chapters 2 and 3, I showed that training exercises should reinforce the type of passing and movement that creates fluid football. Such exercises are the starting point for Guardiola's approach: he was the coach of the Barcelona team that covered the pitch with passing triangles on their way to a Champions League title in 2010–11. But Guardiola also expects players to think one level above triangles: to understand the geometry of positioning and formation over the whole pitch. He trains his team on a pitch overlaid with a positional grid, as shown in Figure 15.1. The grid provides a guide to allow the players to space themselves, and to know which direction they should face.

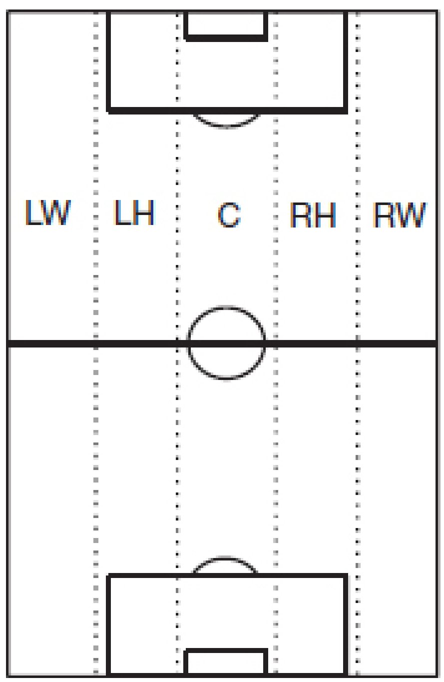



**Figure 15.1** Pitch broken down in to five spaces; left wing (LW), left half-space (LH), centre (C), right half-space (RH) and right wing (RW).

One of Guardiola's innovations was to recognize the importance of the 'half-space', the spaces marked LH and RH in Figure 15.1. In earlier grid systems, such as the 6x3 grid adopted by Louis Van Gaal, just three vertical spaces were marked: one central space, a left wing and a right wing.[2] Guardiola realised that it was halfway between the wing and the central point that was important in attacks, and added the further division to his training pitch. These half-spaces were perfectly suited for players like Lionel

Messi, Xavi and Andrés Iniesta to work their triangular magic (as in Figures 2.5 and 2.6). Other teams, playing with two defenders in a central space and two out on the wings, would be unable to cope with these three coming directly at them.

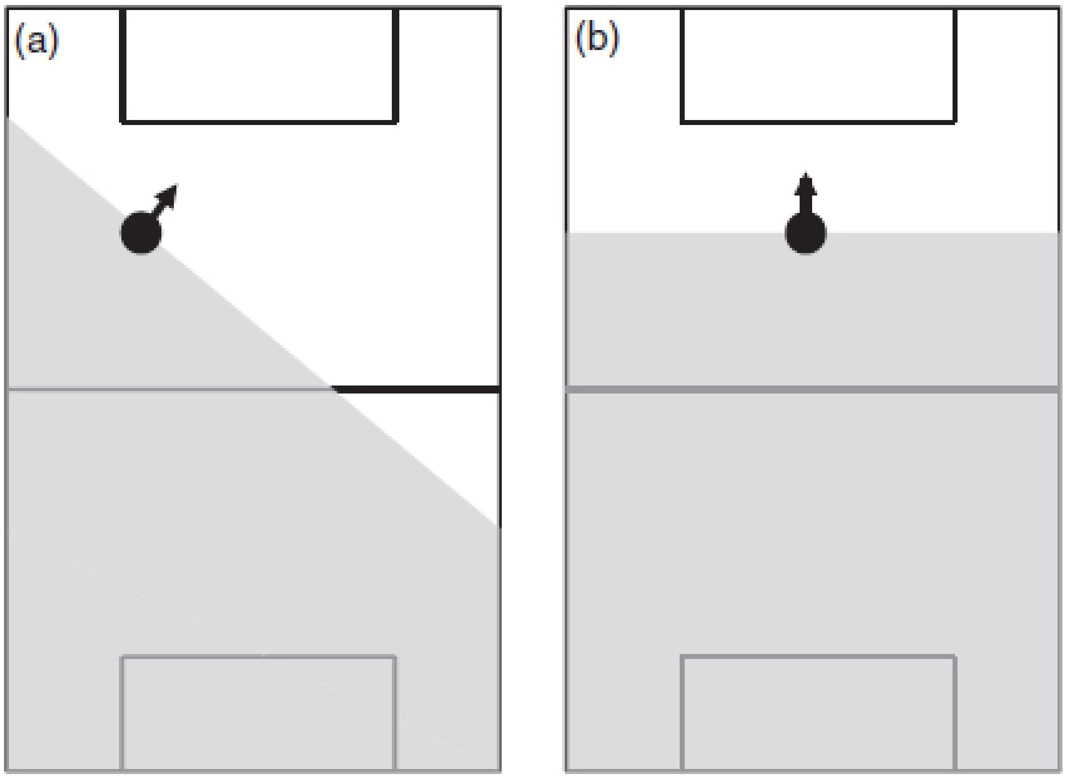



**Figure 15.2** What a player sees when facing towards goal when (a) positioned in the half-space and (b) when in the middle of the pitch. Shaded area is 'blind' area. White area is visible. The arrow is the direction of orientation.

The maths behind half-spaces has been more fully developed by a young Austrian qualitative analyst, René Maric´. René was a talented junior player, but he was unfortunate. While still a teenager he broke his hip and had four knee operations, ending any chance of a playing career. But René loved the game, and spent his time while recuperating watching old matches. He started with Brazil in the 1970s, worked through Dutch football, and ended up with modern-day Barcelona and Bayern Munich. He told me that even as a young player, he was always thinking about positioning and movement, and how he could play better. Now he took the time to teach himself how the game was structured, and to learn how the best teams played together.

In order to have a forum to talk about the tactics he had learnt, René and his friends set up a blog, Spielverlagerung. It wasn't long before his writing attracted the attention of Thomas Tuchel, who at the time was manager of German Bundesliga team Mainz 05. Tuchel was so impressed with René's match analysis that he asked him to prepare opposition reports before their matches, focusing on patterns in play and positioning. From there, René was offered a range of consultancy and analysis roles. Now, at the age of 24, he is working for the Austrian team RB Salzburg, as the academy's qualitative analyst, and is assistant coach for the under-18 team. Tuchel didn't do so badly for himself either, taking over as manager of Borussia Dortmund in 2015.

What impressed me in René's analysis is the use of geometry to describe why tactics work. He explains that one key advantage of playing in the half-space is that it allows a player to see more of the pitch. This point is illustrated in Figure 15.2. A player orientated towards goal in the half-space has a wider view. There is a trade-off involved in half-space positioning. Standing on the touchline gives an even wider view of the pitch, but is further from goal and not as dangerous to the opposition. René explains that 'the half-space is the ideal intersection of "I have enough space" and "what I can't see doesn't matter anyway."'[3]

The second advantage to playing in the half-space is that when the opposition adopts a 4–4-2 formation their players are placed on the lines running from the top to the bottom of Guardiola's training pitch (Figure 15.1). The half-space positions can allow a team to get in between a 4–4-2 opposition. An example of this is shown in Figure 15.3. The open circles are positions of the defenders and midfielders of the defending team. By responding with a 4–3-3 formation, five of the attacking team's forwards and midfielders are placed at maximum distances from the defending team (black circles). The final attacking striker (ST) looks to run on to balls in the box. In theory, this gives each attacking player around 7m to each defending player, a little bit of extra space to create an opportunity with.

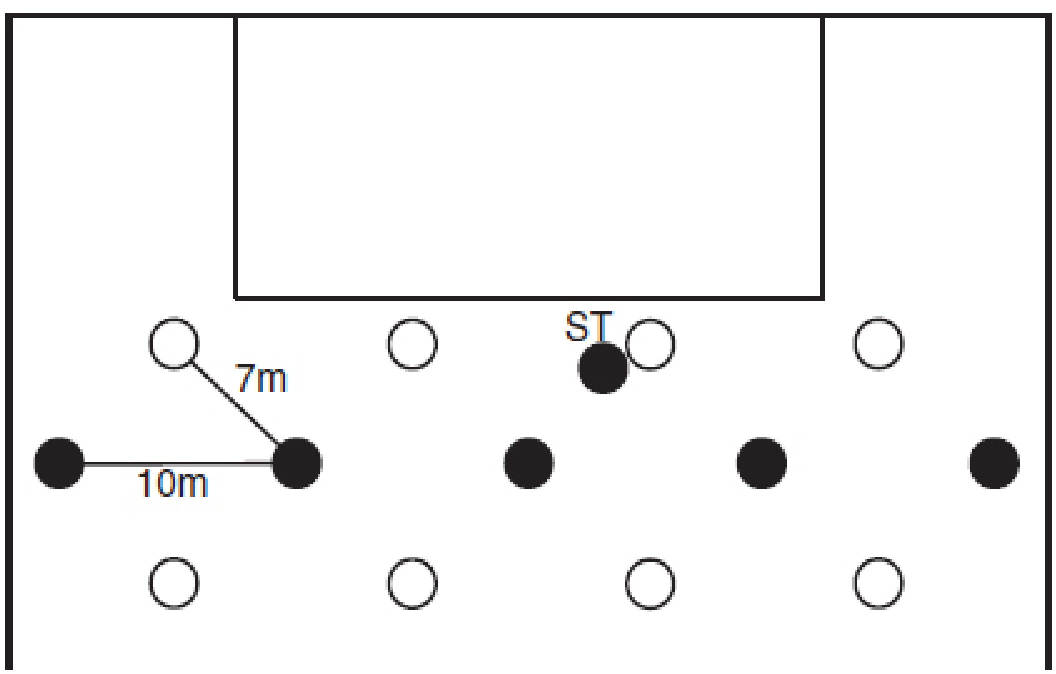



**Figure 15.3** Guardiola's use of the half-space to maximize distance from a team defending with 4-4-2. White team is defending with 10m between the players, with 4 defenders and 4 midfielders. Black team is attacking with six players. Striker (ST) approaches the box. The other five attacking players are equally spaced within white's defence. The distance between the black and the white players is about 7 metres.

Figure 15.3 should not be interpreted as an exact positioning of each of the players during attack. Instead, it is the starting point for the players to find their own solution to the problem of moving the ball forward. When the Under-18 coach at RB Salzburg, Marco Rose, talks to his young players about positioning, he explains how their movements will affect their teammates and the opposition's positioning. Marco and René set up situational training exercises using a pitch marked out with the same spaces used by Guardiola. During the exercises, players are given guidelines about which area of the pitch to move into depending on the ball position and they adjust to playing with the spacing imposed by the markings.

Like Marco and René, Pep Guardiola also makes sure his players understand how to use space creatively during build-up. Talking about his time playing for Barcelona, Thierry Henry explains his positioning using examples of play very similar to those in Figure 15.3.[4] His role during build-up was to occupy the left wing in order to draw out the defending team, opening up the half-space for Iniesta or Messi. Guardiola told his players to 'stay in your position, trust your teammate on the ball, and wait for the ball'. Up until the final third of the pitch the players had to follow the grid, but once in the final third they were allowed to leave the grid and attack freely.

## Where to Shoot and Where to Tumble

Geometry lessons should start early in a football player's career. One of the first thing players learn, when they are as young as seven or eight, is that the more of the goal you can see when you shoot, the better your chances of scoring. Kids start noticing that if they overrun the ball in the box, they end up hitting the side-netting. Shooting directly in front of the goal is a much better option. Ten-year-olds can learn to defend by reducing the angles, by showing the attacking player the way out to the goal line where the shot-angle is reduced. Even primary school kids can do the maths of scoring and defending goals.

Although children can understand the importance of seeing the face of goal, if we want to gain tactical insight then we have to do the mathematics. Figure 15.4 shows three different shooting positions and the angles between the goal posts. In Figure 15.4a the angle is 38°, giving a good chance of a goal. In figure 15.4b and c, the angles are both 17°, providing narrower chances.

