THE INFLUENCE OF SOCIAL FORCES: EVIDENCE FROM THE BEHAVIOR OF FOOTBALL REFEREES

THOMAS J. DOHMEN*

Analyzing the neutrality of referees during 12 German premier league (1. Bundesliga) soccer seasons, this paper documents evidence that social forces influence agents' decisions. Referees, who are appointed to be impartial, tend to favor the home team by systematically awarding more stoppage time in close matches in which the home team is behind. They also favor the home team in decisions to award goals and penalty kicks. Crowd composition affects the size and the direction of the bias, and the crowd's proximity to the field is related to the quality of refereeing. (JEL J00)

I. INTRODUCTION

Do social forces affect decisions or actions of individuals? And if so, under what circumstances do social influences determine socioeconomic behaviors? These are important questions for economists to answer. Sociologists and social psychologists widely acknowledge that individuals' decisions are not only governed by their material payoffs but also influenced by nonmaterial social payoffs that arise in the decision makers' social environment, for example, in the form of social approval or social sanctions, as argued, for example, by Asch (1951) or Coleman (1990). This type of social pressure can cause individuals to make decisions that accommodate the preferences of a social group even if they are not in accordance with the decision maker's own interest. I refer to this view as the "social pressure hypothesis."

Economists have built on the idea that social forces might affect individual behavior and have developed models of social interaction in which the quest for social rewards or the avoidance of social sanctions can explain adherence to social custom as, for example, in Akerlof (1980) as well as the evolution and persistency of social norms as in Bernheim (1994).[1] In these models, social payoffs, which comprise social approval or sanctions, become an argument in the utility function in addition to intrinsic consumption utility, which captures material payoffs. Several other scholars (e.g., Becker and Murphy 2000) have used such a framework that generalizes an individual's utility function to internalize social payoffs

Dohmen: Professor of Economics, Research Centre for Education and the Labour Market (ROA), Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands. Phone +31-43 388 3647, Fax +3143 388 4914, E-mail dohmen@iza.org

1. Akerlof (1980) argues that social sanctions can induce an individual to adhere to a social custom even at the cost of forgoing pecuniary gains from breaking it. In Bernheim's (1994) model, individuals care about social rewards in addition to their intrinsic consumption utility, and conform to a single homogenous standard of behavior if social rewards are sufficiently relative to their intrinsic utility. Adherence to a norm and persistency of a norm are determined by the distribution of intrinsic preferences.

ABBREVIATION

DFB: German Football Association

411

to analyze how social influences affect a variety of socioeconomic behaviors, consumption, status, as well as the evolution of norms, fads, and fashions.[2]

Despite much theoretical progress, there is little empirical evidence that convincingly assesses the role of social influences, not least because data that meet the necessary requirements are scarce.[3] A useful type of data would record how a decision maker behaves in different social environments, where different social groups have well-defined, and potentially conflicting, interests that are not aligned with the interests of the individual decision maker. Garicano, Palacios-Huerta, and Prendergast (2005) exploit the fact that soccer matches offer such a setting and provide one of the few pieces of empirical evidence. In a soccer match, it is in the referee's private interest to be impartial while fans in each camp derive utility from their team's success and therefore have an interest to work toward their common goal by sanctioning referee decisions that do not favor their preferred team and by approving favorable decisions. Analyzing data from two seasons of the Spanish premier soccer league, the *Primera Division*, Garicano, Palacios-Huerta, and Prendergast (2005) detect that Spanish referees favor the home team by prolonging the match by almost 2 min when the home team is one goal behind at the end of regulation time compared to the situation in which the home team is leading by one goal. They also investigate whether crowd size and the ratio of attendance-to-capacity matters and find that a one standard deviation increase in attendance increases the bias by about 20%, while a higher attendance-to-capacity ratio reduces the bias. They conclude that nonmonetary incentives, in particular social pressure from the crowd, cause the pref-

erential treatment.[4] Sutter and Kocher (2004) report corroborative findings based on data from the 2000 to 2001 season in the German premier soccer league (*1. Bundesliga*), although they do not assess whether crowd attributes affect the magnitude and significance of home-biased refereeing.

