The Beautiful Game: A Mathematical Perspective

Akachi Anyanwu

March 2024

"Go deep down enough into anything and you will find mathematics." *- Dean Schicler*

1 Introduction

Dubbed the "beautiful game", the sport of football holds a universal appeal, enjoyed by billions worldwide. The exciting sport, where teams of eleven aim to outscore their opponents with a ball at their feet, transcends any cultural or geographical boundary. From stadiums in Europe to courts in Brazil, football unites people of all ages and backgrounds. Sadly, though, in recent times, football and academia have become two somewhat juxtaposing ideas, as footballers have garnered a certain stereotype over the years, which to the best of my knowledge does not encompass mathematical brilliance, or vague competence for that matter. However, I am a keen footballer and maths enthusiast who aims to break this stereotype by showing that maths and football do go hand in hand. In fact, one could go as far as to say maths is behind almost everything that happens in realm of football! By taking you on a journey through different aspects of the footballing world and using examples to show the mathematical principles intertwined within them, I hope to share with you some insight on how maths keeps the world's biggest sport thriving.

2 The Transfer Market

Throw yourself straight into the best football club in history: Real Madrid. You are the owner and its summer 2023, meaning the transfer market is wide open. Open just for two months of the year, the transfer market allows you to strengthen your squad through purchase of players from all over the world. You have plenty of money at your disposal, and what is better to get the fans on your side after a poor Champions League and La Liga campaign than a successful, statement signing? Paris Saint German's Kylian Mbappe is the best player in the world, and he is fresh from scoring a hattrick in the final of the 2022 World Cup. However, there is only one thing standing in your way: the £213 million fee for his signature and services. You consult your staff and come to two possible conclusions: pay the fee now or wait 1 year until his contract runs out and sign him for free (You can sign players for free when their contracts run out, but it is riskier as many other clubs will also try and do the same with ludicrous wage offers). But how can we use maths to determine if the former option is worth it?

It is important to remember that football clubs are businesses, and the value of a footballer to a club is defined as the sum of their present value based on different streams of revenue they generate for the club (Not how many goals they score in a season). To begin, we can assume salaries are financed by match-day sales. A player's most notable stream of revenue generated from the club comes from shirt sales – which is reliant on their popularity. Essentially, the more popular they are to the fans, the higher value they will be. To model a player's value, we will start by applying Schneider's equilibrium formula for player *a*:

$$x_a = (y_a - c)z_a \tag{1}$$

where x_a is the player's value – their potential to increase sales for the club through monetization of positive externalities that come about from their popularity. $(y_a - c)$ corresponds to profit margin where y_a is the sales price of shirt sales and sold merchandise directly due to player a and c is the cost of production (of merchandise). z_a is the number of potential fans who would buy the merchandise.

However, we don't know the number of potential fans, z_a, so we must establish it recursively. We can do this with the following functional relationship:

$$x_a = \alpha + \mu_1(position) + \mu_2(age) + \mu_3(performance) + \mu_4(z_a) \tag{2}$$

This therefore defines the player's value as a combination of multiple factors: age, position (dummy variables for position are used in this model), and performance. Age has a significant bearing on value as the older a player is, the more prone they are to experience performance deterioration (unless you are Cristiano Ronaldo). Performance is defined as $\frac{\sum total\ minutes\ played\ in\ a\ season}{\sum minutes\ the\ player\ plays}$; the closer this number is to 1, the better. The coefficients were determined by testing the model with data from thousands of players and their actual transfer market values, by inputting the relevant parameters into equation (2). Since z_a is not available, we can instead use social media following to obtain a similar coefficient as the number of potential customers is closely linked to popularity. We can then manipulate equation (1) and equation (2) to obtain an expression for the potential number of customers/fans, z_a, as follows:

given that $x_a = (y_a - c)z_a$, we can rewrite equation (2) as:

$$(y_a - c)z_a = \alpha + \mu_1(position) + \mu_2(age) + \mu_3(performance) + \mu_4(z_a)$$

By rearranging and factorising we will obtain equation (3) in the following steps:

$$(y_a - c)z_a - \mu_4(z_a) = \alpha + \mu_1(position) + \mu_2(age) + \mu_3(performance)$$

$$z_a((y_a - c) - \mu_4) = \alpha + \mu_1(position) + \mu_2(age) + \mu_3(performance)$$

$$z_a = \frac{\alpha + \mu_1(position) + \mu_2(age) + \mu_3(performance)}{((y_a - c) - \mu_4)} \tag{3}$$

Now we have a valid model to assess the number of customers, we can consider a model for a fair transfer fee that Real Madrid should pay:

$$t_a = z_a v_a w_a (y_a - c) \tag{4}$$

where t_a is fair transfer value, z_a is number of potential customers as before, v_a is a sensitivity factor on a scale from 0.0 to 1.0 that assesses the player's ability to attract new customers, w_a is length of contract in years, and $(y_a - c)$ corresponds to the profit margin.

The sensitivity factor is crucial because it considers the fact that a goalkeeper will be less attractive to fans than an attacker, as an example, and it also takes performance and form into account. So now that we have established the necessary equations, let's consider if we, Real Madrid, want to pay the hefty fee for the French superstar Mbappe.

The relevant coefficients for equations (2) and (3) are simplified in the table below[1]:

Table 1: Shows the relevant coefficients which were calculated by using metrics from the past 5 seasons in Europe's top 5 leagues.

Constant	Value
α	36,321,636
μ_1	0
μ_2	-0.109
μ_3	0.233
μ_4	0.487

Since equation (2) is the model for an attacking player, μ_1 is set to 0 to "switch off" the dummy variable

Consider Equation (3) with the new coefficients:

$$z_a = \frac{36{,}321{,}636 + 0 - 0.109(age) + 0.233(performance)}{((y_a - c) - 0.487)}$$

Mbappe was 24 years old in summer 2023, and he played 3,564 minutes[2] out of a possible 4590 minutes[3] that PSG played in 22-23. Therefore, his performance is $\frac{3564}{4590}$ = 0.776. Now let's consider the profit margin for shirt sales. PSG sells Mbappe's shirts for £143 [4], of which they keep 33% [5]. The cost of production of a shirt is about £12 [6] so the profit margin, $(y_a - c)$, is $(143 \times 0.33) - 12$ = 47.19. Now we have the relevant data, we can calculate z_a, Mbappe's potential number of fans and customers:

$$z_a = \frac{36{,}321{,}636 + 0 - 0.109(24) + 0.233(0.776)}{(47.19) - 0.487}$$

= 777715.2124 fans/customers

Note: The number of customers would be higher if the shirts were not priced so highly.