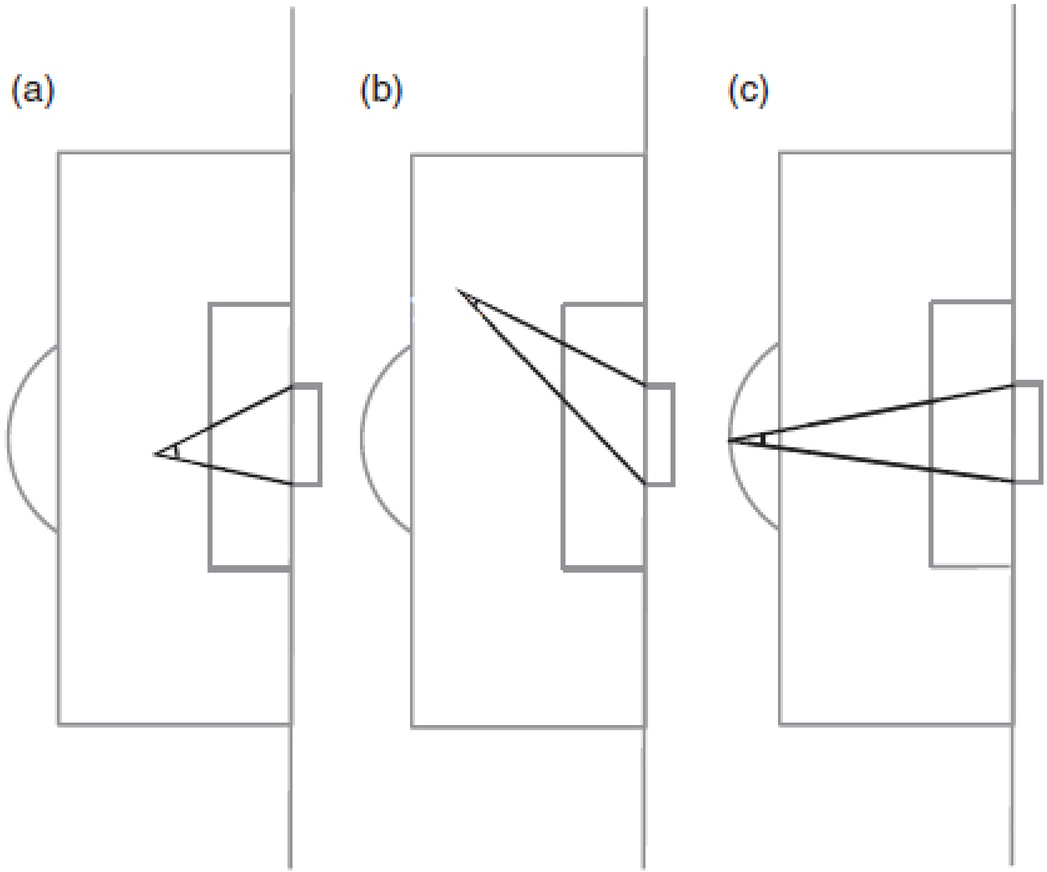



**Figure 15.4** How between-post angle depend on shot position. (a) A shot close to the goal has a large between-post angle, in this case 38°. Shots from a less central position (b) or further out (c) have smaller between-post angles (in both these cases the angle is 17°).

This angle between the goal posts is the single most important factor in whether a shot will be a goal. The bigger the angle at the point a shot is taken, the more likely it is that the shot will go in. Figure 15.5a shows a series of circles emanating from the goal. Along each of the points on these circles the angle between the posts is the same. The outer circle marks all the positions where the angle is 17°, the middle circle has all the 38° angle shots, and the inner circle contains 55° angle shots.

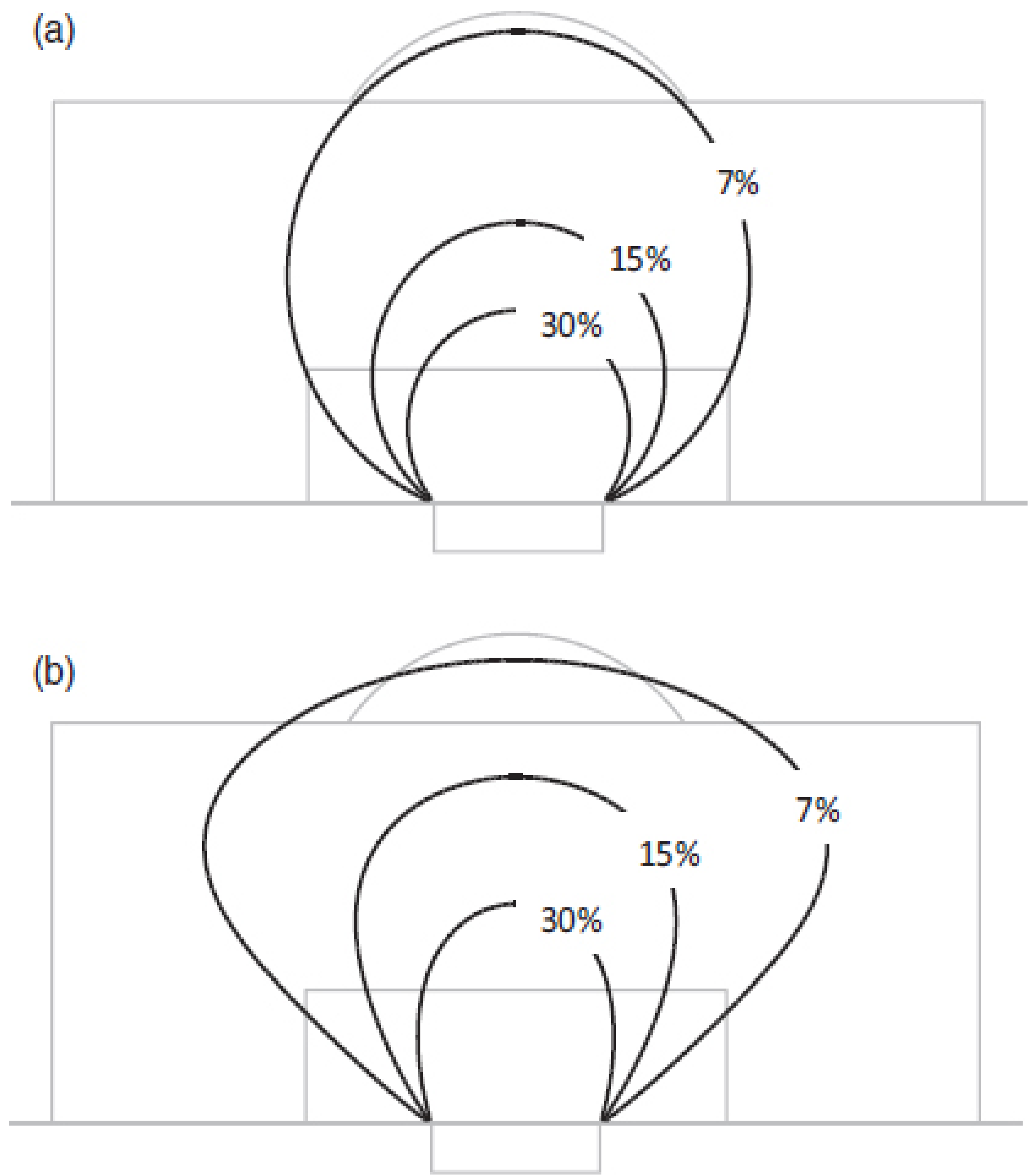



**Figure 15.5** Contours where the probability of scoring is 7%, 15% and 30% for two models. In the simplest model (a) each contour line corresponds to positions where the between-goal post angles are equal. In the more accurate model (b), the contour lines also account for the distance from the centre of the pitch and from the goal line.

The probability of scoring a goal is also similar for the different points on these circles. Around 15% of shots taken with an angle of 38° and 7% of shots with an angle of 17° result in a goal. There is a remarkable reliability in this geometrical football rule. I have looked at shooting data from the Premier League, Bundesliga, La Liga and Champions League, and worked out the probability of a goal from different points. The results are always similar: the angle between the goal posts, as seen by the player taking the shot, is roughly proportional to the probability of scoring.

I like the angle-based rule because it is simple and it relates to something coaches can explain to players: the more of the face of the goal you can see, the higher the probability that you'll score a goal. The rule can be further refined.[5] A circle isn't quite the right shape to describe the probability of scoring from different points. The best shape is a little bit more squashed out at the sides. Figure 15.5b shows one example where I made a statistical fit based on Opta's shot data from the last two years. The contours show how the probability of scoring increases as the shot location becomes closer to goal.

Based on this observation, Venezuelan engineer Cesar Morales has proposed that the shape of the penalty area should be redrawn to reflect contours similar to those in Figure 15.5a[6]. Instead of the traditional rectangles, the penalty area should be some form of squashed circle, reflecting the probability of scoring from different positions. Cesar's argument is that the shape of the current box encourages diving. If a player has less than 7% chance of scoring from the outside edges of the box then there is an incentive for him to throw himself to the ground. One example is the penalty awarded to Holland against Mexico in the 2014 World. Arjen Robben tumbled over after a tackle on the goal line, just outside the six-yard box, a position that it is almost impossible to score from. The penalty, awarded in the 94th minute of the match, proved decisive. Holland won the game 2–1 and Mexico went out of the World Cup.

Some Mexican fans and professional commentators labelled Robben as a diver. This is debatable. Robben was awarded 4 penalties in the 2013–14 Champions League, more than any other player, and he did appear to go down easily when tackled by Rafael Marquez. But there was clear physical contact between him and the defender. Robben is often in the box and is very fast. He gets himself in positions where it is easy for the opposition to make mistakes. The more important question is not about Robben as a player, but whether it is fair that a player in a position with almost no chance of scoring should be awarded a penalty. The probability of scoring a penalty kick is around 75%, while the chance that Robben, who was surrounded by three Mexico defenders, would have scored was close to zero. Cesar's redrawn penalty area would have meant a free kick to Holland instead of a penalty.

I very much doubt that FIFA will redraw the penalty area. Instead, referees already attempt to compensate for its poor current design. 61% of penalties are awarded in the 18 yard by 20 yard area found by extending forward from the six-yard box to the edge of the penalty box.[7] The other 39% are awarded in the two 18 yard by 12 yard areas on either side. That makes the probability per square yard of being awarded a penalty in a central area 2.1 times greater than the probability of being awarded one on the outer edges. Penalties on the edges of the box are exceptions rather than the rule. Unfortunately for Mexico fans, the Robben penalty is one exception that came at exactly the wrong time.

## Arsenal's Expected Win

During the 2015–16 season, Arsenal manager, Arsène Wenger mentioned ‘expected goals’ in interviews on several occasions. Arsenal were consistently ahead in the Premier League in terms of this measurement. They were making shots centrally, in front of goal, in positions that we would usually have expected to produce goals. In other words, they had more good quality chances than any other team.

The problem for Wenger was that, while Arsenal might have won more ‘expected’ matches during the second half of the 2015–16 season, Leicester City won more real matches. And it is real matches that put you in a title race. After Christmas, Arsenal started to fall behind both Leicester and Tottenham Hotspur, and although they eventually caught back up with Spurs, they ended the season 10 points behind Leicester.

In February 2016 I met several analysts in a pub in London for an afternoon chat about analytics. At the end of our meeting, one of them, Neil Charles, took a quick poll of the participants, asking them who they thought would win the league. Five went for Arsenal, three for Spurs, one for Manchester City and one for Leicester. I would like to claim I went for Leicester. I didn’t. I went with the numbers and backed Arsenal.

One explanation of Arsenal’s failure, which Wenger might agree with, is simply bad luck. This is plausible. There was nothing in their underlying numbers to suggest that the team collapsed or didn’t have the nerve for the fight, as many newspaper articles framed it. There is a great deal of randomness in football, and Arsenal may have fallen victim to it.

There are, however, other explanations. Arsenal took a much smaller proportion of shots from outside the box than any other team in the Premier League. They focused on getting the ball within the contours of Figure 15.5. It is possible that these tactics, focused on maximising expected goals, might have been ‘found out’ by other teams. By packing more defenders into the box, opposition sides were able to reduce the quality of Arsenal’s chances.

A related problem is that if the players know that the manager judges them on creating a certain type of chance, then they might tune their style of play to these requirements. This doesn’t mean that strikers would deliberately miss the chance to score a real goal in order to generate better *expected* goals, but if an attacking midfielder becomes more likely to be selected for the team by making short successful passes in to the box then this will be in the back of his mind while playing. During 2015–16, Mesut Özil made more successful passes into the box than any other player in the Premier League. In the first half of the season this resulted in 16 assists, but in the second half the number dropped to three. It is plausible that Özil became predictable.

Arsenal and Özil illustrate an important point about expected goals. While they can be used as a guide for evaluating and improving a player’s decision-making, teams should be careful in how they use them as a tactical tool. Expected goal models that fit to past data show the average outcome of a large number of different footballing tactical battles. They don't tell managers what worked in specific situations. Each manager still has to fight his or her own tactical battle with next Saturday's opposition.

## Have a Go

In terms of holding his position, Liverpool's Philippe Coutinho is a player that Pep Guardiola would approve of: the majority of his passes are received and made from the left half-space in front of the box. But Coutinho also loves to take a long distance shot at goal. During an attack down the left, he takes a step inside, increases the tempo of his movement, lines the ball up on his right foot and shoots. During the 2015–16 Premier League season he scored four goals like this from outside the box.

Despite the goals, it is reasonable to criticise Coutinho for his shooting. These four goals came from 69 shots taken from outside the box. Shots from outside the box have a typical conversion rate of around 3%, so the tally of four goals could be considered slightly lucky. There will always be a few players who get a little bit of luck in a particular season and are able to convert more chances, but this is not usually sustainable from one season to the next.

In Coutinho's defence, he also made more successful passes into the box per 90 minutes than any almost any other attacking midfielder in the Premier League. The only player well ahead of him on that statistic is Özil. Coutinho seems to have found a good balance between long-distance shooting and passing.

