Journal of Economic Behavior and Organization 236 (2025) 107112

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

The effect of social pressure when judging favorites and underdogs

Carlos Varela-Quintana [a,*], Luis Carlos Sánchez [b,c], Julio del Corral [d]

[a] Department of Applied Economics, University of Oviedo, School of Economics and Business, Avda. del Cristo, s/n, Oviedo 33006, Spain
[b] Department of Economics, University of Oviedo, Av. del Cristo s/n, 33006 Oviedo, Spain
[c] Faculty of Business and Economics, Universidad Internacional de la Rioja (UNIR), Av. de la Paz 137, 26006 Logroño, Spain
[d] Department of Economic Analysis and Finance, University of Castilla-La Mancha, Spain

ARTICLE INFO

JEL classification:
D91
L83
M50
Z2

Keywords:
Crowd support
Home advantage
Referee bias
Prestige effect
Social pressure

ABSTRACT

Numerous studies have shown that social pressure and the reputation of others exert a significant influence on individuals' decision-making processes. Analyzing how these two factors interact simultaneously in real-world situations poses, however, a formidable challenge. This study aims to shed light on this issue by exploiting a natural experiment in European football involving matches played in empty stadiums over more than two decades. The research examines whether the reduction in refereeing bias in stadiums without a crowd differs depending on whether the home team is the favorite or the underdog, and how this behavior influences match outcomes. Utilizing comprehensive data from 62,923 matches across eight major European leagues (England, France, Germany, Italy, the Netherlands, Portugal, Russia, and Spain) between 1998/99 and 2021/22, a causal model reveals two key results. First, it finds that, under social pressure, referees tend to favor stronger home teams more than weaker home teams. Second, the study observes that, although weaker local teams receive less preferential treatment from referees, social pressure is more important to achieve better results for them than for stronger home teams. This latter result is likely due to the fact that the stronger local teams would have won anyway without the referees' assistance. The evidence presented strongly supports the idea that the effect of social pressure is moderated by the reputation of the agent being evaluated.

1. Introduction

Evaluating the behavior of others is a critical task in key areas like education, justice, or management. In these contexts, systematic errors in evaluators' thinking pose an enormous challenge to the credibility of the reward system, potentially reducing the incentive for effort (Prendergast & Topel, 1993). Systematic errors can arise when heuristics—shortcuts typically applied quickly, automatically, and unconsciously—are used to reduce complex tasks (Kahneman, 2011; Tversky & Kahneman, 1982). For example, others' reputation can systematically bias our judgments through phenomena such as the performance cue effect (Baltes & Parker, 2000; Murphy & Jones, 1993; Staw, 1975) or the halo/horn effect (Burton et al. 2015; Landy & Sigall, 1974; Thorndike, 1920). Another notable source of systematic error is the predisposition to be influenced by social pressure (Akerlof, 1980; Becker & Murphy, 2009; Bénabou & Tirole, 2006; Bernheim, 1994; Freeman, 1997).

A substantial body of literature has examined the role of social pressure and reputation in prosocial behavior. Most studies indicate

* Corresponding author.
E-mail addresses: varelacarlos@uniovi.es (C. Varela-Quintana), luiscarlos@uniovi.es, luiscarlos.sanchezmartinez@unir.net (L.C. Sánchez), julio.corral@uclm.es (J. del Corral).

https://doi.org/10.1016/j.jebo.2025.107112
Received 8 July 2024; Received in revised form 3 June 2025; Accepted 13 June 2025

that donations tend to be higher in the presence of others, as individuals seek social recognition or feel pressured to act altruistically (e.g., Ariely et al., 2009; DellaVigna et al., 2012; Reinstein & Riener, 2012). This paper takes a different approach. Rather than focusing on the reputational motivations of those making the decision (e.g., donors, voters, or managers), we examine the reputation of those being evaluated (e.g., politicians, workers, or students). Unfortunately, although it is reasonable to hypothesize that the social pressure experienced by the evaluator is moderated by the appraisee's reputation, empirical tests in real-world settings analyzing the joint effect of both factors on evaluators are hard to conduct. Moreover, to the best of our knowledge, there is no convincing evidence on the importance of such an interaction for those being evaluated.

For instance, scholars have examined social influence on voting behavior, but findings remain mixed. While some studies support the bandwagon effect—where voters align with leading candidates (e.g., Fredén et al., 2022; Morton et al., 2015; Rothschild & Malhotra, 2014)—others highlight the underdog effect, in which voters gravitate toward less popular candidates (e.g., Ceci & Kain, 1982; Chatterjee & Kamal, 2021). Beyond these inconsistencies, distinguishing between social pressure and rational information diffusion in this context is challenging. In cases of pure information diffusion, individuals adjust their behavior after receiving new information that leads them to reevaluate the optimal decision. In contrast, in sports, the role of social pressure from fans in the stadium is more evident (DellaVigna, 2009). For example, a referee reviewing the same play in private, free from crowd influence, might reach a different decision.[1] This suggests that social pressure, rather than new information, plays a key role in decision-making, rendering sports an ideal setting to study this phenomenon.

Characterized by experienced participants, high stakes, and fixed, transparent rules, sports provide a valuable laboratory for analyzing human behavior in real competitive environments. These settings often allow researchers to exploit unique sources of exogenous variation, where the mechanisms governing exposure resemble random assignment, thereby creating favorable conditions for causal inference (Bar-Eli et al., 2020; Palacios-Huerta, 2023). Here, we use association football as a particularly well-suited laboratory to empirically test how social pressure on the appraiser is moderated by the appraisee's reputation. Our hypothesis is that the social pressure exerted by the crowd on the referee is more effective when the home team has higher credibility.[2] To address this question, we analyze matches played in stadiums without spectators (closed doors) and with spectators (open doors), while using the Elo rating system commonly employed in chess to determine the relative strength of teams. This approach helps to mitigate potential endogeneity problems present in previous studies, as the absence of spectators is exogenous to any factor affecting team strength (Scoppa, 2021), and the measurement of team strength is unaffected by match outcomes. Moreover, the large number of countries and seasons reduces the risk of selection bias, while the large sample size minimizes the possibility of committing a type II error.[3]

This line of inquiry is particularly relevant for designing fairer appraisal systems, reducing biases, and promoting more equitable decision-making across various domains. Our findings contribute to the literature on decision-making behavior, particularly in the areas of reputational influence (Boone & van Ours, 2006; Williams & Jones, 2005) and social influence (Bassanini et al., 2017, 2021; Tan et al., 2016). They also expand research on the bandwagon and underdog effects (Combrink & Lew, 2020; Dahlgaard et al., 2016), as well as identity-driven bias, including discrimination (Caselli et al., 2023, 2024; Chowdhury et al., 2024). Finally, our study contributes to sports economics by examining the occurrence of biased contests (Chowdhury, 2023) and the dynamics of home advantage (Reade et al., 2022), helping to foster more balanced competitive environments (Forrest et al., 2005).

The remainder of the paper is organized as follows. Section 2 provides an overview of the institutional framework of association football, examines the role of social pressure in this context, and discusses the events leading to the natural experiment. Section 3 outlines the data and empirical strategy employed. Sections 4 and 5 present and discuss the estimation results, while Section 6 concludes.

2. The game of football association and the social pressure

Association football, commonly referred to as soccer, has become an ideal setting for exploring the subtle dynamics of social pressure on the field (Palacios-Huerta, 2014). Originated in England in the 19th century, soccer became an organized sport following the establishment of standardized rules, known as the Laws of the Game, by the Football Association in 1863. The number of teams participating in top-level leagues varies across countries but generally ranges from 16 to 20. The league champion is determined by the total number of points accumulated at the end of the season, which are awarded based on match results: three points for a win, one point for a draw, and none for a loss.

Referees play a central role in maintaining order and fairness during matches by enforcing rules related to fouls, misconduct, and game time. For more serious infractions, referees issue cards as a disciplinary measure. A yellow card serves as an official warning, while a second yellow card or a straight red card results in the player or coach being ejected from the game. A major technological advancement in officiating was the introduction of the Video Assistant Referee (VAR) system, which made its FIFA World Cup debut at the 2018 tournament. The VAR system aims to improve the accuracy and consistency of refereeing decisions by providing a review of key incidents such as goals, penalty decisions and potential red card infringements.

[1] The classic experiment by Nevill et al. (2002) showed that referees officiating in a noisy environment were more uncertain in their decisions and awarded significantly fewer fouls against the home team compared to those officiating in silence.

[2] The underlying mechanism could involve a combination of the halo effect, where referees may have a positive perception of top teams or players that influences their decisions, and confirmatory bias, which makes them more susceptible to social pressure that aligns with their beliefs.

[3] A type II error, also known as a false negative, occurs when the hypothesis test fails to reject a null hypothesis that is actually false in the population.

Leagues typically follow a round-robin format, with each team playing all other teams twice during the season - once at its home stadium and once at the opponent's stadium. Since the influential article by Schwartz and Barsky (1977) pointed out that home teams won more than 50% of their games, home advantage has been the subject of intense research in Behavioral Economics. Among the four explanatory factors proposed by Courneya and Carron (1992)—crowd, learning, travel, and rules—social pressure has emerged as the primary suspect to explain the crowd factor, especially after studies by Downward and Jones (2007) and Garicano et al. (2005) showed that referees favor home teams when stadium attendance is higher. Using a large sample of football competitions, Page and Page (2010) found that some referees were more susceptible to crowd influence than others, possibly due to demographic variables, differences in expertise, or different abilities to cope with social pressure. This aspect was further explored by Sors et al. (2019), who found that referees with higher levels of anxiety were more easily influenced by crowd noise.

Unfortunately, these estimates may be questioned because higher-quality teams also tend to attract larger crowds (Peeters & van Ours, 2021), making it difficult to distinguish the effect of a stronger team from the effect of larger crowds (Ponzo & Scoppa, 2018). The COVID-19 crisis, however, provided a natural experiment that helped address this issue. On March 11, 2020, the WHO declared the outbreak a pandemic, prompting European countries to impose strict lockdowns and travel restrictions to curb the spread of the virus. While Ligue 1 ended the 2019/20 season in April, other leagues, such as La Liga and the Premier League, resumed play in June but without spectators. From autumn 2020, some countries gradually allowed fans back into stadiums, albeit with limited capacity, social distancing, and mandatory mask-wearing. Travel restrictions between regions also affected teams and supporters. As vaccination campaigns progressed and health conditions improved, restrictions were gradually lifted during the 2021/22 season, allowing stadiums to return to full capacity by late 2021 (BBC, 2021).

Numerous articles have provided evidence that the absence of spectators in stadiums led to fairer refereeing decisions and a reduction in home advantage (e.g., Bilalić et al., 2021; Destefanis et al. 2022; Pettersson-Lidbom & Priks, 2010; Scoppa, 2021).[4] For instance, Kocsoy (2025) reported that referees added less stoppage time in behind-closed-doors matches when the home team was trailing by one goal, compared to matches played with spectators. For their part, Cohen et al. (2024) provided evidence that referees' bias is greatest when their decisions cannot be unambiguously identified as erroneous and when social pressure is perceived as understandable or reasonable. Taken together, these findings offer field evidence that social pressure influences decision making beyond controlled laboratory experiments; however, the different impact that social pressure may exert on favorite and underdog home teams is still not entirely clear (Reade et al., 2022). The aim of this paper is to fill this gap.

Here, we build on previous studies showing that perceptions can be biased by brands (e.g., Allison & Uhl, 1964; Audrin et al., 2017; Billett et al., 2014; Koenigs & Tranel, 2008; Schuldt & Schwarz, 2010). Although the field of sports provides evidence for this type of bias for athletes (Barrett, 2021; Findlay & Ste-Marie, 2004) and for teams (Audrino, 2020; Dawson & Dobson, 2010; Erikstad & Johansen, 2020; Lago-Peñas & Gómez-López, 2016), there is little research on how it interacts with social pressure in the stadium. Results to date are mixed and inconclusive. While Scoppa (2008) found no evidence of more favoritism towards home teams suspected of connections with referees after the 2006 Serie A scandal, Buraimo et al. (2010) found evidence that favorites were less likely to receive yellow cards when playing at home than away in the Premier League.