Now using the fair transfer value equation, $t_a = z_a v_a w_a (y_a - c)$, we can calculate the reasonable value we should pay for Mbappe. We would likely want to sign him on a long term 5-year contract ($w_a = 5$), and owing to his popularity and unreal performances week in week out, he would have a high sensitivity factor, v_a, at 0.95, as he would attract many new fans. Therefore, his fair transfer value is as follows:

777715.2124 × 0.95 × 5 × 47.19

=£174,326,809

It is important to note that this value should be higher if all merchandise sales related to Mbappe is considered; I only considered shirt sales with this model.

Therefore, PSG's asking price is about £40 million more expensive than our fair calculated value. Other factors will influence our decision whether to buy him in the summer or wait until his contract runs out the year after, such as which clubs will battle for his signature once he is a free agent. However, he has expressed interest in Real Madrid for the past two years, so it would be better to wait it out as supposed to overpaying a ludicrous fee which won't pay your investment back. This is actually what Real Madrid ended up doing last summer. Nevertheless, what would you have done in this situation if you were the owner, Florentino Perez?

3 The Starting XI

Another club that can't resist a big money transfer is my favourite team, Chelsea Football Club. However, after spending more than £1 billion on players in the past 18 months, it seems that we have developed a fond appreciation of the middle of the Premier League table, standing a grand 11th place out of 20 teams as of March 2024: the worst form we have seen for a very long while. The issue is, we just can't seem to put out a starting line-up that can find any consistency, resulting in the Chelsea fan experience over the past 2 years being a rollercoaster of (mostly negative) emotions.

The solution to Chelsea's problem? Maths, of course. Imagine you are an assistant coach and Pochettino (Chelsea's current manager) gives you the huge responsibility of choosing the starting XI for the next game based on recent form. He gives you a choice of the two formations we have been playing this season: 4-2-3-1 and 4-3-3, shown below.

Figure 1: Shows the formations which will be examined; there are 10 possible positions across both formations: Goalkeeper (GK), Centre Back (CB), Left Back (LB), Right Back (RB), Defensive Midfielder (DM), Central Midfielder (CM), Attacking Midfielder (AM), Right Wing (RW), Striker (ST), and Left Wing (LW)

In order to find which formation is better, and where each player should play, we need a way of assessing their recent performances. For this, I used SofaScore[7], a reputable website and app which rates a player's performance on a scale of 1-10 in each game. I examined all of the ratings from Chelsea's past 5 matches and collated the data in the following table:

Player Name	Match and Date (2024)					Average Rating		
	Vs Liverpool (25/02)	Vs Leeds (28/02)	Vs Brentford (02/03)	Vs Newcastle (11/03)	Vs Leicester (17/03)	Primary Position	Secondary Position	Tertiary Position
Petrovic (GK)	8.90	DNP	6.40	6.20	DNP	7.17	N/A	N/A
Sanchez (GK)	DNP	6.40	DNP	DNP	5.90	6.15	N/A	N/A
Gusto (RB)	7.40	6.60	7.40	7.00	7.60	7.20	N/A	N/A
Gilchrist (RB)	DNP	6.50	DNP	DNP	DNP	6.50	N/A	N/A
Cucurella (LB)	DNP	DNP	DNP	7.00	7.50	7.25	N/A	N/A
Disasi (CB)	7.00	6.70	7.60	6.90	5.50	6.74	N/A	N/A
Chalobah (CB)	6.70	6.90	6.50	6.50	6.40	6.60	N/A	N/A
Colwill	7.60 (CB)	6.50(RB)	6.90 (CB)	DNP	DNP	7.25(CB)	6.50(RB)	N/A
Chilwell (LB)	7.10	6.90	6.60	DNP	6.60	6.80	N/A	N/A
Casadei (AM)	DNP	DNP	DNP	6.70	DNP	6.70	N/A	N/A
Chukwuemeka (AM)	DNP	DNP	DNP	6.60	8.20	7.40	N/A	N/A
Caicedo	7.60 (CM)	7.70 (DM)	7.10 (DM)	7.30 (DM)	7.50 (CM)	7.55 (CM)	7.37 (DM)	N/A
Fernandez	7.50(CM)	7.70(DM)	7.50 (CM)	7.00 (DM)	DNP	7.50 (CM)	7.35 (DM)	N/A
Palmer	7.40 (RW)	6.80 (AM)	7.70 (ST)	8.30 (AM)	8.40 (RW)	7.90 (RW)	7.55 (AM)	7.70 (ST)
Madueke	7.00 (RW)	7.00(AM)	DNP	DNP	7.20 (RW)	7.10 (RW)	7.00(AM)	N/A
Gallagher	7.40 (AM)	7.30(AM)	7.40 (CM)	7.20 (AM)	6.90 (CM)	7.30 (AM)	7.15 (CM)	N/A
Sterling	7.00 (LW)	7.20 (AM)	6.50 (LW)	6.30 (AM)	6.70 (LW)	6.73 (LW)	6.75 (AM)	N/A
Mudryk	6.60(LW)	7.30 (AM)	6.70 (AM)	7.80 (AM)	7.00 (AM)	6.60 (LW)	7.20 (AM)	N/A
Jackson (ST)	7.90	7.00	7.20	7.90	7.90	7.58	N/A	N/A
Nkunku (AM)	6.60	DNP	DNP	DNP	DNP	6.60	N/A	N/A

Table 2: Shows player ratings over past 5 matches and the calculate averages. DNP means they did not play.