Long-distance goals produce some of the best footballing highlights and it is these shots we tend to remember years afterwards. France and West Ham player Dimitri Payet also likes to shoot from further out, and occasionally scores to spectacular effect. And Zlatan Ibrahimovich scored an impressive six out of 52 shots taken from outside the box during his last season in Ligue 1 for Paris Saint-Germain.

Numbers are not everything in football, though. There are moments of magic conjured up by Coutinho, Ibrahimovic, Payet and other players. It would be petty, and indeed wrong, of me to label them as lucky because they defy the expected goals model. But neither is football just about the magic of long-distance shooting. There is a structure to the events on the pitch, and it is often the manager who creates that structure. Wenger created a good structure in 2015–16 – it just didn't happen to win Arsenal the title. The challenge for every football team is getting that balance between magic, randomness and structure right.

## Long Ball's Back Again

*The* story of Premiership season 2015–16 was Leicester City. They were a perfect example of a team becoming more than the sum of their parts. After Leicester beat Watford 1–0, during a remarkable run of 1–0 victories, the Watford striker Troy Deeney told the BBC that the Leicester defence spent the entire game talking to each other. Deeney said ‘if Danny Drinkwater did not hear the right call then he was always going back from midfield and asking what was what.’[8]

There was a shared understanding within Leicester about how to turn defence in to attack. In one post-match interview, after a particularly impressive pass to Jamie Vardy, Drinkwater said, ‘The majority of times you don’t need to look. You just know he [Vardy] is going to be on the move.’

It wasn’t until halfway through the season that I started to realise there must be something really special about Leicester. They weren’t just lucky. I looked in more detail at how they performed. I broke down the football pitch in to $x$ and $y$ co-ordinates, and looked at how far each team’s passes travelled across ($x$-coordinate) and up ($y$-coordinate) the pitch during the build-up to a shot. Arsenal, Manchester United, Liverpool and Manchester City tend to pass a lot of times during build-ups, moving the ball back and forth and sideways to try to confuse the opposition. In 2015–16, Arsenal and Manchester City made around five or six passes during the 30 seconds leading up to a shot. Each pass travelled an average of less than 5m forward in the y-coordinate.

Leicester made fewer build-up passes than any other team in the Premier League: less than four passes in the 30 seconds leading to a shot. And their passes were also much longer than any other team, travelling an average of 9m up the pitch per pass. Their style of football was very different from the rapid passing style that is typically a statistical indicator of good football.[9]

Leicester weren’t the only team in the league hitting longer balls. Southampton, Crystal Palace, West Ham United and Aston Villa all passed longer distances than Leicester. What was unique about Leicester was the directness of their play. The other longer-ball teams were primarily playing crosses in to the box. This meant the ball travelled further in the $x$-coordinate across the pitch. Leicester’s passes travelled 2m less on average in the $x$-coordinate than these other long-ball team’s passes, but took the ball 2m further up the pitch in the $y$-coordinate than any other team in the Premier League.

Blogger Will Gürpınar-Morgan[10] and analyst Dan Altman[11] conducted their own statistical studies that reached similar conclusions to mine. Leicester City were defending more centrally, and using fast, long-ball counter-attack to surprise the opposition, in a way that no other team was able to pull off to the same degree.

Analysing Leicester’s play was watching tactics evolve, and it wasn’t just in the Premier League that this evolution was happening. In the Champions League during the same season, Atlético Madrid let opponents pass the ball, and waited for a mistake.

When Atlético knocked Pep Guardiola's Bayern Munich out of the Champions League they defended deep and made long forward passes. It was one such pass from defence to Antoinne Griezmann that sent him free for the vital away goal against Bayern and a place in the Champions League final. During Euro 2016, Iceland also showed that direct play can be highly effective. With just 32% possession they scored the only two open play goals in a 2–1 win over England.

## Tactical Evolution

The tactical evolution goes on. There is no one ultimate mathematical analysis or a single winning equation. Tactics change, and new analyses are needed all the time to understand the direction the game is taking.

While Guardiola's use of the attacking half-space is a beautiful example of how rigorous tactics in the build-up help players create in the final third, it is by no means infallible. In the Champions League semi-finals of 2009–10, Barcelona played Inter Milan in what proved to be one of the most important tactical matches of recent history. The manager of Inter Milan at that time was José Mourinho.

Mourinho realized that the key to defending against Guardiola's Barcelona team was to concentrate defensive efforts on the area in front of the box, and let Barcelona have the ball out at the wings. This style of defence is shown in Figure 15.6. Against most opposition teams this would be risky, since it allows crosses into the box. In this match, however, most of Barcelona's attacking players were shorter than the Inter defenders, making it harder for them to convert crosses. Moreover, playing five in the middle provided an extra defender who could concentrate on dealing with these crosses.

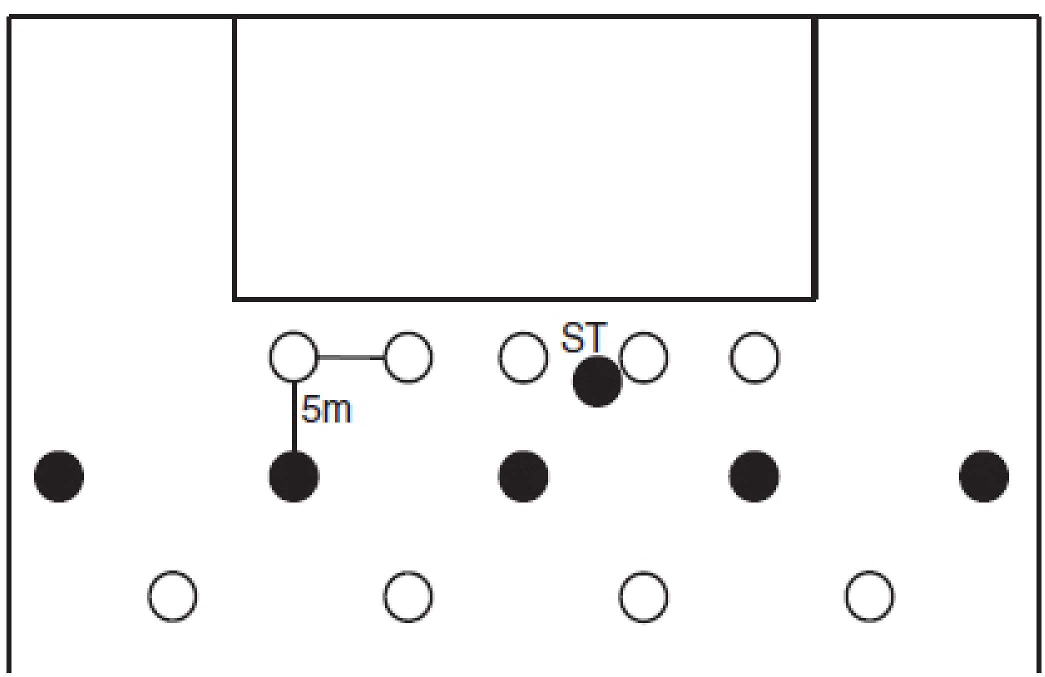


**Figure 15.6** Mourinho's use of five defenders to prevent attacks from the half-space. White team defending with 5m between the five defenders, and 10m between the four midfielders. Black team attacking with six players, as described in Figure 15.3. The distance between the black and the white players is now five metres.

Always thinking one step ahead, Guardiola realised what Mourinho was thinking, and played Zlatan Ibrahimovich as striker. At 1.92m, Zlatan would have no problem getting on the end of crosses. But there were problems with this plan. Zlatan wasn't match fit, and didn't fully suit the Barcelona style of play. Barcelona lost the first leg in Milan 1–3.

While the home leg was an important victory for Inter, it was at the return leg at Camp Nou where Mourinho was able to prove the dominance of his tactics. Inter midfielder Thiago Motta was sent off after 27 minutes, leaving Inter to defend their lead for the remainder of the match. Although Barcelona scored one goal, 10-man Inter kept the 11 men of Barcelona in check for over an hour. Mourinho had found a weakness in Guardiola's tiki-taka tactics.

The Inter-Barcelona encounter was to have a deep effect on the tactical development of both men. Returning as Chelsea manager in 2013, Mourinho perfected the box-crowding technique against Liverpool in the 2013–14 season to ensure they couldn't win the Premier League. He did the same again a year later against Arsenal to kill off their title challenge. Chelsea went on to win the league in 2014–15.

Working at Bayern Munich between 2013 and 2016, Pep Guardiola focused his energy on trying to solve the problem of breaking down 5–4-1 formations. And he found a new geometrical solution. This is shown in Figure 15.7. By surrendering the half-spaces, Bayern Munich were able to stretch the opposition out again. They overwhelmed teams on the wing before sending balls in to Robert Lewandolski in the middle, who scored goal after goal.

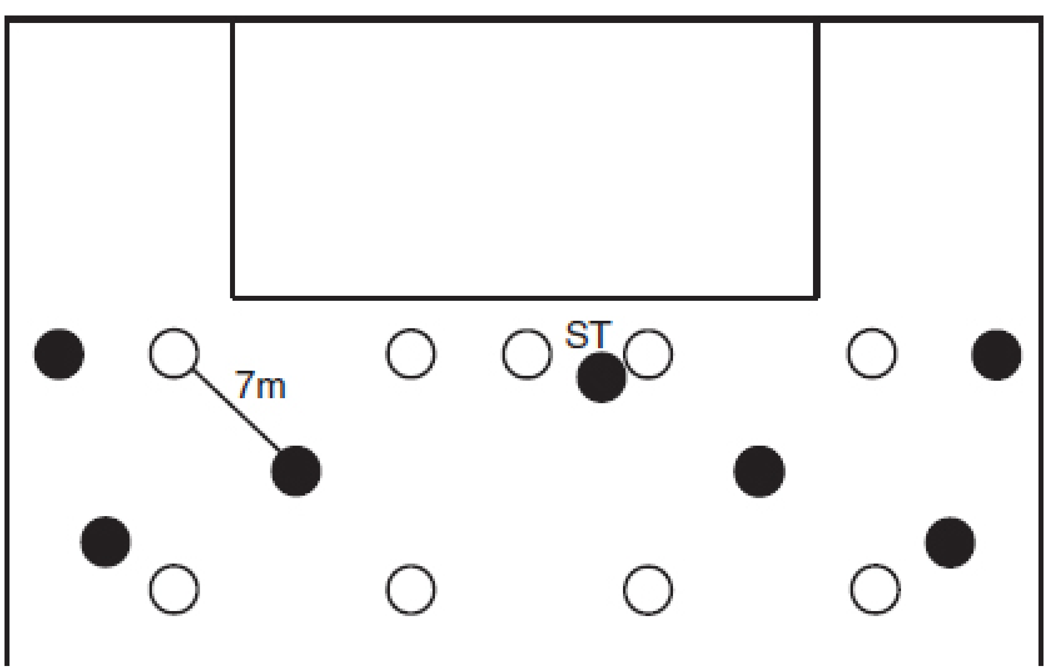


**Figure 15.7** Guardiola's answer to the use of five defenders. Black team is attacking with seven players (striker marked ST). White team is defending, and with 5m between the three middle defenders. The outer defenders are drawn out by the threat of two attacking players on both the left and right sides. The attacking player in the half-space again has seven metres to the nearest defender.

Guardiola even has an answer to Leicester's counter-attacking tactics. The formation proposed in Figure 15.7 places seven players a long way up the pitch, potentially leaving even more space for a direct long ball counter-attack. Guardiola's solution was to again use the half-space, but this time in defence. One or both of the left and right backs would move in to the defensive half-space.[12] The 'false' full backs then block teams like Leicester from playing the long balls down the middle. They are again forced out to the touchlines, where their attacks become less dangerous.

When Guardiola and Mourinho arrived in the Premier League for the 2016–17 season their tactical rivalry recommenced. It is really a rivalry between two different forms of mathematical thinking. Pep is the embodiment of the footballing mathematician. The club analysts I have spoken to referred to Guardiola as the chess player, and there was a clear respect for the way he has developed football's basic theory. José is more of an engineer, getting his hands dirty and solving problems. The respect from analysts for Mourinho was given more grudgingly, but they acknowledged his deep understanding of the game and his ability to think through tactical changes. Mourinho is often referred to as a motivator, but he motivates players by showing them that he knows what he is doing. This is true of all of the coaches who have moved to England to manage the big teams. Jürgen Klopp, Mauricio Pochettino and Antonio Conte are, above all, managers with a deep understanding of tactics and the structure of the game. Players are intelligent, and they only accept leadership of someone who is at least as smart as they are.

## Blue Sky Hacking

Until recently, the calculations and diagrams used by coaches were made on paper, and the movements were shown on the tactics board. At the Opta Pro Forum in London in 2016, I saw several analysts and researchers present Voronoi diagrams, like the one I used in Chapter 2. During the summer of 2016, the Voronoi diagram started to find its place in the collective consciousness of football analysts, who were talking about how they could be used to measure how teams created and controlled space.

