Judging Under Public Pressure[1]

Florian Auferoth[2] Alma Cohen[3] Zvika Neeman[4]

July 19, 2020

[1]Acknowledgments to be added.

[2]Friedrich-Alexander-Universität, Erlangen-Nürnberg

[3]Harvard Law School, Tel Aviv University, CEPR, ECGI, and NBER

[4]Tel Aviv University

Abstract

Individuals who engage in "judging" – that is, render a determination in a dispute or contest between two parties – might be influenced by public pressure to favor one of the parties. Many rules and arrangements seek to insulate such individuals from public pressure or to address the effects of such pressure. We study this subject empirically, investigating the circumstances in which public pressure is more and less likely to affect judging.

Using detailed data from the Bundesliga, Germany's top soccer league, our analysis of how crowd pressure affects the decisions of referees yields two key insights. First, we show that crowd pressure biases referee's decisions in favor of the home team for those decisions that cannot be unambiguously identified as erroneous but not for those decisions that can. In particular, referees exhibit a bias in favor of the home team with respect to more subjective decisions such as the showing of yellow cards (cautions), which is based on the referee's judgment call, but not with respect to more objective decisions such as validating goals and awarding penalty kicks, where live TV coverage often allows for objective identification of errors.

Second, we show that the effect of crowd pressure on referee decisions depends on the extent to which such pressure is viewed by the referee as understandable or reasonable (or even justified). Specifically, a referee's bias in favor of the home team in yellow card issuance is strengthened after the referee makes an objectively identifiable error against the home team and thus might view crowd heckling as understandable. This effect is stronger when the referee's error is costlier to the home team because the game is more important or the error is more consequential due to the closeness of the game at the time of the error.

The introduction of VAR (Video Assisted Referee) technology in 2017 and Covid-19, which caused games to be played without crowds for the second part of the 2019-20 season allows us to test our results under three different regimes (pre-VAR, post-VAR, and post-VAR but without any crowd). Inspection of the results under these three different regimes serves to reinforce them. As expected, VAR reduces the number of referee errors, but the pattern of no bias with respect to errors is preserved. VAR has no effect on the number of yellow cards. Once the crowd disappears, so does the home advantage in field goals. Referee errors are unaffected, but the home bias with respect to yellow cards disappears as well. This confirms the effect that the crowd has on referee's more subjective decisions.

JEL CODES: K40, Z20

KEYWORDS: Judging, judicial decisions, public pressure, sub judice, make up call.

1 Introduction

In many settings, an independent decision maker is required to make one or more determinations in a dispute or contest between two or more parties with competing interests with respect to these determinations. We refer to any such determinations as “judging.” Defined as such, judging is very broad: it includes decisions by judges and jurors in court cases, decisions by arbitrators in arbitration proceedings, determinations by independent fact-finding commissions on issues that are publicly or politically contested, and decisions by referees in sports events. In his nomination hearings, Chief Justice John Roberts famously compared the job of a Supreme Court Justice to that of a baseball referee; the analogy drawn between these two types of activities reflects a view that both roles involve what we refer to as "judging.” [1]

Even when the individual who engages in judging is independent from the relevant parties, her determinations could be subject to and affected by public pressure. In the case of judges, this is especially relevant for elected state court judges who expect to come up for reelection,[2] as well as for other judges who have career concerns and expect the conformity of their decisions with those favored by the general public to affect their career prospects. Even judges who are appointed for life and thus presumably are free of such career concerns might be influenced by public demands in various direct and indirect ways. As to the case of jurors, there is a long-standing recognition and concern that pretrial publicity, and the resulting widespread public views of the "correct" outcome, might have a significant effect on jurors’ decisions.

In recognition of the distorting effects of pubic pressure on judging, some rules and arrangements have been devised in order to insulate judging decisions from public pressures, completely or partly. In many common law jurisdictions, there have been long standing limitations on public commentary for cases that are *sub judice* (Latin for "under a judge," namely, under trial or otherwise under consideration by a judge or court). In the US, the First Amendment guarantee of free speech has pre-

[1]Opening Statement of Judge John Roberts before the Senate Judiciary Committee, Monday, September 12, 2005.

[2]An earlier study by two of us (Cohen et al., 2018) shows how the prospect of elections affects the decisions of federal judges. For a wider perspective on this issue, see Shepherd (2011) and the references therein.

vented tight restrictions on comments regarding matters that are *sub judice*, but State Rules of Professional Conduct governing attorneys often place restrictions on the out-of-court attorney statements regarding ongoing cases.

Furthermore, in jury cases, judges often seek to minimize as much as possible the exposure of jurors to outside public pressure. It is common for judges to instruct jurors not to read newspaper articles about the trial. In addition, the law enables criminal defendants to move the trial venue to a different state when pretrial publicity would make it difficult to find jurors that would not have been exposed to the publicity and to the public pressures produced by it. And some criminal convictions have been overturned due to the exposure of jurors to media coverage or to atmosphere of a “media circus." (See Phillipson, 2008, for an extensive discussion of these issues).

In this paper, we investigate empirically whether and when public pressure affects judging. We study the subject in a setting where it is possible to obtain rich and detailed data that facilitates such an investigation. In particular, we use comprehensive and detailed data from the Bundesliga – the premiere soccer league in Germany. We investigate whether and when crowd pressure influences the decisions of referees and the extent to which those decisions are biased in favor of the home team, which is favored by the (great) majority of the crowd. We are fortunate to be able to compare the results under three regimes: without VAR (Video Referee Assist) technology, with VAR technology, and with VAR technology, but no crowd. As explained below, this comparison reinforces our results.

Our analysis yields two key insights. First, we hypothesize that crowd pressure should be expected to influence some referee decisions significantly more than others, and we find empirical evidence that is consistent with our hypothesis. In particular, we conjecture that crowd pressure has less influence on referees’ decisions when errors are indisputably observable because in such situations the countervailing force of referees’ concern for their reputation is strong. In contrast, we expect crowd pressure to have more influence on referees’ decisions when errors are not indisputably observable, because such situations make it easier for referees to rationalize and defend their decisions without incurring large costs to their professional reputation.

Our data enables us to test this conjecture because it includes two types of decisions: (1) referees' decisions whether to validate and whether to award penalty kicks (which in turn are converted to goals with a high probability), and (2) referees' decisions whether to award disciplinary sanctions to players in the form of yellow cards (cautions). The correctness of referees' decisions with respect to goals and penalty kicks is (relatively) indisputably observable: live TV coverage provides an immediate replay of all player movements during the seconds before the actions that are the subject of the referee's decisions, often from several different angles, and this generally enables an objective assessment of whether the referee made a correct or erroneous call. As a result, information on whether the referee made or did not make an error spreads quickly among all spectators of the game.

In contrast, referees' decisions on whether to award a yellow card are (relatively) more subjective. Such decisions typically involve a judgment call on which reasonable people may often disagree. As a result, a referee can often make a decision that benefits the home team to award (or fail to award) a yellow card through an easily defensible decision that does not impose a significant reputational cost on the referee, if at all.

We provide evidence that, as predicted by our hypothesis, errors made in referee decisions regarding validating goals and regarding award of penalty kicks do not exhibit a bias in favor of the home team. In contrast, we find evidence that decisions to issue yellow cards are biased in favor of the home team. In particular, we find that referees award more yellow cards to the away team relative to the home team, and that this effect is economically and statistically significant.

Our second key finding concerns the circumstances that enhance the magnitude of the effect of crowd pressure on those decisions that are amenable to such pressure (in our setting, the issuance of yellow cards). We hypothesize that crowd pressure is likely to have a stronger effect if it is viewed by the referee as relatively or partly understandable and reasonable. In particular, we conjecture that when a referee makes an error against the home team in deciding whether to validate a goal or award a penalty kick, the referee will be more inclined to view subsequent crowd pressure and heckling as understandable or reasonable, at least to some extent.

Consistent with our hypothesis, we find that when a referee make an error decision against the home team in decisions regarding goals or penalty kicks, the referee subsequently exhibits an increased bias in favor of the home team in decisions regarding the issuance of yellow cards. Moreover, the increase in bias in favor of the home team is magnified when the referee's erroneous decision occurred (i) during a game that is more important, or (ii) at the point of time in the game in which the score was close and so the error was likely to be more consequential.