Players can play in multiple positions; for example, Colwill can be both a right back and a centre back, so their average rating for each position is taken, and specified in the table. Think of these average ratings as a judgement of how well a player as played in each position in the past 5 games. But how will all these numbers help us find the best possible formation?

We can treat the optimal formation as an objective function, Z, and use binary integer programming (BIP) to maximise the function. Essentially, it will calculate the formation with the highest combined rating across all 11 players. You can then compare the total rating of both formations to determine which one is better. The model is usually written[8]:

$$Z = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n \tag{5}$$

where $x_1, x_2, \ldots x_n$ is binary (Yes [1], or no [0]). To enable application of BIP modelling, I simplified the table to only show average rating and position which was played. For example, Palmer's rating as an attacking midfielder is 7.00, as a right winger is 7.90, and as a striker is 7.70.

In this objective function, x_n is a variable that represents each cell in the table with a number rating. x_1 corresponds to Petrovic as a goalkeeper with a rating of 6.98, x_2 corresponds to Sanchez as a goalkeeper with a rating of 6.15, x_3 is Cucurella as a left back with a rating of 7.25, and x_{29} is Jackson as a striker with a rating of 7.58. This fits the criteria for BIP, where 0 means "no" and 1 means "yes", as "yes" (1) will be attributed to cells with players in the starting lineup, and "no" (0) will represent players not in the team (the starting lineup at least)

Player	Position									
	GK	LB	CB	RB	DM	CM	AM	LW	RW	ST
Petrovic	6.98									
Sanchez	6.15									
Gusto				7.20						
Gilchrist				6.50						
Cucurella		7.25								
Disasi			6.74							
Chalobah			6.60							
Colwill			7.25	6.50						
Chilwell		6.80								
Casadei							6.70			
Chukwuemeka							7.40			
Caicedo					7.37	7.55				
Fernandez					7.35	7.50				
Palmer							7.00		7.90	7.70
Madueke							7.00		7.10	
Gallagher						7.15	7.30			
Sterling							6.75	6.73		
Mudryk							7.20	6.60		
Jackson										7.58
Nkunku							6.60			

Table 3: A simplification of Table 2

Based on equation (5), we can assign the ratings in each cell as the x coefficients (c) as follows:

$$Z = 6.98x_1 + 6.15x_2 + 7.25x_3 + 6.80x_4 \dots + 7.70x_{28} + 7.58x_{29} \quad (6)$$

Before we can get to computation of the maximum objective function, we have to set multiple constraints for each formation as follows, which we will introduce into the code.

Constraints on 4-2-3-1:

1. $x_1+x_2+x_3\dots+x_{27}+x_{28}+x_{29}=11$ *As there can only be eleven players*
2. $x_1+x_2=1$ *As there can only be one goalkeeper*
3. $x_2+x_3=1$ *As there can only be one left back*
4. $x_5+x_6+x_7=2$ *As there can only be two centre backs*
5. $x_8+x_9+x_{10}=1$ *As there can only be one right back*
6. $x_{11}+x_{12}=2$ *As there can only be two defensive midfielders*
7. $x_{16}+x_{17}+x_{18}+x_{19}+x_{20}+x_{21}+x_{22}+x_{23}=3$ *As there can only be three attacking midfielders*
8. $x_{28}+x_{29}=1$ *As there can only be one striker*
9. $x_7+x_{10} \leq 1$ *As Colwill can't play in two different positions simultaneously.*
10. $x_{18}+x_{28} \leq 1$ *As Palmer cannot play in two different positions simultaneously.*

It doesn't matter, for example, that Sterling can play both attacking midfielder and right winger because there is no right winger in a 4-2-3-1 formation.

There are also additional constraints to ensure the function works properly:

11. $x_1+x_2+x_3+x_4+x_5+x_6+x_7+x_8+x_9+x_{11}+x_{12}+x_{16}+x_{17}+x_{19}+x_{20}+x_{21}\ x_{22}+x_{23}+x_{29} \leq 18$
12. $x_{13}+x_{14}+x_{24}+x_{25}+x_{26}+x_{27}=0$ *(As these positions cannot be played in a 4-2-3-1)*

I then used R studio to optimize the starting lineup. First, I set the objective function's coefficients as in equation (5):

```
> f.obj <- c(6.98,6.15,7.25,6.80,6.74,6.60,7.25,7.20,6.50,6.50,7.37,7.35,7.55,7.50,
7.15,6.70,7.40,7.00,7.00,7.30,6.75,7.20,6.60,6.73,6.60,7.90,7.10,7.70,7.58)
```

Secondly, I set a matrix to correspond to the constraints, entering in about 300 1s and 0s to map out each of the 12 constraints by typing 1 if a coefficient is involved and 0 if it is not. For example, in the first constraint, every coefficient is involved, meaning I inputted 29 1s. In the second constraint, only x_1 and x_2 are involved meaning my input was 1, 1, 0, 0, 0, 0, 0, 0, etc, as is highlighted below.

```
> f.con <- matrix(c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,0,0,1,0,0,1,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,
0,0,0,0,1,0,1,1,1,1,1,1,0,1,1,0,1,1,0,0,0,1,1,0,1,1,1,1,1,0,0,0,0,0,1,0,0,0,0,0,0,0,
0,0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,1,1,1,1,0,0) ,nrow = 12, byrow = TRUE)
```

Below is the same matrix, highlighted in each 29 number segment with a different colour to demonstrate how they correspond to each constraint (Green → 1, Yellow → 2, Cyan → 3, etc).

```
> f.con <- matrix(c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,0,0,1,0,0,1,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,
0,0,0,0,1,0,1,1,1,1,1,1,0,1,1,0,1,1,0,0,0,1,1,0,1,1,1,1,1,0,0,0,0,0,1,0,0,0,0,0,0,0,
0,0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,1,1,1,1,0,0) ,nrow = 12, byrow = TRUE)
```