This paper provides complementary evidence of referee bias based on data from 3,519 games of the *1. Bundesliga*, which supports the view that the social environment can affect individual's decisions. The empirical analysis that professional referees, who are appointed and paid by the German Football Association (DFB) and are expected to be impartial, in fact systematically favor the home team. Favoritism is manifested in stoppage time decisions and in decisions to award goals and penalty kicks. The data also provide new evidence that crowd characteristics such as crowd composition and distance to the soccer field impair referees' decisions in a way that is consistent with the social pressure hypothesis, that is, that social forces influence individual behavior. I find that the size of the bias depends on the composition of the crowd: the home bias tends to be smaller when more supporters of the visiting side attend the match. This is consistent with the idea that social approval and social sanctions have countervailing effects on net social rewards. We expect supporters of each side, who have the common interest that their preferred team achieves success, to work toward this common goal by acclaiming favorable decisions of the referee and by expressing dissatisfaction with unfavorable referee decisions. Referees' decisions hence evoke social approval from supporters of the favored team and social sanctions from the opponent side. A referee who is not inherently biased, that is, who does not derive intrinsic utility from a particular match outcome and values social payoffs, is expected to weight the social costs and benefits.

---

2. Dufwenberg and Lundholm (2001) argue that social rewards mitigate moral hazard problems in the context of unemployment insurance as the desire for social respect induces higher search effort; Austen-Smith and Fryer (2005) formulate a theory of peer effects; and Prendergast and Topel (1996) show theoretically how social influences can create distortions in an agency model.

3. Empirical work on social interactions is plagued by identification problems, which arise due to the complex relationships in interaction-based models. Consequently, data requirements for empirical testing are high. Brock and Durlauf (1999) review the empirical work on interaction-based models and provide a detailed discussion of the problems that arise in this literature.

4. Other studies have provided evidence that reveals innate bias, that is, intrinsic preferences for discrimination, among arbitrators who should be impartial. For example, Goldin and Rouse (2000) analyze data from auditions of symphony orchestras and find suggestive evidence for sex-biased hiring decisions. Knowles, Persico, and Todd (2001) derive a test that allows to distinguish between statistical discrimination and prejudice (i.e., a taste for discrimination); their empirical analysis uncovers clues for prejudiced racial discrimination if particular assumptions about individuals' utility functions are made.

Strikingly, home team favoritism is found to be stronger when the match takes place in a stadium without a running track, that is, when the crowd is physically closer to the field and to the referee, in which case the intensity of social pressure is arguably higher.[5] This finding lends support for the conjecture that social forces influence the referees' decision, be it because social pressure from the crowd directly triggers-biased refereeing or because of a more oblique channel, in which, for example, the crowd creates an atmosphere that encourages the players on the field to exert pressure on the referee. Since the nature of biased refereeing is affected by the crowd's proximity to the field, we can dismiss an alternative mechanism that leads to home-biased decisions, namely that the DFB condones the preferential treatment of home teams or even instructs its referees to favor home teams. Even a soccer association with strong preferences for nondiscriminatory competition might rationally accept home-biased refereeing as long as the referees' preferential treatment was exactly the same for all home teams such that the resulting home advantage would balance out over the season and therefore would not affect the outcome of the championship. This could be an optimal policy to maximize gate revenues if attendance was boosted when the home team is more likely to win, as Garicano, Palacios-Huerta, and Prendergast (2005) discuss. However, given the fact that teams who play in stadiums with an athletics track are affected differently than teams who play in a stadium without a track, this mechanism is implausible.

In fact, the DFB monitors the quality of refereeing and sacks a referee if he is detected to be biased so that being partial lowers reappointment probabilities.[6] Since remuneration amounted to more than 3,000 Euros (~US$4,000) per match at the end of the observation period, biased refereeing entails substantial expected pecuniary losses for umpires. This suggests that referees are induced to favoritism by social forces, although they have a strong intrinsic motivation for impartiality. In summary, the findings indicate that social groups can influence individuals to work toward achieving an outcome desired by that group (and to adhere to the groups' social norm) even if individuals do not derive utility from the outcome and adherence to the norm is diametrically opposed to their own private interest.

The remainder of the paper is organized as follows. Section II provides some institutional background and describes the data. Section III presents the empirical results and Section IV concludes.