A closely related work is that of Békés et al. (2024), who observed a significant bias in the stoppage time added by referees in close matches in favor of top teams when they lost at home to a minnow. Our paper differs in three ways. First, they did not test an interaction with lack of stadium attendance, which prevents them from properly attributing their finding to social pressure. A second difference with their work is that we do not use the stoppage time added by referees in close matches (i.e., matches with a single goal difference) as a response variable. Our reason for not using it is that the composition of close matches played behind closed doors is likely to be different from those played behind open doors, which could lead to biased estimates.[5] Finally, to measure team quality, the authors used variables that are potentially endogenous (the team's position at the end of the season, a dummy variable identifying the top 20 revenue-generating teams over the last 10 years, the team's annual market value as estimated by Transfermarkt, and the home minus away odds difference). We solve this problem by using pre-match measures of team quality.

In addition to analyzing referee behavior, a second strand of literature has examined whether crowd support has a different impact on match outcomes depending on whether the home team is weak or strong. Here, too, the evidence is limited. While Correia-Oliveira and Andrade-Souza (2022) found a significant positive relationship between home field advantage and team quality when analyzing English, German, Italian and Spanish football leagues, Allen and Jones (2014) and Ramchandani et al. (2021) found that home advantage was higher for teams in lower positions in English professional football. Moreover, the methodology used to analyze this issue is controversial. First, in these studies, strength could be correlated with stadium attendance, so it is not possible to make a proper inference. Second, they used the popular *home advantage index* proposed by Pollard (1986), which, as the same author points out, is inadequate for making comparisons between teams, since the best teams are likely to win the match regardless of whether they play at home or away (Pollard, 2006). Both problems are avoided in the present work.

As notable exceptions, some studies provide evidence on the different role of crowd support for favorites and underdogs by examining kick-off times and days of the week. Goller and Krumer (2020) and Krumer (2020) document a smaller home advantage for underdogs compared to favorites when playing at later start times or on infrequent days, suggesting that underdogs may rely more on crowd support than favorites. Given that, as the authors themselves suggest, other variables—such as psychological factors—may be

[4] For a systematic review of the literature, see Leitner et al. (2023).

[5] In open-door games, an even match might involve a lower-quality team playing at home against a higher-quality team handicapped by playing away. Conversely, in closed-door games, an even match would involve teams of similar quality because the compensating effect of crowd support is absent.

involved in these situations, the use of closed-door matches could provide a more conclusive analysis.

3. Materials and methods

3.1. Data and summary statistics

This section presents the data and summary statistics used in the analysis. The dataset contains a total of 62,923 matches from the eight major European football leagues played between 1998/99 and 2021/22, spanning 24 seasons. These competitions are the Premier League (England), Ligue 1 (France), Bundesliga (Germany), Serie A (Italy), Eredivisie (the Netherlands), Primeira Liga (Portugal), Premier Liga (Russia), and La Liga (Spain). The selection of these leagues is based on the average of the Union of European Football Associations (UEFA) country coefficients,[6] while the choice of the 1998/99 season as the starting point is explained by UEFA's structural change, which required all its competitions to be played in all-seater stadiums from the summer of 1998.

Table 1 presents descriptive statistics for the sample, covering refereeing, performance, and contextual factors. Since red cards constitute a small portion of the data, we aggregate them with yellow cards into a single variable (total cards), based on the standard equivalence that one red card equals two yellow cards. As can be seen, home teams won, drew, and lost 46.2%, 25.9%, and 27.9% of the time, respectively; they scored 0.400 more goals per game than their opponents, and received 0.551 more points per game in their favor. These results are usually attributed to social pressure on referees, since, as can also be observed, away teams received 0.422 more cards per match than home teams.

The variable strength captures the quality of the home team relative to the visiting team as perceived by the referee. It is estimated using the standard logistic formula commonly employed in the literature, $P_{win} = 1/\left(10^{-\Delta r/400} + 1\right)$, where Δr is the difference in Elo ratings between the two teams, obtained from the website http://clubelo.com.[7] This probability implicitly treats draws as half wins and half losses. The variable is designed to exclude the effect of home-field advantage, ensuring that, on average, home and visiting teams have the same probability of winning. Additionally, a t-test with unequal variances confirms that there are no significant differences in the home team's win probability between closed-door and open-door matches (diff. = 0.001, p-value = 0.815), ruling out any potential bias in this variable. To ensure comparability, values have been normalized by season and country to have a mean of zero and a standard deviation of one. A value of zero indicates that both teams have an equal chance of winning the match, while negative values correspond to a relatively weaker home team and positive values to a relatively stronger one.

A total of 2,930 matches (4.7% of the sample) were played behind closed doors.[8] Most of the matches played behind closed doors were the result of social distancing measures implemented in the 2019/20–2021/22 seasons to prevent the spread of the SARS-CoV-2 coronavirus, while only 72 matches were played before that. Match results, referee decisions (yellow and red cards) and attendance were collected from https://www.worldfootball.net. As previously stated, the extensive inclusion of countries and years diminishes the risk of selection bias, while the substantial sample size mitigates the potential for committing a type II error.

Figure 1 illustrates the evolution of football matches played behind closed doors from the 1998/99 to 2021/22 seasons in eight countries. Before the pandemic, closed-door matches were relatively rare, typically linked to disciplinary or security reasons. Italy (28), France (20) and Russia (17) had the highest numbers, with notable increases observed before the pandemic. In Italy, the 2006/07 season saw 14 closed-door matches due to fan violence, while France peaked at four in 2016/17, largely due to security concerns. Russia experienced a rise in 2014/15, with six matches attributed to fan misconduct and safety issues.

During the pandemic, Spain and Italy recorded the highest numbers, with 495 and 465 closed-door matches, respectively, while Russia had the lowest with only 21. The peak of the pandemic occurred from March to May 2020, leading to strict lockdowns and the suspension of most football competitions. Once leagues resumed in mid-2020, matches were predominantly played behind closed doors. This trend continued until the first half of 2021, particularly in Spain, Italy, and England, where high infection rates and restrictions persisted. As vaccination efforts advanced by August 2021, many countries—notably England, Italy, and Spain— gradually allowed spectators to return to stadiums, leading to a decline in closed-door matches. However, sporadic closed-door games continued into the 2021/22 season, especially in the Netherlands, Germany and France, due to local outbreaks and regulatory measures.

Table 2 presents the home teams' winning probabilities (derived from Elo ratings), the quality difference between home and away teams (calculated as the difference in the logarithms of the market value of their starting lineups), and the average annual attendance at the home team's stadium stadiums in matches played with spectators. These values, normalized by league and season, are shown for the full sample and for the four quartiles into which the home teams are divided based on the strength variable: *Clear underdogs, Slight underdogs, Slight favorites*, and *Clear favorites*. As expected, there is a moderate linear correlation between the normalized win probability and the normalized average annual attendance (0.49), and a very strong linear correlation between the Strength variable and the normalized player market value difference (0.91).

[6] The UEFA country coefficients can be found at https://kassiesa.net/uefa/ (accessed in June 2022).

[7] The Elo rating system, adopted by the World Chess Federation (FIDE) in 1970 and informally used in sports such as American football, baseball, basketball, or association football, is a method of ranking the strength of players or teams. In this system, winning teams take points from losing teams based on the difference in skill. An advantage of this method is that it allows the probability of winning to be estimated by a simple logistic function using the Elo rating difference between two teams.

[8] Eighteen decided matches and 175 unplayed matches were not considered in the dataset. All unplayed matches belong to the Dutch and French leagues, which were suspended in March 2020 after the COVID-19 outbreak.

Table 1
Descriptive statistics.

	Mean	Std. dev.	Min.	Max.
Home wins	0.462	0.499	0	1
Home draws	0.259	0.438	0	1
Home losses	0.279	0.448	0	1
Point difference (Home-Away)	0.551	2.523	-3	3
Goal difference (Home-Away)	0.400	1.767	-13	10
First yellow card difference (Away-Home)	0.364	1.678	-7	8
Second yellow / red card difference (Away-Home)	0.044	0.483	-3	4
Total Card difference (Away-Home)	0.422	1.860	-8	10
Strength	0.000	0.998	-2.914	2.988
Closed doors	0.047	0.211	0	1
Days of the week				
Monday	0.037	0.188	0	1
Tuesday	0.022	0.147	0	1
Wednesday	0.061	0.239	0	1
Thursday	0.011	0.106	0	1
Friday	0.054	0.225	0	1
Saturday	0.421	0.494	0	1
Sunday	0.394	0.489	0	1

Notes: Number of observations: 62,923. Total cards=YC+2RC, where YC are yellow cards and RC are red cards. The strength variable is the probability of winning at home (calculated using Elo ratings) normalized by season and country.

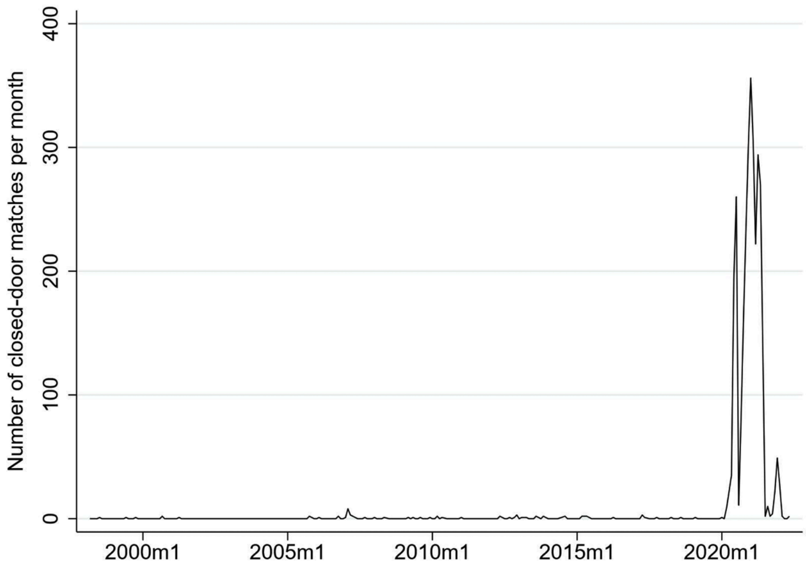

Fig. 1. Evolution of closed-door matches (1998/99-2021/22).

Table 2
Attendance, market value and winning probability (mean values).

	Strength	Win probability	Market value diff. (normalized)	Avg. Annual Attend. (normalized)
Clear underdog	-1.290 (15,771)	0.266 (15,771)	-1.153 (11,745)	-0.407 (15,552)
Slight underdog	-0.341 (15,697)	0.438 (15,697)	-0.275 (11,700)	-0.301 (15,466)
Slight favorite	0.343 (15,769)	0.562 (15,769)	0.275 (11,758)	-0.075 (15,514)
Clear favorite	1.294 (15,686)	0.735 (15,686)	1.159 (11,685)	0.798 (15,224)

Notes: Number of observations in parentheses. Win probability, market value differences and average annual attendance were normalized by league and season. Calculations are based on data obtained from https://www.worldfootball.net and http://clubelo.com.

3.2. Hypothesis testing and econometric model specification

To examine the interaction between crowd support and team quality, we compare disciplinary sanctions and match outcomes for both open-door and closed-door games, controlling for the home team's relative strength. Hypothesis testing is conducted to assess the statistical significance of mean differences across the entire sample and within each subgroup. A *p*-value of less than 0.05 is considered statistically significant, and all tests are two-tailed.

While this initial analysis provides useful insights, it does not establish causality, as other unobserved factors may influence the results. To draw causal inferences, we employ a regression model that controls for observed confounding variables and incorporates individual fixed effects to account for unobserved confounders that remain constant over time (Angrist & Pischke, 2009). Thus, our model is defined as:

$$y_{it} = \beta_0 + \beta_1 \text{ Closed }_{it} + \beta_2 \text{ Strength }_{it} + \beta_3 (\text{ Strength } \times \text{ Closed })_{it} + \varsigma_{\mathrm{d}} + \nu_{\mathrm{k}} + \delta_{\mathrm{w}} + \rho_{\mathrm{r}} + \lambda_{\mathrm{t}} + \alpha_{\mathrm{i}} + \varepsilon_{it}$$

where the subscript *i* denotes matchups played by the same pair of home and away teams (e.g., Liverpool vs. Everton, Liverpool vs. Manchester City, Everton vs. Liverpool, and so on), while the subscript *t* denotes seasons. The dependent variable, y_{it}, captures the outcome of the match (point difference, goal difference, and match win) and the referee's decisions (cards issued to the home team, cards issued to the away team, and difference in cards).