I then set inequality/equality signs for each of the 12 constraints, as well as the right-hand side coefficients:

```
> f.dir <- c("=", "=", "=", "=", "=", "=", "=", "=", "<=", "<=", "<=", "=")
> f.rhs <- c(11, 1, 1, 2, 1, 2, 3, 1, 1, 1, 18, 0)
```

Finally, I entered the command for the maximum objective function, as well as the solution:

```
> lp("max", f.obj, f.con, f.dir, f.rhs, int.vec = 1:16, all.bin = TRUE)
Success: the objective function is 79.74
>
> lp("max", f.obj, f.con, f.dir, f.rhs, int.vec = 1:16, all.bin = TRUE)$solution
 [1] 1 0 1 0 1 0 1 1 0 0 1 1 0 0 0 0 1 0 0 1
[21] 0 1 0 0 0 0 0 1 0
```

I repeated all the steps for the 4-3-3 formation (including new constraints) as follows.

```
> f.con <- matrix(c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,
1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,0,0,0,0,1,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,1,0,1,1,1,1,1,1,0,
1,1,0,0,0,1,1,1,0,0,0,0,0,0,0,0,1,1,0,1,0,1,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,1,1,1,1,1,
1,1,1,0,0,0,0,0,0) ,nrow = 13, byrow = TRUE)
> f.dir <- c("=", "=", "=", "=", "=", "=", "=", "=", "=", "<=", "<=", "<=", "=")
> f.rhs <- c(11, 1, 1, 2, 1, 3, 1, 1, 1, 1, 1, 15, 0)
> lp("max", f.obj, f.con, f.dir, f.rhs, int.vec = 1:16, all.bin = TRUE)
Success: the objective function is 79.83
> lp("max", f.obj, f.con, f.dir, f.rhs, int.vec = 1:16, all.bin = TRUE)$solution
 [1] 1 0 1 0 1 0 1 1 0 0 0 0 1 1 1 0 0 0 0 0
[21] 0 0 0 1 0 1 0 0 1
```

I then used the solutions and the value for the maximum objective function to produce the following table and formations:

	4-3-3 Formation				4-2-3-1 Formation			
X_n	Binary	Player	Position	Rating	Binary	Player	Position	Rating
1	1	Petrovic	GK	6.98	1	Petrovic	GK	6.98
2	0				0			
3	1	Cucurella	LB	7.25	1	Cucurella	LB	7.25
4	0				0			
5	1	Disasi	CB	6.74	1	Disasi	CB	6.74
6	0				0			
7	1	Colwill	CB	7.25	1	Colwill	CB	7.25
8	1	Gusto	RB	7.20	1	Gusto	RB	7.20
9	0				0			
10	0				0			
11	0				1	Caicedo	DM	7.37
12	0				1	Fernandez	DM	7.35
13	1	Caicedo	CM	7.55	0			
14	1	Fernandez	CM	7.50	0			
15	1	Gallagher	CM	7.15	0			
16	0				0			
17	0				1	Chukwuemeka	AM	7.40
18	0				0			
19	0				0			
20	0				1	Gallagher	AM	7.30
21	0				0			
22	0				1	Mudryk	AM	7.20
23	0				0			
24	1	Sterling	LW	6.73	0			
25	0				0			
26	1	Palmer	RW	7.90	0			
27	0				0			
28	0				1	Palmer	ST	7.70
29	1	Jackson	ST	7.58	0			
				Max Z = 79.83				Max Z = 79.74

Table 3: shows the results of the BIP model, showing the best players in each position, and the max Z.

Figure 2: Solved formations with players in each position

So, what can we glean from these results? We can see the 4-2-3-1 is the better formation as it has a higher max Z, but only by a margin of 0.09, so it's not a comprehensive superior. However, it does give us a clear insight of where each player should fit in the formation, and it is a useful method of putting out lineups based on recent form. Teams may also do this using average ratings from a whole season, which would include more players, and test out more formations to see which has the highest max Z to try out new tactics. Obviously, football is more complex than the optimum solution for Z, but what could possibly stop this 4-2-3-1 starting XI from winning the Premier League next season?

4 The Perfect Shot

Figure 3: Akachi Anyanwu misses an absolute sitter.

"He's through on goal, he has the chance to put his team in the lead, HE SHOOTS!"

"Oh no, he's made an absolute mess of a great opportunity, no striker's instinct there, unfortunately."

That's what I imagine a commentator would have said about my miss, if I played at a level high enough to have my games broadcasted on television. Yes, the goalkeeper saved this shot but without a doubt, if I had been thinking about the cosine rule and projectile motion formulae in that moment, I would have been sure to bury the ball in the back of the net. Let's look into how maths could have given me the assist here.

In my mind, I had two options here, a power shot into the bottom left corner, or a chip shot (lobbing the ball over the keeper), as the keeper was off his line. I opted for the former, and the keeper made a great save, so I will explore why the latter was a better option.

First, I will briefly explain why my in-game decision was nearly impossible to execute. The initial image is slightly deceptive as it doesn't accurately portray how narrow the shot was. The following camera angle does it more justice. Let's use this to explore my chances of scoring.

Figure 4: Elevated view of Akachi Anyanwu missing an absolute sitter

You just can't write about maths and football and not mention expected goals, or xG as it is known. It is one of the most popular and most used statistics in football as it is an easy way for fans to assess how likely a player is to score a given shot, with a quantifiable number between 0 and 1, where 1 is a guaranteed goal. xG is an incredibly complex statistic which calls on heaps of data from matches and shots taken to train and test the model. While each xG model has its own characteristics, they tend to consider variables such as body part the shot was taken from (e.g. header), the previous action (e.g. dribble, receiving from a cross, through-ball, etc.), and most importantly, distance and angle to goal.

For a low driven shot, my conditions were mostly favourable: I was close to goal, I had dribbled in behind through their defence, so I didn't have to deal with the ball awkwardly bouncing, and I took the shot on my dominant right foot. However, it was the angle that made my xG drop. Below are two images that show how tight the angle to shoot I had was. I am assuming that I shot the ball powerfully enough such that the keeper would not have time to react to make a diving save, given that I placed it in the window that angle θ gives me.