The Voronoi diagram, and other automated methods for measuring space, perfectly suits the way modern managers see the game. Several analysts I talked to referred to how Pep Guardiola was constantly hunting for new opportunities to exploit the areas on the pitch that the opposition left open. These computational tools could help him find new spaces.

In moving to Manchester City, Guardiola has joined an organisation that is a blueprint for the future of football. Manchester City's chief executive officer, Ferran Soriano, came to the club from Barcelona with a philosophy of 'looking to play good football and to win, in [that] order.'[13] City have taken a highly scientific analytical approach: they have adopted the same style of play, based on 4–3-3, throughout their academy; they release data sets for statistical enthusiasts to analyse; they have one of the best training facilities in the world; and they have invested properly in their women's team, who won the 2016 Women's Super League. Manchester City are Football 2.0.

During the summer of 2016, Manchester City ran a football analytics hackathon. City released detailed player positional and passing data from 10 of their matches to 60 participants, with a wide range of backgrounds, chosen from over 400 applicants. The participants worked within City's indoor training facilities, and were provided with sleeping 'igloos'. These igloos were hardly used as the hackers worked non-stop in small groups for 48 hours to get the most out of the data. An added incentive was a prize of £7,000 for the best project.

One hackathon participant, Rob Suddaby, told the Opta Pro podcast that he had seen 'Voronoi diagram after Voronoi diagram' produced by participants.[14] He himself was trying to use convex hulls, the shapes I used to build tactical maps in Chapter 7, to find team structure. The hackers were using detailed tracking data to find the underlying geometry of the team. The winning group developed a machine-learning algorithm to evaluate player decision-making.

Most of the club analysts I have spoken to see the possibilities for using computerised tools for controlling space on the pitch. They want to find which areas the opposition are leaving open, identify the most common pass sequences, and measure compactness of defences. At present, the analyst's approach starts with videos. Pre-match, the analysis team looks at how the opposition plays. When they find something interesting, they try to work out a way of measuring it using statistics. This allows them to check whether what they saw in the video was a one off, or part of a larger pattern of play.

'Managers and coaches are inquisitive people,' one analyst told me. 'If you put something in front of them, be it a passing statistic or a Voronoi diagram, they nearly always take an interest. But they are also short on time, so the next thing they ask is "so what?"'

The question from managers is always how a statistic helps win matches. This is the continual challenge for those using numbers within football clubs. They have to produce answers that make footballing sense, have a firm statistical grounding, can be communicated within a few minutes, and ultimately improve results.

In the future, calculations made from data collected live during matches may well be used to make tactical decisions. But I don't think it will be Pep Guardiola who develops computerised tactics. It is more likely to be younger coaches who have grown up reading and writing analytics blogs, or who have participated in hackathons. At the RB Salzburg academy, René Marić and his colleagues collect data from every training exercise. This data is used to analyse both individual performance and team positioning. René believes that it is possible to use algorithms to study the geometry created during training, and to improve match tactics. But when I asked him for more specific details he became tight-lipped. ‘All I can tell you is that there are lots of possibilities, and we are analysing the data.’

Premier League analysts were even more cautious than René when they spoke to me. As I talked to them about mathematical methods for analysing player tracking data in order to improve tactics they were very keen to hear my ideas, but they were not so keen to talk on the record about what they were currently doing. However, my overall impression remained the same as I described at the end of Chapter 9. The main reason that clubs don’t let their analysts talk in detail about using player-tracking data isn’t because they are worried their secrets will be revealed. Instead, they are worried that the opposition will find out that they don’t have *any* secrets to reveal. While maths is increasingly used in scouting, its potential uses in tactical development remain largely unexploited.

This will change. After signing confidentiality agreements, I have started to talk to some clubs about how they could better exploit the data they collect. I may well continue to work with clubs in the future, and hopefully I will be able to write about it again, but as far as *Soccermatics* is concerned this is where my journey into the beautiful game has to finally end.

As for the evolution of football tactics – this will *never* end. No single manager has all the answers, nor does one mathematical method solve all of a team’s problems. I have shown how mathematical calculations can start to be combined with the tactical analyses already performed by coaches. This analysis shouldn’t be seen as an end point, nor should this book be seen as the definitive manual for mathematical football. The philosophy behind *Soccermatics* is that there are many different models, all of which help us in slightly different ways to understand the game. Putting these models together to improve team performance is a massive challenge.

These are exciting times for football, and I look forward to seeing young coaches and analysts rise to the challenge. I wish you all the best of luck.

# The Full-Time Whistle

It was the former Liverpool manager Bill Shankly who spoke the immortal words, ‘Football is not a matter of life and death . . . It’s more important than that.’ These words are often interpreted as conveying how strongly fans feel about their team, or to explain the obsessions of players and managers. But they can be read in other ways. When Shankly spoke them in a TV interview in 1981, he was partly expressing regret that he was unable to properly enjoy life beyond football. He was describing an addiction to football that had clouded other parts of his life.

Something else that Shankly referred to as more important than life and death was the rivalry between Liverpool and Everton. He was talking about how, in some Merseyside families, half the members supported Liverpool and the other half Everton. Rival fans from the same family would tease one another, even when walking to the ground together on derby day. For Shankly, the fans saw in football something that transcended their day-to-day troubles, something that allowed them, for 90 minutes, to set aside the mundane problems posed by life and death.

Going through some old papers recently, I found a letter from my grandad, written when I was in hospital in Kirkcaldy when I was eight years old. It ended like this: ‘Do you think your rash could have been caused by wearing that RED Track Suit with L-------L on it? Perhaps you’d better throw it away and have one with EVERTON instead.’ If I could focus on the battle between Everton and Liverpool, then I could use that to keep my spirits up in my own battle to get better. That’s how he probably saw it. For my part, I simply couldn’t understand why anyone in his right mind would support Everton.

My grandad died of cancer the next year. Our family went down to Liverpool during the last weeks of his life. He was confined to the front room, and the adults formed a cordon to keep the children from disturbing him. I knew better. I knew that however ill he was, he would love to have a laugh and a joke about the Reds and the Blues. He’d written to me in hospital, and now it was him who needed cheering up. When most of the adults went out, I snuck in and managed to say something about Liverpool being best. He smiled and replied that they could never be as good as the Blues. The exchange didn’t last long, as I was caught by my Auntie May and removed before I could bother him any more, but I could still see him smiling as I was dragged from the room.

There is a pathos in this anecdote of a kid and his dying granddad, but to me it was much more than that. My granddad, and the rest of my family, have always been ready to talk and argue about anything and everything. Our Liverpool v Everton banter may not have been highbrow intellectual discussion, but it served an important purpose. Football gives people a subject to argue about and discuss, a common frame of reference to develop their thinking. Seeing my gran now, still a true Blue (strictly in footballing terms) at the age of 94 and having this same banter with my son Henry, who is 10 and has a signed Stevie G portrait on his bedroom wall, reminds me how football helps us communicate in so many different ways.

It was this love of discussion, of banter and of argument that led me to become a mathematician. A fascination with numbers was part of my interest in maths, but more important than that was a desire to communicate about the world. Football has helped me communicate about mathematics. And mathematics can help us all communicate about football. From gambling to strategy, from synchronised attack to six-second pressing, from team geometry to player statistics – mathematics can be found in every part of football, just as it can be found in every part of life.

It is that view of mathematics that I want to leave you with. We often hear that football and life can't be reduced to numbers, that 'number crunchers' can't replace common sense and knowledge. This I entirely agree with. Life is not mathematics, but maths gives us those small surprising insights into life that we may otherwise never have. In football, maths gives us a proper appreciation of how players create and narrow down space, the physics of goals, the nuances of tactics, team cooperation and passing distributions. It also allows us to make a few quid at the bookies, and to understand why sometimes we get lucky and other times we don't.

Mathematics should be grounded in everyday life, just as football already is. If mathematics sets itself up as superior to the everyday, then it becomes an easy target when the theory starts to go wrong. But if maths is integrated into football, into society and into science, it no longer makes sense to say that mathematicians are out-of-touch number crunchers. If we accept that we should approach football using our heads as well as our hearts, then it is obvious that mathematical thinking is part of the game.

I believe that mathematics, like football, is more than a game. It is a way of reasoning that will live on long after you and I are gone. But maths is not just an abstraction; it is about putting reasoning into practice. Maths has theory, it has application and it has passion. It is only when we combine all three that we get results.

So the next time you see a well-executed passing triangle, take a few seconds to think about how it works. Admire both the technical skill and the abstract shape it generates. See the emotion it arouses in the fans. These are all ways of looking at football, and they are all part of maths. It is when we put them together that we truly understand the beautiful game.

# Notes

## Chapter 1: I Never Predict Anything and I Never Will

1 Subbuteo appears to have all but disappeared from our culture, although I have heard that it might be about to enjoy a revival. It is a table football game, where you flick miniature playing figures with metal bases around a cloth mat.

2 The model matches pretty well, but there are differences between reality and the model. The chi-squared test statistic is based on

$$X^2 = \sum_{i=0}^{10} \frac{(O_i - E_i)^2}{E_i}$$

where $O_i$ is the number of games in which $i$ goals were scored and $E_i$ is the prediction of the model. The sum $X^2 = 26.3$, which is statistically significant at a 0.5% level with 10 degrees of freedom. This high value of $X^2$ occurs mainly because of the two 10-goal games seen during the season. According to the model, 10-goal games should occur only once every four years. If we group together all games with 9 or more goals in comparing model and data, then the sum $X^2 = 14.6$, which is significant at a 10% level with 9 degrees of freedom.

The other deviation between the model and the data is for 0–0 draws. I return to this later in the text.

3 The statistic for the NHL data is $X^2 = 19.6$, which is not statistically significant for the 13 degrees of freedom of the data.

4 A comprehensive history of the work of Bortkiewicz can be found on the StatProb Encyclopedia (statprob.com/encyclopedia/LadislausVonBortkiewicz.html). His book on the 'law of small numbers' and the application of the Poisson distribution is available in the original German at the California Digital Library (archive.org/details/dasgesetzderklei00bortrich).

5 Some of these examples are listed in more detail in Letkowski, J. 2012. Applications of the Poisson probability distribution. In *Proceedings of the 2012 Academic and Business Research Institute Conference, San Antonio*.

6 Tomasetti, C. & Vogelstein, B. 2015. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* 347(6217): 78–81.

7 In this model I use four parameters for each team: the average numbers of goals scored at home ($S_H$), goals conceded at home ($C_H$), goals scored away ($S_A$) and goals conceded away ($C_A$). These are estimated from the goals scored in the 2012/13 season. When two teams meet in the league in my simulated 2013/14 season, I first generate goals for the home team. These are Poisson-distributed with a mean equal to ½ ($S_H + C_A$), which takes into account both the attacking strength of the home side and the defensive strength of the visitors. The visitors' goals are determined by a Poisson distribution with a mean equal to ½ ($C_H + S_A$). The same procedure is repeated home and away for all teams to give a simulated season.

## Chapter 2: How Slime Moulds Built Barcelona

1 To connect up all 11 players requires at least 10 links between the players. A network connecting all the players together using exactly 10 links is called a spanning tree. To build the network shown, I first find the spanning tree which has the smallest total length. This is known as the minimal spanning tree, and it connects all the players using the shortest possible total distance. At a second stage, I calculate a new minimum spanning tree that doesn't include any links from the first spanning tree. The network shown is both these trees combined.

2 All the positions used here are adapted from Jonathan Wilson's excellent book *Inverting the Pyramid*: *The History of Football Tactics* (Orion Books, London, 2008). The book covers these and many other formations used through the history of football.

3 Finding the length here involves a few steps of trigonometry, but nothing more advanced than the SOHCAHTOA rule you learnt at school. First notice that the angles involved are 120˚. This means that the length of each of the four branches connected to the suburbs is

$$\frac{1/2}{\sin 60°} = 1/\sqrt{3}$$

applying SOH: sin(angle) = opposite/hypotenuse

Applying Pythagoras's rule, the middle length is then

$$1-2\sqrt{\left(\frac{1}{3}-\frac{1}{4}\right)} = 1-1/\sqrt{3}$$

Finally, adding up the middle lengths and the four branches gives

$$1-(1/\sqrt{3})+(4/\sqrt{3}) = 1+\sqrt{3}$$

4 Here are the solutions:

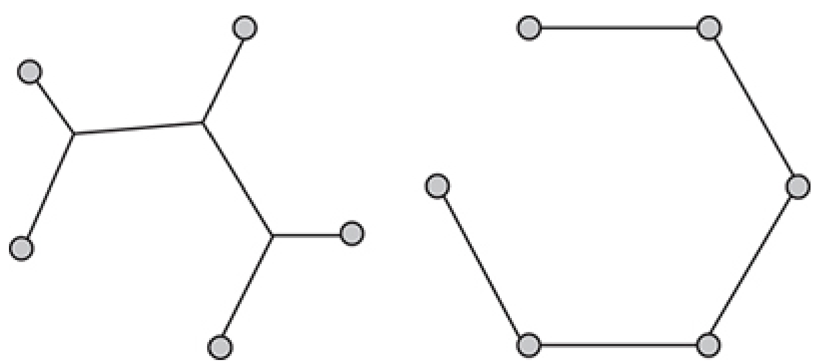


**Figure N.1** Optimal solution to the problem of connecting up five suburbs on a pentagon (left) and six suburbs on a hexagon (right) using the least possible track.

