Mathematics and Sports

A Myth or Fact?
Are Foul Calls in Soccer Influenced by Jersey Colors

Kwabena Duku[1], Olcay Akman[1], Kedai Cheng[2,*]

[1]Illinois State University, Normal, IL, [2]University of North Carolina Asheville, Asheville, NC

[*]Corresponding Author E-mail:
kcheng@unca.edu

Abstract

Color has psychological influence on human feelings and perception of aggression. Team sports are complex and combating. Some violations are subtle. Prior research studied aggression perception on specific colors, say red, white or black. In this study, authors use foul calls received to study aggression perception on light and dark colored jersey using English Premier League (EPL) data from 2009 to 2020. Authors found that home teams receive fewer foul calls than away teams. Meanwhile, home teams also tend to wear light colored jersey. A further study, conditioned on home and away teams, denies any significant statistical difference in foul calls between light and dark colored jersey teams. A more comprehensive study at referee level also confirmed that result that home teams tend to receive fewer foul calls, but not because of jersey colors.

Keywords: English Premier League (EPL), Soccer, Jersey colors, Foul calls, Aggression

1 Introduction

Team sports, such as volleyball and basketball, typically involve two teams competing against each other. The team whose headquarters is located at the stadium hosting the match is designated as the home team, while the opposing team is referred to as the away team. Each team customarily designs and selects its jerseys based on factors such as financial sponsorships and team culture. Furthermore, a team may possess multiple jersey kits, providing options for various occasions. Despite variations in the complexity of design and color schemes, regulations govern the selection of these kits.

Team sports is generally held in the stadium associated with the home team's headquarters, with the home team playing in their designated stadium, and the opposing team serving as the away team. It is common practice for teams to possess multiple jersey kits, and they are typically required to have at least two distinct colored versions of their jerseys. It is not practically feasible to investigate all team sports in one study. In this paper, we study soccer games as an example. As outlined by the Fédération Internationale de Football Association (FIFA), the colors of the team's jerseys should be sufficiently contrasting to avoid confusion during the match ([7] and [8]). These contrasting jerseys are generally categorized into light-colored and dark-colored versions. Although FIFA does not provide an official definition of light and dark colors, for the purposes of this paper, light colors are defined as those that are predominantly bright, while dark colors are characterized by deep hues.

In team sports, referees are tasked with enforcing the rules of the game. When a player violates these rules, the referee may issue a violation call and, depending on the severity of the infraction, a warning card may be issued to the player involved. Violations can occur in a variety of forms, such as aggressive physical contact with an opponent, which constitutes a breach of the game's regulations. However, violations are not limited to physical conduct; verbal confrontations may also result in fouls. Details of fouls and misconduct is specifically referred in Law 12 by The Football Association ([2]). A summary can also be found in [1]. *Caution* (Yellow Card) and *Expulsion* (Red Card) are severer in the degree of violation and more obvious than fouls and misconduct. In professional leagues, both athletes and referees are professionally trained, and cautions and expulsion are rarely issued by mistake. On the contrary, foul calls are made in response to misdemeanors such as tripping, pushing, holding, or subtle kicking opponents.

In certain ball games, such as volleyball, teams are positioned on opposite sides of the playing field, and each team is confined to its designated area. While teams switch sides during the halftime break, players do not typically engage in direct physical interaction with the opposing team. In contrast, soccer requires close physical interaction between players, necessitating that referees make judgment calls from a distance. When physical contact occurs, the referee must decide whether it is foul or not.

Despite their rigorous training and adherence to professional ethical standards, referees may still be influenced by unforeseen factors that affect their decision-making. One of these attributions is jersey color. Color is proved to have psychological influence ([15] and [6]) as such human feelings and human behaviors. [13] studied wall color and students' aggression in a classroom setting and asserted that red colored wall increase students' aggression. Aligning with this result, [16] showed that red clothing is perceived as dominance, anger, and aggression. However, a softened color tone reduces psychological association with aggression ([11]). Not only in a classroom or social setting, colors also deliver psychological indications, especially when contrast is present. The study conducted by [5] indicated that soccer teams in red jersey may have an advantage over their opponents. When both teams wear dark jerseys, say black and red, there is no evidence to support that either team is related to a higher level of aggression ([3]). However, when color contrast emerges, teams in dark colored jersey tend to receive more fouls than teams in light colored jersey. For example, a quasi-experimental study for National Hockey League (NHL) matches done by [14] showed that teams wear black jerseys were penalized more than they did not, and when teams switched to colored-jerseys at home games, they receive more penalties when wearing white jerseys when playing as homes team.

