

JAMES TIPPETT

with clubs of far greater wealth? I hope the reader takes as much joy in reading *The Expected Goals Philosophy* as I did in writing it.

1

EXPECTED GOALS

*What to Expect
When You're Expecting Goals*

WHAT IS THE EXPECTED GOALS METHOD?

Fans make several subconscious judgements whenever they watch a football match. Most routinely, they form an opinion of which team is playing better. Does the scoreline accurately reflect the performance level of each side? Should one team have scored more goals than they actually have? In other words, has luck had a considerable impact on the outcome of the match?

Football is a sport riddled with randomness. Bad teams will often defeat good teams, bad players will often go through patches of playing well and bad managers will often end up in charge of high profile clubs. The damaging effects of chance make football an incredibly hard sport to understand. The Expected Goals method is a tool which we can use to separate the skilful from the lucky and more accurately gauge performance levels.

Put simply, xG tells us the quantity and quality of chances that each team creates from a match. When we look back on results, we often have strong opinions over which side played better. *'If we had scored that penalty we would have won'.* *'They were so lucky to have scored that deflected shot from long-range'.* *'We created so many good chances, how did we lose?'* We base our view of who should have won on the scoring opportunities that were created during the match. Expected Goals data is simply a way of quantifying these scoring chances.

Football is heavily centred around goals. Match reports generally focus on the opportunities that each team created. Highlight packages centre around the openings that each team carved out. Commentators get most excited when a team is about to score. Football revolves around goals, and goals can only occur when teams create scoring opportunities. Indeed, every tactic ever created represents a coach's attempt to develop his team's ability to create chances, whilst at the same time improving their ability to not concede chances. A football match is essentially a series of attacks from each team on the other team. The sides who have the highest quantity *and highest quality* of attacks are clearly the best sides. This is what the Expected Goals method measures.

However, goals are almost as rare as they are important. Thousands of actions take place over the course of each match – in fact, Opta's data collectors suggest that an average of 3,000-4,000 events such as passes, tackles, duels, saves and so on happen over a ninety-minute period. Only a handful of these actions will be shots that result in goals (the average number of goals per match is around 2.7). Each one of the thousands of in-play events is geared towards one thing: chance creation. Assessing the nature of the chances that each team created will reveal which side has deserved to win, regardless of the *actual* scoreline.

The Expected Goals method cuts right through to the core of football thinking. The ability of sides to threaten the opposition goal, whilst simultaneously preventing danger to their own goal, is what separates good teams from bad teams.

HOW IS EXPECTED GOALS CALCULATED?

Essentially, xG indicates how many goals a team could have expected to score based on the quantity and quality of chances that they created in a match. Fans often come away from football matches thinking, "We created much better chances than the opposition, we definitely should have won". The Expected Goals metric is a way of quantifying these scoring opportunities, allowing a better insight into the ability of teams than the *actual* scoreline does.

Teams are often judged by the *quantity* of shots that they have in a match, or indeed in a season. Media companies will show the stats for how many attempts at goal each side has taken in a game. The central premise of xG is that the *quality* of those shots is of equal importance as the quantity. Analysts can work out the number of goals that a team would have expected to score from a certain amount of shots of a certain quality. Similarly, analysts can work out which players have scored more chances than they would be expected to.

Expected Goals data is collected by several different data companies, football clubs and betting firms. The main provider of xG stats to media companies is Opta Sports, who claim to collect the most complete dataset for the Premier League, English Football League, Scottish Professional Football League and many other divisions across the globe. Opta's data experts have analysed over 300,000 shots to calculate the likelihood of an attempt being scored from a specific position on the pitch during a particular phase of play.

Expected Goals works by measuring the likelihood of each shot resulting in a goal. Each effort at goal which takes place in a match has a "Shot Probability" value. For instance, a shot from 30 yards out through a crowd of players may only have a 2% chance of hitting the back of the net, giving it a value of 0.02(xG). On the other hand, a shot into an open goal from six yards out might have a 95% chance of being scored, resulting in a value of 0.95(xG).