5 The full account of the work is in Tero, A. *et al.* 2010. Rules for biologically inspired adaptive network design. *Science* 327(5964): 439–442.

6 The zones I calculate here, as I explain later in the main text, are the sets of points that are closest to each player. So all the points in a player's zones are those that are closer to that player and no other. This partitioning is known as a Voronoi diagram, after the Ukrainian mathematician Georgy Voronoy, who made the first general study of them.

7 To calculate triangulations, we first use the Voronoi diagram to calculate zones. We then take the centre points of all the zones in the Voronoi diagram (*i.e.* the players) and draw links between them if they have neighbouring zones, to create a Delaunay triangulation. For the Barcelona network, the first and the second minimum-spanning tree between them contain most of the edges of the Delaunay triangulation. The Delaunay triangulations tend to maximise the angles in connecting networks, while the Voronoi diagram maximises zone sizes. We can switch interchangeably between the two: every Voronoi diagram has an equivalent Delaunay triangulation, and vice versa. So when we maximise angles, we maximise zones, and vice versa.

8 For a review see Sumpter, D. J. *et al.* 2012. The modelling cycle for collective animal behaviour. *Interface Focus* 2(6): 764–773.

9 Alvarez, G. A. & Franconeri, S. L. 2007. How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision* 7(13): 14.1–14.10.

10 Michels, R. 2001. *Teambuilding: The Road to Success*. Reedswain Publishing, Spring City, PA. Quoted from p. 88.

## Chapter 3: Check My Flow

1 The technical work in this experiment was done by Emil Rosen, a Masters student in my research group. Emil went on to write a thesis studying the movement patterns of my team: Rosen, E. 2016. Analysis of collective motion in youth football using GPS and visualization of collective motion in professional football. Master's Thesis, DiVA, Uppsala University.

2 Moussaïd, M. *et al*. 2009. Experimental study of the behavioural mechanisms underlying self-organization in human crowds. *Proceedings of the Royal Society B: Biological Sciences*. DOI: 10.1098/rspb.2009.0405.

3 The arrows in Figure 3.4 show only the effect of the stationary student, and not the general tendency of the moving student to walk forward. In an empty corridor, a moving student walks forward at a relatively constant rate. To create the figure, this forward motion in the absence of a stationary student is subtracted from the movement measured in the presence of the stationary student, thus giving the overall effect of the stationary student.

4 In 148 experimental trials, the walking student passed to the left 60 times (40.6%) and to the right 88 times (59.4%).

5 'Badstuber on defending one-on-ones', UEFA training ground videos, www.uefa.com/trainingground/skills/video/videoid=1654613.html.

6 Pan, S. *et al*. 2012. Pursuit, evasion and defense in the plane. In *American Control Conference*. IEEE, New York, 4167–4173.

7 Stander, P. E. 1992. Cooperative hunting in lions: The role of the individual. *Behavioral Ecology and Sociobiology* 29(6): 445–454.

8 The paper detailing the method is Ranganathan, S. *et al*. 2014. Bayesian dynamical systems modelling in the social sciences. *PloS one* 9(1): e86468. The basic method is to fit the direction of the pass as a polynomial function of the position on the pitch. Both the $dx$ and $dy$ components of pass direction are fitted as a function of the $x$ and $y$ position. This gives a good idea of the average direction of the passes, although it fails to capture variation in pass direction.

## Chapter 4: Statistical Brilliance

1 The figure of 265 million comes from FIFA's 'Big Count' in 2007: www.fifa.com/mm/document/fifafacts/bcoffsurv/emaga_9384_10704.pdf. FIFA admit that this number is only a rough estimate, and I'm not even sure how it could be reliably measured. But, given the worldwide popularity of football, it sounds reasonable enough.

2 There are a few exceptions to this. The 1986/87 season ended with a round-robin tournament between the top teams, who consequently each played 44 matches in total, and the 1995/96 and 1996/97 seasons each involved 22 teams. Otherwise, since the 1986/87 season La Liga has involved 20 teams.

3 Playing guessing games isn't foolproof. A good example of where they fall down is in calculating the probability that I'll drop dead tomorrow. I am alive today, I was alive yesterday, and I have been alive for roughly 15,000 days. So, following the guessing-game logic, the chance that I'll drop dead tomorrow is less than 1 in 15,000. When I'm 100 years old, this probability will be only 1 in 36,525. The longer I live, the less likely I am to die! All models have their limitations, and the modeller must be aware of them.

4 Whether a player has scored a particular goal is not always black and white. La Liga's official statistics award 40 goals to Ronaldo. But the experts who judge the Pichichi Trophy decided that a goal that had been attributed to Pepe should have been credited to Ronaldo – hence the 41st goal.

5 The form presented here is the Gumbel distribution. The probability of the top goalscorer scoring $G$ goals or fewer is

$$\exp\left\{\exp\left(-\left(\frac{G-b}{a}\right)\right)\right\}$$

where $a = 5.44$ and $b = 26.9$ are parameters estimated from the data. The Gumbel distribution is the most common form of extreme value distribution and the one that typically fits data best, because it fits the case where the sample distribution has an exponential tail.

6 Here we calculate the probability of the top scorer scoring 50 or more goals in a season, rather than the probability of him scoring exactly 50 goals.

7 From Sterl, A. *et al.* 2009. An ensemble study of extreme storm surge related water levels in the North Sea in a changing climate. *Ocean Science* 5(3): 369–378.

8 These and other predictions are made in Van den Brink, H. W. & Können, G. P. 2011. Estimating 10000-year return values from short time series. *International Journal of Climatology* 31(1): 115–126.

9 The full report by the IPCC is entitled 'Special report on managing the risks of extreme events and disasters to advance climate change adaptation (SREX)' and can be found at ipcc-wg2.gov/SREX/report/.

10 The most comprehensive study of weather extremes is Alexander, L. V. *et al.* 2006. Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research: Atmospheres* 111(D5). DOI: 10.1029/2005JD006290.

11 www.premierleague.com/en-gb/players/ea-sports-player-performance-index/what-is-the-ea-sports-ppi.html.

12 The full paper is: McHale, I. G. *et al.* 2012. On the development of a soccer player performance rating system for the English Premier League. *Interfaces* 42(4): 339–351.

13 www.premierleague.com/en-gb/players/ea-sports-player-performance-index.html.

14 Lewis, M. 2004. *Moneyball: The art of winning an unfair game*. W. W. Norton & Company, London.

15 Anderson, C. & Sally, D. 2013. *The Numbers Game: Why everything you know about football is wrong*. Penguin, London.

16 www.optasportspro.com/about/optapro-blog/posts/2015/film-optapro-forum-beyond-shots/.

17 www.theguardian.com/football/2015/jul/27/how-fc-midtjylland-analytical-route-champions-league-brentford-matthew-benham.

## Chapter 5: Zlatan Ibrah Rocket Science

1 See for yourself: www.youtube.com/watch?v=yzvQCbdAIZQ.

2 Here are the equations of motion. Zlatan is nearly 2 metres tall, and when he turns upside down he rotates pretty much through 180˚, so the initial height of the ball when kicked is also about 2 metres. The ball is kicked with an initial velocity $v$ at an angle $\theta$. The initial upward velocity is then $v \sin \theta$, and the height of the ball over time $t$ is

$$z(t) = 2 + vt\sin\theta - \frac{1}{2}gt^2$$

where $g = 9.8\ \text{m/s}^2$ is the acceleration due to gravity. Assuming no air resistance, the distance from the goal is determined by

$$x(t) = 27 - vt\cos\theta$$

where 27 metres is the initial distance to the goal.

3 The ball reaches the goal when $x(t) = 0$, which happens at time

$$t = \frac{27}{v \cos \theta}$$

The height of the ball at this point is

$$z\left(\frac{27}{v\cos\theta}\right) = 2 + 27\frac{\sin\theta}{\cos\theta} - \frac{729g}{2v^2\cos^2\theta}$$

The question is, for which velocity $v$ will this final height be between 0 and 2.44 metres (the height of the goal)? Rearranging this equation we find that

$$\sqrt{\frac{729g/2}{2\cos^2\theta + 27\sin\theta\cos\theta}} < v < \sqrt{\frac{729g/2}{(2-2.44)\cos^2\theta + 27\sin\theta\cos\theta}}$$

is the condition for the ball to go in. This is the basis for the plots in Figure 5.3 on page 93. Even though there are cosines, sines and square roots flying around everywhere, all this is still just standard secondary-school geometry. The difficulty in solving this problem lies in keeping track of what you are trying to find out. Then it is just a matter of moving the symbols around to solve the problem.

4 www.grc.nasa.gov/WWW/k-12/airplane/soccercode.html.

5 For a more detailed review of football motion and analysis of Gerrard's free-kicks, see Goff, J. E. 2010. Power and spin in the beautiful game. *Physics Today* 63(7): 62–63.

6 Hong, S. & Asai, T. 2014. Effect of panel shape of soccer ball on its flight characteristics. *Scientific Reports* 4: 5068. DOI: 10.1038/srep05068.

## Chapter 6: Three Points for the Bird-brained Manager

1 This discussion is based on two papers:

Smallegange, I. M. & Van Der Meer, J. 2007. Interference from a game theoretical perspective: Shore crabs suffer most from equal competitors. *Behavioral Ecology* 18(1): 215–221.

Smallegange, I. M. *et al*. 2007. Assessment games in shore crab fights. *Journal of Experimental Marine Biology and Ecology* 351(1): 255–266.

2 Nagy, M. *et al*. 2013. Context-dependent hierarchies in pigeons. *Proceedings of the National Academy of Sciences* 110(32): 13049–13054.

3 I do a Wilcoxon rank sum test with null hypothesis that there is no difference between the number of draws before and after the change to 3 points. The $p$-value is 0.0108.

4 We set up this problem as follows. Assume that if both teams attack, then the probability of your team winning is $w$, the probability of drawing is $d$ and the probability of losing is $l$. If one team defends (it doesn't matter which), then the probability of winning and losing is multiplied by $p$, which is the effectiveness of the defence. This gives the following probabilities of winning, drawing and losing:

$$pw,\ d + (1-p)w + (1-p)l,\ pl$$

Under an $x$-point system, the weaker team should defend whenever

$$xpw + d + (1-p)w + (1-p)l > xw + d$$

Rearranging, we get

$$(1-p)w + (1-p)l > xw\,(1-p)$$

The $(1-p)$ terms cancel, and we are left with the condition

$$l > (x-1)\,w$$

For a 2-point system, $x = 2$, and the condition for defending is $l > w$. So the weaker team should always defend. For a 3-point system $x = 3$, and the condition for defending is $l > 2w$. So the weaker team should defend only if their probability of losing when playing attack football is twice that for defending.

A similar argument can be made for the case where both teams defend. The algebra is messier, but the conclusion is the same.

5 Yasukawa, K. & Bick, E. I. 1983. Dominance hierarchies in dark-eyed juncos (*Junco hyemalis*): A test of a game-theory model. *Animal Behaviour* 31(2): 439–448.

6 Kianercy, A. *et al.* 2014. Critical transitions in a game theoretic model of tumour metabolism. *Interface Focus* 4(4): 20140014. DOI: 10.1098/rsfs.2014.0014

# Chapter 7: The Tactical Map

1 Grund, T. U. 2012. Network structure and team performance: The case of English Premier League soccer teams. *Social Networks* 34(4): 682–690.

2 When teams have the ball they usually pass faster than three or five times a minute. These passing rates look low because the measure of possession used in the study includes time when the ball is out of play. For a full definition of passing rate, see http://optasports.com/news-area/blog-optas-event-definitions.aspx.

3 http://espn.go.com/espnw/athletes-life/the-buzz/article/13173839/uswnt-teammates-unleash-their-abby-wambach-impressions.

4 There are more connecting lines in this network than those made for Italy, England and Spain above. That is because in my plots I only showed passing pairs with a threshold of 11 or more passes for Euro 2012, and here the threshold is 5 passes.

5 For full details of these definitions and the rest of the study, see Bearman, P. S. *et al.* 2004. Chains of affection: The structure of adolescent romantic and sexual networks. *American Journal of Sociology* 110(1): 44–91.

6 www.fourfourtwo.com/statszone/.

7 More specifically, the convex hull is the smallest shape for which we can draw a straight line from any of the black circles to any other without leaving the shape itself.

8 ‘Reasonably near’ means within one standard deviation in both the direction left/right across the pitch and up/down the pitch.