The aforementioned research studied psychological indications of either a specific color (red or Baker-Miller pink) ([11]), or compared penalty received by two specific colors (black versus red, black versus non-black, or white versus non-white) in NHL games ([14]). In soccer games, jersey colors are generally classified as light or dark. Little research have been conducted to study the effect between light and dark colored jerseys. This paper examines the hypothesis that foul calls may be influenced by the color of a team's jerseys, a topic that has fueled consistent speculation among fans. The structure of the paper is as follows: Section 2 presents and describes the dataset used in the analysis. Sections 3 and 4 outline the methodology and analysis, respectively. Finally, Section 5 offers a discussion of the findings.

2 Data

Various soccer data sets are available. Some datasets are brief, while others cover a broad spectrum of statistics for each match. Our paper focuses on the effect of jersey color on foul calls. Due to the limitation of available datasets, we use data from English Premier League (EPL) from DataHub.io ([4]) as it is available and also provides a comprehensive list of variables that we look for.

Our data set comprises 3,800 matches from the English Premier League (EPL) spanning 10 seasons, from 2009–2010 to 2018–2019. It includes the match date, the names of the home and away teams, and the referee assigned to each match. In addition, the data set records foul calls and the jersey colors of both teams for analysis. Jersey colors are categorized as 'LC' (light-colored) and 'DC' (dark-colored) based on their RGB values, using the relative luminance formula, where a luminance threshold of 0.5 distinguishes light from dark. Mathematically, the luminance threshold, L, is defined as

$$L = 0.2126R + 0.7152G + 0.0722B$$

where R, G, and B are red, green, and blue values, respectively, of the standardized RGB values. RGB values are obtained from the image color picker ([9]). If $L < 0.5$, the color is defined as dark color, and if $L \geq 0.5$, it is a light color. A snippet of the dataset is presented in Figure 1.

3 Methodology

A $r \times c$ contingency table is used for distributional and frequency studies of two categorical variables, which has r and c factor levels, respectively. A χ^2-test is implemented to justify the independence between two qualitative variables. The null and alternative hypotheses of such test are stated as

$$\begin{aligned} \mathrm{H}_0 &: \text{Factor A is independent of Factor B} \\ \mathrm{H}_A &: \text{Factor A is not independent of Factor B} \end{aligned}$$

A 2×2 contingency table is demonstrated in Table 1. The test statistic of a χ^2-test is obtained as

$$\chi^2 = \frac{(ad-bc)^2(a+b+c+d)}{(a+b)(c+d)(b+d)(a+c)}$$

	Date	HomeTeam	AwayTeam	Referee	HF	AF	HomeJersey	AwayJersey
1	2010	Aston Villa	Wigan	M Clattenburg	15	14	LC	DC
2	2010	Blackburn	Man City	M Dean	12	9	LC	DC
3	2010	Bolton	Sunderland	A Marriner	16	10	LC	LC
4	2010	Chelsea	Hull	A Wiley	13	15	LC	LC
5	2010	Everton	Arsenal	M Halsey	11	13	LC	DC
6	2010	Portsmouth	Fulham	M Atkinson	11	18	LC	DC
7	2010	Stoke	Burnley	S Bennett	15	10	LC	LC
8	2010	Wolves	West Ham	C Foy	9	5	LC	DC
9	2010	Man United	Birmingham	L Mason	13	7	LC	DC
10	2010	Tottenham	Liverpool	P Dowd	14	16	LC	DC
11	2010	Sunderland	Chelsea	S Bennett	14	10	LC	DC
12	2010	Wigan	Wolves	M Jones	8	21	LC	LC
13	2010	Birmingham	Portsmouth	L Probert	11	20	LC	LC
14	2010	Burnley	Man United	A Wiley	8	12	LC	DC
15	2010	Hull	Tottenham	C Foy	23	13	LC	DC
16	2010	Liverpool	Stoke	P Walton	10	9	LC	DC
17	2010	Arsenal	Portsmouth	S Bennett	9	10	LC	LC
18	2010	Birmingham	Stoke	C Foy	8	13	LC	DC
19	2010	Hull	Bolton	M Jones	19	19	LC	DC
20	2010	Man City	Wolves	L Mason	7	11	LC	LC

Figure 1: Data Snippet

where a, b, c, and d are count for each cell. A p-value is calculated based on a χ^2-distribution with degree of freedom of $(r-1)(c-1)$, where r and c are number of rows and columns of the contingency table, respectively. Data available to us present foul calls for home

Table 1: 2×2 Contingency Table

		Factor 1	
		Level 1	**Level 2**
Factor 2	**Level 1**	a	b
	Level 2	c	d

teams and away teams wearing both light and dark colored jerseys. Unfortunately, two layers of information, being home or away team and wearing light or dark jersey, is mixed. We use the χ^2-test to determine whether jersey color is independent of a team playing at home or away.