At the end of any given match, the Shot Probabilities from either side are added up to reveal the Expected Goals scoreline from a match. For example, suppose that Arsenal play against Manchester City. The London side have six shots over the course of the match, but they are all long shots from distance with a Shot Probability value of 0.1(xG). The Gunners will have amassed a total Expected Goals score of 0.6(xG). Over the course of the same match, Man City only have two shots at goal,

but they are both from close range. Suppose that one shot is worth $0.3(xG)$ and the other is worth $0.4(xG)$. Man City's Expected Goals score over the course of the ninety minutes is $0.7(xG)$. Thus, the xG scoreline from the match would be *Arsenal* $0.6(xG)$ – $0.7(xG)$ *Man City*¹. The scoreline would reflect the fact that Manchester City performed narrowly better than their London counterparts.

There is a natural question which follows: how do you determine the probability of a shot's success? The *location* of a shot has a large bearing on how likely it is to result in a goal. A shot which is taken from a wide position, thirty yards out from goal, will only have a small chance of going in. On the other hand, a close-range shot from a central position will have a high probability of scoring.

An analyst could look at a large sample of past shots taken from an exact position and find how many beat the goalkeeper. Say an analyst looked at 1,000 shots taken from the exact position at the right-hand corner of the penalty area (we are assuming the analyst has a large database of shots spanning across several divisions

¹ As we will see, this xG scoreline actually occurred between the two teams in November 2017. Manchester City actually won the game by three goals to one, prompting Arsène Wenger to come under fire in his post-match press conference. The Frenchman referenced the Expected Goals scoreline (*Arsenal* $0.6(xG)$ – $0.7(xG)$ *Man City*) in an attempt to outline how close the game had actually been. The two teams had been evenly matched, but the visitors had had luck on their side.

and several seasons. Thousands of shots are taken each season, each one being recorded by companies like Opta. These companies can draw upon vast databases of past shots in order to determine Expected Goals probabilities). Suppose only 50 of these shots ended up in a goal. He could conclude that future shots from this location have a 5% chance of beating the goalkeeper (as $50/1000$ previous shots from this position hit the back of the net). Thus, the Expected Goals value from this position is $0.05(xG)$.

Whilst the location of a shot forms the main basis of its danger level, other factors also play their part. Shots which come from crosses are considerably harder to convert than shots which take place when the ball is standing still. Whether the shot is headed, volleyed or hit from the ground also affects its chance of success. So too does it matter whether the effort is taken on a player's weaker foot. Analysts account for a whole range of such factors in their Expected Goals models.

The above description is a very brief introduction into how to collect Expected Goals data, but hopefully the reader can already get a sense of how it can be used to analyse football. Teams who are consistently creating high value chances are clearly dangerous opposition, whilst those who are only clocking low xG Shot Probabilities evidently lack potency going forward. Similarly, an analyst can work out how many goals a player would have expected to score based on the quality of

chances that he has received. If a striker is scoring lots of goals from difficult positions, we might applaud him. Conversely, if a player is drastically underperforming his Expected Goals output, we might question his ability.

WHY IS THE EXPECTED GOALS METHOD IMPORTANT?

Every judgement, opinion or prediction that we make about the beautiful game is grounded on our assessments of past performances. But how can we expect to accurately analyse the sport when it is so considerably determined by luck? This is where the Expected Goals method steps in, providing an antidote to the disease of randomness which permeates football. It is at the forefront of a smarter, more analytical football philosophy.

In recent years, Expected Goals has been used by various establishments in order to facilitate a better understanding of the sport. The more innovative football clubs have begun using xG in several regards. Most obviously, they use it to measure the performances of their own team and players. Has their side been achieving the results expected of them? Do they actually deserve to be where they are in the league table? How efficient have they been at creating chances and preventing the opposition from creating scoring opportunities?