$Closed_{it}$ is a binary variable indicating whether or not the match was played behind closed doors (i.e., without spectators). It includes both matches played without spectators to mitigate the spread of the SARS-CoV-2 coronavirus pandemic (2019/20–2021/22 seasons) and matches played behind closed doors due to other factors, mainly disciplinary sanctions.

$Strength_{it}$ measures the pre-game probability that the home team will win based on Elo rating. Although the Elo rating system provides better predictive performance than the official FIFA ranking (Lasek et al., 2013), it is not as accurate as methods based on market odds (Hvattum & Arntzen, 2010). However, betting odds may introduce an endogeneity problem. Some studies have found evidence that, at least in the long run, betting markets have adjusted to reflect the reduced home advantage in games played behind closed doors (see Fischer & Haucap, 2021; Hegarty, 2021; Hegarty & Whelan, 2023). Annual team market values—provided by www.transfermarkt.com after the end of the season—also suffer from an endogeneity problem, as club performance is one of the factors determining player market values. An alternative, which we use in the robustness check, is to use the market values of the starters of each match.

The interaction between *Strength* and *Closed*—$(Strength \times Closed)_{it}$—is the variable of interest in the model. A significant value of the coefficient β_3 would provide evidence that the effect of playing behind closed doors differs across the relative strength of the home teams. A positive sign would indicate that the strongest teams exhibit higher values of the dependent variable when playing behind closed doors compared to the weakest teams.

Our model accounts for potential confounding factors. For instance, attendance was prohibited in the second half of the 2019/20 season, a period when games were likely perceived as more decisive (Di Mattia & Krumer, 2023). At the same time, to ensure the tournament concluded on time, a large proportion of games were scheduled on weekdays and at unusual kick-off times, potentially affecting home advantage (see Goller & Krumer, 2020; Krumer, 2020; Krumer & Lechner, 2018). To address these factors, we include ς_{d}, ν_{k}, δ_{w}, and ρ_{r}, which represent vectors of coefficients for day of the week, kick-off time[9], match week, and referee dummies, respectively.

Finally, λ_{t} captures season-specific effects (e.g., the presence of exceptionally talented players or the use of VAR), while α_{i} accounts for matchup-specific effects. The latter helps control for unobserved heterogeneity related to factors such as travel distance or historical rivalries.

4. Results

4.1. *The effect of closed doors on the refereeing of underdogs and favorites*

This section examines whether referees reacted differently to lack of attendance depending on whether they officiated for strong or weak home teams. We used the variables "Total cards", "Total card difference" (calculated as cards of the away team minus cards of the home team), "Total cards received by home teams" and "Total cards received by away teams". The values of all these variables are expressed per match.

Table 3 shows the values for the entire sample, as well as the four quartiles into which the dataset is divided based on the relative strength of the home team compared to the away team, using our *Strength* variable (see Panel (a) of Fig. A1 in the Appendix for a visual reference). These quartiles are labeled as *Clear favorite, Slight favorite, Slight underdog,* and *Clear underdog*. The analysis of the full sample shows a significant decrease in "total cards" from 4.484 with open doors to 4.121 with closed doors. This decrease was exclusively observed for away teams (-0.400), while home teams experienced a non-significant increase (0.037). These opposing effects shifted the card difference between away and home teams from 0.442 to 0.005, providing evidence that away teams receive fairer refereeing behind closed doors due to less social pressure on officials.

The effect is not the same in all quartiles. As can be seen, as the relative strength of the home team increases, so does the card difference in open-door games (from -0.004 for clear underdogs to 0.814 for clear favorites). However, it is difficult to determine whether these differences are unfair. They could be due to a higher level of foul play by lower-quality teams, greater social pressure in large stadiums, or a bias by referees in favor of quality teams.

An analysis of the effect of closed doors may shed some light on this issue. As shown, the reduction in the card difference was

[9] Since the most common kick-off time may vary across countries and years, the values of the discrete variable that generates the dummy variables were centered by league and season.

Table 3
Total cards according to the relative strength of the home teams.

	All matches (N=62,923)	Clear underdog (N=15,771)	Slight underdog (N=15,697)	Slight favorite (N=15,769)	Clear favorite (N=15,686)
Total cards					
Closed (N=2,930)	4.121	4.021	4.392	4.214	3.853
Open (N=59,993)	4.484	4.538	4.687	4.586	4.125
Difference	-0.363***	-0.518***	-0.295***	-0.372***	-0.272***
p-value	0.000	0.000	0.002	0.000	0.003
Home cards					
Closed (N=2,930)	2.059	2.199	2.154	2.095	1.780
Open (N=59,993)	2.021	2.271	2.158	1.999	1.656
Difference	0.037	-0.073	-0.004	0.096*	0.124***
p-value	0.187	0.213	0.942	0.083	0.019
Away cards					
Closed (N=2,930)	2.063	1.822	2.238	2.120	2.074
Open (N=59,993)	2.463	2.267	2.529	2.587	2.469
Difference	-0.400***	-0.446***	-0.291***	-0.468***	-0.396***
p-value	0.000	0.000	0.000	0.000	0.000
A-H card difference					
Closed (N=2,930)	0.005	-0.377	0.084	0.025	0.294
Open (N=59,993)	0.442	-0.004	0.370	0.589	0.814
Difference	-0.438***	-0.373***	-0.287***	-0.564***	-0.519***
p-value	0.000	0.000	0.000	0.000	0.000

Notes: ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively, for two-sample t-tests assuming equal or unequal variances as appropriate. Period of analysis: seasons 1998/99 to 2021/22. H means home, A means away. Data sources: www.worldfootball.net and http://clubelo.com.

greater for the favorites (-0.564 and -0.519) than for the underdogs (-0.373 and -0.287), which shows that part of the card difference in favor of the better teams was due to the nepotism of the referees caused by the support of the crowd. Another piece of evidence can be found by analyzing the cards received by the home teams; notably, the increase in Total cards was concentrated in the clear favorites (0.124), while the non-favorites did not experience a significant change.

Although the t-test provided suggestive results, we need to control for potential confounders using a fixed-effects specification. Table 4 presents the results in two stages: first, the Pooled models without control variables, and second, the fixed-effects (FE) models that include all control variables. Models 1, 2, 5, 6, 9, and 10 report results without the interaction variable between *Closed* and *Strength*. As can be observed, the relative strength of home teams is negatively related to the number of cards received by home teams and positively related to the number of cards received by away teams. Consequently, the strength of the home team is positively related to the card difference in its favor. These results are in line with expectations, as the weaker away teams have to commit more fouls to compensate for their lack of talent.

The negative coefficient of the variable *Closed* shows that the absence of spectators at the stadium reduced the total number of cards received by both home and away teams, but the effect was much greater for away teams. This hurt home teams by reducing the card difference per game by 0.385. This result provides additional evidence to the recent literature (e.g., Bilalić et al., 2021; Scoppa, 2021) on how the absence of spectators contributes to reducing referee bias.

Models 3, 4, 7, 8, 11, and 12 test our hypothesis by including the interaction (*Strength* × *Closed*). The negative and significant value of the coefficient β_3 in Model 4 shows that the decrease in the card difference behind closed doors was not uniform but was larger when the home team was stronger. Models 8 and 12 show that the reduction in the card differential caused by closed doors was due to an increase in the number of cards received by stronger home teams, rather than a decrease in the total number of cards received by visiting teams playing stronger home teams.[10]

It is important to note that the effect of social pressure may differ depending on the importance of the decision. Therefore, in Table 5, we re-estimate our main specification (A-H card difference) separately for the first yellow cards (i.e., cautions) and for the second yellow cards or red cards (i.e., expulsions). The results hold only for yellow cards, supporting the idea that agents are able to insulate themselves from social pressure when they are aware that their decision has a large impact on the game, but are very permeable for the smaller decisions. Table A1 in the Appendix displays the estimates for each country based on Model (16).

4.2. *The effect of closed doors on the performance of underdogs and favorites*

In this section, we analyzed whether the different referee responses to team quality identified in the previous subsection had an impact on match results. To do this, we analyzed point difference, wins, and goal difference.

Table 6 presents the match results with and without attendance (see also panels (b) and (c) in Fig. A1 in the Appendix for a visual reference). Analyzing all matches, home teams performed significantly worse in empty stadiums, regardless of the metric used,

[10] Given the count nature of the cards, we also used a fixed-effects Poisson estimator for models 6, 8, 10, and 12, obtaining very similar results.

Table 4
The effect of closed doors on refereeing.

	Away - Home card difference				Home cards received				Away cards received			
	Pooled	FE	Pooled	FE	Pooled	FE	Pooled	FE	Pooled	FE	Pooled	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Strength	0.314***	0.281***	0.318***	0.286***	-0.234***	-0.186***	-0.238***	-0.190***	0.080***	0.095***	0.080***	0.096***
	(0.007)	(0.015)	(0.007)	(0.015)	(0.006)	(0.012)	(0.006)	(0.012)	(0.006)	(0.013)	(0.006)	(0.013)
Closed	-0.434***	-0.385***	-0.435***	-0.387***	0.035	-0.144***	0.036	-0.142**	-0.399***	-0.529***	-0.400***	-0.529***
	(0.035)	(0.069)	(0.035)	(0.069)	(0.028)	(0.056)	(0.028)	(0.056)	(0.030)	(0.055)	(0.030)	(0.055)
Strength × Closed			-0.080**	-0.104***			0.073***	0.095***			-0.007	-0.010
			(0.035)	(0.036)			(0.028)	(0.028)			(0.030)	(0.028)
Constant	0.442***	0.440***	0.442***	0.440***	2.021***	2.029***	2.021***	2.029***	2.463***	2.469***	2.463***	2.469***
	(0.007)	(0.003)	(0.007)	(0.003)	(0.006)	(0.003)	(0.006)	(0.003)	(0.006)	(0.003)	(0.006)	(0.003)
Matchup FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Season FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Referee FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Day of week FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Kick-off time FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Match week FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	62,923	60,662	62,923	60,662	62,923	60,662	62,923	60,662	62,923	60,662	62,923	60,662
R^2	0.031	0.196	0.031	0.196	0.025	0.283	0.025	0.283	0.005	0.258	0.005	0.258

Notes: Robust standard errors, clustered at the matchup level, are reported in parentheses in the FE models (2,261 singleton observations were dropped). The variable *Strength* is derived from pre-match Elo ratings.
*** $p<0.01$ ** $p<0.05$ * $p<0.1$.

Table 5
The effect of closed doors on refereeing for minor and major decisions.

	A-H caution card difference (A-H first yellow card diff.)				A-H expulsion card difference (A-H 2^{nd} yellow / red card diff.)			
	Pooled (13)	FE (14)	Pooled (15)	FE (16)	Pooled (17)	FE (18)	Pooled (19)	FE (20)
Strength	0.280***	0.249***	0.284***	0.254***	0.025***	0.025***	0.025***	0.026***
	(0.007)	(0.014)	(0.007)	(0.014)	(0.002)	(0.004)	(0.002)	(0.004)
Closed	-0.403***	-0.345***	-0.404***	-0.347***	-0.029***	-0.030*	-0.029***	-0.030*
	(0.031)	(0.061)	(0.031)	(0.061)	(0.009)	(0.018)	(0.009)	(0.018)
Strength × Closed			-0.091***	-0.107***			-0.001	-0.007
			(0.031)	(0.033)			(0.009)	(0.009)
Constant	0.383***	0.381***	0.383***	0.381***	0.045***	0.045***	0.045***	0.045***
	(0.007)	(0.003)	(0.007)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
Matchup FE	No	Yes	No	Yes	No	Yes	No	Yes
Season FE	No	Yes	No	Yes	No	Yes	No	Yes
Referee FE	No	Yes	No	Yes	No	Yes	No	Yes
Day of week FE	No	Yes	No	Yes	No	Yes	No	Yes
Kick-off time FE	No	Yes	No	Yes	No	Yes	No	Yes
Match week FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	62,923	60,662	62,923	60,662	62,923	60,662	62,923	60,662
R^2	0.030	0.195	0.030	0.195	0.003	0.148	0.003	0.148

Notes: Robust standard errors, clustered at the matchup level, are reported in parentheses in the FE models (2,261 singleton observations were dropped). The variable *Strength* is derived from pre-match Elo ratings.
*** $p<0.01$ ** $p<0.05$ * $p<0.1$.

providing compelling evidence for the link between home advantage and social pressure. However, the effect seems to be unevenly distributed when we take into account the relative strength of the home team. While the performance of the clear favorites remained largely unchanged, the rest of the home teams experienced a significant decline.