A two-sample t-test is a commonly used statistical test when sample size is large and the population variances of two underlying distributions are believed to be equal. A two-sample Welch's t-test releases the homogeneity assumption and grants a more flexible test environment. A two-sided Welch's t-test is conducted under the null hypotheses that two population means (μ_1 and μ_2) are equal, mathematically,

$$\begin{aligned} \mathrm{H}_0 &: \mu_1 = \mu_2 \\ \mathrm{H}_A &: \mu_1 \neq \mu_2 \end{aligned}$$

The test statistic of the Welch's t-test is obtained as

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$\overline{X}_i$ is the sample mean, s_i^2 stands for the sample variance, and n_i represents sample size, for $i = 1, 2$. The degree of freedom, ν, of the underlying t-distribution is

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}}$$

The data provided to us includes 3,800 matches across 10 seasons, with a total of 82,658 foul calls recorded. Among these, 40,351 fouls were called on the home side, while 42,307 fouls were called on the away side, all of which are available for analysis. Given this considerably large sample size, we use Welch's t-test for testing the difference of foul calls between home and away teams, and light and dark jerseys as a whole.

Unlike Welch's t-test, which assumes an underlying distribution, the Wilcoxon rank sum test is a nonparametric two-sample test that compares two independent population medians when sample sizes are small or underlying population distributions do not suit in a symmetric and bell-shaped curve. A two-tail Wilcoxon test also justifies the equality of two population means, such that

$$\begin{aligned} \mathrm{H}_0 &: \text{median}_1 = \text{median}_2 \\ \mathrm{H}_A &: \text{median}_1 \neq \text{median}_2 \end{aligned}$$

The test statistic of Wilcoxon rank sum test is obtained as

$$W = R_1 - \frac{n_1(n_1+1)}{2}$$

where R_1 is the sum of rank of all observations in sample 1, and n_1 is the corresponding sample size. Details of the rank sum test and handling of ties are beyond the scope of this paper, but they can be referred to [17], [18], and [10].

In this paper, test statistics and corresponding p-values are computed under software R 4.3.2. More specifically, functions of `chisq.test`, `t.test`, and `wilcox.test` in the `stats` package are used. Details can be found in [12].

4 Analysis

4.1 Dependency between Jersey Color and Home Team

In our dataset, jersey color is recorded as a binary qualitative variable of either light or dark for home and away teams. Provided multiple jersey color options, a team may have either a dark colored jersey kit or a light colored jersey kits regardless of being the home or away team. A contingency table, Table 2, summarizes the jersey colors in each match. Among all 3,800 matches analyzed, home teams wore light-colored jerseys in 2,717 matches (72.5%), while away teams wore dark-colored jerseys in 1,729 matches (45.5%).

Table 2: Contingency Table for Jersey Color

	Home Team	**Away Team**
Light Jersey	2717	1729
Dark Jersey	1083	2071

A χ^2-test is used to justify whether a team not wearing a light colored or dark colored jersey is independent of being the home or away team in a match. Although there is no restriction of color choice for being the home or away home, that said, it is possible that both home and away teams choose light colored jersey, or both choose dark colored jersey, the negligible p-value ($< 2.2e^{-16}$) shows that home teams tend to wear a light colored jersey more often than away teams. Given the dependency between being the home and away team and jersey color, analyses need to be practiced with caution.

4.2 Foul Calls for Home and Away Teams

Look at the aggregated number of foul calls for home and away teams as a whole. Among 3800 matches, home teams receive an average of 10.62 foul calls with a standard deviation of 3.40 for each match, while away teams receive more foul calls in general, with an average of 11.13 and standard deviation of 3.58. A Shapiro tests, with both p-values $< 2.2e^{-16}$, are performed to affirm the normality of the distributions of foul calls for both home and away teams. The empirical distribution of foul calls for home and away teams are presented in Figure 2 and descriptive statistics summarized in Table 3. A two-sample Welch's test is followed to test the equality of foul calls

Table 3: Foul Calls Received by Home and Away Teams

	Home Teams	**Away Teams**
Mean	10.62	11.13
S.D.	3.40	3.58
95% **C.I**	(10.51 , 10.73)	(11.02 , 11.25)
P-value for comparison is $1.369e^{-10}$.		

between home and away teams as a whole. The p-value of $1.3690e^{-10}$ shows the significant statistical difference in the number of foul calls between home and away teams.