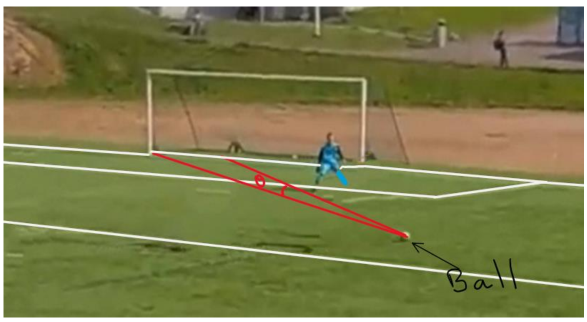


Figure 5: Two representations of the angle θ which represents the angle I had to strike the ball at to score.

Already we can visually see how tight the margin for error is. But what is the exact angle? All we need to do is use the cosine rule, $cos(A) = \frac{b^2+c^2-a^2}{2bc}$, for the following triangle using estimated lengths in meters from the (small) pitch:

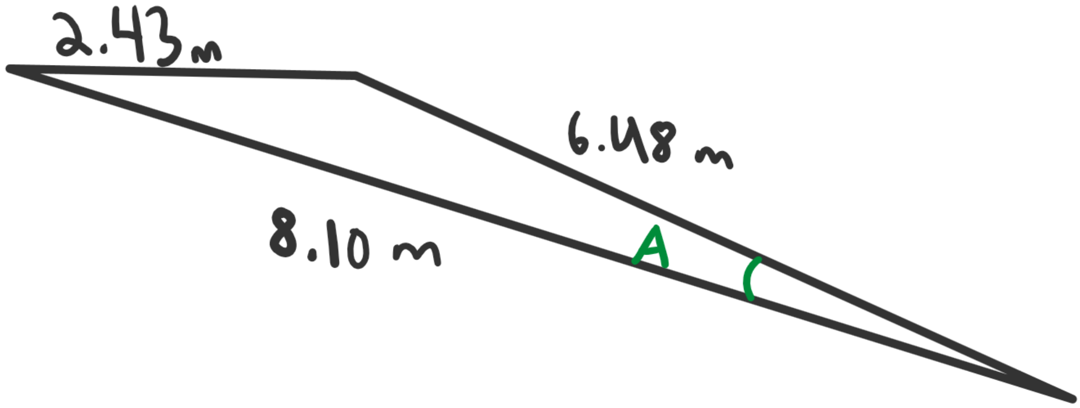


Figure 6: Lengths required for angle calculation. A = θ from figure 5.

Therefore, with the cosine rule, we obtain the following:

$$A = cos^{-1}\left(\frac{(8.10)^2 + (6.48)^2 - (2.43)^2}{2(6.48)(8.10)}\right)$$

= 14.4 degrees

This tiny angle would invariably greatly decrease the xG on any given model, rendering my probability of scoring very low, hence highlighting my incorrect choice of shot. I must begrudgingly also admit that me missing was somewhat due to a lack of skill, as although the angle is small, there is nonetheless still an angle for me to have scored,

The reason I didn't attempt the chip shot may also have been to a lack of skill, as I didn't fancy myself to have the incredible precision and accuracy required for such a delicate strike. Before we explore the chip, we have to take a brief look at forces which impact a trajectory of a football in the below diagram[9]:

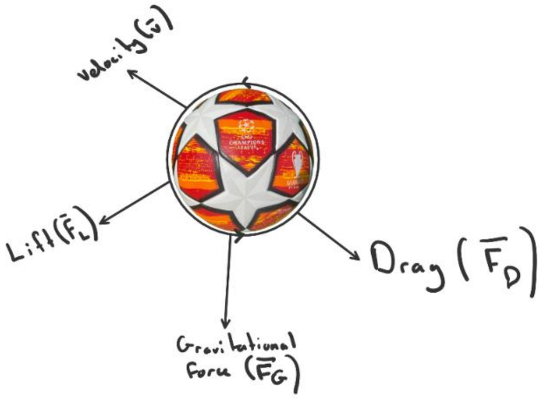


Figure 6: Forces acting on a football in motion.

In this diagram we can see three different forces acting on the ball: lift, drag, and gravitational force. However, the essence of the chip shot is being able to ignore drag and lift and focus on gravitational force, $\bar{F}_G$. We can make $\bar{F}_G$ dominant by satisfying the conditions $\bar{F}_G > \bar{F}_L$, and $\bar{F}_G > \bar{F}_D$. It is also crucial that an initial velocity of less than 20 m/s[10] is applied to the ball. Essentially, the air surrounding the ball will have no impact on the ball's motion, and spin would not be high enough to induce other forces. The final condition is that drag, $\bar{F}_D$, must be slightly greater than the effect of spin.

If these are satisfied, the other forces are ignored and we can manipulate Newton's second law, $f = ma$ as follows:

We start with $f = ma$, of course. Since we are considering solely the motion under gravity, we can obtain an expression for, $\bar{F}_G$ in vector notation:

$$\bar{F}_G = -mg\hat{q} \tag{7}$$

Where $\hat{q}$ is a unit vector in the vertical direction. Since we are considering one dimensional vertical motion, we can express acceleration (a) as the second derivative of the vertical position, y, with respect to time as follows:

$$a = \frac{d^2y}{dt^2}$$

We can then equate $\bar{F}_G$ to mass (m) and vertical acceleration, ($\frac{d^2y}{dt^2}$) to get:

$$-mg\hat{q} = m\left(\frac{d^2y}{dt^2}\right)\hat{q}$$

Since force and acceleration are scalar multiples of unit vector $\hat{q}$, we can drop the unit vector and simplify the equation to:

$$m\left(\frac{d^2y}{dt^2}\right) = -mg \tag{8}$$

Hence, the second derivative of the vector position with respect to time gives the acceleration due to gravity. Now this has been established, we can model the chip shot as having a clean parabolic trajectory:

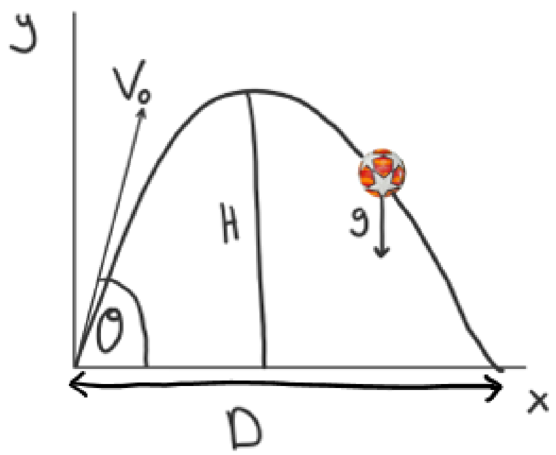


Figure 7: Parabolic motion of a chip shot

Where V_0 is the initial velocity, D is horizontal distance the ball travels before hitting the ground, H is height, and θ is the angle of projection. We can now apply the useful equation for a projectile thrown in a parabolic path:

$$y = x\,tan\,\theta - \frac{gx^2}{2(V_0)^2\,cos^2\,\theta} \tag{9}$$

From this equation, we can get equations to express D, H, and time of ball flight (t) respectively[11]:

$$D = \frac{(V_0)^2}{g} sin(2\theta) \tag{10}$$

$$H = \frac{(V_0)^2\,sin\,2\theta}{2g} \tag{11}$$

$$t = \frac{2V_0 sin\theta}{g} \tag{12}$$

Now let's take things back to the real world and examine what I needed to do to execute a world class finish. I was about 7 meters away from goal as the following diagram shows.

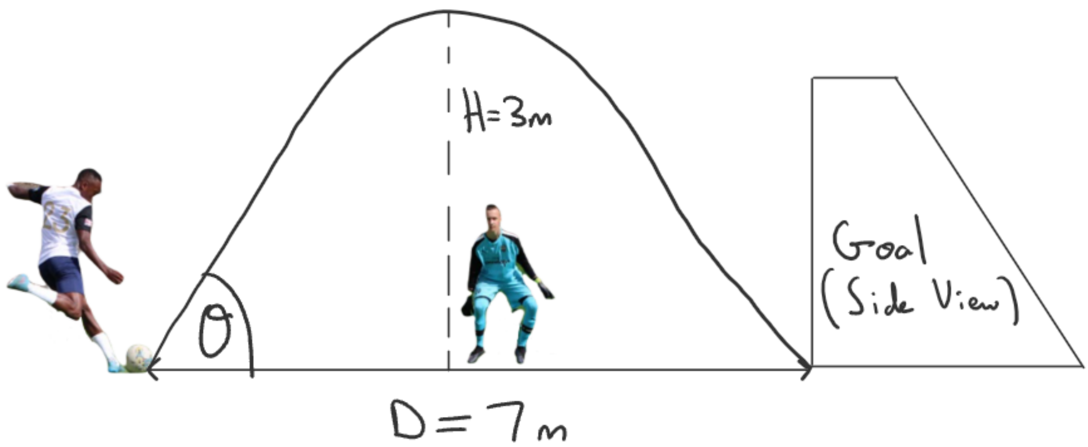


Figure 8: Side view of the chip shot

As the keeper is about 1.8 meters tall, he can increase the height he covers by raising his hands and jumping to make a save, so to be safe I would have to shoot the ball such that it would be 3 meters high when it passes over the keeper. To calculate the velocity, I need to hit the ball (which must be less than 20 m/s), the angle of projection, and the time for the ball to reach the goal we can use the above equations with our values for H and D. After crunching some numbers in equations 10, 11, and 12, I reached the values:

$$V_0 = 8.53ms^{-1}$$

$$\theta = 45^o$$

$$t = 1.07s$$

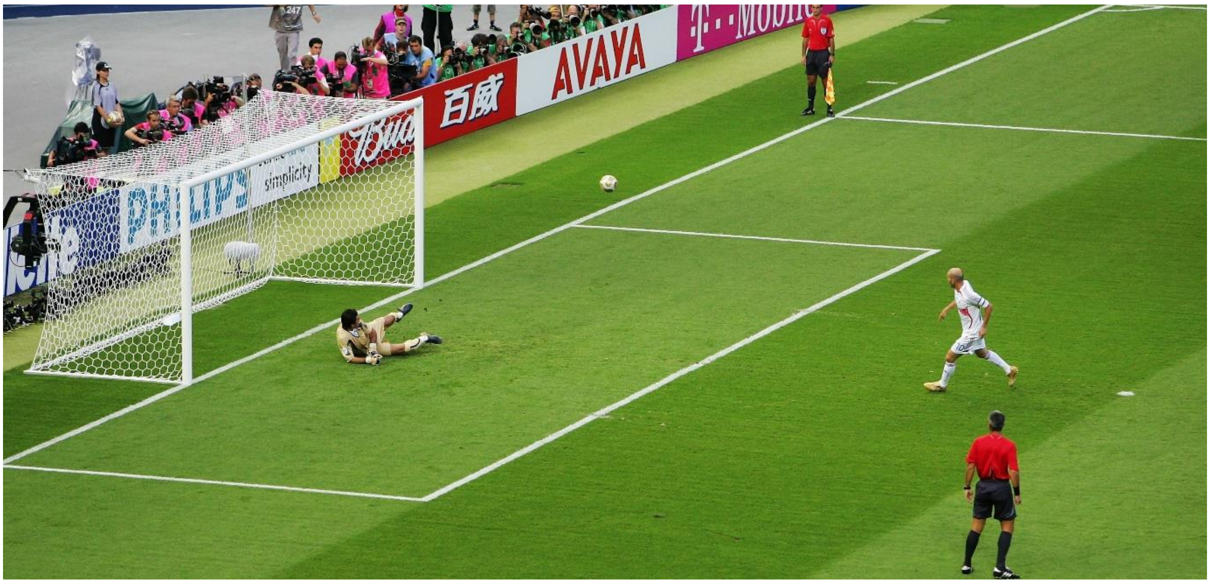
So, to have executed the perfect chip shot in this scenario, I would have had to strike the ball at a speed of $8.53ms^{-1}$ at an angle of projection of 45 degrees (this projection angle will always be optimal), and the ball would have travelled in a clean parabolic path over the keeper's head into the back of the net. Any nearby defender would have only had 1.07

seconds to prevent the goal, which makes the chip shot so beautiful and effective. The conclusion here is simple; remember your trigonometry and projectile motion when you're through on goal, and you will be sure to score!