# Chapter 8: Total Cyber Dynamo

1 The model I use here comes from evolutionary game theory, first proposed by John Maynard Smith in *Evolution and the Theory of Games* (Cambridge University Press, 1982). Following his approach. I assume that the population of shirkers has a growth rate proportional to the pay-off for shirking minus the average pay-off in the population. This gives the following replicator equation:

$$\frac{dx}{dt} = x(1-x)\left[2x + 0.5(1-x) - 3x\right]$$

2 Assume you would read 30 minutes for your own daughter (who shares half your genes). Then 30 divided by 16 is 1.875 minutes.

3 West, S. A. *et al*. 2011. Sixteen common misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior* 32(4): 231–262.

4 Quoted from an article written by Jonathan Wilson for the *Guardian*: www.theguardian.com/football/blog/2011/may/ 12/valeriy-lobanovskyi-dynamo-kyiv.

5 The three curves link performance, $p$, with effort, $x$. The linear curve is $p = 10x$, the sub-linear curve is $p = 32\sqrt{x}$, and the super-linear curve is $p = x^2$.

6 Beekman, M. *et al*. 2001. Phase transition between disordered and ordered foraging in Pharaoh's ants. *Proceedings of the National Academy of Sciences* 98(17), 9703–9706.

7 The star who plays with 90% effort will now get (10.9 × 10.9/11) – 0.9 = 9.9, which is also less than the 10 obtained with 100% effort.

8 Kormelink, H. and Seeverens, T. 2003. *The Coaching Philosophies of Louis van Gaal and the Ajax Coaches*. Reedswain Publishing, Spring City, PA. Quoted from p. 5.

9 Michels, R. 2001. *Teambuilding: The Road to Success*. Reedswain Publishing, Spring City, PA. Quoted from p. 117.

10 A good starting point for the scientific study of cooperation is: Nowak, M. A. 2006. Five rules for the evolution of cooperation. *Science* 314(5805): 1560–1563. What I have called cooperation in families is called 'kin selection' by Nowak, and my model of super-linear teams he calls group selection. But this paper is controversial in its categorisation, and if you are interested in the subject you should read widely about different theories.

# Chapter 9: The World in Motion

1 Duarte, R. *et al*. 2012. Sports teams as superorganisms: implications of sociobiological models of behaviour for research and practice in team sports performance analysis. *Sports Medicine* 42(8): 633–642.

2 Sumpter, D. J. 2006. The principles of collective animal behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences* 361(1465): 5–22.

3 Balague, N. *et al*. 2013. Overview of complex systems in sport. *Journal of Systems Science and Complexity* 26(1): 4–13.

4 Mutschler, C. *et al*. 2013. The DEBS 2013 grand challenge. *Proceedings of the 7th ACM International Conference on Distributed Event-Based systems*. Association for Computing Machinery, New York.

5 The process of working out the average positions involves smoothing out the data. Details can be found in: Bialkowski, A. *et al*. 2014. Identifying team style in soccer using formations learned from spatiotemporal tracking data. *2014 IEEE International Conference on Data Mining Workshop*. Institute of Electrical and Electronics Engineers, New York, 9–14.

6 Lucey, P. *et al*. 2015. 'Quality vs quantity': Improved shot prediction in soccer using strategic features from spatiotemporal data. *9th Annual MIT Sloan Sports Analytics Conference*.

7 Vicsek, T. *et al*. 1995. Novel type of phase transition in a system of self-driven particles. *Physical Review Letters* 75(6): 1226–1229.

8 Dyer, J. R. G. *et al*. 2009. Leadership, consensus decision making and collective behaviour in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364(1518): 781–789.

9 I was part of a team that did one of the first such tests. The paper we wrote on this is Buhl, J. *et al*. 2006. From disorder to order in marching locusts. *Science* 312(5778): 1402–1406.

10 Folgado, H. *et al*. 2014. Competing with lower level opponents decreases intra-team movement synchronization and time-motion demands during pre-season soccer matches. *PLoS One*: e97145.

11 You may be wondering who the 'Portuguese first team' and the 'Premier League team' are. They aren't specified in the paper, but there are a few clues. The coach, Pedro Caixinha, is thanked for the Portuguese team and a Manchester City analyst is a co-author on the Premier League paper. See Folgado, H. *et al*. 2015. The effects of congested fixtures period on tactical and physical performance in elite football. *Journal of Sports Sciences* 33(12): 1238–1247.

12 Biro, D. *et al*. 2006. From compromise to leadership in pigeon homing. *Current Biology* 16(21): 2123–2128.

13 Pettit, B. *et al*. 2013. Interaction rules underlying group decisions in homing pigeons. *Journal of the Royal Society Interface* 10(89): 20130529.

14 Nagy, M. *et al*. 2010. Hierarchical group dynamics in pigeon flocks. *Nature* 464(7290): 890–893.

15 www.fourfourtwo.com/performance/tactics/luis-enrique-how-play-pressing-game.

16 https://vimeo.com/124521118.

17 Rosenthal, S. B. *et al*. 2015. Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion. *Proceedings of the National Academy of Sciences* 112(15): 4690–4695.

18 There is also a wide range of excellent blogs to read on this. Check out 'Statsbomb', 'Analytics FC', 'Differentgame' and 'Scoreboard Journalism' for just some examples.

## Chapter 10: You'll Never Walk Alone

1 In reality, bacteria require at least an hour to get going with their reproduction, and they may find it hard to spread within the steak. But once they do get going, then if we expect them to double every 20 minutes, there will be 8 times as many within an hour, and after 7 hours one bacterium will have become 8 × 8 × 8 × 8 × 8 × 8 × 8 = 2,097,152.

2 Edelstein-Keshet, L. 1988. *Mathematical Models in Biology.* Society for Industrial and Applied Mathematics, Philadelphia, p. 152.

3 Here I give a more detailed derivation of the logistic growth curve. Under the two assumptions, the total rate at which singing starts on the next round is proportional to $2X(N - X)$, where $N$ is the number of fans and $X$ is the number who have already started to sing. This rate of increase is plotted below.

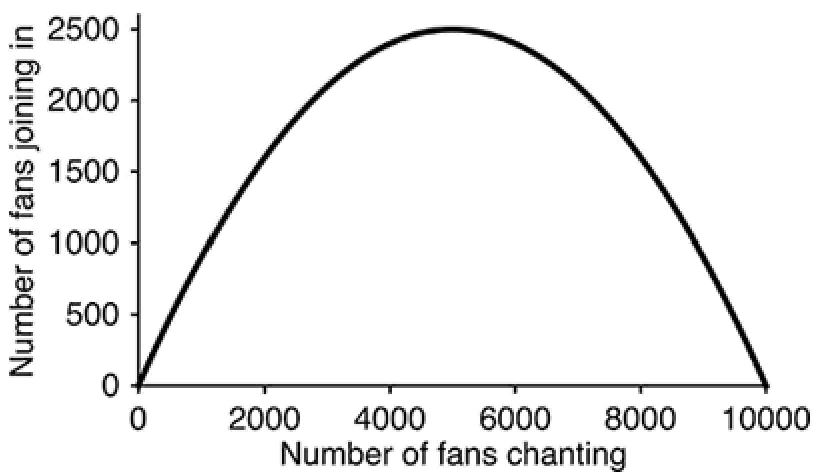



**Figure N.2** Rate of increase of chanting as a function of those already chanting for logistic growth.

When $X$ is small, singing spreads slowly because there are fewer singers to copy. In addition, when $X$ is near to $N$ the singing spreads slowly, because there are fewer people available to join in. We can also see from the curve that the maximum increase occurs at the point $N/2$, when half the fans are singing and the other half haven't started yet.

4 Mann, R. P. *et al*. 2013. The dynamics of audience applause. *Journal of the Royal Society Interface* 10(85): 20130466.

5 The search data is taken from www.google.co.uk/trends. I collected the newspaper data myself from the *Guardian* website.

6 Farkas, I. *et al*. 2002. Social behaviour: Mexican waves in an excitable medium. *Nature* 419(6903): 131–132.

7 Herbert-Read, J. E. *et al*. 2015. Initiation and spread of escape waves within animal groups. *Royal Society Open Science* 2(4): 140355.

8 Farkas, I. J. & Vicsek, T. 2006. Initiating a Mexican wave: An instantaneous collective decision with both short- and long-range interactions. *Physica A: Statistical Mechanics and Its Applications* 369(2): 830–840.

9 I would like to reassure the research-granting bodies who fund our research that we did not fly to Sydney just to watch the cricket. We did lots of other experiments out there. Teddy's parents paid for the tickets (thank you, Trevor and Nicola), and all beer consumption was privately funded.

10 Silverberg, J. L. *et al*. 2013. Collective motion of humans in mosh and circle pits at heavy metal concerts. *Physical Review Letters* 110(22): 228701.

11 See Matt Bierbaum's simulator at mattbierbaum.github.io/moshpits.js/.

12 Silverberg *et al.* go on to discuss more advanced moshing behaviour such as the wall of death, where moshers separate into a ring and then come running into the middle to crash into one another. These can't be explained in the simple version of the model.

13 For a review see: Auf der Heide, E. 2004. Common misconceptions about disasters: Panic, the 'disaster syndrome', and looting. In *The First 72 Hours: A Community Approach to Disaster Preparedness*. iUniverse, Bloomington, Indiana, p. 337.

14 This description only touches on the full range of pedestrian behaviours identified by Mehdi and his collaborators. A full description of the experiments and the results can be found at his webpage, www.mehdimoussaid.com/archives/53.

15 These stop/start waves and the subsequent 'turbulence' are detailed in Helbing, D. *et al*. 2007. Dynamics of crowd disasters: An empirical study. *Physical Review E* 75(4): 046109.

16 See www.pedestrian-dynamics.com/pedestrian-dynamics/pedestrian-dynamics-features.html for details of the methods used.

17 Helbing, D. & Mukerji, P. 2012. Crowd disasters as systemic failures: Analysis of the Love Parade disaster. *EPJ Data Science* 1(1): 1–40.

18 Sadly, after I wrote this, a similar disaster did happen again at the Hajj, in September 2015, and this time over 700 people lost their lives.

## Chapter 11: Bet Against the Masses

1 The average is the sum of all the guesses divided by the number of guesses, while the median is the guess in the middle. So of the 19 students, 9 of them guessed that there were fewer than 90 sweets and 9 of them guessed that were more than 90. The one student who guessed 90 is the median guess.

2 Surowiecki, J. 2005. *The Wisdom of Crowds*. Anchor, New York.

3 Surowieki's book starts with a description of an experiment conducted by Francis Galton at a country fair, where the average of 787 guesses of the weight of a living ox was very close to the correct value. The original paper is Galton, F. 1907. Vox populi (the wisdom of crowds). *Nature* 75(1949): 450–451.

4 The maximum guess was set in advance at 1,500.

5 King, A. J. *et al*. 2012. Is the true wisdom of the crowd to copy successful individuals? *Biology Letters* 8(2): 197–200.

6 In Andrew King's experiment there was no limit to the size of the guess allowed. So one participant guessed over 10,000, and another two guessed over 5,000. These made the average (mean) guess unreliable, but it didn't affect the median. The difference between average and median is made clear in note 1 to this chapter.

7 The average (mean) is calculated as

$$\frac{0+1+2+3+4+5+6+7+8+9+10+11+12+13+14+15+16+17+18+19+20+21+22}{23}=11$$

and the median is 11.

8 Berg, J. E. *et al*. 2008. Prediction market accuracy in the long run. *International Journal of Forecasting* 24(2): 285–300.

9 http://prosoccertalk.nbcsports.com/2014/08/14/pst-writers-predict-the-2014–15-premier-league-standings-do-you-agree/.

10 You can read Simon's analysis and results on his blog: https://scoreboardjournalism.wordpress.com.

11 That is

$$\frac{\frac{0+0+0+0+0+0+3+5+1+2+4+2}{+5+5+6+0+2+7+1+3}}{20} = 2.3$$

Note that absolute differences are taken here: e.g. a difference of –4 is taken to be 4.

12 I assume that promoted teams are ranked 18, 19 and 20 depending on their position in the Championship the year before.

13 Together with Dutch advice bureau Hypercube.

14 We assume for simplicity that the odds are the same for both teams.

15 As above, the probability of both making the same prediction and getting it wrong is

$$\frac{0.3\times 0.3}{(0.3\times 0.3)+(0.7\times 0.7)} = 0.155$$

## Chapter 12: Putting My Money Where My Mouth Is

1 You are welcome to go and buy this masterpiece from Princeton University Press. I still get a small percentage on every book sold! Sumpter, D. J. T. 2010. *Collective Animal Behavior*, Princeton University Press, Princeton, NJ.

2 For Chelsea, (0.25 × 3.4) + (0.75 × 0) = 0.85, and for Arsenal, (0.3375 × 3.1) + (0.6625 × 0) = 1.046.

3 The equation for how much cash I have left is $10 \times 0.98^{w}$, where $w$ is the number of weeks I have been betting.