The scouting teams at clubs also use xG data to uncover and sign hidden gems; players who are undervalued

by the rest of the footballing world. Certain English clubs have managed to consistently sign great players for low prices because of their Expected Goals tools. This analytical style of recruitment has allowed teams to enjoy great success on shoe-string budgets. We will study the methods of these teams later in the book.

A different type of institution has also utilised the immense predictive ability of the Expected Goals method. Professional gambling syndicates have used xG to calculate accurate probabilities of events occurring. These companies use Expected Goals data to generate odds, which they compare to the bookmakers' odds. The success of their businesses depends on their ability to make more accurate forecasts than the bookies. The Expected Goals method has allowed them to do this. Later, we will more closely examine the top secret gambling cohorts who have turned over millions of pounds through utilising xG.

Fans have finally begun to pay an increasing amount of attention to Expected Goals stats. Slowly but surely, the media have taken notice of this increase in interest. Football supporters are insatiable consumers of facts. In a sport where knowledge is power, Expected Goals is slowly emerging as the most authoritative form of data. In later sections, we will see how the media have entered the early stages of the xG revolution, what problems they have faced in incorporating it into their broadcasts and where the future may lie for supporter interaction. I am

confident that, in a few years' time, it will be impossible to read a match report without finding a reference to the Expected Goals scoreline from the game.

The increased exposure which the metric has seen over the last couple of years is simply a drop in the ocean of what is to come. The reason why is simple: *Every footballing judgement ever made is based on an analysis of the performance of teams or players.* And the Expected Goals method offers by far the most advanced, profound and accurate gauge of performance.

Soon, those who do not understand or pay attention to xG data will be left behind. The Expected Goals method allows you to speak about football in a more intelligent language.

“Isn't X an incredible manager?!”

No, xG shows that his team have incredibly lucky to get the results that they have.

“How could Y's defence play so badly?!”

Actually, the Expected Goals data shows that they played very well.

“Why does Z keep missing absolute sitters?!”

Sorry to correct you, but xG suggests that this player is actually scoring more goals than would be expected of him.

Too often in football, the result dictates the narrative. A team who plays badly and wins has “a great mentality” and is able to “grind out results even when not playing well”. However, a team who plays badly and loses will be deemed to have obvious flaws. Both of these teams have performed at the same level (i.e. badly), but notice how our analysis has been changed dependent on their result. In order to avoid being fooled by randomness, we should direct more attention to the Expected Goals totals amassed from each game. This will allow us to assess *performances*, rather than *results*.

Only when we fully embrace the Expected Goals method can pundits begin to more accurately comment of football. Only then can managers give more reasonable post-match interviews. Only then can the fans select the best players for their fantasy teams. Only then can we haul football out of the dark ages and into a more intelligent era of analysis.

2

FOOTBALL PHILOSOPHERS

Traditionalists Versus Revolutionaries

**“THE MOST USELESS STAT
IN THE HISTORY OF FOOTBALL”**

Jeff Stelling is one of the most popular television pundits worldwide. His charm and charisma make him an affable media personality. However, on one afternoon in November 2017, he went on an uncharacteristic rant that sparked debate throughout the world of football. Manchester City had just defeated Arsenal three-one and Arsène Wenger had cited the Expected Goals data from the match in his post-match press conference. The Gunners manager highlighted that Arsenal had amassed 0.6(xG), whilst Man City had accumulated a marginally higher total of 0.7(xG). This reflected the fact that the Manchester side only performed slightly better than their London counterparts. The xG data suggested that the actual scoreline considerably flattered Man City.

Stelling was belligerent in his disapproval of the statistic. “He’s the first person I’ve ever heard to take notice of Expected Goals, which has to be the most useless stat in the history of football”, the pundit exclaimed. “What does it tell you? The match finished 3-1, why do you show Expected Goals afterwards? It’s absolute nonsense, it really is.” Stelling’s reaction was met with audible laughter from the rest of the *Soccer Saturday* panel, with Soccer AM later sharing the video on Twitter. Stelling is an influential figure within football. His mocking of

the Expected Goals method sparked widespread criticism of the statistic from fans. The incident outlines the scepticism of mathematical methods that exists within the sport.