5 The Penalty Kick

As someone born in the wake of the 2006 FIFA World Cup Final, I sometimes ponder the fact that if I was born just 6 months earlier, I would have been able to witness the legendary match between France and Italy, as a newborn baby of course. While my debut to the world was impending, I unfortunately missed France's Zinedine Zidane's iconic headbutt, and Italy's goalkeeper Gianluigi Buffon making heroic save after save. What I do know is that during that game, before Zidane lost his temper and got sent off the pitch, he took what in my opinion was the greatest penalty of all time.

When most players take on the penalty kick, an unopposed shot on goal from 11 metres away awarded for a foul in the penalty box, they aim to shoot to the left or the right to beat the goalkeeper. On average, the shot takes less than half a second to reach the goal, and as a result the goalkeeper, who starts in the middle, must decide where to dive before the ball is struck. So which option, left or right, did Zidane choose? Neither. Not only did he take the risk of shooting down the middle, but he also did it in outrageous fashion, with a relaxed, cheeky, lofted dink of the ball. Buffon, already committed to a dive to Zidane's left, could only helplessly watch from the ground as the ball flew over him.



But why would he do something so risky in the biggest match of his career with his whole country relying on him scoring? Actually, it wasn't that risky. In fact, the middle may be one of the best places to shoot when taking a penalty. Let's break it down.

First let's look at penalty kick strategy through a game theory lens, the branch of mathematics that covers analysis of strategy in competitive situations. What makes penalties so interesting is the fact that the goalkeeper has to guess a side to dive before the ball is struck; the ball is simply hit too fast in most cases to decide where to dive after the shot is taken. Therefore, the player has three options: shoot right, left, or down the middle. The keeper also has the same three options: dive left, dive right, or stay in the middle. The player's objective is to strike the ball in either of the two positions that the keeper fails to occupy, and the goalkeeper's objective is to match the side that the player chooses. This very common type of game can be represented in a payoff matrix, but we must make a few important assumptions first. First, let's assume that both players are superhuman. That is to say, if the goalkeeper dives the same way as the kicker shoots the ball, the keeper will always save the shot, but the player will never miss the target and thus will score as long as he shoots the ball away from the keeper. This represented in the payoff matrix:

Goalkeeper

Player	Dive left	Stay Middle	Dive right
Kick Left	0, 0	1, -1	1, -1
Kick Middle	1, -1	0,0	1, -1
Kick Right	1, -1	1, -1	0,0

In this matrix, the rows represent the player's choices, while the columns represent the keeper's choices. The numbers in the table represent the payoffs of the outcomes. For example, if the player kicks left, and the goalkeeper dives left, the player has a payoff of 0 as he failed to score, and the keeper gets 0 payoff as he did not concede a goal. If the players kicks left and the goalkeeper stays in the middle, the player is successful and scores a goal with a payoff of 1, but since the goalkeeper conceded a goal, he has a payoff of -1. The goalkeeper cannot have a payoff of 1 because he cannot score a goal in this scenario (you may notice that he technically can get a payoff of 1 between -1 when conceding and 0 when saving; this is owing to the fact that if we set all keeper payoffs to 0, it seems as though he is indifferent to saving the ball or not, which is not correct).

Game theory dictates two types of strategies that can be employed when modelling this type of competitive scenario: Pure Nash Equilibrium Strategy and mixed strategy. In simple terms, a pure strategy is where a competitor sticks to the same outcome every time. For example, this would be a striker shooting to the right every time he gets a penalty. This is problematic because it can be easily exploited; a goalkeeper defending this sort of player will always dive right, and this reduces striker payoff. Therefore, we cannot apply this to a penalty shootout. Instead, we must look to a mixed strategy, where you do not choose the same outcome every time. For example, this would be an attacker shooting right some of the time, left some of the time, and down the middle as well. This is impossible to exploit as the keeper essentially has to guess every time where to dive. We can calculate the mixed strategy to see how often a player should shoot in a certain location to optimize their output.

First, let's denote the probabilities of the player kicking left, middle, or right as x_l , x_m, and x_r respectively. We can set up equations for the expected payoff for kicking left, middle, and down the right as follows:

$$E(P_{left}) = (x_l \cdot 0) + (x_m \cdot 1) + (x_r \cdot 1)$$

$$E(P_{middle}) = (x_l \cdot 1) + (x_m \cdot 0) + (x_r \cdot 1)$$

$$E(P_{right}) = (x_l \cdot 1) + (x_m \cdot 1) + (x_r \cdot 0)$$

Because the expected payoffs for each strategy should be equal in the mixed strategy, we can set these equations equal to each other:

$$E(P_{left}) = E(P_{middle}) = E(P_{right})$$

$$(x_l \cdot 0) + (x_m \cdot 1) + (x_r \cdot 1) = (x_l \cdot 1) + (x_m \cdot 0) + (x_r \cdot 1)$$

$$(x_l \cdot 0) + (x_m \cdot 1) + (x_r \cdot 1) = (x_l \cdot 1) + (x_m \cdot 1) + (x_r \cdot 0)$$

Simplifying the equations we then get

$$x_m = x_l \text{ and } x_r = x_l$$

Therefore, each probability has the same value. Let's call this value *j*. Since all the probabilities must add up to 1, then:

$$x_l + x_m + x_r = 1$$

$$3j = 1$$

$$j = \frac{1}{3}$$

Therefore, the mixed strategy equilibrium is $x_m = x_r = x_l = \frac{1}{3}$ which means a player should randomly choose to kick in each direction with an equal probability of $\frac{1}{3}$ to maximise their expected payoff. Essentially choose randomly which part of the goal you will kick the ball to, but make sure to evenly distribute this random choice.