II. BACKGROUND AND DATA

The data were made available for scientific use by *IMP—Innovative Medientechnik und Planung AG*, a company that maintains the official soccer database of the *DFL Deutsche Fußball Liga GmbH*, the German soccer league association, and cover all 3,519 matches that were played in the German premier soccer league (*1. Bundesliga*) since the start of the season 1992/1993 until the first half of the season 2003/2004. IMP AG sends several observers to each league match who record about 2,000 actions per match, including all goals, shots on goal, tackles, passes, corner kicks, every single ball contact, yellow and red cards, as well as the number of injury treatments on the field. Various match statistics are provided separately for both periods of the match, including the amount of stoppage time in each half. The data also record the date, destination, and outcome of the match, the number of spectators, and the referee's identity.

The *1. Bundesliga* consists of 18 teams that compete for the national soccer championship. Teams play each other twice a season, once during rounds 1–17 and once during rounds 18–34 when the status of home and visiting team is reversed.[7] The outcome of a match (i.e., a win, a draw, or a loss) determines the number of points that are allocated to the teams: no points are allocated to the loser of a match, while each team receives one point in case of a draw. The reward for winning a match was raised from two to three points at the start of the season 1995/1996. The accumulated number of points during a season

---

5. Festinger, Schachter, and Back (1950) find that physical distance is a key determinant of the intensity of social interaction.

6. This was confirmed by the DFB upon request. The DFB official was reluctant to state why particular referees were not reappointed in the past.

7. A round consists of nine matches, which are typically played on weekends.

determines the league ranking, and the team that finishes the season at the top of the ranking (i.e., the team that has accumulated most points) wins the German soccer championship.[8] The three lowest ranking clubs of the championship table are relegated to the second division (*2. Bundesliga*) and are replaced by the three highest ranking teams from the second division.[9]

The referees for *1. Bundesliga* matches are appointed by the DFB and receive a piece rate for refereeing. This piece rate has risen over time: from July 1992 (i.e., at the start of the observation period) until July 1997, referees received 2,500 German Marks (DM) per match; from August 1997 until July 2000, they were paid 4,000 DM per match; and since August 2000, the reward is 6,000 DM or 3,067.75 Euros per match. In addition, travel expenses (including hotel and transportation) have been fully covered during the entire period since 1992. *Bundesliga* referees are experienced and have been selected in sequential promotion tournaments. After having passed a written and a physical test, referees typically start in the lowest division. Once they have been promoted to referee in the *Landesliga*, the sixth division, they can be promoted at most one division each year if judged as qualified by official observers. In total, 73 umpires have refereed at least 1 of the 3,519 games in our sample. The number of referees in the *1. Bundesliga* has been limited to 22 since 1995. Previously, the DFB had appointed up to 36 referees each season.

The performance of referees is monitored and judged by an official observer of the DFB referee committee, who attends the match in the stadium and fills a performance evaluation form afterward. The first item in this evaluation addresses the referee's appearance during the game, in particular whether the referee was decided, secure, and had the courage to take unpopular decisions, or whether he was impressed by complaining players.[10] The second and third points of the evaluation form concern how well the referee interpreted and implemented the *Laws of the Game* and additional instructions. This covers the adequacy of the referee's allowance for time lost (see Linn 2003).

Information on the frequency of match interruptions for injury treatments on the field makes it possible to control for the "appropriate" amount of stoppage time, which is a key determinant of stoppage time according to the official rules, the *Laws of the Game*, which are authorized by the International Football Association Board and established by the Fédération Internationale de Football Association. According to Law 7 of the *Laws of the Game* (Fédération Internationale de Football Association 2002, 19), "[a]llowance is made in either period for all time lost through: substitution(s), assessment of injury to players, removal of injured players from the field of play for treatment, wasting time, any other cause. The allowance for time lost is at the discretion of the referee." Data on the number of corner kicks, goal kicks, and yellow or red cards allow to control for events that according to Linn (2003) might result in "wasting time" by impeding a quick restart of play.