To further verify the result, a two-sample Wilcoxon's test is utilized for each condition when both the home and away teams wear the same type of color. Distributions of number of foul calls are shown in Figure 3 and respective statistics are included in Table 4. When

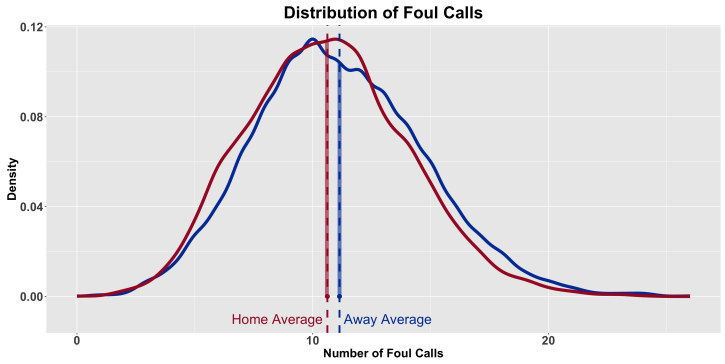

Figure 2: Distribution of Foul Calls
(Shaded Areas are 95% Confidence Intervals)

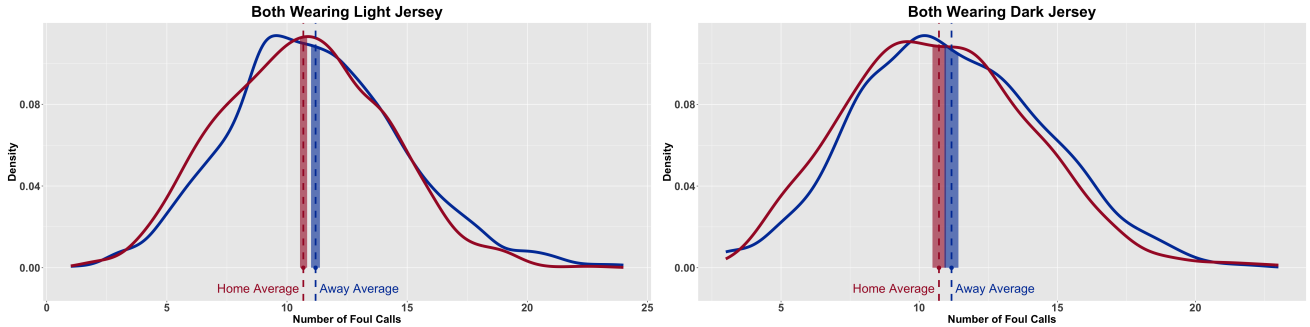

Figure 3: Distribution of Foul Calls When Wearing Same Color
(Shaded Areas are 95% Confidence Intervals)

Table 4: Foul Calls Received by Home and Away Teams
(Both Teams Wearing Same Colored Jerseys)

	Both Wearing Light Jerseys		**Both Wearing Dark Jerseys**	
	Home	**Away**	**Home**	**Away**
Mean	10.68	11.19	10.72	11.17
S.D.	3.36	3.59	3.32	3.40
95% **C.I**	(10.49 , 10.87)	(10.99 , 11.39)	(10.45 , 10.98)	(10.90 , 11.44)
P-value for comparison is 0.0020.			**P-value for comparison is** 0.0183.	

both teams wear light jerseys, home teams receive fewer foul calls with mean of 10.68 and standard deviation of 3.36, while away teams get 11.19 foul calls on average with standard deviation of 3.59. The p-value of 0.0020 shows the significance in difference between number of foul calls between home and away teams.

Similarly, when both teams wear dark jerseys, home teams receive an average of 10.72 foul calls with standard deviation of 3.32, while away teams have an average of 11.17 foul calls in each match with standard deviation of 3.40. The consistent small p-value of 0.0183 reassures the statistical difference in the number of foul calls between home and away teams.

4.3 Foul Calls for Light and Dark Jerseys

Despite home teams receiving fewer foul calls during each match, it is still unclear to us whether jersey color is the primary cause of the discrepancy of the number of foul calls. In this section, we compare foul calls between light and dark jerseys conditioned on being the home or away team. Explicitly, we first look at number of foul calls for different jersey color at the aggregated level. Then we create two pools of matches. One of the groups are home teams but wearing different colored jerseys. The other group are all away teams with different jersey colors. Number of foul calls are compared within each group. In such a way, the noise caused by being the home and away team is filtered. Distributions of foul calls for each group are shown in Figure 4, and descriptive statistics are listed in Table 5.