This result contrasts with what might be expected when looking at changes in refereeing decisions. As seen earlier, the presence of fairer refereeing in unattended matches negatively affected all home teams, with favorites being the most affected. However, the loss of crowd support in the stadium seems to have had an irrelevant effect on the results of the clear favorites.

Table 7 presents the estimates from the Pooled and fixed-effects (FE) models. Similar to the previous section, we present two specifications for each dependent variable: one excluding the interaction variable and the other including it. The dependent variable "Victory" is a binary variable that takes the value one if the home team wins and zero otherwise.

As expected, the stronger the home team is relative to its opponent, the higher its point difference, probability of winning and goal difference. The closed doors dummy variable is negative and significant in all models, providing evidence that the presence of crowd support is one of the causes of home field advantage. Our variable of interest, the interaction between closed doors and team strength, is positive and significant in FE models 24, 28 and 32, indicating that the higher the relative strength of the home team, the lower the negative impact of playing without crowd support. All these results provide evidence that crowd support is a relevant input for lower-quality teams, but negligible for the highest-quality teams.[11] Table A2 in the Appendix presents the country-specific estimates for Model (24).

4.3. Robustness checks

In the Appendix, we provide several robustness checks to ensure that our results are not driven by the assumptions made in our estimates. Specifically, we examine the robustness of our findings in five key aspects: (i) alternative ways of measuring the *Strength* variable (Models A17-A24), (ii) replacing matchup and season fixed effects with team-season and referee-season fixed effects (Models A25-A48), (iii) assessing whether the effect of playing without spectators was also significant during the pre-COVID and COVID-reopening periods (Models A49-A56), (iv) analyzing the linearity of the effect (Models A57-A60), and (v) restricting the sample to the VAR era (Models A61-A68). Finally, we conduct two permutation tests: a resampling test and a placebo test (Table A9).

Regarding the specific variable chosen to measure the relative strength of the home teams, we first address the fact that the Elo methodology implicitly treats draws as half wins and half losses. To refine this, we explicitly model the possibility of a draw using a multinomial logit model, where the dependent variable includes three possible outcomes (win, draw, loss) while the independent variables include the Elo ratings. This approach allows us to compute the expected number of points, which we use as an alternative measure of *Strength*, normalized by country and season.[12]

Second, we introduce an additional metric for *Strength* based on player market values estimated prior to the match, ensuring their exogeneity. The *Strength* variable is defined as the log difference between the market value of the home team's starters and that of the

[11] We checked our results using different estimators: for models 23 and 24, we used a random-effects ordered probit with match outcome (win, draw, and loss) as the dependent variable; and for models 26 and 28, we employed a fixed-effects logit model. All these approaches yielded results similar to those reported in Table 7.

[12] We thank the anonymous reviewers for this suggestion.

Table 6
Match results according to the relative strength of the home teams.

	All matches (*N*=62,923)	Clear underdog (*N*=15,771)	Slight underdog (*N*=15,697)	Slight favorite (*N*=15,769)	Clear favorite (*N*=15,686)
Victories (%)					
Closed (*N*=2,930)	0.400	0.164	0.348	0.413	0.682
Open (*N*=59,993)	0.466	0.259	0.411	0.511	0.683
Difference	-0.066***	-0.095***	-0.064***	-0.098***	-0.001
p-value	0.000	0.000	0.001	0.000	0.961
H-A point difference					
Closed (*N*=2,930)	0.160	-1.289	-0.021	0.379	1.611
Open (*N*=59,993)	0.570	-0.646	0.337	0.890	1.704
Difference	-0.411***	-0.642***	-0.357***	-0.512***	-0.093
p-value	0.000	0.000	0.000	0.000	0.265
H-A goal difference					
Closed (*N*=2,930)	0.148	-0.960	-0.011	0.299	1.296
Open (*N*=59,993)	0.412	-0.495	0.201	0.600	1.345
Difference	-0.264***	-0.464***	-0.211***	-0.301***	-0.050
p-value	0.000	0.000	0.001	0.000	0.448

Notes: ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively, for two-sample t-tests assuming equal or unequal variances as appropriate. Period of analysis: seasons 1998/99 to 2021/22. *H* means home, *A* means away. Data sources: www.worldfootball.net and http://clubelo.com.

away team's starters, normalized by country and season. Market value data are available only from the 2004/05 season onward. Table A3 presents the estimates for models 14, 16, 22, and 24 using the new *Strength* variables. As shown, the sign and significance of our variable of interest (*Strength* × *Closed*) remain unchanged.

A second concern is the possibility that the absence of spectators can influence contests in ways beyond social pressure on referees. For instance, racial harassment decreases when there is no crowd, leading to improved performance among Black players (Caselli et al., 2023; 2024). As a result, our findings may be driven by reduced social pressure on minority players rather than on referees. For example, if a home team signs several Black players in the same season that matches are played without spectators, the model might mistakenly attribute a reduction in home advantage to decreased pressure on referees, when it is actually due to reduced racial harassment toward players. To address this potential confounding factor, we estimate team-season fixed effects, which allow us to control for changes in team composition from one year to another for both home and away teams.[13] As shown in models A25 to A32 in Table A4, the signs of the coefficients remain unchanged, but their significance decreases, particularly for the interaction term in Model A26. However, when we re-estimate the model using the pre-match market values of the starters, the interaction term remains significant at the 1% level (see Model A18 in Table A5 in the Appendix).

Additionally, referees may exhibit individual trends over time that could also confound our results. For instance, referees might become less susceptible to social pressure as they gain experience, independent of the absence of fans in the stadium. To account for these time-varying referee-specific factors, we include referee-season fixed effects (see models A33 to A36 in Table A4). The effect of closed matches remains significant and is not absorbed by these fixed effects, suggesting that the impact of playing without spectators on refereeing behavior was consistent across referees and not driven by a small subset of individuals.

Another potential issue is that most of the closed-door matches occurred during the pandemic, which may have introduced confounding factors. For instance, movement restrictions, mask mandates, and social distancing measures could have unevenly influenced teams' training routines. To address these concerns, we conduct two robustness checks. In the first, we examine whether the effect of playing without spectators was also significant during the pre-COVID era, as outlined in models A49 to A52 in Table A6. As shown, the coefficients of the variables of interest during the pre-COVID period exhibit the same signs as those observed during the COVID era. However, none of these coefficients is statistically significant, likely due to the limited number of closed-door matches prior to the pandemic (72). Despite this, the test for the equality of coefficients does not reject the null hypothesis, providing support for our results.

To assess whether performance changes were influenced by differences in adaptation to training routines, we adopted the approach proposed by Chowdhury et al. (2024), introducing a second treatment group comprising matches played after the pandemic under open-door conditions (see models A53 to A56). Although the Covid-reopen effect is not significant and the null hypothesis of coefficient equality is rejected, the interaction term remains significant. This suggests that other factors may have continued to influence home advantage unevenly after the closed-door period. Possible explanations include regional disparities in the timeline of lifting restrictions, internal movement limitations affecting fan attendance, and the prolonged use of masks and social distancing measures, which may have continued to diminish social pressure within stadium. These findings reinforce the idea that home advantage is shaped by multiple contextual factors, some of which persisted beyond the closed-door period. While our results highlight the role of social pressure, further research is needed to disentangle its interaction with other post-pandemic effects.

In the following robustness check, we examine whether the interaction between team strength and the closed-door effect follows a

[13] Home team × Season FE and away team × Season FE are included in the model separately to prevent them from absorbing the variable of interest *Strength*.

Table 7
Effect of closed doors on home team performance.

	Home - Away point difference				Home win				Home - Away goal difference			
	Pooled	FE	Pooled	FE	Pooled	FE	Pooled	FE	Pooled	FE	Pooled	FE
	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)
Strength	0.923***	0.528***	0.916***	0.521***	0.166***	0.096***	0.165***	0.094***	0.728***	0.420***	0.722***	0.416***
	(0.009)	(0.020)	(0.010)	(0.020)	(0.002)	(0.004)	(0.002)	(0.004)	(0.006)	(0.014)	(0.007)	(0.014)
Closed	-0.402***	-0.338***	-0.400***	-0.335***	-0.064***	-0.065***	-0.064***	-0.065***	-0.257***	-0.233***	-0.256***	-0.231***
	(0.044)	(0.088)	(0.044)	(0.088)	(0.009)	(0.017)	(0.009)	(0.017)	(0.030)	(0.061)	(0.030)	(0.061)
Strength × Closed			0.163***	0.147***			0.025***	0.024***			0.137***	0.092***
			(0.044)	(0.044)			(0.009)	(0.008)			(0.031)	(0.036)
Constant	0.570***	0.568***	0.570***	0.568***	0.465***	0.466***	0.465***	0.466***	0.412***	0.411***	0.412***	0.411***
	(0.010)	(0.004)	(0.010)	(0.004)	(0.002)	(0.001)	(0.002)	(0.001)	(0.007)	(0.003)	(0.007)	(0.003)
Matchup FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Season FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Referee FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Day of week FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Kick-off time FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Match week FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	62,923	60,662	62,923	60,662	62,923	60,662	62,923	60,662	62,923	60,662	62,923	60,662
R^2	0.135	0.272	0.135	0.272	0.111	0.251	0.112	0.251	0.170	0.303	0.171	0.303

Notes: Robust standard errors, clustered at the matchup level, are reported in parentheses in the FE models (2,261 singleton observations were dropped). The variable *Strength* is derived from pre-match Elo ratings.
*** $p<0.01$ ** $p<0.05$ * $p<0.1$.

non-linear relationship. To do this, instead of modeling team strength as a continuous variable, we categorize it into four quartile dummies and interact them with the closed-door variable. The estimates in models A57 to A60 in Table A7 suggest that the relationship is not entirely linear, particularly in the case of performance, where the impact appears to be more pronounced in closer games. To facilitate interpretation, figures showing the conditional marginal effects are provided in the Appendix (see Figure A2).

Another potential confounding factor is the long-term decline in home advantage over the past 45 years (Peeters & van Ours, 2021), as well as the introduction of the VAR system in the later years of our sample (Gasparetto & Loktionov, 2023). To address this, we conduct a robustness check by limiting the dataset to recent seasons in which VAR was implemented: England (from the 2019/20 season), Russia (from February 28, 2020), Spain, France, and the Netherlands (from the 2018/19 season), and Germany, Italy, and Portugal (from the 2017/18 season). As shown in Models A61–A64 in Table A7, the sign of the Closed variable and its interaction with Strength remain unchanged. The most notable difference is the reduction in the magnitude and significance of the interaction term in Model A64, likely because matchup fixed effects absorb much of the variation in Strength. In Table A8 of the Appendix, we re-estimate the model using home team fixed effects and the pre-match market values of the starters, finding that the interaction term remains significant (see Model A68).

Finally, we performed two resampling-based inference procedures. The first is a subsampling test, in which we randomly select 40% of the pre-pandemic matches from the sample and repeat this process 5,000 times. As shown in Table A9, we do not reject the null hypothesis that the mean coefficients from the resampled datasets are equal to the coefficient estimated from the full sample in Models 16 and 24. This finding, together with the rejection of the null hypothesis that the coefficients are zero, further reinforces the robustness of our results.

The second test is a randomized placebo test, in which we randomly assign the treatment (closed-door matches) to 2,930 matches—the same number as in the actual treated group—repeating this process 5,000 times. The resulting empirical *p*-values indicate that we cannot reject the null hypothesis that the mean effect of this false treatment is zero, whereas we do reject the null hypothesis that it equals the actual estimated coefficient. This further supports the validity of our findings, suggesting that the observed effects are not the result of random chance or spurious correlations.