Our analysis reveals an asymmetric pattern in what follows errors in referee decisions against the home team regarding goal validation or award of penalty kicks. When such an error is made against the home team, referees tend to "make up" for the error by increasing the number of yellow cards awarded to the away team. However, when such an error is made against the away team (and thus in favor of the home team), referees do not "make up" for the error by subsequently awarding more yellow cards against the home team. This lack of symmetry is consistent with the view that crowd pressure plays a key role in inducing referees to make up for errors and, in particular, induces referees to make up for errors that upset the crowd.

The introduction of VAR (Video Assisted Referee) technology in 2018 and Covid-19, which caused games to be played without crowds for the second part of the 2019-20 season reinforces our results. As expected, VAR reduces the number of referee errors. VAR does not completely eliminate referee errors because a referee can still err by not consulting the VAR system when it should. However, the pattern of no bias with respect to referee errors is preserved. VAR has no effect on the number of yellow cards. This is also to be expected because referees do not usually consult the VAR system with respect to relatively minor offences such as yellow cards decisions.

Surprisingly, once the crowd disappears, so does the home advantage in field goals. This suggests that the crowd has a strong effect on the players. Referee errors are unaffected, but the home bias with respect to yellow cards disappears as well. In fact, once the crowd disappears, referees seem to also compensate the away team by giving more yellow cards to the home team after they make an error. The number of games played with no crowd is small so this finding is not statistically significant, but it is not surprising because with VAR, those errors that are made by referees become even more glaring, and so referees may experience a stronger compulsion to make up for them. And when there is no crowd, there is no one to militate against compensation of the away team.

Literature Review

Social influence, of which public pressure is a part, is a formidable force. Much of the research in the field of sociology is devoted to the study of its determinants and many manifestations. However, the study of social influence in the context of judging is more limited.

There is substantial literature on the desirability of *sub judice* restrictions on public commentary on pending court cases. In particular, there is a widely shared concern that the ubiquity of media coverage makes fair trials impossible because of the difficulty of preventing jurors from being influenced by media coverage (Phillipson, 2008; Marder, 2014). Relatedly, media coverage of the work of commissions of inquiry in Canada has been called "concerning" by the head of several such committees (Gomery, 2006).

However, the empirical literature on the effect of pubic pressure on judging has been much more limited. Some work has investigated empirically the subject by examining how sentencing decisions by state court judges are influenced by their proximity to reelection (Huber and Gordon, 2004; Berdejó and Yuchtman, 2013; Cohen et al., 2018). There is also experimental evidence and surveys that shows that pretrial publicity affects jurors (Steblay et al., 1999). We seek to contribute to this limited body of empirical work on the subject.

Due to the widely shared interest in sports and the availability of rich datasets regarding sport events, there exists a substantial literature on the decisions of sport referees (see, e.g, Boeri and Severgnini, 2011; Buraimo et al. 2010; Garicano et al., 2005; Nevill at al., 1996; Sutter and Kocher, 2004; and the references therein). This literature has documented the home bias in the award of yellow cards that we also find (see, e.g., Dawson et al., 2007). However, this literature has not examined the two key issues regarding the effect of public pressure on judging that are the focus of our analysis: the difference between the type of decisions that are more and less affected by crowd pressure (and in this connection the importance of the extent to which decisions can be objectively assessed by outside observers); and how the impact of crowd pressure is affected by the extent to which this pressure is viewed as understandable or reasonable by the decision-maker (and in this connection the power of crowd pressure to produce make-up calls).[3]

The rest of the paper proceeds as follows. In the next section, we describe the institutional setting and the categorization of errors. The data is described in Section 3. In Section 4, we discuss our empirical methodology and present our results with respect to referee errors. In Section 5, we examine the subject of yellow cards. In Section 6, we describe the effect of the introduction of VAR technology and no crowds. Finally, Section 7 offers a brief conclusion.

2 Institutional Background

The data we use is from the German premier soccer league, the Bundesliga. Soccer is the most popular sport in Germany. A Bundesliga game has an average number of 30,000 spectators present in the stadium, and many others watch or otherwise follow games through various media.

The Bundesliga consists of two divisions. The first Bundesliga is the highest soccer division in Germany and consists of 18 teams, which face each other twice every season. Each team is the home team in one such game and the visiting or away team in the other game. This generates 34 match days per season with nine matches per match day. The second Bundesliga has the same number of teams and the same number of match days per season. At the end of each season, the two lowest ranked teams in the first Bundesliga are relegated to the second Bundesliga, and the first two ranked teams of the second Bundesliga are promoted to the first Bundesliga. The team that is ranked third from last in the first Bundesliga (16th place) plays two matches against the third ranked team of the second Bundesliga to determine the team that will play in the first Bundesliga during the next season.

In each game, the objective of each team is to win the game by scoring more goals than the other team. The interactions between the two teams are regulated by

[3]Interestingly, this literature has documented that the effect of the crowd on referees' decisions is weaker in stadiums where the field is surrounded by a running track, supposedly because this implies that the distance between the crowd and the referee is a little larger. Being able to compare referees' decisions in games with and without crowds as we do here provides a more convincing way to test this hypothesis.

the laws of soccer, which specify the playing time and permitted actions. Each team is permitted to use all body parts except for arms and hands in order move the ball in order to score goals. Physical tackles between players are regulated.

Each game is refereed by a referee who is responsible for ensuring that both teams adhere to the rules of the game. Bundesliga referees are experienced and have been selected in sequential promotion tournaments. After passing a written and physical test, they typically start in the lowest division. Once they have been promoted to the sixth division, they can be promoted at most one division each year if judged as qualified by official observers. The performance of referees is monitored and judged by an official observer of the German Football Association (DFB) who attends the game, and evaluates the referee for being "decided, secure, with the courage to take unpopular decisions, and unimpressed by complaining players" as well as for how well the referee interprets and implements the laws of soccer. Referees who are found to be biased are dismissed (Dohmen, 2008).[4]

Two assistants support the referee. The assistants' task is to indicate offside calls and whether the ball was out of bounds. The referee's task is to detect (and sanction) violations of the rules, to stop play if the rules are violated, and to ensure that play then continues according to the rules of the game. Failure of the referee to detect violations or incorrect findings of violations can have a critical effect on the outcome of the game. The referee has the final authority to decide whether the rules of the game have been violated. Specifically, the referee determines the followings: whether tackles between players are illegal (a "foul"); whether a player illegally touched the ball with his hand or arm (a "hand ball"); whether a player was in an illegal position ("offside"); whether the ball crossed the perimeter line; and whether players or officials in other ways violated the rules of the game.

Violations of the rules lead to stoppages of play. In the case of fouls or handballs, play then restarts with a penalty kick or free kick for the team that did not commit the violation. A penalty kick is given if the violation was committed inside the penalty box and provides an excellent opportunity to score a goal by giving the team a shot

[4]Referees are paid 3,800 and 2,000 Euros per match in the first and second Bundesliga, respectively, on top of an annual base salary of 35,000 Euros in the second Bundesliga, and twice as much in the first Bundesliga. They are also compensated for their travel expenses, including hotel and transportation. In 2019, the average wage in Germany was a little over 48,000 Euros.

from 11 meters distance to the goal, which may only be blocked by the goalkeeper. A free kick can also be a good goal scoring opportunity, but any player may block it. The free kick's location is where the offense occurred. Following players' violations of the rules, the referee may also sanction players with a warning (yellow card) or dismissal (Red card), depending on the severity of the violation. Offences justifying a yellow card include unsportsmanlike behavior, persistent infringement of the rules, delaying the restart of the game or dissent by word or action (FIFA, 2018). Red cards are awarded for seriously foul play, illegally denying goal scoring opportunities, violent conduct, insulting behavior or receiving a second yellow card (FIFA, 2018).

The referee may consult with two assistants. Video replays were only introduced in August 2017 and were not available to the referees in the games considered in our dataset. However, as mentioned above, they were generally available for the spectators watching the game, possibly with delay. At any point during the game, the referee needs to make immediate decisions whether the rules were violated and if so, what is the appropriate sanction. Thus, a referee faces a quick succession of situations that demand its attention and consideration, and errors in his assessment of the situation or sanctions, including serious errors that have a large effect on the outcome of the game, are not uncommon. Players, coaches, and sports' crowds alike all hotly debate referees' errors both during and after each game.