Generally, teams wearing light colored jersey receive an average of 10.83 foul calls, with standard deviation of 3.50, in each match, and teams with dark colored jersey receive an average of 10.95 foul calls with standard deviation of 3.49. The p-value (0.1327) of a

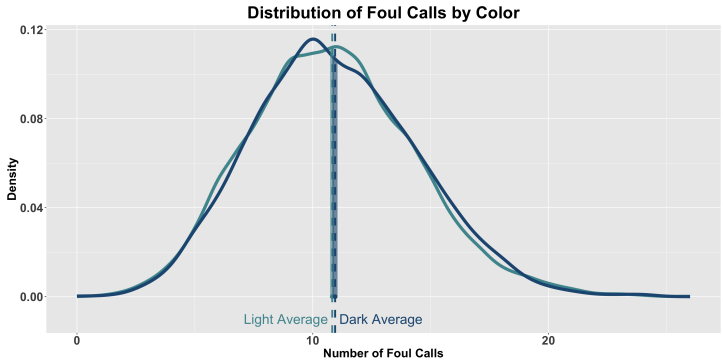

Figure 4: Distribution of Foul Calls by Color
(Shaded Areas are 95% Confidence Intervals)

Table 5: Foul Calls Received by Light- and Dark-Jersey Teams

	Light Jersey	**Dark Jersey**
Mean	10.83	10.95
S.D.	3.50	3.49
95% **C.I**	(10.72 , 10.93)	(10.83 , 11.07)
P-value for comparison is 0.1327.		

two-sample Welch's test shows the insignificance in terms of foul call numbers between different jersey colors.

Within group comparison is further performed. Distributions of foul calls for home teams and away teams in different colors of jerseys are shown in Figure 3, and results are summarized in Table 6. Home teams receive an average of 10.61 foul calls, with standard

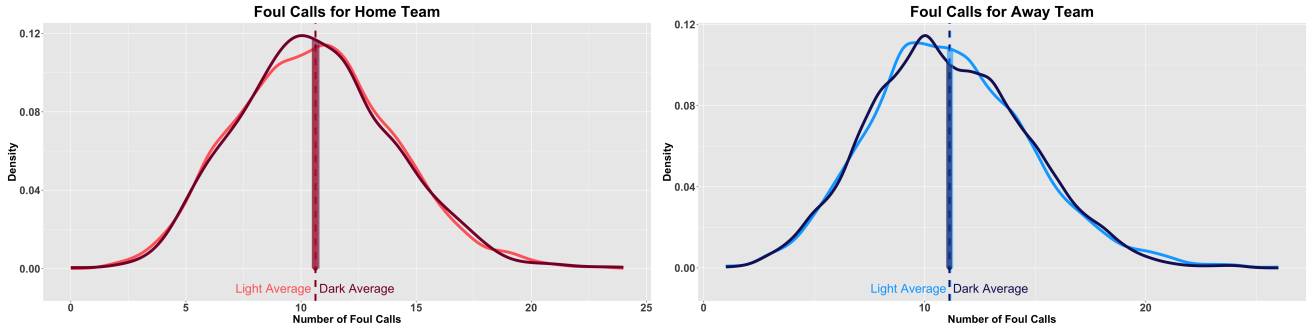

Figure 5: Distribution of Foul Calls for Home and Away Teams
(Shaded Areas are 95% Confidence Intervals)

Table 6: Foul Calls Received by Light- and Dark-Jersey Teams
(Categorized by Home/ Away Teams)

	For All Home Teams		**For All Away Teams**	
	Light Jersey	**Dark Jersey**	**Light Jersey**	**Dark Jersey**
Mean	10.61	10.63	11.16	11.11
S.D.	3.41	3.36	3.62	3.55
95% **C.I**	(10.49 , 10.74)	(10.43 , 10.83)	(10.99 , 11.33)	(10.96 , 11.27)
P-value for comparison is 0.9792.			**P-value for comparison is** 0.9038.	

deviation of 3.41, in each match when wearing a light colored jersey, and receive an average of 10.63 foul calls, with standard deviation of 3.36, when wearing a dark colored jersey. The two-sample Welch's test with a p-value of 0.9792 confirms the insignificance in terms of foul calls between jersey color choices for all home teams.

Similar results apply to away teams. While wearing light jerseys, away teams receive an average of 11.16 foul calls with standard

deviation of 3.62, and approximately 11.11 foul calls with standard deviation of 3.55 when in a dark colored jerseys. The two-sample Welch's test has a p-value of 0.9038, which aligns with our speculation that jersey colors do not have any significant impact on number of foul calls received statistically.