The Expected Goals method has struggled to permeate mainstream football knowledge and conversation for three fundamental reasons. Firstly, there is an underlying ignorance. This stems from the lack of exposure which analytical methods gain in the media. Football fans are not given a chance to engage with clever means of assessing the beautiful game, and thus are ignorant to the revolutionary impacts that they could have.

Secondly, many fans are simply not interested in statistical analysis. Unfortunately, not everyone will care about a more intelligent approach to football. Fans generally support their team for the social aspect, to give themselves common ground with an overwhelmingly large community of worldwide supporters. Many fans care more about drinking pints before the match than they do about examining their striker’s mathematical level of performance. And that’s fine. Not everyone wishes to be intelligible on such matters. Ignorance is bliss. There is nothing wrong with going to football simply to slate the referee and have a fun day out. If you have picked up this book, then you are probably interested in understanding more about the sport.

The final reason why a traditionalist might reject more scientific means of analysis is incompetence. It must be confessed, however, that incompetence is irrefutably linked with the previous two reasons. I strongly believe that *every football fan has the ability to understand the philosophy outlined in this book*. It isn't rocket science. A lot of people, especially football fans, automatically switch off when faced with what could be described as a "mathematical approach". The methods outlined in this book could be understood by a schoolchild. As long as the topics are approached with an open mind, every fan has the capability of grasping the concept of the Expected Goals method.

So, for what reason did Jeff Stelling refuse to engage with xG? He certainly isn't ignorant. The fact that he was talking about the Expected Goals method proves that he has at least heard of it. He certainly shouldn't have been disinterested. It is his job to attempt to understand all footballing matters to the best of his abilities. And he certainly isn't incompetent. He is a very intelligent man and certainly competent enough to understand the workings of xG.

Perhaps there is a fourth reason for football's unwillingness to accept the Expected Goals method. *Fear*.

"THE FIRST THROUGH THE WALL ALWAYS GETS BLOODY"

"There is an epidemic failure within the game to understand what is really happening", says Peter Brend, assistant general manager at the Oakland Athletics baseball team in Aaron Sorkin's film adaptation of Michael Lewis' famous book, *Moneyball*. Oakland's financial disadvantage to other teams in the league meant they had to play by different rules in order to achieve the same success. If they hadn't innovated, adapted and gambled, their final position in the table would have ultimately reflected their financial situation and they would have been rooted to the bottom of the league.

The central hypothesis of *Moneyball* is that the traditional methods used by baseball insiders (in particular those of managers, coaches and scouts) over the past century were subjective and often flawed. Statistics such as stolen bases, runs batted in and batting average, typically used to gauge player ability, were relics of a 19th-century view of the game and the statistics available at that time. Billy Beane, the man in charge of player recruitment for the Oakland A's, believed that the future of baseball lay in a broader statistical approach based on numerical models. He realised that the traditional scouting methods used by every other team were outdated, and played this to the advantage of the Oakland

A's. They developed more analytical gauges of player performance in order to field a team that could better compete against their wealthy opponents.

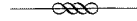
Beane was tasked with recruiting players for the A's, in the same way a Chief Scout might be tasked with recruiting talent for an English football club. He employed the services of Paul DePodesta, who had just graduated from Harvard with a degree in Economics. Together they developed a system of finding undervalued baseball players, whom they could purchase for little money but who could have a large positive impact on the team's performance. At first, Beane's revolutionary and previously unheard of methods were met with rigid opposition; not only from outsiders but also from the other management staff of the Oakland A's. Beane was on the verge of losing his job, before a miraculous turn-around saw the A's embark on a twenty-match winning run.