For the purpose of our analysis we consider referee errors that were recorded on the website www.wahretabelle.de. This website was established in 2006 with the goal of recording the correct result for all Bundesliga games. Accordingly, the website establishes what would have been the final score of the game if the referee had not made any errors. The website allows users to submit photos and video recorded scenes from Bundesliga games for review if they believe that the referee's decision was wrong and potentially affected the result of the game. A panel of experts assembled by the website decides if the referee's decision was correct or not. Wrong decisions are recorded as such only if they are judged by the panel to have had a direct impact on the final score of the game. If the team that was advantaged by the error does not profit from the error, the error is not recorded in this dataset.

The panel of experts consists of twelve contributors to the website, who have each established a reputation for competence and impartiality. The members of the panel represent supporters of many different clubs. Neither the panel nor the website play any official role in the running of the German leagues. Referees are not held accountable to this panel or to the website. The purpose of the panel is only to accurately record events on the website, based on their best judgement.[5]

A referee error is categorized as such by the website if and only if it has a direct and immediate effect on the score of the game. In particular, this is the case if:

- A referee incorrectly approves a goal. For example, if the referee failed to notice that an attacking player fouled a defense player just prior to scoring the goal or touched the ball with his hand.
- A referee sanctions the defending team incorrectly, and this leads to a goal. For example, by incorrectly approving a penalty kick, which is converted into a goal.
- A referee incorrectly denies a goal. For example, by incorrectly calling an offside offense against the attacking team.
- A referee fails to call an offense by the defending team that has a sanction that provides the attacking team an excellent opportunity for scoring a goal. For example, if the referee fails to notice a foul inside the penalty box, which would have implied a penalty kick for the attacking team.

3 Data and Summary Statistics

Our dataset consists of all the matches in the first German Bundesliga in the eight seasons from 2009/10 to 2016/17. We also add all the games in the second German Bundesliga in the four seasons between 2013/14 and 2016/17.[6]

[5] The Bundesliga has an internal referee evaluation system. However, the data used in this system and its deliberations are not publicly available. This system is frequently criticized for its lack of transparency, and for its over reliance on the goodwill of the head referee who is in charge of the system (see https://www.n-tv.de/sport/fussball/collinas_erben/Manipulation-Machtmissbrauch-Mobbing-article20099310.html).

[6] Video assistant referee (VAR) was introduced in the 2017-2018 season. As expected, the availability of VAR greatly reduced the number of referee errors. In addition, the presence of VAR changed the nature of referee errors. With VAR, the referee has the best available information. Therefore, any referee decision that is classified as an "error" is necessarily subjective and a decision on which reasonable people may disagree. For this reason, we exclude these seasons from our sample.

The data we use has been extracted from publicly available match summaries at www.wahretabelle.de using a computer algorithm. The data has been supplemented by data from forum discussions on www.wahretabelle.de in all cases where match summaries are contradictory or incomplete. All reports about possible errors are submitted for consideration through this forum. The forum also exhibits the vote and the panel's ruling on submitted possible errors, as well as all the evidence that was considered. We use the majority ruling by the panel to classify errors. We use the match summaries available at www.wahretabelle.de to extract other relevant information about the game as well as the referee's name.

We supplement our dataset with match-, team- and referee-data extracted from the website www.kicker.de. Kicker is Germany's largest soccer magazine. It records extensive statistics and offers live descriptions of matches. From these descriptions we extract the timing of yellow and red cards awarded to players for all games in the dataset.[7] Team rankings in the league table are also extracted from Kicker. Finally, we also extract the grades referees received for their performance in games from Kicker. For each game Kicker rates the referee based on the accuracy of its decisions and its control over the game. The grades are awarded on the same ranking scale as German school grades (1 is the best grade, 6 the worst) and thus serve as a measure of quality of the referee's performance in each match.

The data contains 3,662 games over the years 2009-2016. Table 1 below shows the following summary statistics:

[7] The number of red cards in our sample is too small so we don't use it as a separate dependent variable; we count red cards as two yellow cards.

Table 1: Summary Statistics

	Average	Away Team	Home Team	diff	P-value
Number of Goals	2.78	1.23	1.55	-0.32	0.00
	(1.69)	(1.18)	(1.3)		
Number of Referee Errors	0.38	0.20	0.18	0.02	0.11
	(0.65)	(0.46)	(0.020)		
Minutes to First Error	47.39	48.64	46.07	2.57	0.09
	(25.03)	(25.49)	(24.49)		
Number of Games with $\geq$ 1 Errors	1,112	573	539		
Number of yellow Cards	3.88	2.13	1.74	0.39	0.00
	(2.13)	(1.43)	(1.33)		
Stadium capacity	35,194				
	(18,891)				
Attendance/capacity Ratio	0.86				
	(0.18)				
Number of Games per Referee	57.84				
	(52.51)				
Number of Referees	64				
Number of Games	3,662				

Note: Standard deviation are in parenthesis.

On average 2.78 goals are scored in each game in our sample: the Away Team scores 1.23 goals, and the Home Team scores 1.55 goals. A t-test indicates that the rather large difference between these two means is statistically significant at the 1% level. The difference in the number of goals scored indicates that the home team generally possesses an advantage over the away team.

Out of the 3,662 games in our sample, in 1,112 games there was at least one referee error. In 573 games the first error was against the Away team and in 539 against the Home Team. Table 1 shows that referees make 0.38 errors per game on average, 0.20 against the Away Team, and 0.18 against the Home Team. This difference is small and not statistically significant. The number of errors has to be considered in relation to the average number of goals in the game, which is 2.78. So when an error occurs, it has a large effect on the final score of the game.

Table 1 shows that the mean time to the first error is 47.39 minutes, with 48.64 minutes if the first error is against the home team, and 46.07 minutes if the first error is against the away team. The p-value for the difference between these values is statistically significant but small in magnitude.

Table 1 also shows that referees issue an average of 3.88 yellow cards per game, with 2.13 yellow cards issued to the Away team, and 1.74 to the Home Team. This difference is large and statistically significant at the 1% level. It is indicative of the presence of a bias in favor of the home team, as we argue below. Finally, Table 1 shows that stadiums have an average capacity of 35,000 spectators. Attendance in games is usually high, with an average attendance to capacity ratio of 86%. There are 64 unique referees in our sample, and on average they each referee 58 games.

4 "Disputable" vs. "Indisputable" Errors

On average, the crowd in each game consists of more than 30,000 spectators, of which about 90% support the home team. This implies that public pressure on the referee is likely to be very strong, especially in favor of the home team.

Nevertheless, as explained in the introduction, we hypothesize that: (1) referees' integrity is sufficiently strong to withstand crowd pressure on those decisions where an error would be "indisputable," such as those decisions regarding the validation or invalidation of field goals and penalty kicks. (2) Referees' integrity is not sufficiently strong to withstand crowd pressure on those decisions where an error would not be "indisputable," such as decisions about sanction by a yellow card.

In order to test this hypothesis, we run the following regression:

$$
\begin{aligned}
Y_{g,i} = \\
&\beta_1 Home\ Team_{g,i} + \beta_2 Mild_g + \beta_3 High_g + \beta_4 Home\ Team_{g,i} \times Mild_g \\
&+ \beta_5 Home\ Team_{g,i} \times High_g + \beta_6 Attendance/Capacity_g + \beta_7 Capacity_g \\
&+ \alpha_h + \gamma_a + \delta_s + \eta_d + \theta_w + \lambda_r + \epsilon_{g,i}
\end{aligned}
$$

where $g = (s, d, w, h, a)$ indexes games (uniquely identified by the season, s, the division, d, the week in the season, w, the home team, h, and the away team, a). The index $i \in \{1, 2\}$ represents the first or second playing team. The dependent variable $Y_{g,i}$ is the number of errors against/yellow cards given to team i in game g. Our main variable of interest is *Home Team* which is a dummy variable that is equal to 1 if the game is a home game for the team, and 0 otherwise. This variable is supposed to capture the bias, if any, created by crowd pressure. The variables *Mild* and *High* are dummies that are equal to one if the game is "mildly" or "highly" important, and zero otherwise. Remaining games are considered "less" important.