4.4 Statistical Tests for Individual Referee

In this section, we break down the number of foul calls received under each condition by individual referees. Due to small sample sizes, a two-sample Wilcoxon test is implemented to test the difference in foul calls under each condition for every referee. Number of games participated for each referee and their names are tabulated in Table 7. In total, we have 33 referees participated during the 10 seasons. The maximum number of games referred by a referee is 291, while the minimum game is 1. On average, each referee participated 115.15 games over 10 season, with standard deviation of 91.99.

Table 8 shows average number of foul calls given by each referee to the home and away team. Table 9 lists average number of foul calls given by each referee to different jersey colors. In cases where the sample size is too small to obtain the statistic, a "NA" is returned. It is worth noting that each referee is assigned a unique numerical ID, and ID's in Table 7, 8, and 9 are consistent. That said, Referee 7 is *S Bennett*, and he participated 28 games (see Table 7). For all 28 matches *S Bennett* participated, he gave 10.93 foul calls to home teams, and 12.46 foul calls to away teams on average during each match (see Table 8). In addition, *S Bennett* gave and average of 11.61 foul calls to teams in light jersey, and an average of 11.74 foul calls to teams in dark jersey for each match he participated (see Table 9). A two-sample Wilcoxon rank sum test is performed for each referee to test the hypothesis that foul calls given to home

Table 7: Number of Games Participated by Referee

Referee ID	Name	Number of Games	Referee ID	Name	Number of Games	Referee ID	Name	Number of Games
1	M Clattenburg	206	12	L Probert	159	23	R East	91
2	M Dean	291	13	P Walton	69	24	C Pawson	134
3	A Marriner	253	14	H Webb	147	25	R Madley	91
4	A Wiley	26	15	S Attwell	89	26	P Tierney	55
5	M Halsey	61	16	K Friend	204	27	K Stroud	3
6	M Atkinson	278	17	St Bennett	1	28	G Scott	50
7	S Bennett	28	18	Mn Atkinson	2	29	S Hooper	10
8	C Foy	135	19	A Taylor	226	30	C Kavanagh	40
9	L Mason	221	20	M Oliver	234	31	L Mason	1
10	P Dowd	159	21	N Swarbrick	132	32	D Coote	12
11	M Jones	190	22	J Moss	200	33	A Madley	2

Table 8: Foul Calls for Home and Away Teams Under Same Jersey Color by Referee

Referee ID	Home Team Mean (S.D.)	Away Team Mean (S.D.)	P-value	Referee ID	Home Team Mean (S.D.)	Away Team Mean (S.D.)	P-value	Referee ID	Home Team Mean (S.D.)	Away Team Mean (S.D.)	P-value
1	10.80 (3.28)	10.86 (3.51)	0.8106	**12**	9.81 (3.45)	10.78 (3.80)	0.0354	**23**	11.09 (3.35)	11.52 (3.23)	0.2822
2	10.57 (3.25)	10.98 (3.43)	0.1407	**13**	9.41 (2.73)	10.09 (2.94)	0.0600	**24**	10.80 (3.90)	11.33 (3.80)	0.3034
3	10.55 (3.35)	10.41 (3.58)	0.5561	**14**	10.72 (3.42)	11.26 (3.65)	0.2473	**25**	10.43 (3.74)	11.75 (3.62)	0.0037
4	12.04 (3.40)	12.50 (4.23)	0.6326	**15**	10.34 (3.73)	10.13 (3.80)	0.6646	**26**	10.65 (3.37)	10.85 (3.55)	0.9425
5	9.28 (3.05)	10.10 (3.68)	0.1634	**16**	10.75 (3.64)	11.33 (3.61)	0.1545	**27**	14.00 (2.00)	12.00 (3.00)	0.5066
6	10.60 (3.43)	11.53 (3.48)	0.0011	**17**	9.00 (NA)	15.00 (NA)	1.0000	**28**	9.38 (3.17)	10.80 (3.25)	0.0190
7	10.93 (3.04)	12.46 (3.97)	0.2470	**18**	12.50 (0.71)	14.50 (0.71)	0.2453	**29**	12.30 (2.98)	11.80 (4.39)	0.7311
8	10.49 (3.58)	10.65 (3.51)	0.5672	**19**	11.02 (3.54)	10.44 (3.74)	0.0336	**30**	9.60 (2.52)	10.25 (2.31)	0.1436
9	10.76 (3.33)	11.41 (3.25)	0.0275	**20**	10.88 (3.35)	11.24 (3.37)	0.3596	**31**	11.00 (NA)	10.00 (NA)	1.0000
10	10.94 (3.17)	12.32 (3.66)	0.0009	**21**	10.13 (3.10)	10.80 (3.48)	0.2083	**32**	10.42 (2.11)	12.42 (3.53)	0.1460
11	11.11 (3.47)	12.25 (3.78)	0.0036	**22**	10.69 (3.28)	11.25 (3.51)	0.1546	**33**	8.50 (2.12)	8.00 (4.24)	1.0000