Beane endured an initial period of difficulty after implementing his radical, mathematical philosophy at Oakland. The A's struggled, and Beane and his analytical team faced ridicule from all corners of the baseball world. It was argued that '*you cannot measure the size of a player's heart*', and that '*baseball is poetry, not mathematics*'. (Incidentally, both these arguments have also been made in order to undermine the progress of statistical analysis within football). Beane came into particular conflict with the manager of the Oakland A's, Art

Howe. Howe refused to play the players that Beane recommended to him because of a scepticism towards the general manager's mathematical methods. For instance, Beane's calculations showed that playing Scott Hatteburg at first base would increase Oakland's chances of victory. However, Howe went with the seemingly obvious choice of Carlos Pena because it would be "easier to explain in job interviews" come the end of the season – the manager was more focused on his reputation than on the performance of his team.

Beane's revolutionary analytical philosophy took time to gain traction. However, after the initial period of struggle, Oakland's innovative approach turned the sport of baseball on its head. Within a handful of seasons, every team in the NBL had adopted the methods outlined in Michael Lewis' *Moneyball*. Beane's success was founded on a previously overlooked statistic: *on-base percentage (OBP)*. His team discovered, following rigorous statistical analysis, that OBP was a far better indicator of offensive ability than the traditional stats used to measure batting talent. Beane realised that a player's ability to get on base was a quality much cheaper to attain in the open market than more historically valued qualities such as speed and contact. This gave the Oakland A's an edge when scouting players, allowing them to compete on a much stricter budget than their competitors. However, it also contradicted conventional baseball

wisdom, disproving the beliefs of Oakland's scouts, Art Howe and the sport's media.



Billy Beane's decision to adopt a mathematical philosophy at Oakland demonstrates how traditionalists and revolutionaries can come into conflict. Just as Jeff Stelling refused to accept the Expected Goals method as a form of analysis in football, Art Howe and his scouts refused to accept on-base percentage data as a form of analysis in baseball. Perhaps the main reason for the rejection of scientific methods within sport is fear, or (more specifically) a fear of being replaced. This was certainly the case amongst Oakland's scouting team. The recruitment department were worried that their skill for spotting talent could just as easily be done by a number-crunching computer, prompting them to try and undermine such an approach.

Stelling's rejection of Expected Goals stems from a different type of fear. The job of a sports pundit, to a certain extent, is to tell the television audience what they want to hear. For Stelling, rejecting the Expected Goals method, a new type of analysis which questions the conventional wisdom of football, is much easier than advocating it. The *Soccer Saturday* host's enraged rant about the metric went viral, being shared thousands of times on social media. A clip of him advocating

and explaining the statistic would not have been nearly as popular. In fact, it probably would have sparked widespread ridicule from Sky's audience.

This is where the football media fails its fans. Jeff Stelling has the influence to make people listen, to evoke change within the sport, but instead he chose to tell people what they wanted to hear. The Expected Goals method is certainly the most profound form of football analysis, as proven by the success it has brought professional clubs and gambling syndicates alike. However, influencers within the sport are more concerned with being popular than in being radical. *Football needs influential people to go against the herd from time to time, in order to promote intelligent debate within the sport.* The sport is currently spoken about in clichés. Nothing new or innovative comes to the fore because pundits, the people who carry authority within the sport, speak in safe and tiresome platitudes.

Meanwhile, those people who truly understand the power of Expected Goals (such as bettors and scouts) actively try to keep their methods secret. It is within their interest to not promote the Expected Goals method, thus maintaining their edge over the competition. If you had the recipe to make the world's tastiest cake, you wouldn't publish it for everyone to see. You would use it to bake and sell as many products as you could. The Expected Goals method is football's tastiest means of analysis, but the media are worried that it is too different

from the current flavours enjoyed by consumers. Fans are still enjoying the current batch of goodies that the media are serving up to them, blissfully unaware of the tastier treats being enjoyed by bettors and professional football clubs.

3

A HISTORY OF FOOTBALL ANALYTICS

How the Beautiful Game Evolved