We define a game to be "less" important it if is not important to either the home or away teams according to our definition; "mildly" important if it is important to the away but not to the home team; and it is considered to be "highly" important if it is important for the home team (regardless of its importance to the away team). [8] The reason we give a bigger weight to the home team is that most of the crowd in the stadium supports the home team (on average, only about 10% of the crowd support the away team), so that if a game is important to the home team, it is also more important to a much larger fraction of the crowd.

The regression, as well as all other regressions below, also controls for division, season, week, teams, and referee dummies. Standard errors are clustered by the referee.[9] Table 2 below presents a basic analysis of home bias.

[8]A game is considered important to a team if the team ranks among those in the top or bottom thirds in the Bundesliga table at the time of the game, and the absolute difference in score between the team and those adjacent to it in the Bundesliga table from above and below is less than or equal to 2. The definition of importance is motivated by the idea that teams in the top or bottom thirds of the Bundesliga table have stronger incentive to do well. Those in the top want to secure their top position, which allows them to compete in European leagues and promise other rewards, and those in the bottom third want to avoid being relegated to a lower division at the end of the season.

[9]Running the regression with different clusterings does not change our results.

Table 2: Number of Errors/Yellow Cards by Importance

Number of:	(1) Errors	(2) Errors	(3) Yellow Cards	(4) Yellow Cards
Home Team	-0.0173	0.0044	-0.3919***	-0.4068***
	(0.0130)	(0.0260)	(0.0308)	(0.0704)
Mild		0.0365*		0.0276
		(0.0202)		(0.0551)
High		0.0304		0.0582
		(0.0242)		(0.0606)
Home Team x Mild		-0.0206		-0.0133
		(0.0333)		(0.0851)
Home Team x High		-0.0291		0.0311
		(0.0297)		(0.0884)
Attendance/capacity	0.0559	0.0539	0.5254***	0.5109***
	(0.0598)	(0.0608)	(0.1902)	(0.1899)
Stadium capacity (per 10,000)	-0.0113	-0.0125	-0.0716	-0.0768
	(0.0197)	(0.0200)	(0.0572)	(0.0576)
N	7,302	7,302	7,302	7,302
R^2	0.0078	0.0077	0.0879	0.0879
Mean Dependent Variable	0.1900	0.1900	1.9404	1.9404

Note: Standard errors are clustered by the referee and are in parenthesis.
Stars denote the level of statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
We control for division, season, week, referee and teams dummies.

Table 2 columns (1) and (3) shows the results of running the regression on each of the dependent variables without controlling for the importance of the game. Columns (2) and (4) provide the results for each of the dependent variables once we control for the importance of the game and its interaction with the variable *Home Team*. Columns (1) and (2) show the results when the dependent variable is the Number of errors. When we do not control for the importance of the game (column (1)), the estimated measure of crowd pressure (the coefficient of *Home Team*) is not statistically significant. Column (2) shows that this is unaffected by adding controls for the importance of the game. The results indicate that there are more errors in games that are mildly and highly important, where this effect is statistically significant at the 10% level for games that are mildly important, but not significant for games that are highly important. However, the interaction of the importance of the game with *Home Team* is not statistically significant. This implies that, even in games that are more important to the home team, where we expect crowd pressure from the home team fans to be stronger, there is no extra bias to the home team. We also find that the ratio of attendance to capacity, which measures how packed the stadium is, and the size of the stadium have no effect of the Number of errors.

Table 2 Columns (3) and (4) perform a similar analysis for the dependent variable number of yellow cards. Contrary to what we got in column (1) and (2), the results show a large and statistically significant home bias at the 1% level. On average the *Home Team* receives 20% less yellow cards in a game. Column (4) shows that the effect of the importance of the game, although positive, is not statistically significant. The statistical insignificant of the interaction of the importance of the game with *Home Team* suggests that the home bias does not increase with the importance of the game. The attendance to capacity ratio and the size of the stadium have large and positive effect (about 26%) on the number of yellow cards given. This is statistically significant at the 1% level. This indicates that players' respond to a bigger and denser crowd by playing more aggressively, or alternatively that referees change their behavior in such games. It could also be due to the fact that larger crowds and long rivalries between teams that induce more aggressive play are positively correlated.

The results in table 2 are consistent with our hypothesis that errors made in referee decisions regarding validating or invalidating goals and penalty kicks do not exhibit a bias in favor of the home team. However, we find evidence that decisions to award yellow cards are biased in favor of the home team. In particular, we find that referees award more yellow cards to the away team, relative to the home team, and that this effect is economically and statistically significant.

5 The Effect of an Intensification in Crowd Pressure Following a Referee's Error

Our second hypothesis is about the way in which the referee responds to an intensification in the degree of crowd pressure as a result of a referee error. We hypothesize that when a referee makes an error against the home team on an indisputable decision, the referee would be more inclined to view subsequent crowd pressure and heckling as understandable or reasonable, at least to some extent.

Before moving to test this hypothesis, we first show that when a referee makes an error with respect to the validation or invalidation of goal and penalty kicks, the referee makes the error in a way that does not bias against the away or the home team. The assertion that the first error is not biased against the away or home team allows us to use the error as an exogenous event for increasing crowd pressure and to test whether referees are more likely to "make up" teams for which they made errors. Since 90% of the crowd is from the Home team, we expect referees to be more likely to "make up" the home team when the first error is against the home team, than to "make up" the away team, when the first error is against the away team.

We therefore run similar regressions to the ones in the Section 4, only that the dependent variable now is, *First error*, which is a dummy variable equal to one if the first error was against team i in game g, and zero otherwise.

Table 3: First Error by Importance

	All Games		Games with ≥ 1 errors	
	(1)	(2)	(3)	(4)
Home Team	-0.0090	0.0162	-0.0296	0.0570
	(0.0116)	(0.0224)	(0.0388)	(0.0820)
Mild		0.0346*		0.0523
		(0.0196)		(0.0478)
High		0.0372*		0.0523
		(0.0214)		(0.0532)
Home Team x Mild		-0.0226		-0.0921
		(0.0316)		(0.0995)
Home Team x High		-0.0346		-0.1105
		(0.0263)		(0.0962)
Attendance/capacity	0.0378	0.0346	-0.0000	0.0011
	(0.0400)	(0.0407)	(0.0000)	(0.0034)
Capacity (per 10,000)	-0.0128	-0.0141	0.0000**	0.0002
	(0.0129)	(0.0129)	(0.0000)	(0.0005)
N	7,302	7,302	2,226	2,226
R^2	0.0278	0.0286	0.0009	0.0026
Mean Dependent Variable	0.1524	0.1524	0.5	0.5

Note: Standard errors are clustered by the referee and are in parenthesis.
Stars denote the level of statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
We control for division, season, week, referee and teams dummies.

Table 3 columns (1) and (2) show the results when all games are included. Columns (3) and (4) repeat the analysis on the subset of games with errors.

When we do not control for the importance of the game, Column (1) shows that the estimated measure of home bias (the coefficient of *Home Team*) is not statistically significant. Column (2) shows that this is unaffected by controlling for the importance of the game. The results indicate that there are more errors in games that are mildly and highly important. This effect is statistically significant at the 10% level. However, the interaction of the importance of the game with *Home Team* is not statistically significant. This implies that there is no extra home bias to the home team that depends on the importance of the game. We also find that the ratio of attendance to capacity and the size of the stadium have no effect of the Number of errors.

Table 3 Columns (3) and (4) perform a similar analysis on the subset of games with errors. The results are similar to those obtained for the general class of games in our sample.

Table 3 is consistent with our hypothesis of no crowd pressure bias on referee decisions with respect to decisions that can unambiguously identifiable as erroneous.

5.1 Yellow Cards

In this subsection, we show that referees do exhibit home bias in the following sense: they award more yellow cards to the away team after an error against the home team, but not vice versa.

To test this idea, we compare the number of yellow cards before and after the first error in games with at least one error, with the number of yellow cards before and after half-time in games with no errors. The result that first errors are not biased against the away or the home team, allows us to make the comparison between games with and without errors.