5. Discussion

Social pressure and reputation are important determinants of individual behavior in many contexts. For example, the worker's negative reputation has been shown to increase the likelihood of dismissal (Boone & van Ours, 2006), and studies using mock juries show that the perceived credibility of lawyers and witnesses can significantly shape jurors' attitudes (Williams & Jones, 2005). On the other hand, decisions to fire workers (Bassanini et al., 2017, 2021) or to adopt decisions about corporate social responsibility (Tan et al., 2016) can be influenced by social pressure from local communities. Similarly, verdicts can be swayed by the majority opinion within the jury (Lively, 2017) or pre-trial exposure to community outrage (Zimmerman et al., 2016).

Using a natural experiment in sports, our research aims to analyze how these two factors interact. Specifically, we seek to determine whether social pressure experienced by evaluators can be amplified by the reputations of the evaluated individuals and whether this amplification has a tangible impact on outcomes. Consistent with previous research, our paper provides evidence that the absence of crowd support leads to fairer refereeing (Bilalić et al., 2021; Scoppa, 2021) and reduces home advantage (Destefanis et al., 2022; Scoppa, 2021), supporting the hypothesis that social pressure plays an important role for real-world competitive environments.

Our study extends these findings by showing that referee bias is not uniform but moderated by team quality. Specifically, we find that higher-quality local teams receive fewer total cards in open-door matches, which suggests that referees are more susceptible to social influence (conformity to the majority) when the home team enjoys a higher reputation. This effect is observed for the first yellow card but not for the second yellow or red card, where decisions tend to be less ambiguous. This pattern suggests that our findings may be driven by implicit discrimination rather than conscious taste-based or statistical discrimination. In other words, referee bias appears to operate at a subconscious, automatic level rather than through deliberate favoritism, which is consistent with the findings of Faltings et al. (2023) and Gallo et al. (2013). These results have important implications for fields such as education, human resource management, and the judicial system, where social pressure and unconscious associations between attributes and social group membership often interact.

As a second contribution, this paper shows that the crowd effect on match outcomes is also moderated by team quality. However, the effect is the opposite of that observed for referee bias. Our estimates provide evidence that although home team favoritism is lower for lower-quality teams, it is precisely these teams that benefit the most, which aligns with the results of Goller and Krumer (2020) and Krumer (2020). This pattern likely occurs because the quality of the players of the clear-favorite teams is so high that the referee's help is irrelevant, as they would have won the match anyway. This finding, which is undoubtedly relevant for achieving more egalitarian sports competitions, may also be relevant in broader organizational contexts. For example, social pressure from coworkers, customers or the media may have a countervailing effect in discrimination cases where an employee is disadvantaged by racial, social, or gender presumptions.

Our study has important implications for designing fairer evaluation systems, reducing bias, and promoting more equitable work environments. Previous research highlights the need to consider social pressure in the assignment, training, and monitoring of referees (Buraimo et al., 2010), as well as other professionals who face various forms of social pressure, such as pilots and police officers (Page & Page, 2010). Our work emphasizes the importance of considering the relationship between social pressure and reputation in the design of such programs. It also prompts reflection on the extent to which social pressure can positively influence evaluations that may be biased by the prior reputation of the person being evaluated.

This work is not without limitations. While the Covid-19 pandemic provided a unique natural experiment to examine the effects of disruptions in social pressure, it also presents challenges—particularly in isolating the absence of social pressure from other pandemic-related factors. Therefore, results should be interpreted with due caution, and further research is needed. The pandemic offers valuable insights for future studies. Differences in the timing of restriction lifts across regions, movement limitations affecting fan attendance, and ongoing public health measures such as mask mandates and social distancing provide opportunities to explore how these factors interacted and the specific mechanisms through which social pressure operates in these settings.

6. Conclusions

There is long-standing interest in studying how individuals' decision-making deviates under social pressure. Our goal is to shed light on this question by investigating the combined effect of social pressure and reputation in a real-world setting. To this end, we analyze the impact of matches played without spectators on refereeing decisions in the eight major European leagues and examine how these effects may vary based on the relative strength of the home teams.

Consistent with existing research, we find that the absence of spectators in stadiums leads to fairer refereeing and a reduction in home advantage. Our work is the first to provide compelling evidence that these effects are moderated by team quality. Specifically, we find that playing behind closed doors reduces referee bias significantly more for high-quality teams, likely because social pressure is more effective when reputation is higher. In addition, we find that playing behind closed doors does not negatively affect the outcomes of the clear-favorite home teams, possibly because their player quality was so high that they would have won anyway. These findings have implications for the selection, evaluation, and training of individuals engaged in judging and subject to various forms of social pressure.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used DeepL and ChatGPT in order to improve readability and language of the work. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no conflicts of interest and have not received any funding for this project.

Acknowledgments

The authors would like to thank the editorial team and the anonymous reviewers for their valuable suggestions and insightful comments, which greatly enhanced the quality of our manuscript. All errors and omissions remain the sole responsibility of the authors.

Appendix

Table A1,A2,A3,A4,A5,A6,A7,A8,A9

Table A1
The effect of closed doors on refereeing. Dependent variable: A-H caution card difference.

	England (A1)	France (A2)	Germany (A3)	Italy (A4)	Netherlands (A5)	Portugal (A6)	Russia (A7)	Spain (A8)
Strength	0.266***	0.222***	0.226***	0.337***	0.252***	0.312***	0.216***	0.256***
	(0.035)	(0.030)	(0.037)	(0.041)	(0.037)	(0.064)	(0.047)	(0.039)
Closed	-0.195	-0.329*	-0.373**	-0.389**	-0.092	-0.724***	-0.003	-0.549**
	(0.162)	(0.190)	(0.161)	(0.172)	(0.167)	(0.255)	(0.301)	(0.218)
Strength × Closed	0.055	-0.137	-0.137	-0.178**	-0.119	-0.028	-0.086	-0.214**
	(0.078)	(0.101)	(0.086)	(0.086)	(0.084)	(0.094)	(0.272)	(0.090)
Constant	0.361***	0.398***	0.382***	0.331***	0.414***	0.500***	0.338***	0.350***
	(0.008)	(0.008)	(0.008)	(0.010)	(0.008)	(0.015)	(0.002)	(0.012)
Matchup FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Season FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Referee FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Kick-off time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Match week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,779	8,414	7,117	8,289	7,190	6,548	5,523	8,771
R^2	0.215	0.185	0.176	0.225	0.182	0.223	0.221	0.195

Notes: The variable *Strength* is derived from pre-match Elo ratings. Robust standard errors clustered at the matchup level are reported in parentheses. Singleton observations were dropped.
*** $p<0.01$ ** $p<0.05$ * $p<0.1$.

Table A2
Effect of closed doors on team performance. Dependent variable: H-A point difference.

	England (A9)	France (A10)	Germany (A11)	Italy (A12)	Netherlands (A13)	Portugal (A14)	Russia (A15)	Spain (A16)
Strength	0.679***	0.398***	0.475***	0.570***	0.697***	0.391***	0.523***	0.457***
	(0.054)	(0.046)	(0.056)	(0.056)	(0.061)	(0.075)	(0.059)	(0.054)
Closed	-0.114	-0.596*	-0.172	-0.060	-0.667***	0.104	0.346	-0.364
	(0.284)	(0.312)	(0.268)	(0.257)	(0.252)	(0.324)	(0.480)	(0.273)
Strength × Closed	-0.004	0.263*	0.186	0.248**	0.163	-0.006	0.165	0.168*
	(0.127)	(0.136)	(0.124)	(0.115)	(0.109)	(0.111)	(0.316)	(0.101)
Constant	0.513***	0.633***	0.533***	0.557***	0.578***	0.517***	0.523***	0.626***
	(0.014)	(0.012)	(0.014)	(0.015)	(0.012)	(0.019)	(0.003)	(0.015)
Matchup FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Season FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Referee FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Kick-off time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Match week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,779	8,414	7,117	8,289	7,190	6,548	5,523	8,771
R^2	0.302	0.229	0.242	0.307	0.288	0.349	0.289	0.247

Notes: The variable *Strength* is derived from pre-match Elo ratings. Robust standard errors clustered at the matchup level are reported in parentheses. Singleton observations were dropped.
*** $p<0.01$ ** $p<0.05$ * $p<0.1$.

Table A3
Robustness check: The effect of closed doors on refereeing and performance using alternative measures of strength.

	Strength variable: Expected points				Strength variable: Pre-match market values			
	A-H caution card difference		H-A point difference		A-H caution card difference		H-A point difference	
	(A17)	(A18)	(A19)	(A20)	(A21)	(A22)	(A23)	(A24)
Strength	0.250***	0.255***	0.530***	0.523***	0.206***	0.213***	0.634***	0.628***
	(0.014)	(0.014)	(0.020)	(0.020)	(0.020)	(0.020)	(0.029)	(0.029)
Closed	-0.345***	-0.347***	-0.337***	-0.334***	-0.350***	-0.353***	-0.362***	-0.358***
	(0.061)	(0.061)	(0.088)	(0.088)	(0.062)	(0.062)	(0.090)	(0.090)
Strength × Closed		-0.110***		0.145***		-0.132***		0.123***
		(0.033)		(0.044)		(0.035)		(0.045)
Constant	0.381***	0.381***	0.567***	0.567***	0.353***	0.353***	0.526***	0.526***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)
Matchup FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Season FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Referee FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Kick-off time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Match week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	60,662	60,662	60,662	60,662	44,736	44,736	44,736	44,736
R^2	0.195	0.195	0.272	0.272	0.213	0.213	0.295	0.295

Notes: The variable *Strength* of models A17 to A20 is derived from the expected points using Elo ratings, while in models A21 to A24 is derived from the pre-match market values of the starters. Robust standard errors, clustered at the matchup level, are reported in parentheses. Singleton observations were dropped.
*** $p<0.01$ ** $p<0.05$ * $p<0.1$.

Table A4
Robustness check: The effect of closed doors on refereeing and team performance, controlling for team composition and referee behavior changes over time.

	Home team x Season FE				Away team x Season FE				Referee x Season FE			
	A-H caution card diff.		H-A point diff.		A-H caution card diff.		H-A point diff.		A-H caution card diff.		H-A point diff.	
	(A25)	(A26)	(A27)	(A28)	(A29)	(A30)	(A31)	(A32)	(A33)	(A34)	(A35)	(A36)
Strength	0.279***	0.282***	0.810***	0.805***	0.284***	0.289***	0.756***	0.750***	0.276***	0.281***	0.826***	0.820***
	(0.010)	(0.010)	(0.015)	(0.015)	(0.009)	(0.010)	(0.013)	(0.014)	(0.010)	(0.010)	(0.014)	(0.015)
Closed	-0.327***	-0.328***	-0.228**	-0.225**	-0.344***	-0.347***	-0.232**	-0.227**	-0.319***	-0.315***	-0.392***	-0.397***
	(0.057)	(0.057)	(0.094)	(0.094)	(0.060)	(0.060)	(0.096)	(0.096)	(0.040)	(0.039)	(0.057)	(0.058)
Strength × Closed		-0.068		0.124**		-0.098**		0.144***		-0.121***		0.164***
		(0.043)		(0.055)		(0.042)		(0.050)		(0.038)		(0.051)
Constant	0.379***	0.379***	0.562***	0.562***	0.380***	0.380***	0.562***	0.562***	0.376***	0.376***	0.577***	0.577***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.002)	(0.002)	(0.003)	(0.003)
Matchup FE	No	No	No	No	No	No	No	No	No	No	No	No
Season FE	No	No	No	No	No	No	No	No	No	No	No	No
Home team x Season FE	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
Away team x Season FE	No	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No
Referee x Season FE	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Home team FE	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Referee FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Kick-off time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Match week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	62,855	62,855	62,855	62,855	62,855	62,855	62,855	62,855	59,349	59,349	59,349	59,349
R^2	0.128	0.129	0.200	0.200	0.125	0.125	0.203	0.203	0.225	0.225	0.290	0.290

Notes: Models A25 to A28 include home-team × season fixed effects, with robust standard errors clustered at the home-team × season level (dropped 68 singleton observations). Models A29 to A32 include away-team × season fixed effects, with robust standard errors clustered at the away-team × season level (dropped 68 singleton observations). Models A33 to A36 include referee × season fixed effects, with robust standard errors clustered at the referee × season level (dropped 3,574 singleton observations). The variable *Strength* is derived from pre-match Elo ratings.
** $p<0.05$ * $p<0.1$.