4 The odds I describe here were based on those provided at www.oddsportal.com, where you can find best closing odds at kick-off for every match for a range of leading bookmakers. I have then adjusted the odds to make them fair, with the probabilities adding up to 100%.

5 1/1.33 = 75% and 1/1.43 = 70%. 1 divided by the odds is the probability the bookmakers assign to a win.

6 This result is statistically significant. The probability that 25 or more wins out of 28 would occur if the probability of a win is 72.5% is 0.01.

7 This technique is described in more detail in Jakobsson, R. & Karlsson, N. 2007. Testing market efficiency in a fixed odds betting market. Working paper No. 12, Department of Statistics, Örebro University. The best-fit model by logistic regression is

$$P(\text{home win}) = \frac{1}{1+0.961\left(\frac{p}{1-p}\right)^{-1.06}}$$

where $p$ is the bookmakers' probability for a home win.

8 This research is reviewed in: Innocenti, A. *et al*. 2012. The importance of betting early. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1999459.

9 Here I define a match as one in which neither team has the upper hand if the difference between the probabilities of each team winning is less than 10 percentage points. There were 72 such matches, and 25 resulted in a draw. This is not a statistically significant difference from the null hypothesis that the odds predict the match. It has a *P*-value of 0.14, but it is a sufficiently interesting difference to test as a strategy.

10 The table below gives three examples from the first day of the 2015/16 season.

| Match | Probability of home win (from odds) | Probability of away win (from odds) | Difference | Bet suggested by strategy |
|---|---|---|---|---|
| Manchester United v Newcastle United | 0.6 (odds: 1.65) | 0.16 (odds: 6.00) | 0.44 | Home win |
| Norwich v Crystal Palace | 0.37 (odds: 2.63) | 0.34 (odds: 2.90) | 0.03 | Draw |
| Leicester v Sunderland | 0.50 (odds: 1.97) | 0.22 (odds: 4.38) | 0.28 | No bet |

11 This result is described in detail in Thomas Grund's paper discussed in Chapter 7: Grund, T. U. 2012. Network structure and team performance: The case of English Premier League soccer teams. *Social Networks* 34(4): 682–690.

12 Based on a study by the performance analysis company Prozone. See www.theguardian.com/football/blog/2014/apr/27/bayern-munich-possession-football.

13 http://www.fourfourtwo.com/statszone.

14 We have already discussed this point in Chapter 7, and the analysis it was based on can be found in Thomas Grund's paper (see note 11 to this chapter).

15 The rankings are calculated as

$$0.13 \times (\text{passing rate}) + 0.76 \log (\text{expected goals for})$$

which was shown by logistic regression to give the best prediction of match outcome.

16 These passing rates are underestimated since they include time when the ball is dead, e.g. out for a throw-in or when a player is down injured. They do, however, reflect the relative passing rates of the teams and can be safely used in fitting the model.

17 We set the odds as follows. Let the bookmakers' probability of one team winning be $p$, and the probability of the other team winning be $q$. Then if $|p - q| > 0.4$, we set the strategy's probability of winning as

$$\frac{1}{1 + 0.961\left(\frac{p}{1-p}\right)^{-1.06}}$$

(see note 7). Similarly, if $|p - q| < 0.15$, then the probability of a draw is set to

$$0.355 - 0.25\,|p - q|$$

All other result probabilities are then taken from the bookmakers' odds and adjusted appropriately to reflect the new draw probability. For

$$0.15 \leq |p - q| \leq 0.4$$

the probabilities of winning are set to be the same for the strategy as the original odds and no bet is placed.

18 www.euroclubindex.com/asp/matchodds.asp.

19 I go on to make a small adjustment, because the Poisson model slightly underestimates draws, based on the calculation in note 16. The other probabilities are adjusted so that they add up to 1.

20 I made both goals-for and goals-against models, and found the following best-fit models using Poisson regression. The average number of goals per match for the home team is estimated to be

$$\exp(-0.7574 + 0.13\, r + 0.76\mathrm{h})$$

where $r$ is the mean passing rate and $h$ is the average number of goals the home team is expected to have scored in previous matches. The average number of goals per match for the away team is estimated to be

$$\exp(-0.0784 + 0.5057\, g + 0.5527 b)$$

where $g$ is average number of goals the home team is expected to have conceded in previous matches, and $b = 1$ if the visiting team is Chelsea, Manchester City, Manchester United, Arsenal, Liverpool or Spurs, and $b = 0$ for all other visitors. The variable $b$ accounts for a kind of statistical magic these teams have of scoring more goals away than might be expected. I don't have an explanation for this, but it does fit the data better.

## Chapter 13: The Results Are In


1 Kelly Jr, J. L. 1956. A new interpretation of information rate. *IRE Transactions on Information Theory* 2(3): 185–189.

2 Spann, M. & Skiera, B. 2009. Sports forecasting: A comparison of the forecast accuracy of prediction markets, betting odds and tipsters. *Journal of Forecasting* 28(1): 55–72.

3 Andersson, P. *et al.* 2005. Predicting the World Cup 2002 in soccer: Performance and confidence of experts and non-experts. *International Journal of Forecasting* 21(3): 565–576.

4 http://fivethirtyeight.com/datalab/nfl-elo-ratings-are-back/

5 Hvattum, L. M. & Arntzen, H. 2010. Using ELO ratings for match result prediction in association football. *International Journal of Forecasting* 26(3): 460–470.

6 www.espnfc.com/fifa-world-cup/story/1873765/soccer-power-index-explained.

7 http://fivethirtyeight.com/datalab/introducing-nfl-elo-ratings/


8 The probability of a Lucky Luke having a capital of £240 or more over 90 matches is 3.6%, giving us 96.4% certainty that the model beats a random betting strategy. The usual benchmark for statistical significance is 95%.


9 Jakobsson, R. & Karlsson, N. 2007. Testing market efficiency in a fixed odds betting market. Working paper No. 12, Department of Statistics, Örebro University.


## Chapter 14: Finding the Talent


1 www.theguardian.com/football/2014/oct/17/arsenal-place-trust-arsene-wenger-army-statdna-data-analysts

2 www.optasportspro.com/about/optapro-blog/posts/2015/blog-inside-leicester-city/

3 www.bbc.com/sport/football/31365153

4 www.telegraph.co.uk/football/2016/02/05/tim-sherwood-its-not-seen-as-sexy-to-sign-players-from-lower-lea/


5 I discuss expected goals in Chapter 12, and will look at them again in Chapter 15.


6 www.espnfc.com/blog/five-aside/77/post/2476427/top-european-league-player-stats-of-the-2014–15-season

7 www.myajc.com/news/sports/pro-sports/meet-the-woman-behind-atlanta-uniteds-data/nq8ym/

8 www.statsbomb.com/2016/04/understand-football-radars-for-mugs-and-muggles/

9 The Kante analysis was shared in a Tweet from Ted’s account @mixedknuts. Follow him for regular player radar updates.

10 I discuss this point in more detail at the end of Chapter 4.

11 See pages 246–247 for an example.

12 www.optasportspro.com/about/optapro-blog/posts/2013/blog-manchester-united-shooting-above-the-norm/

13 My question and Sam’s answer were somewhat prescient. Since the book was completed, and after Villa’s failure to make an impact in the Championship season 2016–17, Sam is now *former* Head of Research and starting a new challenge in football analytics.

14 Sarah’s talk can be seen at www.metacafe.com/watch/ 7337475/2011_nessis_talk_by_sarah_rudd/

15 The probabilities in this model are for illustrative purposes, and slightly overestimate the probability of scoring from each area in a real match.

16 Imagine the following model: the pitch is broken down in to 11 positional states; there could be between 1 and 11 players between an attacker and the goal giving 11 defending states; and an attack is either a counter or a possession giving two states. The total number of states for this model is then 11×11×2=242 states.

17 statsbomb.com/2016/08/unpacking-packing/

18 statsbomb.com/2016/08/towards-a-new-kind-of-analytics/

19 The blog deepxg.com details football Thom’s football work.

20 www.telegraph.co.uk/football/2016/08/24/why-data-analyst-craig-kline-will-have-final-say-at-fulham-and-n/

## Chapter 15: Football’s Intelligent Future

1 Vestberg, T., Gustafson, R., Maurex, L., Ingvar, M., & Petrovic, P. (2012). Executive functions predict the success of top-soccer players. *PloS one*, *7*(4), e34731.

2 www.spox.com/myspox/group-blogdetail/Raumaufteilung-I,176891.html

3 www.spielverlagerung.com/2014/09/16/the-half-spaces/

4 www.youtube.com/watch?v=fB9jpteTgH0

5 For a more detailed analysis, including plots for headers and for shots resulting from crosses, see Ted Knutson’s article: www.statsbomb.com/2016/04/explaining-and-training-shot-quality/

6 Morales, C. A. (2016). A mathematics-based new penalty area in football: tackling diving. *Journal of sports sciences*, 1–5.

7 Statistics collected by Opta. Based on seasons 2014–15 and 2015–16 Champions League, Premier League, La Liga and Bundesliga.

8 www.bbc.com/sport/football/35955616

9 See notes 1 and 2 to Chapter 7.

10 www.statsbomb.com/2016/04/leicester-city-need-for-speed/

11 www.economist.com/blogs/gametheory/2015/12/competitive-balance-football

12 www.spielverlagerung.com/2016/08/14/guardiolas-manchester-city-narrowly-beat-sunderland/

13 www.telegraph.co.uk/sport/football/teams/manchester-city/10077840/Manchester-City-chief-Ferran-Soriano-on-Guardiola-Pellegrini-squad-tensions-and-becoming-best-in-Europe.html

14 www.optasportspro.com/about/optapro-blog/posts/2016/blog-hackmcfc-review/

# Acknowledgements

This book would never have been written without Chris Wellbelove. Thank you, Chris, for the ideas, the feedback throughout the process and your faith in me. Thanks also to the members of my ‘reading group’ – Dora Biro, Teddy Herbert-Read, Simon Gleave, Philip Maini, Shyam Ranganathan, Eileen and John Sumpter and Ingemar Törnblom – for all of your honest feedback. Thanks also to the members of the collective behaviour group in Uppsala, who always have time for random discussions with me about football and everything else.

During the writing process, I discovered a community of football nerds who think in the same way I do. Conversations on Twitter and elsewhere have been invaluable to me. In particular, I would like to thank Ryan Bahia, Andrew Beasley, Michael Caley, Michael Cox, Emma Faldt, Thomas Grund, Johannes Harkins, Sean Ingle, Robin Jacobsson and Oddscraft, Thom Lawrence, Ian McHale, Colm McMullan, Rui Marcelino, Mehdi Moussaid, Mate Nagy, Selina Pan, Devin Pleuler, Paul Power, the guys at Analytics FC, Gregor McGregor and the staff at *FourFourTwo* for useful feedback.

I would like to thank Frank, Elias and Edvin for their time doing training experiments, and also the rest of the boys and trainers at Upsala IF P05 for reminding me each week what football is really about. Tack så mycket!

At Bloomsbury I would like to especially thank Jim Martin for understanding the idea behind the book straight away, and Anna MacDiarmid for all of her work in the later stages.

Thank you to Emil Rosen, first for writing his rather unusual Masters project on football with me and then for all the help with the figures. Thanks to John Woodruff for a very detailed and complete copy-editing. Thank you to the Mathematics Department at Uppsala University for providing such as creative working environment.

There are a few glimpses of my family in this book, which I hope reflect how much I value you. Thanks to my mum and dad, my gran, Ruth and Colin and your families, for being there. Finally, thank you Elise for the figure drawing, Henry for the football playing, and Lovisa for the gambling; thank you all three for the balance you give me in life.