In order for the comparison between the number of yellow cards in games with and without errors to be meaningful, we multiply the number of yellow cards before and after an error by the ratio of the length of the time interval in which they were measured and 45 minutes.[10] For example, if the first error was made in the 30th

[10] A game is played for 90 minutes. Half-time is at 45 minutes.

minute, then a total of 60 minutes still remains in the game after the error. We split the game into two parts: part 1 includes all yellow cards up to the 30th minute; and part 2 includes all yellow cards in the 60 minutes after the error. In order to account for the different interval length and the resulting difference in game events, we adjust the dependent variable. This is achieved by dividing the number of yellow cards before and after the error by the time interval length. To make this measure comparable to the number of yellow cards in a half of the game, we then multiply the measure by 45 (the number of minutes in one half). If the first error occurred in minute 30, we multiply the number of yellow cards before the first error by a 1.5 and multiply the number of yellow cards after the first error by 2/3. This transformation accounts for the remaining playing time after an error.[11] We also winsorize this variable to avoid extreme values in case the first error occurred in the first or last couple of minutes of the game.

Our hypothesis is that in games with errors, the number of yellow cards given to a team following a first error against the other team would be larger than if no error was made. As explained above, we find that this is indeed the case in games where the first error was against the home team, but not in games where the first error was against the away team.

To test this hypothesis we run the following regression:

$$
\begin{aligned}
Yellow\ Cards_{g,i,t} = &\\
& \beta_1 Home\ Team_{g,i} + \beta_2 After_{g,t} + \beta_3 Home\ Team_{g,i} \times After_{g,t}\\
& + \beta_4 Home\ Error_g + \beta_5 Away\ Error_g + \beta_6 Home\ Error_g \times Home\ Team_{g,i}\\
& + \beta_7 Away\ Error_g \times HomeTeam_g + \beta_8 Home\ Error_g \times After_{g,t}\\
& + \beta_9 Away\ Error_g \times After_{g,t} + \beta_{10} Home\ Error_g \times Home\ Team_{g,i} \times After_{g,t}\\
& + \beta_{11} Away\ Error_g \times Home\ Team_{g,i} \times After_{g,t} + \beta_{12} Attendance/Capacity_g\\
& + \beta_{13} Capacity_g + \alpha_h + \gamma_a + \delta_s + \eta_d + \theta_w + \lambda_r + \epsilon_{g,i,t}
\end{aligned}
$$

The dependent variable $Yellow\ Cards_{g,i,t}$ is the number of yellow cards given to team i ($i \in (1,2)$), that played in game g, in part t, where $t = 1$ if it is the second half of

[11] Recall that the mean time to a first error is 47.49 minutes.

the game in a game with no errors or if it is after the first error in a game with errors, and zero otherwise (if it is the first half of the game in a game with no errors or if it is before the first error in a game with errors).

$After_{g,t}$ is a dummy variable that is equal to 1 if it is the second half or after the first error in games with no errors and with at least one error, respectively; $Home\ Error_{g,i}$ and $Away\ Error_{g,i}$ are dummy variables that are equal to 1 if the first error in the game was against the Home and Away teams, respectively. Our main explanatory variables of interest are the triple interaction of ($Home\ Error_{g,i}$ or $Away\ Error_{g,i}$) $\times Home\ Team_{g,t}$ $\times After_{g,t}$ (and all possible other interactions between these variables). As we did previously we also control for fixed effects for division, season, week, teams, and referee dummies.

Table 4 below shows the results. Column (1) includes all games, Columns (2)–(3) exclude games with a first error in the last 5 and 10 minutes of the game, respectively. In such games, the referee's chance of finding itself in a situation where it can compensate a team for an error made previously against the other team is quite limited.

The results show that, on average, away teams receive more yellow cards. This effect is large and statistically significant. We also find that the number of yellow cards in the second part of the game is larger and statistically significant, which suggests that games becomes more intense as they progress. Namely, players play more aggressively and are cautioned more often for aggressive play. However, the fact that the $HomeTeam_{g,i} \times After_{g,t}$ interaction variable is small and insignificant implies that there is no additional home bias in the second half of the game in games with no errors. In games where the first error was against the Home Team, the Away Team receives less yellow cards than in games with no errors or with first error to Away Team.

The results also show that the number of yellow cards after an error was given to either Home or Away Team is larger and statistically significant, which suggests that games becomes more intense after an error was made.

Our main variables of interest are the triple interaction $HomeError_{g,i} \times HomeTeam_{g,i} \times After_{g,t}$ and $AwayError_{g,i} \times HomeTeam_{g,i} \times After_{g,t}$. These triple interaction variables capture referees' tendency to favor the home team by giving the away team more yellow cards after an error against the home team, if such a tendency exists. The regression results confirm that such a tendency indeed exists only after a first error against the home team, but not after a first error against the away team. In all the specifications in Table 4 below, the coefficient of $HomeError_{g,i} \times HomeTeam_{g,i} \times After_{g,t}$ is indeed negative and statistically significant while the coefficient of $AwayError_{g,i} \times HomeTeam_{g,i} \times After_{g,t}$ is small and insignificant. Columns (2) and (3) show that the results are not affected by the last minutes of the game.

Table 4: Yellow Cards Home/Away Team, Before/After

	(1) All Games	(2) Minutes≤85	(3) Minutes≤80
Home Team	-0.156***	-0.156***	-0.156***
	(0.021)	(0.021)	(0.021)
After	0.626***	0.626***	0.626***
	(0.038)	(0.038)	(0.038)
Home Team x After	-0.057	-0.057	-0.057
	(0.044)	(0.044)	(0.044)
Home Error	-0.076**	-0.104***	-0.145***
	(0.032)	(0.033)	(0.035)
Away Error	-0.007	-0.044	-0.057
	(0.043)	(0.048)	(0.050)
Home Error x Home Team	0.092**	0.093**	0.108**
	(0.041)	(0.040)	(0.042)
Away Error x Away Team	-0.043	-0.048	-0.042
	(0.057)	(0.060)	(0.062)
Home Error x After	0.262***	0.304***	0.340***
	(0.071)	(0.068)	(0.074)
Away Error x After	0.223***	0.225***	0.222***
	(0.068)	(0.067)	(0.070)
Home Error x Home Team x After	-0.221***	-0.226***	-0.240***
	(0.077)	(0.077)	(0.086)
Away Error x Home Team x After	-0.009	0.053	0.045
	(0.103)	(0.107)	(0.112)
Attendance/capacity	0.253**	0.231**	0.224**
	(0.106)	(0.101)	(0.107)
Capacity (per 10,000)	-0.018	-0.025	-0.030
	(0.029)	(0.031)	(0.032)
N	13,764	13,496	13,288
R^2	0.1424	0.1475	0.1507
Mean Dependent variable	0.9568	0.9534	0.9487

Note: Standard errors are clustered by the referee and are in parenthesis.
Stars denote the level of statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
We control for division, season, week, referee and teams dummies.

In Table 5, we exhibit the results of a regression that is similar to the one whose results are exhibited in Table 4, except we run the regression separately for games that are more or less "important" to the teams (Columns (1)-(3)) and "close" and "non-close" games (Columns (4)-(5))

Inspection of the sign and statistical significance of the triple interaction term that is our main variable of interest indicates that referees tend to differentially compensate the home team only in games that are important to the home team. Specifically, columns (1) and (2) of Table 5 show that in games that are less important for the home team the coefficients of our main variables of interest are small and insignificant. However, in column (3) of Table 5 that depicts more important games for the home team, these coefficients are larger in magnitude and also statistically significant at the 5% level.

Column (5) in Table 5 includes only games in which the score of the game was tied at the moment of the first error. Column (4) includes all the remaining games. Inspection of the results presented in Table 5 reveals that our main variable of interest, namely the coefficient of the triple interaction term, is negative in a statistically significant way only in "close" games. The magnitude of the effect is larger in comparison to the size of the effect in Table 4, and the effect is statistically significant at the 1% level.