However, in reality, this is much different. For example, a right footed player like me will find it easier to strike the ball to the left of the goal and will be weaker when shooting to the right. Fortunately, we can also model this with the mixed strategy equilibrium. We can use the same matrix as above with a slight tweak. Let's say the kicker misses the goal half of the time he shoots to the right. We would then update the kick right dive left, and the kick right dive middle payoff to (0.5, -0.5) as we now have to write it as a probabilistic move. Using the same exact steps as above, I calculated that the kicker should now shoot to the right 2/3s of the time, and down the middle and to the left for a combined 1/3 of the time. It seems strange that the player should shoot most of the time to a side that he has a 50% chance of missing! The reason for this lies in the goalkeeper's strategy. If a player favours a side that he is better at shooting, then the goalkeeper will plan for this and exploit the players weakness. Therefore, to keep the equilibrium balanced, the player must shoot to his weaker side more to counteract the goalkeeper predicting a shot to the stronger side.

So that is what game theory says on the matter: you should aim to distribute your shots randomly and equally across the three options: left, middle, and right, and if you have a weaker side, be sure to shoot to that side more. In reality, the facts seem to back this random shot selection up as analysed patterns of penalty shot selection in matches seem to purely be random draw[12 13]. However, although the players do seem to follow this random strategy, a pretty big aspect is omitted: shooting down the middle. Players will adopt this strategy for only shooting left or right. Even most resources online which model and explain the penalty kick ignore the middle. I delved deeper into this fascinating blind spot and found that in the top 5 leagues, goalkeepers choose not to dive and stay in the middle only 2 percent of the time, and in the past half century in the Champion's league and the World Cup, goalkeepers stayed in the middle only 3% of the time[14]. There is a logical reason for this: if a keeper chooses not to dive and concedes, it looks as if he did not even attempt to make a save, and a high degree of embarrassment and criticism follows. Likewise, the same goes for players. If a player shoots to a side and the ball is saved, credit goes to the keeper, and not too much criticism to the player. However, if the keeper saves the penalty by standing still, the player will be berated and relieved of his penalty taking duties.

Of course, if every footballer stuck to what the game theory dictates, this blind spot would not exist, but this psychological aspect skews it. However, this was what Zidane used to his advantage with his stunning penalty. The keeper Buffon likely only considered left and right outcomes and guessed randomly, and Zidane was able to exploit this fact and give his country the lead in the World Cup Final in an iconic way. Unfortunately, and ironically, France ended up actually losing this game in a penalty shootout, but if Zidane had imparted his knowledge of game theory and his research on the blind spot (the middle) upon his teammates, they surely would have won the 2006 World Cup.

6 Conclusion

Funnily enough, the part about this essay I found the hardest was choosing what to write about. There are so many fascinating intersections of maths and every single dimension of the football world that I was spoiled for choice (and maybe got a tiny bit carried away). From the maths behind a chip shot to looking at a formation as an objective function, I hope this essay has given an appreciation of a few of the countless ways the beautiful game is backed by a beautiful subject.

References

[1] Užík M, Warias R, Glova J. Management of transfer prices in professional football as a function of fan numbers. Mathematics [Internet]. 2022 Aug 18;10(16):2982. Available from: https://www.mdpi.com/2227-7390/10/16/2982

[2] Kylian Mbappé - Stats 22/23 [Internet]. Transfermarkt. Available from: https://www.transfermarkt.co.uk/kylian-mbappe/leistungsdaten/spieler/342229/saison/2022/plus/1#gesamt

[3] Wikipedia contributors. 2022–23 Paris Saint-Germain F.C. season [Internet]. Wikipedia. 2024. Available from: https://en.wikipedia.org/wiki/2022%E2%80%9323_Paris_Saint-Germain_F.C._season

[4] https://store2.psg.fr/en/paris-saint-germain-kylian-mbappe/t-36200371+a-6735576413+z-99-3381033309

[5] Chris Jones. Full breakdown: The price of a football shirt. Footy Headlines [Internet]. 2023 Jul 3; Available from: https://www.footyheadlines.com/2023/07/full-breakdown-price-of-football-shirt.html#:~:text=The%20Cost%20of%20a%20Football,usually%20taking%20place%20in%20Asia.&text=When%20a%20shirt%20is%20sold,manufacturing%20brand%20receives%20%C2%A323.47.

[6] IBID

[7] https://sofascore.com

[8] Hillier, F. S., & Lieberman, G. J. (1995). Introduction to operations research. New York: McGraw-Hill Science, Engineering & Mathematics.

[9] Sánchez, H.C. (2018) *Bend it like Newton: curves in football*. https://chalkdustmagazine.com/blog/bend-it-like-newton-curves-in-football/.

[10] Sánchez, H.C. (2018b) *The maths behind a chip goal*. https://chalkdustmagazine.com/blog/the-maths-behind-a-chip-goal/.

[11] Leela, J.K., Comissiong, D.M.G (2009) *Modelling football penalty kicks, Lat. Am. J. Phys. Educ.*, pp. 259–260. http://www.lajpe.org/may09/12_Jeffrey.pdf.

[12] Chiappori, P.-A., Levitt, S. and Groseclose, T. (2002). Testing Mixed-Strategy Equilibria When Players Are Heterogeneous: The Case of Penalty Kicks in Soccer. American Economic Review, 92(4), pp.1138–1151. https://doi.org/10.1257/00028280260344678.

[13] Palacios-Huerta, I. and Brown University (2003) *Professionals play Minimax, Review of Economic Studies*, pp. 395–415. http://www.palacios-huerta.com/docs/professionals.pdf.

[14] Woodcock, S. (n.d.). *Aim for the middle: it could be your best shot for a goal in a penalty shootout*. [online] The Conversation. Available at: https://theconversation.com/aim-for-the-middle-it-could-be-your-best-shot-for-a-goal-in-a-penalty-shootout-99679.