teams are statistically significantly different from foul calls given to away teams. Eight (Referee 6, 9, 10, 11, 12, 19, 25, and 28) out of thirty-three referees have a significant result with a p-value < 0.05 for the comparison test. Referee 13 has a p-value of 0.06 for the

Table 9: Foul Calls for Light and Dark Jerseys Under Same Team by Referee

Referee ID	Light Jersey Mean (S.D.)	Dark Jersey Mean (S.D.)	P-value	Referee ID	Light Jersey Mean (S.D.)	Dark Jersey Mean (S.D.)	P-value	Referee ID	Light Jersey Mean (S.D.)	Dark Jersey Mean (S.D.)	P-value
1	11.24 (3.45)	10.57 (3.33)	0.0466	**12**	10.68 (3.71)	10.04 (3.61)	0.2851	**23**	11.25 (3.30)	11.35 (3.30)	0.7943
2	10.66 (3.30)	10.86 (3.38)	0.6936	**13**	10.00 (2.77)	9.56 (2.91)	0.4493	**24**	11.11 (4.21)	11.03 (3.54)	0.9829
3	10.31 (3.39)	10.60 (3.51)	0.3496	**14**	11.04 (3.33)	10.96 (3.65)	0.8067	**25**	11.73 (3.84)	10.71 (3.62)	0.1235
4	11.32 (4.03)	13.15 (3.43)	0.1135	**15**	9.99 (3.93)	10.39 (3.65)	0.3885	**26**	10.29 (3.55)	11.42 (3.22)	0.1023
5	9.96 (3.67)	9.52 (3.21)	0.3862	**16**	10.91 (3.64)	11.13 (3.63)	0.5596	**27**	12.75 (2.99)	13.50 (2.12)	1.0000
6	11.24 (3.53)	10.94 (3.45)	0.4592	**17**	12.00 (4.24)	NA (NA)	NA	**28**	9.75 (3.25)	10.51 (3.29)	0.2661
7	11.61 (3.85)	11.74 (3.51)	0.5472	**18**	15.00 (NA)	13.00 (1.00)	0.3711	**29**	10.44 (3.36)	13.36 (3.50)	0.0656
8	10.72 (3.61)	10.47 (3.49)	0.4274	**19**	10.61 (3.64)	10.80 (3.66)	0.7339	**30**	9.69 (2.25)	10.23 (2.62)	0.3624
9	11.40 (3.25)	10.86 (3.33)	0.0674	**20**	11.17 (3.43)	10.99 (3.32)	0.6838	**31**	10.50 (0.71)	NA (NA)	NA
10	11.67 (3.40)	11.61 (3.55)	0.6953	**21**	10.76 (3.20)	10.26 (3.37)	0.2664	**32**	11.35 (3.16)	11.57 (2.88)	0.7734
11	12.00 (3.34)	11.46 (3.87)	0.1081	**22**	11.25 (3.32)	10.75 (3.45)	0.1600	**33**	9.00 (2.83)	7.50 (3.54)	0.6985

test. In other words, nearly one third (9 out of 33) of the referees tend to make fewer foul calls to the home teams. For the remaining twenty-four referees, although statistical tests indicate insignificance, the average number of foul calls between home and away teams still differ numerically. For example, Referee 7 (*S Bennett*), although the p-value is 0.2470, the average number of foul calls given are 10.93 versus 12.46 to home and away teams, respectively. The insignificance of the test result is led by small number of matches refereed, and comparatively large standard deviation.

The two-sample Wilcoxon rank sum comparison tests of referees with respect to foul calls given to different jersey colors display consistent insignificancy. There is only one referee (Referee 1, *M Clattenburg*) whose test result is significant with a p-value of 0.0466. The actual average numbers of foul calls given by Referee 1 (*M Clattenburg*), to home and away teams are 11.24 and 10.57, respectively. This difference is not practically substantial.

5 Discussion

Color has a psychological influence on human emotions and perceptions of aggression. Soccer is a physically intense, close-contact sport that involves two opposing teams. Teams wear contrast colors of jerseys to be distinguished from each other. This paper comprehensively studies color factor on referees' judgments in soccer games.