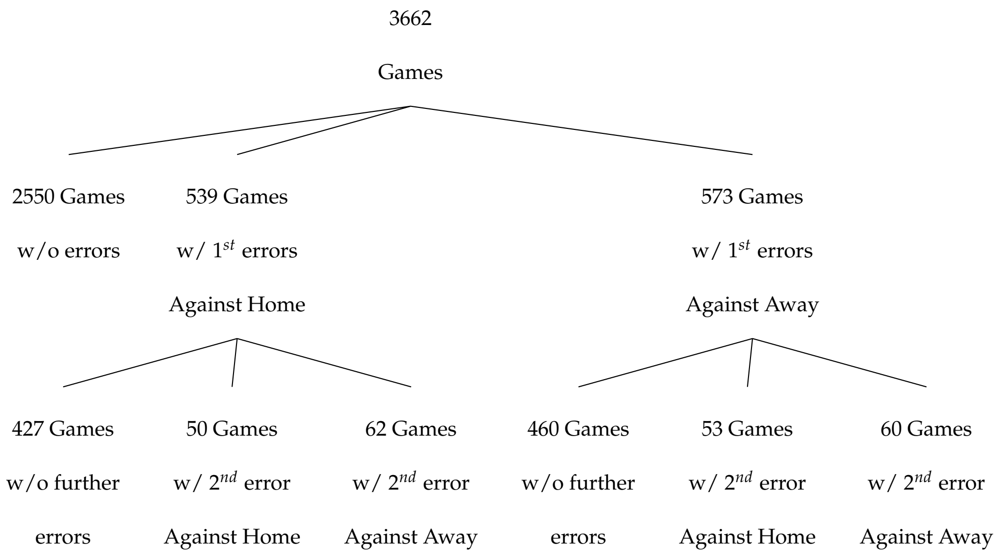
Table 5: Yellow Cards Before/After an Error controlling for Importance/Closeness

	Importance			Closeness	
	Low	Mild	High	No	Yes
Home Team	-0.128***	-0.127***	-0.180***	-0.164***	-0.154***
	(0.043)	(0.047)	(0.028)	(0.045)	(0.024)
After	0.567***	0.667***	0.628***	0.536***	0.651***
	(0.085)	(0.073)	(0.040)	(0.063)	(0.041)
Home x After	-0.120	-0.046	-0.040	-0.049	-0.059
	(0.095)	(0.088)	(0.054)	(0.089)	(0.048)
Home Error	-0.030	-0.077	-0.071*	0.136	-0.117**
	(0.074)	(0.080)	(0.041)	(0.087)	(0.044)
Away Error	-0.031	-0.043	0.003	0.070	-0.041
	(0.131)	(0.087)	(0.059)	(0.077)	(0.053)
Home Error x Home Team	0.096	0.156	0.061	-0.110	0.137**
	(0.083)	(0.127)	(0.071)	(0.104)	(0.055)
Away Error x Away Team	-0.106	-0.005	-0.040	-0.041	-0.043
	(0.141)	(0.106)	(0.081)	(0.131)	(0.057)
Home Error x After	0.186	0.282*	0.283***	0.019	0.315***
	(0.128)	(0.168)	(0.097)	(0.224)	(0.080)
Away Error x After	0.356*	0.287**	0.149	0.248	0.217***
	(0.206)	(0.143)	(0.093)	(0.160)	(0.082)
Home Error x Home Team x After	-0.050	-0.295	-0.250**	0.037	-0.280***
	(0.186)	(0.239)	(0.123)	(0.249)	(0.101)
Away Error x Home Team x After	0.135	-0.055	-0.033	-0.132	0.021
	(0.266)	(0.221)	(0.114)	(0.226)	(0.102)
Attendance/capacity	0.257	0.177	0.062	0.297	0.178^t
	(0.257)	(0.234)	(0.135)	(0.280)	(0.119)
Capacity (per 10,000)	-0.086	-0.030	-0.054	-0.023	-0.019
	(0.074)	(0.072)	(0.034)	(0.066)	(0.035)
N	2,618	3,520	7,626	2,816	10,948
R^2	0.1896	0.1850	0.1527	0.1577	0.1548
Mean Dependent Variable	0.9083	0.9690	0.9679	0.9680	0.9540

Note: Standard errors are clustered by the referee and are in parenthesis.
Stars denote the level of statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
We control for division, season, week, referee and teams dummies.

5.2 Referee Errors

In this subsection, we show that referees do not exhibit home bias in the following sense: they do not make error against the away team following an error against the home team. To test this, we examine whether a referee who made an error against the home team will be more likely to make an error to the away team. The distribution of the order of referees' errors is depicted in the tree-chart below.



Out of the 3662 games in our sample, in 2500 games (68%) there were no errors, in 539 games (15%) the first error was against the Home team, and in 573 games (16%) the first error was against the Away team.

Out of the 539 games in which the first error was against the Home team, in 427 games (79%) [12] there were no additional errors, in 50 games (9%) the second error was against the Home team, and in 62 games (11.5%) the second error was against the Away team. And out of the 573 games in which the first error was against the Home team, in 460 games (80%) there were no additional errors, in 53 games (9%) the second error was against the Home team, and in 60 games (10.5%) the second error was against the Away team. [13]

From the tree-chart above, one can see that errors against home team and against away team distributed quite symmetrically. This pattern seems true both for first and second errors. This suggests that crowd pressure has less influence on referee decisions that are indisputable observable. In particular, out of the 112 games with at

[12] The fact that the proportion of games with no additional errors beyond the first one in the two conditional distributions (79% and 80%) is larger than the proportion of games with no errors at all (68%) is due to the fact that there is less time for the second error to occur, which means that it is less likely to occur.

[13] There are only 39 games with three or more errors. There are four different conditional distributions of the third error (after two errors against the home team, after an error against the home team that is followed by an error against the away team, etc.). The number of games in each one of the relevant categories are small so we do not present them here. However, these games are included in the regression analysis that is performed below.

least two errors, where the First Error was against the home team, in in 55% games the second error was against the away team, and out of the the 113 games with at least two errors, where the First Error was against the away team, in 47% games the second error was against the home team. A simple t-test fail to reject that the null hypothesis that these number are equal to half, suggesting no home bias. However, because a simple t-test cannot control for additional variables, we run the following regression:

$$
\begin{aligned}
2^{nd}Error_{g,i} = \\
&\beta_1 Home\ Team_{g,i} + \beta_2 1^{st} Error\ Against\ Home_g \\
&+ \beta_3 Home\ Team_{g,i} \times 1^{st} Error\ Against\ Home_g \\
&+ \beta_4 Attendance/Capacity_g + \beta_9 Capacity_g \\
&+ \alpha_h + \gamma_a + \delta_s + \eta_d + \theta_w + \lambda_r + \epsilon_{g,i}
\end{aligned}
$$

The dependent variable, *2^{nd} Error*, is a dummy variable equal to one if there was a second error and zero otherwise. *1st Error Against Home* is a dummy variable equal to one if the first error in the game was given against the *Home Team*.

The idea is that public pressure from the crowd on the referee is likely to be larger after an error against the home team because most of the crowd is a crowd of the home team, and that this pressure would be larger if the game is important, and or close (a game is defined as "close" if the difference in game score between the teams is less than or equal than 1 at the time of the first error). We expect that the referee would exhibit a bigger bias when it is subject to stronger pressure. Not finding an indication for such bias even in close and important games indicates that no such bias exists.

Table 6: Second Error by Importance/Closeness

			Importance		Closeness	
	(1)	(2)	(3)	(4)	(5)	(6)
	All	Low	Mild	High	No	Yes
Home Team	-0.0140	-0.0659	0.0154	-0.0195	-0.0463	-0.0065
	(0.0175)	(0.0481)	(0.0362)	(0.0249)	(0.0412)	(0.0219)
1st error to Home	0.0138	0.0073	-0.0098	0.0055	0.0080	0.0188
	(0.0227)	(0.0706)	(0.0384)	(0.0287)	(0.0465)	(0.0271)
Home Team x 1st error to Home	-0.0101	0.0169	0.0022	-0.0097	-0.0037	-0.0117
	(0.0290)	(0.0731)	(0.0386)	(0.0332)	(0.0590)	(0.0373)
Attendance/capacity	0.0336	0.2587	0.1697	-0.1300	0.3344**	-0.0006
	(0.0733)	(0.2532)	(0.1482)	(0.1092)	(0.1379)	(0.0856)
Capacity (per 10,000)	0.0205	0.0398	0.1517***	-0.0488*	0.0772	0.0150
	(0.0253)	(0.0687)	(0.0551)	(0.0264)	(0.0870)	(0.0262)
N	2,226	386	564	1,265	416	1,810
R^2	0.0635	0.3548	0.2845	0.1311	0.3667	0.0832
Mean Dependent Variable	0.1020	0.1166	0.1099	0.0949	0.0529	0.1133

Note: Standard errors are clustered by the referee and are in parenthesis.
Stars denote the level of statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
We control for division, season, week, referee and teams dummies.