Table A5
Robustness check: The effect of closed doors on refereeing and team performance, controlling for team composition and referee behavior changes over time (Alternative strength measure).

	Home team x Season FE				Away team x Season FE				Referee x Season FE			
	A-H caution card diff.		H-A point diff.		A-H caution card diff.		H-A point diff.		A-H caution card diff.		H-A point diff.	
	(A37)	(A38)	(A39)	(A40)	(A41)	(A42)	(A43)	(A44)	(A45)	(A46)	(A47)	(A48)
Strength	0.265***	0.272***	0.956***	0.949***	0.270***	0.277***	0.905***	0.898***	0.261***	0.270***	0.924***	0.914***
	(0.012)	(0.012)	(0.017)	(0.017)	(0.011)	(0.012)	(0.015)	(0.016)	(0.012)	(0.012)	(0.017)	(0.018)
Closed	-0.322***	-0.325***	-0.233**	-0.231**	-0.340***	-0.344***	-0.220**	-0.216**	-0.319***	-0.314***	-0.384***	-0.389***
	(0.058)	(0.058)	(0.096)	(0.096)	(0.061)	(0.061)	(0.098)	(0.098)	(0.040)	(0.040)	(0.058)	(0.058)
Strength × Closed		-0.112***		0.117**		-0.112***		0.117**		-0.154***		0.172***
		(0.042)		(0.056)		(0.042)		(0.052)		(0.038)		(0.051)
Constant	0.352***	0.352***	0.516***	0.516***	0.353***	0.353***	0.515***	0.515***	0.351***	0.350***	0.532***	0.532***
	(0.004)	(0.004)	(0.006)	(0.006)	(0.004)	(0.004)	(0.006)	(0.006)	(0.002)	(0.002)	(0.004)	(0.004)
Matchup FE	No	No	No	No	No	No	No	No	No	No	No	No
Season FE	No	No	No	No	No	No	No	No	No	No	No	No
Home team x Season FE	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
Away team x Season FE	No	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No
Referee x Season FE	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Home team FE	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Referee FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Kick-off time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Match week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	46,847	46,847	46,847	46,847	46,847	46,847	46,847	46,847	44,075	44,075	44,075	44,075
R^2	0.129	0.129	0.222	0.222	0.124	0.124	0.224	0.224	0.225	0.225	0.305	0.305

Notes: Models A37 to A40 include home-team × season fixed effects, with robust standard errors clustered at the home-team × season level (dropped 41 singleton observations). Models A41 to A44 include away-team × season fixed effects, with robust standard errors clustered at the away-team × season level (dropped 41 singleton observations). Models A45 to A48 include referee × season fixed effects, with robust standard errors clustered at the referee × season level (dropped 2,813 singleton observations). The variable *Strength* is based on the pre-match market values of the starters.
*** $p<0.01$ ** $p<0.05$ * $p<0.1$.

Table A6
Robustness check: The effect of pre-Covid closed doors and Covid restrictions during the post-closure period.

	Pre-Covid closed effect				Covid-reopen effect			
	A-H caution card diff.		H-A point diff.		A-H caution card diff.		H-A point diff.	
	(A49)	(A50)	(A51)	(A52)	(A53)	(A54)	(A55)	(A56)
Strength	0.249***	0.254***	0.528***	0.521***	0.249***	0.256***	0.528***	0.516***
	(0.014)	(0.014)	(0.020)	(0.020)	(0.014)	(0.014)	(0.020)	(0.020)
Covid closed (γ_1)	-0.607***	-0.617***	-0.331	-0.301				
	(0.214)	(0.234)	(0.327)	(0.316)				
Pre-Covid closed (γ_2)	-0.323***	-0.324***	-0.338***	-0.337***				
	(0.063)	(0.063)	(0.091)	(0.091)				
Strength $\times$ Covid closed (γ_3)		-0.109***		0.147***				
		(0.034)		(0.045)				
Strength $\times$ Pre-Covid closed (γ_4)		-0.061		0.172				
		(0.264)		(0.304)				
Closed (γ_5)					-0.396***	-0.399***	-0.310***	-0.306***
					(0.078)	(0.078)	(0.114)	(0.114)
Covid-reopen (γ_6)					-0.095	-0.097	0.053	0.054
					(0.095)	(0.095)	(0.142)	(0.142)
Strength$\times$ Closed (γ_7)						-0.111***		0.157***
						(0.034)		(0.045)
Strength $\times$ Covid-reopen (γ_8)						-0.043		0.104**
						(0.034)		(0.045)
Constant	0.380***	0.380***	0.568***	0.568***	0.388***	0.388***	0.564***	0.564***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.008)	(0.008)	(0.011)	(0.011)
p-value (H_0: $\gamma_1 = \gamma_2$)	0.202	0.227	0.982	0.912				
p-value (H_0: $\gamma_3 = \gamma_4$)		0.858		0.934				
p-value (H_0: $\gamma_5 = \gamma_6$)					0.000	0.000	0.001	0.001
p-value (H_0: $\gamma_7 = \gamma_8$)						0.138		0.367
Matchup FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Season FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Referee FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Kick-off time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Match week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	60,662	60,662	60,662	60,662	60,662	60,662	60,662	60,662
R^2	0.195	0.195	0.272	0.272	0.195	0.195	0.272	0.272

Notes: Robust standard errors, clustered at the matchup level, are reported in parentheses in the FE models (2,261 singleton observations were dropped). The variable *Strength* is derived from pre-match Elo ratings.
*** $p<0.01$ ** $p<0.05$ * $p<0.1$.

Table A7
Robustness check: The effect of closed doors on refereeing and performance: linearity and VAR era.

	Linear effect				VAR Era			
	A-H caution card diff.		H-A point diff.		A-H caution card diff.		H-A point diff.	
	(A57)	(A58)	(A59)	(A60)	(A61)	(A62)	(A63)	(A64)
Closed	-0.349***	-0.231***	-0.347***	-0.588***	-0.409***	-0.409***	-0.446***	-0.445***
	(0.061)	(0.083)	(0.088)	(0.119)	(0.079)	(0.079)	(0.115)	(0.115)
Quartile 2	0.218***	0.220***	0.394***	0.380***				
	(0.025)	(0.026)	(0.040)	(0.040)				
Quartile 3	0.371***	0.379***	0.669***	0.661***				
	(0.029)	(0.029)	(0.043)	(0.044)				
Quartile 4	0.508***	0.520***	1.062***	1.038***				
	(0.034)	(0.034)	(0.050)	(0.051)				
Quartile 2 × Closed (β_1)		-0.043		0.307**				
		(0.099)		(0.144)				
Quartile 3 × Closed (β_2)		-0.179*		0.157				
		(0.097)		(0.141)				
Quartile 4 × Closed (β_3)		-0.254***		0.514***				
		(0.097)		(0.129)				
Strength					0.349***	0.374***	0.182**	0.160*
					(0.061)	(0.062)	(0.092)	(0.093)
Strength × Closed						-0.090**		0.080
						(0.042)		(0.055)
Constant	0.107***	0.101***	0.038	0.049	0.305***	0.305***	0.483***	0.483***
	(0.019)	(0.020)	(0.029)	(0.030)	(0.022)	(0.022)	(0.032)	(0.032)
p-value (H_0: $\beta_1 = \beta_2$)		0.170		0.314				
p-value (H_0: $\beta_2 = \beta_3$)		0.446		0.009				
Matchup FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Season FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Referee FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Kick-off time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Match week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	60,662	60,662	60,662	60,662	9,407	9,407	9,407	9,407
R^2	0.193	0.194	0.268	0.269	0.388	0.389	0.464	0.464

N*otes:* Robust standard errors are clustered at the matchup level. Models A57–A60 and A61–A64 exclude 2,261 and 1,343 singleton observations, respectively. The variable *Strength* is derived from pre-match Elo ratings.
*** $p<0.01$ ** $p<0.05$ * $p<0.1$.

Table A8
Robustness check: The effect of closed doors on refereeing and performance: VAR era.

	VAR Era			
	A-H caution card diff.		H-A point diff.	
	(A65)	(A66)	(A67)	(A68)
Strength	0.220***	0.241***	1.003***	0.980***
	(0.025)	(0.027)	(0.036)	(0.039)
Closed	-0.349***	-0.350***	-0.364***	-0.363***
	(0.056)	(0.056)	(0.101)	(0.101)
Strength × Closed		-0.085**		0.090**
		(0.041)		(0.045)
Constant	0.289***	0.289***	0.447***	0.447***
	(0.015)	(0.015)	(0.027)	(0.027)
Matchup FE	No	No	No	No
Home team FE	Yes	Yes	Yes	Yes
Season FE	Yes	Yes	Yes	Yes
Referee FE	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes
Kick-off time FE	Yes	Yes	Yes	Yes
Match week FE	Yes	Yes	Yes	Yes
Observations	10,723	10,723	10,723	10,723
R^2	0.088	0.088	0.210	0.210

Notes: Robust standard errors are clustered at the home team level. The variable *Strength* is based on the pre-match market values of the starters (singleton observations were dropped).
*** $p<0.01$ ** $p<0.05$ * $p<0.1$.

Table A9
Subsampling inference and randomized treatment placebo.

	Mean($\widehat{\beta}^*$)	SE($\widehat{\beta}^*$)	$\widehat{\beta}^*_{0.025}$	$\widehat{\beta}^*_{0.975}$	Emp. p-value ($H_0 : \overline{\widehat{\beta}}^* = \widehat{\beta}$)	Emp. p-value ($H_0 : \overline{\widehat{\beta}}^* = 0$)
Subsampling inference						
Dep. Var.: A-H yellow card diff. (Model 16)						
Closed	-0.349	0.030	-0.409	-0.289	0.966	0.000
Strength × Closed	-0.110	0.014	-0.137	-0.083	0.775	0.000
Dep. Var.: H-A point diff (Model 24)						
Closed	-0.332	0.042	-0.415	-0.249	0.954	0.000
Strength × Closed	0.122	0.019	0.084	0.160	0.168	0.000
Randomized treatment placebo						
Dep. Var.: A-H yellow card diff. (Model 16)						
Closed	0.001	0.034	-0.066	0.066	0.000	0.989
Strength × Closed	-0.001	0.033	-0.065	0.065	0.001	0.977
Dep. Var.: H-A point diff. (Model 24)						
Closed	0.000	0.048	-0.094	0.094	0.000	0.992
Strength × Closed	0.000	0.044	-0.090	0.087	0.001	0.992

Note: Based on 5,000 resamples. The first empirical *p*-value tests whether the mean coefficient from the subsampling process or the randomized treatment placebo equals the estimated coefficient from Models 16 and 24. The second empirical *p*-value tests whether this mean equals zero.