We use number of foul calls as a measure to test whether referees show any bias on teams wearing light or dark colored jerseys. Results based on our study support existing research that teams wearing light colored jersey receive fewer foul calls in general. However, our study also shows that home teams tend to wear light colored jersey more often than away teams. This mixture of information adds complexity to the analysis. We therefore conducted a thorough exploration under different conditions.

We further studied foul calls given to different jersey colors at the overall level, and discovered that teams wearing dark colors do not necessarily receive more foul calls than teams wearing light colored jersey. To confirm the result, we also partitioned matches into two pools, which one group contains only home teams, and the other has exclusively away teams. Within both groups, there is no strong statistical evidence to show the substantial difference in foul calls between different jersey colors.

In addition to studying foul calls data from matches under various conditions, we also look into the referees attribution. Analysis on referee level support results based on matches. Nearly a third of the referees made fewer foul calls to the home teams than to away teams. Conditioned on home or away team, referees give consistently similar number of foul calls to teams wearing either light or dark colored jerseys.

The environment of a soccer match is dynamic and complex. Judgments made by a referee are affected by multiple concurrently ongoing factors. Our study presents a statistical difference on the number of foul calls made by referees for home and away teams.

More explicitly speaking, home teams seem to receive one fewer foul calls than away teams. However, statistical significance does not completely equate to practical significance. Here, we leave the question "Does this discrepancy mean anything practically?" to our readers. Thanks to rapid technology development, Video Assistant Referee (VAR) is now commonly adapted in many sports including soccer. The impacts of VAR system to reduce subtle and subjective psychological human factors are prospective topics.

References

[1] 2025. URL https://www.ducksters.com/sports/soccer/rules_fouls.php.

[2] The Football Association, 2025. URL https://www.thefa.com/football-rules-governance/lawsandrules/laws/football-11-11/law-12---fouls-and-misconduct.

[3] David F Caldwell and Jerry M Burger. On thin ice: Does uniform color really affect aggression in professional hockey? *Social Psychological and Personality Science*, 2(3):306–310, 2011.

[4] Inc Datopian, 2024. URL https://datahub.io/.

[5] Andrew J Elliot. Color and psychological functioning: a review of theoretical and empirical work. *Frontiers in psychology*, 6: 127893, 2015.

[6] Andrew J Elliot and Markus A Maier. Color psychology: Effects of perceiving color on psychological functioning in humans. *Annual review of psychology*, 65(1):95–120, 2014.

[7] FIFAEquipmentRegulations, 2025. URL https://digitalhub.fifa.com/m/7474d3addab97747/original/FIFA-Equipment-Regulations_2021_EN.pdf.

[8] IFAB, 2025. URL https://www.theifab.com/laws/latest/the-players-equipment/#colours.

[9] Inc imagecolorpicker, 2024. URL https://imagecolorpicker.com.

[10] John W Pratt. Remarks on zeros and ties in the wilcoxon signed rank procedures. *Journal of the American Statistical Association*, 54(287):655–667, 1959.

[11] Alexander G Schauss. The physiological effect of color on the suppression of human aggression: Research on baker-miller pink. *International Journal of Biosocial Research*, 7(2):55–64, 1985.

[12] R Core Team. *R: A Language and Environment for Statistical Computing, package "stats"*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL http://www.R-project.org/. ISBN 3-900051-07-0.

[13] Hedayatollah Vakili, Mohammad Hadi Niakan, and Najmeh Najafi. The effect of classroom red walls on the students' aggression. *International Journal of School Health*, 6(1):1–4, 2019.

[14] Gregory D Webster, Geoffrey R Urland, and Joshua Correll. Can uniform color color aggression? quasi-experimental evidence from professional ice hockey. *Social Psychological and Personality Science*, 3(3):274–281, 2012.

[15] TW Whitfield and Travis J Wiltshire. Color psychology: a critical review. *Genetic, social, and general psychology monographs*, 116(4):385–411, 1990.

[16] Diana Wiedemann, D Michael Burt, Russell A Hill, and Robert A Barton. Red clothing increases perceived dominance, aggression and anger. *Biology letters*, 11(5):20150166, 2015.

[17] Frank Wilcoxon. Probability tables for individual comparisons by ranking methods. *Biometrics*, 3(3):119–122, 1947.

[18] Frank Wilcoxon. Individual comparisons by ranking methods. In *Breakthroughs in statistics: Methodology and distribution*, pages 196–202. Springer, 1992.