# Index


AC Milan here
Accrington Stanley here
Agbonhalor, Gabby here
Agüero, Sergio here
Ajax here, here
Alba, Jordi here
Alonso, Xabi here, here
Altman, Dan here
American football here, here
analytics here
Ancelotti, Carlo here, here, here
Anderson, Chris *The Numbers Game* here
Anelka, Nicolas here
animal movement here, here, here, here
ant colonies here
Argentina World Cup 2014 here
Arntzen, Halvard here
Arsenal here, here, here, here, here, here, here
  betting on here, here, here, here, here, here
  Luis Suárez non-transfer here
Aston Villa here, here, here, here, here, here
Atlético Madrid here
Attack/Defend model here

Ba, Demba here
Badstuber, Holger here, here, here
Baines, Leighton here
Balague, Natalie here
Bale, Gareth here, here, here
ball aerodynamics here
Balotelli, Mario here
Balzaretti, Federico here
Barcelona here, here, here, here, here
  Barcelona 2010/11 here, here
  La Masia here, here, here
  Panathinaikos 2010 here, here
  tactics here
  tiki-taka here, here
Barry, Gareth here
baseball here, here
Bayern Munich here, here, here, here, here, here
  Champions League 2013 here
  Champions League 2015 here
  tactics here
Beckham, David here, here
Beekman, Madeleine here, here
Benham, Matthew here, here

Benzema, Karim here, here, here
betting here, here, here
  around the corner here
  Barcelona here, Kilmarnock 1,093 here
  beating the crowd here
  betting channels here
  bore draws here
  career advice here
  challenge to make money here
  Cup Final rumour day here
  expert's advice here
  how good do I have to be? here
  indexical decline here
  Lucky Luke and Calamity Jane here
  not so smart here
  odd probabilities here
  partner's advice here
  Premier predictions here
  satisfied with a draw here
  strong favourites, well-matched draws here
  unexpected goals here
  when should I place a bet? here
  wise crowd here
Bialkowski, Alina here, here
bicycle lob here
  luck, structure and magic here
  Newton explains here
  sent it into space here
Bielsa, Marcelo here
Biro, Dora here, here, here, here
Bolt, Usain here, here
Bonucci, Leonardo here
Bortkiewicz, Ladislaus here
Borussia Dortmund here
Bournemouth here, here
Brazil World Cup 2014 here, here
Brentford here, here
brilliant players here
  objective rankings here
  once in every lifetime here
Burnley here, here, here
Busquets, Sergio here, here, here

Caley, Michael here
Campbell, Sol here
cancer here, here, here
Carragher, Jamie here, here
Carroll, Andy here, here
Cassano, Antonio here
Cavagna, Andrea here
Celtic here
  Euro 1967 here, here

Chelsea here, here, here, here, here, here, here, here, here
betting on here, here, here, here, here, here, here, here, here
Chiellini, Giorgio here, here, here
clapping here, here
climate change here
collective individualism here
Collymore, Stan here, here
cooperative behaviour here
Couzin, Iain here
Cowdenbeath here
crabs here
Creus, Xavi here, here, here, here
cricket here
Croatia
Euro 2012 here
crowd disasters here
Cruyff, Johan here
Crystal Palace here, here, here, here, here, here
betting on here, here, here, here, here, here
cybernetics here

Džeko, Edin here
Da Silva, Marta here
defenders here
Delaunay triangulation here
Derby County here
Disney here, here
Du Sautoy, Marcus here
Duarte, Ricardo here, here
Dunfermline Athletic here
Dussutour, Audrey here
Dynamo Kiev here, here

Edelstein-Keshet, Leah here
Elo here
England here
England 1872 here
England 1953 here
Euro 2012 here
Women's World Cup 2015 here
World Cup 1986 here
Enrique, Luis here, here
Euro Club Index here, here, here, here, here, here, here, here, here, here
Everton here, here, here, here, here, here, here, here
Evra, Patrice here, here
extreme-value theory here

Faria, Jolyon here
Farkas, Illés here
FC Nürnberg here, here, here
Fekir, Nabil here
Ferdinand, Rio here
Ferguson, Alex here, here

Fermat’s Last Theorem here
fish movement here, here, here
flow fields here
picturing Pirlo here
right social convention here
unbeatable back here
why the piggy always wins here
Folgado, Hugo here
football here, here
*Football Manager* here, here
formations here, here
Fulham here, here

gambling simulations here
Gascoigne, Paul here, here, here, here
Germany
Euro 2012 here
World Cup 2014 here, here
Gerrard, Steven here, here, here, here, here, here, here
Giardina, Irene here
Giggs, Ryan here, here
Gleave, Simon here, here, here, here
Gonzalez Blanco, Raúl here
Gourcuff, Yoann here
Graham, Ian here
Greene, Maurice here
Grund, Thomas here
Guardiola, Pep here, here, here, here, here, here, here
guessing games here, here, here
Gullit, Ruud here

Hamilton, William here
Hamilton’s Rule here
Hansen, Alan here, here
Hart, Joe here, here, here
Hazard, Eden here
Helbing, Dirk here, here
Hemelrijk, Charlotte here
Herbert-Read, Teddy here
Hidegkuti, Nándor here
hierarchies here
Hill, Jimmy here, here, here, here
Hull City here
Hungary 1953 here, here
Hvattum, Lars Magnus here

Ibrahimovic´, Zlatan here, here
Iheanacho, Kelechi here
Infostrada Sports here
Iniesta, Andrés here, here, here, here, here, here, here, here, here
Inter Milan 1960 here, here, here
Intergovernmental Panel on Climate Change here
International Football Association Board here

intuitive design here
Iowa Electronic Market here
Italy
  Euro 2012 here, here

Jakobsson, Robin here, here
James, David here
Japan here, here, here
Johnson, Ben here
Junco, Dark-eyed here
Juventus here, here, here
  tactics here

Keane, Roy here
Kelly, John, Jr here
  Kelly criterion here
Kilmarnock here
King, Andrew here, here
Klopp, Jurgen here, here, here
Koeman, Ronald here
Krause, Jens here, here, here, here, here
Krause, Stefan here, here, here
Kroos, Toni here
Kuyt, Dirk here

Lacazette, Alexandre here
Lampard, Frank here
Langdon, Mark here, here
leadership here
Ledesma, Pedro here, here
Leicester here, here, here, here, here, here
Lennox, Bobby here
Lichtsteiner, Stephan here, here, here
line-ups here
Lineker, Gary here, here, here
lionesses here
Liverpool here, here, here, here, here, here, here, here, here
  betting on here, here, here, here, here, here, here, here
  late 1970s here, here
Ljungberg, Hanna here
Lobanovskyi, Valeriy here, here
locusts here
Lyon here

Manchester City here, here, here, here, here, here, here
  betting on here, here, here, here, here, here, here
Manchester United here, here, here, here, here, here, here, here
  betting on here, here, here, here, here, here, here, here
Mann, Richard here, here
Maradona, Diego here, here
Marchisio, Claudio here, here, here
Marseilles here
Mascherano, Javier here

*Match of the Day* here, here
maths here, here
career advice here
McHale, Ian here
McMullan, Colm here
Messi, Lionel here, here, here, here, here, here, here, here, here, here, here
goals here, here, here, here, here
Mexican waves here
Michels, Rinus here, here, here
Midtjylland here, here
*Moneyball* here
Montolivo, Riccardo here
Morata, Álvaro here, here, here
mosh pit here
Mourinho, José here, here
Moussaïd, Mehdi here, here, here, here
movement here, here
data overflow here
line-up here
pressing power here
putting it together here
right social convention here
rules of motion here
schools of fish here, here, here
subtle leader here
Moyes, David here
Mukerji, Patrik here

Nagy, Mate here, here
Nakagaki, Toshi here
Narey, David here
NASA here
natural selection here
Neuer, Manola here, here
Newcastle United here, here, here, here, here, here, here, here, here, here, here
Newtonian physics here
Neymar here, here, here, here
Niemeyer, Michael here
Norwich here, here, here, here
Nottingham Forest here

odds here
odds-bias strategy here, here, here, here, here
OddsCraft here
Oliviera, Rui here
Olympiacos here
opinion polls here
Opta here, here, here, here, here, here, here, here, here
Özil, Mesut here, here

Pan, Selina here
Panathinaikos 2010 here, here
Paris Saint-Germain here

passing exercises here
patterns here
  explained by randomness here
  kicked by the horse here
  random Subbuteo here
  real story here
Pelé, Edson Arantes do Nascimento here, here
performance curves here, here
performance indicators here, here
*Physarum polycephalum* here
Pichichi Trophy, Spain here, here
pigeons here, here
piggy-in-the-middle here
Pirlo, Andrea here, here, here, here, here
Pleuler, Devin here, here
points system here, here
  breaking up the pecking order here
  expanding point pies here
  given half a chance here
  intuitive design here
Poisson distributions here, here, here, here, here
Poisson, Siméon Denis here
Portsmouth here
positioning here
Powell, Asafa here
Power, Paul here, here
predicting results here, here
  challenge to make money here
  expert's advice here
  how good do I have to be? here
  odd probabilities here
  partner's advice here
  performance indicators here
  Premier predictions here
  results index here
  strong favourites, well-matched draws here
  when should I place a bet? here
prediction here
  lack of randomness here
PredictIt here
Premier League Performance Index here
pressing here
Prince-Wright, Joe here, here, here, here, here
probabilities here
  odd probabilities here
Prozone here, here, here, here
PSV Eindhoven here
Pythagoras here

QPR (Queens Park Rangers) here, here

randomness here
  gambling simulations here

lack of randomness here
unexpected goals here
Real Madrid here, here, here
Ronaldo, Cristiano here
relationship networks here
results here
betting channels here
bore draws here
career advice here
indexical decline here
Lucky Luke and Calamity Jane here
satisfied with a draw here
unexpected goals here
Robben, Arjen here
Rodríguez, James here
Ronaldo, Cristianio here, here, here, here
goals here, here, here, here, here, here
Rooney, Wayne here, here, here
Royal Netherlands Meteorological Institute here

Sally, David *The Numbers Game* here
Sánchez, Hugo here
Scarf, Philip here
Schwarzer, Mark here
Schweinsteiger, Bastian here, here, here, here
Scotland here
Scotland 1872 here
Shankly, Bill here
shirk or work? here
Silver, Nate here
Silverberg, Jesse here
singing here
slime moulds here, here, here
Smallegange, Isabel here
Soccer Power Index here
social contagion here
1,2,4,8,16 . . . here
clapping contagions here
Luis Suárez transfer to Arsenal here
mosh pit here
people don't panic here
stand up, sit down and turn around here
social conventions here
Southampton here, here, here, here, here
Spain
Euro 2012 here
stadium disasters here
Stander, Philip 'Flip' here
star commitment here
statistical brilliance here
10,000-year storm here
bolt from the blue here

FC analytics here
game-changers here
guessing game here
new extremes? here
objective rankings here
once in every lifetime here
Statszone here
Steinlein, Claus here
Stoke City here, here, here, here, here, here
storm surges here
structure of football here
demise of one-two-seven here
flexi-zones here
rules of motion here
tiki-taka tesselation here
train-track triangles here
Sturaro, Stefan here, here
Suárez, Isco here, here
Suárez, Luis here, here, here
transfer (not) to Arsenal here
Subbuteo here
Sumpter, John here
Sumpter, Lovisa here, here, here, here, here
Sunderland here, here, here
super-linearity here, here, here
superorganisms here, here
Surowiecki, James *The Wisdom of Crowds* here
Swansea City here, here, here, here, here, here, here, here
Sweden here, here
synchronisation here

tactical map here, here
Barcelona here
Bayern Munich here
Juventus here
Ronaldo, Cristiano here
route-one England here
special relationships here
Wambach, Abby here
team formations here, here
teamwork here
collective individualism here
evolving lazy workers here
rallying the ants here
star commitment here
super-linear dynamo here
tribal hunting here
Tevez, Carlos here, here, here
Théraulaz, Guy here
*tiki-taka* here, here
Tomasetti, Cristian here
Tottenham Hotspur here, here, here, here, here, here, here

train track networks here, here
training exercises here
trajectories here, here
tribalism here

Umeå here
USA here
  Women's World Cup 2015 here

Van Basten, Marco here
Van Gaal, Louis here, here
Van Nistelrooy, Ruud here
Vicsek, Tamás here, here, here, here
Vidic´, Nemanja here
Vieira, Marcelo here
Vogelstein, Bert here

Wambach, Abby here
Watford here, here, here
Wenger, Arsene here, here, here, here, here, here
West Bromwich Albion here, here, here, here, here, here, here, here, here
West Ham here, here, here, here
Wiles, Andrew here
Williams, Serena and Venus here
Wilson, Jonathan *Inverting the Pyramid: The History of Football Tactics* here
wise crowd here, here, here, here
  around the corner here
  beating the crowd here
  not so smart here
work or shirk? here

Zarra, Telmo here

**Also available in the Bloomsbury Sigma series:**

*Sex on Earth by Jules Howard*
*p53: The Gene that Cracked the Cancer Code by Sue Armstrong*
*Atoms Under the Floorboards by Chris Woodford*
*Spirals in Time by Helen Scales*
*Chilled by Tom Jackson*
*A is for Arsenic by Kathryn Harkup*
*Suspicious Minds by Rob Brotherton*
*Herding Hemingway's Cats by Kat Arney*
*Electronic Dreams by Tom Lean*
*Sorting the Beef from the Bull by Richard Evershed and Nicola Temple*
*Death on Earth by Jules Howard*
*The Tyrannosaur Chronicles by David Hone*
*Big Data by Timandra Harkness*
*Goldilocks and the Water Bears by Louisa Preston*
*Science and the City by Laurie Winkless*
*Bring Back the King by Helen Pilcher*
*Furry Logic by Matin Durrani and Liz Kalaugher*
*Built on Bones by Brenna Hassett*

Bloomsbury Sigma
An imprint of Bloomsbury Publishing Plc

50 Bedford Square London WC1B 3DP UK
1385 Broadway New York NY 10018 USA

www.bloomsbury.com



First published 2016. Paperback edition 2017.



British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Library of Congress Cataloguing-in-Publication data has been applied for.

ISBN (paperback) 978-1-4729-2414-8
ISBN (ebook) 978-1-4729-2415-5

Figures by David Sumpter