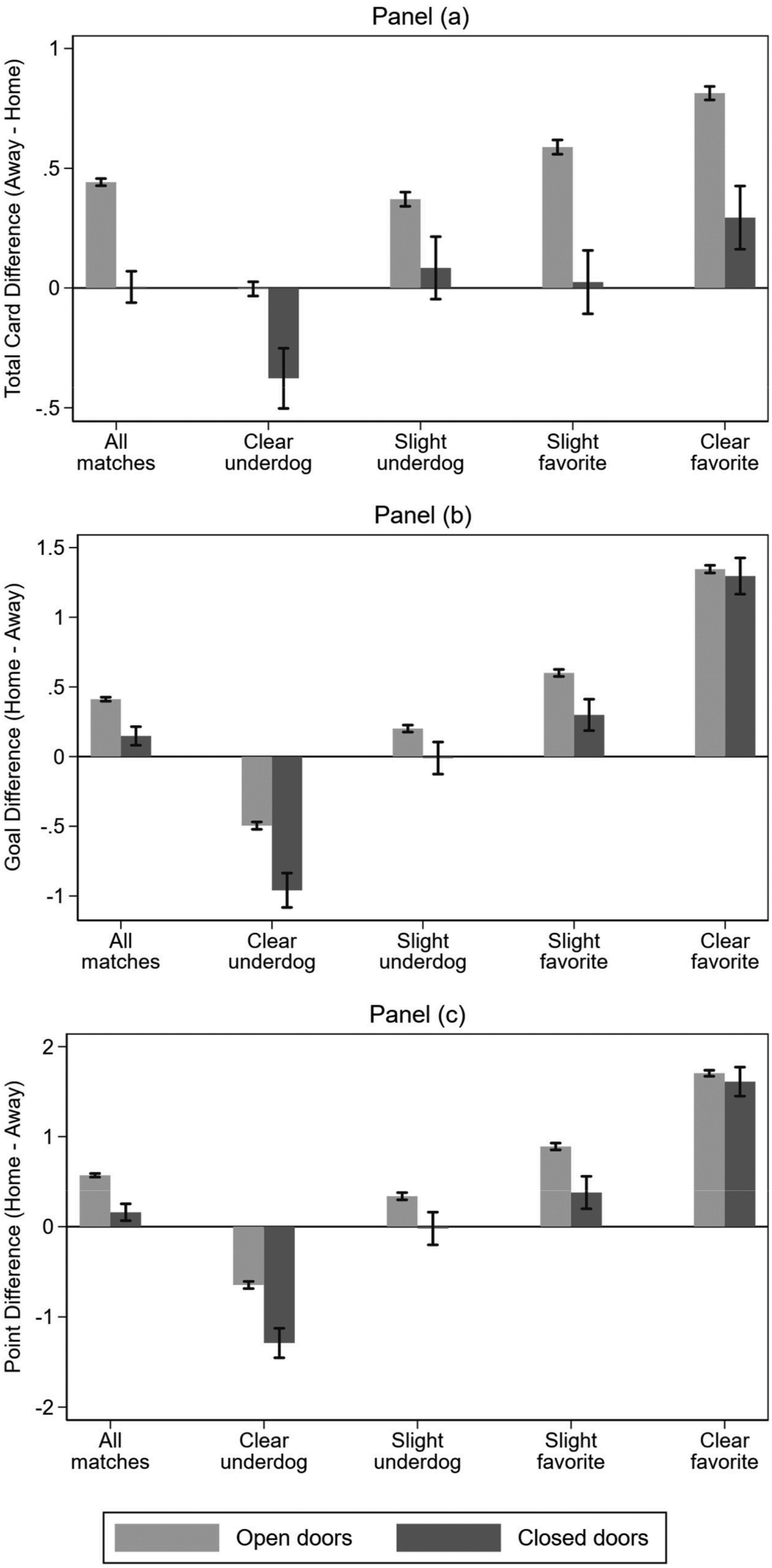

Fig. A1. Average per-game values of the main dependent variables behind closed and open doors.

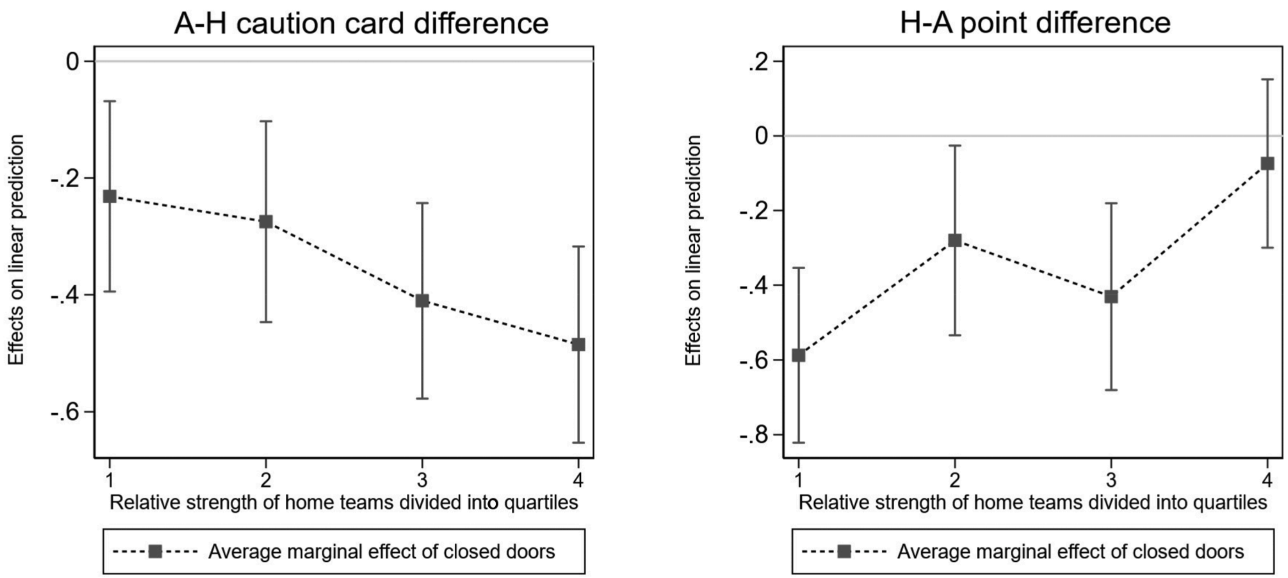

Fig. A2. Conditional marginal effects (95% confidence interval).

Data availability

Data will be made available on request.

References

Akerlof, G.A., 1980. A theory of social custom, of which unemployment may be one consequence. Q. J. Econ. 94 (4), 749–775. https://doi.org/10.2307/1885667.

Allen, M.S., Jones, M.V., 2014. The home advantage over the first 20 seasons of the English Premier League: effects of shirt colour, team ability and time trends. Int. J. Sport Exerc. Psychol. 12 (1), 10–18. https://doi.org/10.1080/1612197X.2012.756230.

Allison, R.I., Uhl, K.P., 1964. Influence of beer brand identification on taste perception. J. Market. Res. 1 (3), 36–39. https://doi.org/10.2307/3150054.

Angrist, J.D., Pischke, J.S., 2009. Mostly Harmless Econometrics: an Empiricist's Companion. Princeton University Press.

Ariely, D., Bracha, A., Meier, S., 2009. Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. Am. Econ. Rev. 99 (1), 544–555. https://doi.org/10.1257/aer.99.1.544.

Audrin, C., Brosch, T., Chanal, J., Sander, D., 2017. When symbolism overtakes quality: materialists consumers disregard product quality when faced with luxury brands. J. Econ. Psychol. 61, 115–123. https://doi.org/10.1016/j.joep.2017.04.001.

Audrino, F., 2020. Strongest team favoritism in European national football: myth or reality? Int. J. Sport Finance 15 (3), 137–151. https://doi.org/10.32731/IJSF/153.082020.04.

Baltes, B.B., Parker, C.P., 2000. Understanding and removing the effects of performance cues on behavioral ratings. J. Bus. Psychol. 15, 229–246. https://doi.org/10.1023/A:1007787010800.

Bar-Eli, M., Krumer, A., Morgulev, E., 2020. Ask not what economics can do for sports - ask what sports can do for economics. J. Behav. Exp. Econ. 89, 101597. https://doi.org/10.1016/j.socec.2020.101597.

Barrett, D., 2021. Star player bias in the NBA: a quantile regression approach. Empiric. Econ. Bull., Undergrad. J. 14 (1), 9. https://digitalcommons.bryant.edu/eeb/vol14/iss1/9.

Bassanini, A., Brunello, G., Caroli, E., 2017. Not in my community: social pressure and the geography of dismissals. J. Labor. Econ. 35 (2), 429–483. https://doi.org/10.1086/689285.

Bassanini, A., Caroli, E., Fontaine, F., Reberioux, A., 2021. Escaping social pressure: fixed-term contracts in multi-establishment firms. J. Econ. Behav. Organ. 188, 697–715. https://doi.org/10.1016/j.jebo.2021.05.027.

BBC, 2021. Capacity crowds return for first Premier League Saturday of 2021-22 season. BBC. https://www.bbc.com/sport/football/58214370.

Becker, G.S., Murphy, K.M., 2009. Social Economics: Market Behavior in a Social Environment. Harvard University Press.

Békés, G., Borza, E., Fleck, M., 2024. Favoritism under multiple sources of social pressure. Econ. Inq. 62 (4), 1748–1769. https://doi.org/10.1111/ecin.13245.

Bénabou, R., Tirole, J., 2006. Incentives and prosocial behavior. Am. Econ. Rev. 96 (5), 1652–1678. https://doi.org/10.1257/aer.96.5.1652.

Bernheim, B.D., 1994. A theory of conformity. J. Polit. Econ. 102 (5), 841–877. https://doi.org/10.1086/261957.

Bilalić, M., Gula, B., Vaci, N., 2021. Home advantage mediated (HAM) by referee bias and team performance during covid. Sci. Rep. 11 (1), 21558. https://doi.org/10.1038/s41598-021-00784-8.

Billett, M.T., Jiang, Z., Rego, L.L., 2014. Glamour brands and glamour stocks. J. Econ. Behav. Organ. 107, 744–759. https://doi.org/10.1016/j.jebo.2014.03.014.

Boone, J., van Ours, J.C., 2006. Are recessions good for workplace safety? J. Health Econ. 25 (6), 1069–1093. https://doi.org/10.1016/j.jhealeco.2006.03.002.

Buraimo, B., Forrest, D., Simmons, R., 2010. The 12th man?: Refereeing bias in English and German soccer. J. R. Stat. Soc.: Ser. A (Stat. Soc.) 173 (2), 431–449. https://doi.org/10.1111/j.1467-985X.2009.00604.x.

Burton, S., Cook, L.A., Howlett, E., Newman, C.L., 2015. Broken halos and shattered horns: overcoming the biasing effects of prior expectations through objective information disclosure. J. Acad. Market. Sci. 43, 240–256. https://doi.org/10.1007/s11747-014-0378-5.

Caselli, M., Falco, P., Mattera, G., 2023. When the stadium goes silent: How crowds affect the performance of discriminated groups. J. Labor. Econ. 41 (2), 431–451. https://doi.org/10.1086/719967.

Caselli, M., Falco, P., Somekh, B., 2024. Inside the NBA bubble: how black players performed better without fans. J. Popul. Econ. 37 (2), 39. https://doi.org/10.1007/s00148-024-01021-y.

Ceci, S.J., Kain, E.L., 1982. Jumping on the bandwagon with the underdog: the impact of attitude polls on polling behavior. Public Opin. Q. 46 (2), 228–242. https://doi.org/10.1086/268715.

Chatterjee, S., Kamal, J., 2021. Voting for the underdog or jumping on the bandwagon? Evidence from India's exit poll ban. Public Choice 188 (3), 431–453. https://doi.org/10.1007/s11127-020-00837-y.

Chowdhury, S.M., Esteve-Gonzalez, P., Mukherjee, A., 2023. Heterogeneity, levelling the playing field, and affirmative action in contests. South. Econ. J. 89 (3), 924–974. https://doi.org/10.1002/soej.12618.

Chowdhury, S.M., Jewell, S., Singleton, C., 2024. Can awareness reduce (and reverse) identity-driven bias in judgement? Evidence from international cricket. J. Econ. Behav. Organ. 226, 106697. https://doi.org/10.1016/j.jebo.2024.106697.

Cohen, A., Neeman, Z., Auferoth, F., 2024. Judging under public pressure. Rev. Econ. Stat. 106 (1), 151–166. https://doi.org/10.1162/rest_a_01141.

Combrink, S., Lew, C., 2020. Potential underdog bias, overconfidence and risk propensity in investor decision-making behavior. J. Behav. Finance 21 (4), 337–351. https://doi.org/10.1080/15427560.2019.1692843.

Correia-Oliveira, C.R., Andrade-Souza, V.A., 2022. Home advantage in soccer after the break due to COVID-19 pandemic: does crowd support matter? Int. J. Sport Exerc. Psychol. 20 (4), 1245–1256. https://doi.org/10.1080/1612197X.2021.1934716.

Courneya, K.S., Carron, A.V., 1992. The home advantage in sport competitions: a literature review. J. Sport Exerc. Psychol. 14 (1), 13–27. https://doi.org/10.1123/jsep.14.1.13.

Dahlgaard, J.O., Hansen, J.H., Hansen, K.M., Larsen, M.V., 2016. How are voters influenced by opinion polls? The effect of polls on voting behavior and party sympathy. World Polit. Sci. 12 (2), 283–300. https://doi.org/10.1515/wps-2016-0012.

Dawson, P., Dobson, S., 2010. The influence of social pressure and nationality on individual decisions: evidence from the behaviour of referees. J. Econ. Psychol. 31 (2), 181–191. https://doi.org/10.1016/j.joep.2009.06.001.