Table 6 Column (1) shows the results when all games are included. The results show that our coefficient of interest, namely the coefficient of the interaction term, is small in magnitude and insignificant. Columns (2)-(4) reveal similar results when only "less", "mildly", and "highly" important games are included. In Columns (5) and (6) we rerun the regression separately for games that we define as "close" and not close. Columns (5) and (6) show that regardless of whether the games are "close" or not, our main variable of interest is small in magnitude and insignificant.

The coefficients of our main independent variable of interest, which is the interaction of *Home Team* and *1st Error Against Home* is small and insignificant in all the regressions above, suggesting that referees do not compensate for their errors by making additional errors against the other team. Our results are consistent with our hypothesis that crowd pressure has less influence on referees' decision that are indisputable.

6 VAR and No Crowds

VAR technology was introduced to the Bundesliga in 2017. The table below provides summary statistics for the 1,061 games that were played under VAR, before March 13, 2020, which is the date after which all games were played under closed doors due to Covid 19.

Table 7: Summary Statistics

	Average	Away Team	Home Team	diff	P-value
Number of Goals	3.02	1.34	1.68	-0.34	0.00
	(1.70)	(1.24)	(1.36)		
Number of Referee Errors	0.22	0.12	0.10	0.02	0.21
	(0.47)	(0.34)	(0.32)		
Minutes to First Error	47.80	48.10	47.44	0.66	0.86
	(26.51)	(25.94)	(27.32)		
Number of Games with $\geq$ 1 Errors	205	112	93		
Number of yellow Cards	3.79	2.06	1.73	0.33	0.00
	(2.02)	(1.31)	(1.31)		
Number of Games	1,061				

Note: Standard deviation are in parenthesis.

The table above shows that the presence of VAR decreased the percentage of games with errors to 0.22 compared with 0.38 in the pre-VAR period. Similar to the pre-VAR period referees do not seem to make more errors against the home compared to the away team. The average number of errors to the away team is 0.12, and 0.10 to the home team. A t-test indicates that these two means are not different.

With respect to number of goals and the issuance of yellow cards by referees, the table exhibits bias in favor of the home team. The magnitude of the bias is similar to that observed in the pre-VAR period.

On average there are 3.02 goals in a game, with the away team scoring 1.34 goals, and the Home Team scoring 1.68 goals. A t-test indicates a large and statistically significant difference between these two means (-0.34 with a p-value at the 1% level)

On average, referees issue more yellow cards to the away team than to the home team. Referees issue on average 3.79 yellow cards per game, with 2.06 yellow cards issued to the Away team, and 1.73 to the Home Team. Again this difference is large in magnitude and statistically significant at the 1% level.

On March 13, 2020 with the outbreak of the Covid-19 epidemic, all games played in the Bundesliga were played under closed doors. A total of 163 games were played under these conditions.

The table below shows the summary statistics for the 163 games played in this period:

Table 8: Summary Statistics

	Average	Away Team	Home Team	diff	P-value
Number of Goals	3.03	1.48	1.55	-0.15	0.29
	(1.73)	(1.36)	(1.35)		
Number of Referee Errors	0.21	0.09	0.12	-0.02	0.47
	(0.44)	(0.29)	(0.32)		
Minutes to First Error	53.16	52.54	53.58	-1.04	0.90
	(45.45)	(42.998)	(39.62)		
Number of Games with $\geq$ 1 Errors	32	13	19		
Number of yellow Cards	4.33	2.09	2.24	-0.06	0.71
	(2.05)	(1.22)	(1.47)		
Number of Games	163				

Note: Standard deviation are in parenthesis.

The table above shows that in games that were played under closed doors the home bias disappears. The table shows that on average there are 3.03 goals in a game, with the away team scoring 1.48 goals, and the home team scoring 1.55 goals. A t-test indicates that in games played with no crowd, the rather large difference between these two means in games played with a crowd (regardless of VAR), is not longer large in term of magnitude and significance.

The table also shows that referees do not favor the home team by issuing less Yellow cards to home team compared to the away team. Referees issue an average of 4.33 yellow cards per game, with 2.09 yellow cards issued to the Away team, and 2.24 to the Home team. Again, unlike in games with a crowd, in games played under closed doors the difference between the number of yellow cards issued to the home team and the number of yellow cards issued to the away team is small in magnitude and is insignificant. Note: in games with no crowd the average number of yellow cards in a game is larger than in games with crowd (4.33 compared to 3.88 (without VAR) and 3.79 (with VAR)). A t-test comparing between these mean can reject the null hypothesis that these means are the same in 1% significant level.

In the appendix we rerun most the regressions that we ran in the pre-VAR period, once for the VAR period and once for no crowd period. The results in Table A1 shows that in the VAR period the results are similar, both in magnitude and significance to the pre-VAR period. We get no bias with respect to errors but large and significant home bias with respect to yellow cards. However, in the no crowd period we get that the home bias that existed previously with respect to yellow cards disappears.

Table A2 looks at first error. The table shows that there is not home bias with respect to awarding the first error, not in the VAR period, where errors are fewer, and, as expected, in games with no crowds.

In Table A3 we rerun the regressions that appear in Table 4 for the two periods: VAR and No Crowd. The results are somewhat different. In both the VAR period and the No Crowd period the coefficients on our main variable of interest, which are the triple interaction of ($Home\ Error_{g,i} \times Home\ Team_{g,t} \times After_{g,t}$) and ($Away\ Error_{g,i} \times Home\ Team_{g,t} \times After_{g,t}$), suggest that referee might compensate the team for which they made an error against.

In all of the regression the coefficients on the triple interaction of ($Home\ Error_{g,i} \times Home\ Team_{g,t} \times After_{g,t}$) is negative and larger in magnitude compared to the pre-VAR period (which has a value that is about half the size). In the VAR period these coefficients are significant and in the No Crowd period they lose significant (we note that the number of games played during the No Crowd period is much smaller and under the VAR technology the number of games with error is even smaller).

The table also shows that the coefficients on the triple interaction $Away\ Error_{g,i}$) $\times Home\ Team_{g,t} \times After_{g,t}$) is positive and large in magnitude compared to the pre-VAR period (0.164-0.251 in the VAR period and 0.303 in the No Crowd period as compared to -0.009-0.053 in the pre-VAR period). However, none of the coefficients is statistically significant. It is important to keep in mind the fact that the number of games in this period is much smaller compared to the pre-VAR period. We view this increase in magnitude as suggestive that there is some compensation happening.

7 Conclusion

There is a long standing concern about the effects of public pressure on judging. Our objective in this paper has been to contribute to the empirical investigation of this subject. In particular, we have focused on investigating the circumstances in which public pressure is more and less likely to affect judging. To this end, we have used detailed and rich data from Germany's top soccer league.

A key insight of our analysis is that the extent to which public pressure affects judging depends on the extent to which judging decisions seeking to placate public pressure can be indisputably identified as erroneous by outside observers and thereby impose a reputational cost on the decision maker. Another key insight is that with respect to those decisions where public pressure affects judging, the strength of the effect depends on the extent to which such pressure is viewed by the decision maker as understandable or reasonable. We hope that future empirical work will further study the circumstances in which public pressure is more and less likely to affect judging.

References

[1] Berdejó, Carlos, and Noam Yuchtman (2013), "Crime, Punishment, and Politics: An Analysis of Political Cycles in Criminal Sentencing," *Review of Economics and Statistics*, 95(3), 741–756., doi:10.1162/rest_a_00296.

[2] Boeri, T., Severgnini, B. (2011), "Match rigging and the career concerns of referees," *Labour Economics*, 18(3), 349-359.

[3] Buraimo, B., Forrest, D., Simmons, R. (2010),"The 12th man?: refereeing bias in English and German soccer," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173(2), 431-449.

[4] Cohen, A., Klement, A., Neeman, Z. (2014), "Judicial Decision Making: A Dynamic Reputation Approach," *Journal of Legal Studies* 44, 133-159.

[5] Dawson, P., Dobson, S. (2010),"The influence of social pressure and nationality on individual decisions: Evidence from the behaviour of referees," *Journal of Economic Psychology*, 31(2), 181-191.

[6] Dawson, P., Dobson, S., Goddard, J., Wilson, J. (2007), "Are football referees really biased and inconsistent?: evidence on the incidence of disciplinary sanction in the English Premier League," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(1), 231-250.

[7] Dohmen, T. J. (2008), "The influence of social forces: Evidence from the behavior of football referees," *Economic Inquiry*, 46(3), 411-424.

[8] Garicano, L., Palacios-Huerta, I., Prendergast, C., 2005. Favoritism under social pressure," *Review of Economics and Statistics*, 87(2), 208-216.

[9] Gomery, John H. (2006), "The Pros and Cons of Commissions of Inquiry, " *McGill Law Journal*, 51(4), 783-798.

[10] Huber, Gregory A., and Sanford C. Gordon (2004), "Accountability and Coercion: Is Justice Blind When It Runs for Office?" *American Journal of Political Science*, 48(2), 247-263., doi:10.2307/1519881.

[11] Marder, Nancy S. (2014), "Jurors and Social Media: Is a Fair Trial Still Possible," *Southern Methodist University Law Review* , 67(3), 617-668.

[12] Nevill, A. M., Newell, S. M., Gale, S. (1996), "Factors associated with home advantage in English and Scottish soccer matches," *Journal of Sports Sciences*, 14(2), 181-186.

[13] Gavin Phillipson (2008), "Trial by the Media: The Betrayal of the First Amendment's Purpose," *Law & Contemporary Problems*, 71(15), 15-29.

[14] Shepherd, Joanna M. (2011), "Measuring Maximizing Judges: Empirical Legal Studies, Public Choice Theory, and Judicial Behavior," *University of Illinois Law Review*, 2011, 1753-1766.

[15] Steblay, Nancy Mehrkens, Jasmina Besirevic, Solomon M. Fulero, and Belia Jimenez-Lorente (1999), "The Effects of Pretrial Publicity on Juror Verdicts: A Meta-Analytic Review," *Law and Human Behavior*, 23(2), 219–235. doi:10.1023/a:1022325019080.

[16] Sutter, M., Kocher, M. G. (2004), "Favoritism of agents–the case of referees' home bias," *Journal of Economic Psychology*, 25(4), 461-469.

Appendix

Table A1: Number of Errors/Yellow Cards by Importance

	VRA		No Crowd	
	Errors	Yellow Cards	Errors	Yellow Cards
Home Team	0.0064	-0.3718**	0.0270	-0.0811
	(0.0319)	(0.1585)	(0.0947)	(0.3826)
Mild	0.0214	-0.0190	-0.0165	0.3167
	(0.0276)	(0.1531)	(0.0793)	(0.3556)
High	0.0405	0.0358	0.0144	0.3877
	(0.0257)	(0.1500)	(0.0830)	(0.4113)
Home Team x Mild	0.0122	0.1238	0.0187	0.3597
	(0.0383)	(0.1888)	(0.1420)	(0.4944)
Home Team x High	-0.0375	-0.0330	-0.0140	0.2607
	(0.0343)	(0.1831)	(0.1321)	(0.4380)
Crowding	0.0792	0.0910		
	(0.1106)	(0.3908)		
Stadium Size (per 10,000)	0.0000	0.0000		
	(0.000)	(0.000)		
N	1,994	1,994	326	326
R^2	-0.0040	0.0850	-0.0137	0.1308
Mean Dependent Variable	0.1098	1.8921	0.1043	2.1718

Note: Standard errors are clustered by the referee and are in parenthesis.
Stars denote the level of statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
We control for division, season, week, referee and teams dummies.

Table A2: First Error by Importance

	VAR		No Crowd	
	All Games	Games with $\geq$ 1 errors	All Games	Games with $\geq$ 1 errors
Home Team	-0.0000	-0.0000	0.0270	0.1429
	(0.0365)	(0.2308)	(0.0947)	(0.7978)
Mild	0.0077	0.0438	-0.0201	-0.2857
	(0.0265)	(0.1833)	(0.0881)	(0.9192)
High	0.0329	0.0380	0.0204	0.0000
	(0.0225)	(0.1113)	(0.0853)	(.)
Home Team x Mild	0.0249	-0.0717	0.0332	0.2338
	(0.0376)	(0.2642)	(0.1515)	(1.4384)
Home Team x High	-0.0340	-0.0827	0.0043	-0.0000
	(0.0388)	(0.2288)	(0.1337)	(1.1665)
Crowding	0.0651	0.0008		
	(0.0962)	(0.0588)		
Stadium Size (per 10,000)	0.0000	0.0000		
	(0.000)	(0.000)		
N	1,994	390	326	64
R^2	0.0625	0.0053	0.3311	0.0519
Mean Dependent Variable	0.0978	0.5000	0.0982	0.5000

Note: Standard errors are clustered by the referee and are in parenthesis.
Stars denote the level of statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
We control for division, season, week, referee and teams dummies.

Table A3: Yellow Cards Home/Away Team, Before/After

	VAR			No Crowd		
	All Games	Minutes≤85	Minutes≤80	All Games	Minutes≤85	Minutes≤80
Home Team	-0.163***	-0.163***	-0.163***	-0.038	-0.038	-0.038
	(0.031)	(0.031)	(0.031)	(0.111)	(0.112)	(0.112)
After	0.608***	0.608***	0.608***	0.405**	0.405**	0.405**
	(0.043)	(0.043)	(0.043)	(0.154)	(0.154)	(0.154)
Home x After	0.011	0.011	0.011	0.122	0.122	0.122
	(0.067)	(0.067)	(0.067)	(0.215)	(0.215)	(0.215)
Home Mistake	-0.138^t	-0.177*	-0.200**	-0.363	-0.465*	-0.465*
	(0.084)	(0.090)	(0.097)	(0.292)	(0.266)	(0.266)
Away Mistake	-0.053	-0.085	-0.118	-0.677***	-0.701***	-0.701***
	(0.083)	(0.089)	(0.094)	(0.225)	(0.218)	(0.218)
Home Mistake x Home Team	0.073	0.091	0.063	0.570*	0.608*	0.608*
	(0.134)	(0.142)	(0.153)	(0.286)	(0.312)	(0.312)
Away Mistake x Away Team	0.027	0.007	0.047	0.102	0.102	0.102
	(0.098)	(0.100)	(0.108)	(0.189)	(0.189)	(0.189)
Home Mistake x After	0.531***	0.640***	0.546***	0.702^t	0.911**	0.911**
	(0.155)	(0.164)	(0.173)	(0.421)	(0.404)	(0.404)
Away Mistake x After	0.063	0.067	0.101	0.369	0.369	0.369
	(0.154)	(0.153)	(0.147)	(0.422)	(0.422)	(0.422)
Home Mistake x Home Team x After	-0.412^t	-0.589**	-0.480*	-0.443	-0.767	-0.767
	(0.250)	(0.244)	(0.261)	(0.646)	(0.577)	(0.577)
Away Mistake x Home Team x After	0.176	0.251^t	0.163	0.303	0.303	0.303
	(0.178)	(0.171)	(0.169)	(0.454)	(0.454)	(0.454)
N	4,208	4,124	4,084	650	640	640
R^2	0.1683	0.1724	0.1708	0.2982	0.3067	0.3067
Mean Dependent Variable	0.9427	0.9411	0.9347	1.0839	1.0851	1.0851

Note: Standard errors are clustered by the referee and are in parenthesis.

Stars denote the level of statistical significance:t $p < 0.15$ * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

We control for division, season, week, referee and teams dummies.

Table A4: Second Error in Games with At Least One Error

	(1) VAR	(2) No Crowd
Home Team	0.0192	0.0000
	(0.0360)	(0.0000)
1st Mistake to Home	0.0119	0.0000
	(0.0401)	(.)
Home Team x 1st Mistake to Home	-0.0522	-0.1111
	(0.0419)	(0.1310)
Crowding	-0.3138	
	(0.2876)	
Stadium Size (per 10,000)	0.0000	
	(0.000)	
N	390	62
R^2	0.3227	0.5407
Mean Dependent Variable	0.0590	0.0323

Note: Standard errors are clustered by the referee and are in parenthesis.
Stars denote the level of statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
We control for division, season, week, referee and teams dummies.