DellaVigna, S., 2009. Psychology and economics: Evidence from the field. J. Econ. Lit. 47 (2), 315–372. https://doi.org/10.1257/jel.47.2.315.

DellaVigna, S., List, J.A., Malmendier, U., 2012. Testing for altruism and social pressure in charitable giving. Q. J. Econ. 127 (1), 1–56. https://doi.org/10.1093/qje/qjr050.

Destefanis, S., Addesa, F., Rossi, G., 2022. The impact of COVID-19 on home advantage: a conditional order-m analysis of football clubs' efficiency in the top-5 European leagues. Appl. Econ. 54 (58), 6639–6655. https://doi.org/10.1080/00036846.2022.2074361.

Di Mattia, A., Krumer, A., 2023. Fewer teams, more games, larger attendance? Evidence from the structural change in basketball's EuroLeague. Eur. J. Oper. Res. 309 (1), 359–370. https://doi.org/10.1016/j.ejor.2023.01.002.

Downward, P., Jones, M., 2007. Effects of crowd size on referee decisions: analysis of the FA Cup. J. Sports Sci. 25 (14), 1541–1545. https://doi.org/10.1080/02640410701275193.

Erikstad, M.K., Johansen, B.T., 2020. Referee bias in professional football: favoritism toward successful teams in potential penalty situations. Front. Sports Act. Living. 2, 19. https://doi.org/10.3389/fspor.2020.00019.

Faltings, R., Krumer, A., Lechner, M., 2023. Rot-Jaune-Verde: on linguistic bias of referees in Swiss soccer. Kyklos. 76 (3), 380–406. https://doi.org/10.1111/kykl.12332.

Findlay, L.C., Ste-Marie, D.M., 2004. A reputation bias in figure skating judging. J. Sport Exerc. Psychol. 26 (1), 154–166. https://doi.org/10.1123/jsep.26.1.154.

Fischer, K., Haucap, J., 2021. Does crowd support drive the home advantage in professional football? Evidence from German ghost games during the COVID-19 pandemic. J. Sports Econom. 22 (8), 982–1008. https://doi.org/10.1177/15270025211026552.

Forrest, D., Beaumont, J., Goddard, J., Simmons, R., 2005. Home advantage and the debate about competitive balance in professional sports leagues. J. Sports Sci. 23 (4), 439–445. https://doi.org/10.1080/02640410400021641.

Fredén, A., Rheault, L., Indridason, I.H., 2022. Betting on the underdog: the influence of social networks on vote choice. Polit. Sci. Res. Methods 10 (1), 198–205. https://doi.org/10.1017/psrm.2020.21.

Freeman, R.B., 1997. Working for nothing: the supply of volunteer labor. J. Labor. Econ. 15 (1, Part 2), S140–S166. https://doi.org/10.1086/209859.

Gallo, E., Grund, T., Reade, J., 2013. Punishing the foreigner: implicit discrimination in the Premier League based on oppositional identity. Oxf. Bull. Econ. Stat. 75 (1), 136–156. https://doi.org/10.1111/j.1468-0084.2012.00725.x.

Garicano, L., Palacios-Huerta, I., Prendergast, C., 2005. Favoritism under social pressure. Rev. Econ. Stat. 87 (2), 208–216. https://doi.org/10.1162/0034653053970267.

Gasparetto, T., Loktionov, K., 2023. Does the Video Assistant Referee (VAR) mitigate referee bias on professional football? PLoS. One 18 (11). https://doi.org/10.1371/journal.pone.0294507.

Goller, D., Krumer, A., 2020. Let's meet as usual: do games played on non-frequent days differ? Evidence from top European soccer leagues. Eur. J. Oper. Res. 286 (2), 740–754. https://doi.org/10.1016/j.ejor.2020.03.062.

Hegarty, T., 2021. Information and price efficiency in the absence of home crowd advantage. Appl. Econ. Lett. 28 (21), 1902–1907. https://doi.org/10.1080/13504851.2021.1883525.

Hegarty, T., Whelan, K., 2023. The wisdom of no crowds: the reaction of betting markets to lockdown soccer games. J. Predict. Markets 17 (2), 123–138. https://doi.org/10.5750/jpm.v17i2.2089.

Hvattum, L.M., Arntzen, H., 2010. Using ELO ratings for match result prediction in association football. Int. J. Forecast. 26 (3), 460–470. https://doi.org/10.1016/j.ijforecast.2009.10.002.

Kahneman, D., 2011. Thinking, Fast and Slow, 1st edition. Farrar, Straus and Giroux.

Kocsoy, A., 2025. Referee bias in football: actual vs. expected additional time. Sports Econ. Rev. 9, 100047. https://doi.org/10.1016/j.serev.2025.100047.

Koenigs, M., Tranel, D., 2008. Prefrontal cortex damage abolishes brand-cued changes in cola preference. Soc. Cogn. Affect. Neurosci. 3 (1), 1–6. https://doi.org/10.1093/scan/nsm032.

Krumer, A., 2020. Testing the effect of kick-off time in the UEFA Europa League. Eur. Sport Manag. Q. 20 (2), 225–238. https://doi.org/10.1080/16184742.2019.1598456.

Krumer, A., Lechner, M., 2018. Midweek effect on soccer performance: evidence from the German Bundesliga. Econ. Inq. 56 (1), 193–207. https://doi.org/10.1111/ecin.12465.

Lago-Peñas, C., Gómez-López, M., 2016. The influence of referee bias on extra time in elite soccer matches. Percept. Mot. Skills. 122 (2), 666–677. https://doi.org/10.1177/0031512516633342.

Landy, D., Sigall, H., 1974. Beauty is talent: task evaluation as a function of the performer's physical attractiveness. J. Pers. Soc. Psychol. 29 (3), 299–304. https://doi.org/10.1037/h0036018.

Lasek, J., Szlávik, Z., Bhulai, S., 2013. The predictive power of ranking systems in association football. Int. J. Appl. Pattern Recognit. 1 (1), 27–46. https://doi.org/10.1504/IJAPR.2013.052339.

Leitner, M.C., Daumann, F., Follert, F., Richlan, F., 2023. The cauldron has cooled down: a systematic literature review on home advantage in football during the COVID-19 pandemic from a socio-economic and psychological perspective. Manag. Rev. Q. 73, 605–633. https://doi.org/10.1007/s11301-021-00254-5.

Lively, C., 2017. How social influence factors might impact the jury. Law Foundation of Newfoundland and Labrador Legal Research Awards for Students of Memorial University: 2015/2016, pp. 1–6.

Morton, R.B., Muller, D., Page, L., Torgler, B., 2015. Exit polls, turnout, and bandwagon voting: evidence from a natural experiment. Eur. Econ. Rev. 77, 65–81. https://doi.org/10.1016/j.euroecorev.2015.03.012.

Murphy, M.R., Jones, A.P., 1993. The influences of performance cues and observational focus on performance rating accuracy. J. Appl. Soc. Psychol. 23 (18), 1523–1545. https://doi.org/10.1111/j.1559-1816.1993.tb01046.x.

Nevill, A.M., Balmer, N.J., Williams, A.M., 2002. The influence of crowd noise and experience upon refereeing decisions in football. Psychol. Sport Exerc. 3 (4), 261–272. https://doi.org/10.1016/S1469-0292(01)00033-4.

Page, K., Page, L., 2010. Alone against the crowd: individual differences in referees' ability to cope under pressure. J. Econ. Psychol. 31 (2), 192–199. https://doi.org/10.1016/j.joep.2009.08.007.

Palacios-Huerta, I., 2014. Beautiful Game Theory: How Soccer Can Help Economics. Princeton University Press. https://www.jstor.org/stable/j.ctt6wq05z.

Palacios-Huerta, I., 2023. The Beautiful Dataset. December 15. SSRN. https://doi.org/10.2139/ssrn.4665889.

Peeters, T., van Ours, J.C., 2021. Seasonal home advantage in English Professional Football; 1974–2018. Economist. (Leiden) 169, 107–126. https://doi.org/10.1007/s10645-020-09372-z.

Pettersson-Lidbom, P., Priks, M., 2010. Behavior under social pressure: empty Italian stadiums and referee bias. Econ. Lett. 108 (2), 212–214. https://doi.org/10.1016/j.econlet.2010.04.023.

Pollard, R., 1986. Home advantage in soccer: a retrospective analysis. J. Sports Sci. 4 (3), 237–248. https://doi.org/10.1080/02640418608732122.

Pollard, R., 2006. Worldwide regional variations in home advantage in association football. J. Sports Sci. 24 (3), 231–240. https://doi.org/10.1080/02640410500141836.

Ponzo, M., Scoppa, V., 2018. Does the home advantage depend on crowd support? Evidence from same-stadium derbies. J. Sports Econom. 19 (4), 562–582. https://doi.org/10.1177/1527002516665794.

Prendergast, C., Topel, R., 1993. Discretion and bias in performance evaluation. Eur. Econ. Rev. 37 (2–3), 355–365. https://doi.org/10.1016/0014-2921(93)90024-5.

Ramchandani, G., Millar, R., Wilson, D., 2021. The relationship between team ability and home advantage in the English football league system. German J. Exerc. Sport Res. 51 (3), 354–361. https://doi.org/10.1007/s12662-021-00721-x.

Reade, J.J., Schreyer, D., Singleton, C., 2022. Eliminating supportive crowds reduces referee bias. Econ. Inq. 60 (3), 1416–1436. https://doi.org/10.1111/ecin.13063.

Reinstein, D., Riener, G., 2012. Reputation and influence in charitable giving: an experiment. Theory Decis. 72, 221–243. https://doi.org/10.1007/s11238-011-9245-8.

Rothschild, D., Malhotra, N., 2014. Are public opinion polls self-fulfilling prophecies? Res. Polit. 1 (2). https://doi.org/10.1177/2053168014547667.

Schuldt, J.P., Schwarz, N., 2010. The "organic" path to obesity? Organic claims influence calorie judgments and exercise recommendations. Judgm. Decis. Mak. 5 (3), 144–150. https://doi.org/10.1017/S1930297500001017.

Schwartz, B., Barsky, S.F., 1977. The home advantage. Soc. Forces 55 (3), 641–661. https://doi.org/10.1093/sf/55.3.641.

Scoppa, V., 2008. Are subjective evaluations biased by social factors or connections? An econometric analysis of soccer referee decisions. Empir. Econ. 35, 123–140. https://doi.org/10.1007/s00181-007-0146-1.

Scoppa, V., 2021. Social pressure in the stadiums: do agents change behavior without crowd support? J. Econ. Psychol. 82, 102344. https://doi.org/10.1016/j.joep.2020.102344.

Sors, F., Tomé Lourido, D., Parisi, V., Santoro, I., Galmonte, A., Agostini, T., Murgia, M., 2019. Pressing crowd noise impairs the ability of anxious basketball referees to discriminate fouls. Front. Psychol. 10, 2380. https://doi.org/10.3389/fpsyg.2019.02380.

Staw, B.M., 1975. Attribution of the "causes" of performance: A general alternative interpretation of cross-sectional research on organizations. Organ. Behav. Hum. Perform. 13 (3), 414–432. https://doi.org/10.1016/0030-5073(75)90060-4.
Tan, A., Benni, D., Liani, W., 2016. Determinants of corporate social responsibility disclosure and investor reaction. Int. J. Econ. Financ. Issues 6 (4S), 11–17.
Thorndike, E.L., 1920. A constant error in psychological ratings. J. Appl. Psychol. 4 (1), 25–29. https://doi.org/10.1037/h0071663.
Tversky, A., Kahneman, D., 1982. Judgment under uncertainty: heuristics and biases. In: Kahneman, D., Slovic, P., Tversky, A. (Eds.), Judgment Under Uncertainty: Heuristics and Biases, 1st ed. Cambridge University Press, pp. 3–20.
Williams, K.D., Jones, A., 2005. Trial strategy and tactics. In: Brewer, N., Williams, K.D. (Eds.), Psychology and Law: an Empirical Perspective. The Guilford Press, pp. 276–321.
Zimmerman, D., Rodriguez, D., Bergold, A., Penrod, S., 2016. The influence of pretrial exposure to community outrage and victim hardship on guilt judgments. Psychol., Crime Law 22 (5), 435–454. https://doi.org/10.1080/1068316X.2015.1120870.