Information on the correctness of decisions to award penalty kicks is available for all 857 penalty kicks that were called since the start of the season 1993/1994 (i.e., when IMP started to collect this information), while there is information on decisions to award goals for all but 3 of the 10,166 goals that had been awarded since the start of the season 1992/1993. Data on decisions not to award a goal have only been collected since 1993 and cover all but one of the 463 situations in which a goal was not accepted. Information on the decision not to award a penalty kick is only recorded since the start of the season 1998/1999. Until January 2004, 892 critical situations, in which suspicious action took place in one of the penalty areas, had been evaluated. Specialists of IMP AG assess the correctness of these decisions after each match, relying on video recordings and considering about 30 pieces

of detailed information in their judgment. A decision is classified as "correct" if the evaluators consent that it was correct. Likewise, a decision is labeled "wrong" if they agree that it was wrong, and the decision is classified as "disputable" if their verdict is not unanimous.

III. RESULTS

The empirical analysis starts with an investigation of stoppage time decisions, which proceeds in three steps. Using data on 12 soccer seasons and including additional control variables for match interruptions like the number of fouls and the number of actual injury treatments on the field, I first replicate the analysis by Garicano, Palacios-Huerta, and Prendergast (2005) to assess whether there is evidence for referee bias in the German data. If the social pressure hypothesis is correct, we expect that referees, in the pursuit of social rewards from the home crowd, award more stoppage time when the home team is behind in score. We also expect this bias to be stronger when more is at stake, for example, when the score margin is close or when second half rather than first-half stoppage time is concerned. In these cases, the expected marginal effect of additional time on the match outcome is larger, and consequently, crowd pressure is likely to be stronger. A model of social interaction, in which material consumption and social payoffs are substitutes (e.g., Bernheim 1994), predicts that deviation from intrinsic goals is less pronounced when social rewards become relatively less important. In a second step, I analyze whether the composition of the crowd matters, an issue that Garicano, Palacios-Huerta, and Prendergast (2005) also addressed. Since supporters of the home team and the visiting team have conflicting goals, we should expect a larger share of visiting team supporters to attenuate the home bias if the social pressure hypothesis is correct because one group's social approval countervails the other group's social sanctions.

In the third step, I study a new aspect of crowd attributes, namely whether physical proximity affects the size of the bias. Given the presumption that the intensity of crowd pressure is higher if the distance between the crowd and the referee is smaller, we hypothesize that the home bias becomes stronger when the crowd is closer to the field. Since stoppage time decisions account for only one way among many others in which the referee can favor a particular team, I investigate referees' impartiality concerning another crucial set of decisions in the game, namely decisions to approve goals and to award penalty kicks.

A. Stoppage Time Decisions

The first piece of evidence for biased stoppage time decisions is presented graphically in Figure 1, which plots kernel density estimates of second-half stoppage time distributions conditional on the score difference (defined as the number of goals scored by the home team minus the number of goals scored by the visiting side).[11] It is apparent that the amount of stoppage time awarded at the end of 90 min regulation time depends on the score difference and on the identity of the team. Most stoppage time is awarded in close matches when the score margin is 1. Referees also seem to use their discretionary power to systematically award more stoppage time to the home team in close matches as the estimated stoppage time distribution conditional on the home team being one goal behind stochastically dominates all other estimated conditional stoppage time distributions.[12]

When the match is decided at the end of regulation time, that is, when one team is ahead by two goals or more, the amount of stoppage time does not depend on the identity of the leading team and less time is added. Interestingly, supporters have much weaker incentives to influence the referee in decided matches in which the ultimate match outcome is unlikely to change during stoppage time. Hence, the raw data are consistent with the hypothesis that referee's decisions are affected by social forces. In the following analysis of stoppage time decisions, I concentrate on close games to further investigate the forces that lead referees to make biased decisions.

11. Garicano, Palacios-Huerta, and Prendergast (2005) proposed this classification of matches. Duggan and Levitt (2002) follow a similar approach of looking at the "relevant margin" in their study of corruption in Sumo wrestling tournaments.

12. The results documented in Figure 1 are largely in accordance with Garicano, Palacios-Huerta, and Prendergast (2005)'s findings except for the fact that Spanish referees seem to award more stoppage time when a match is drawn than when the visiting team is behind by one goal; the opposite is true in the German data.

FIGURE 1

Kernel Density Estimates of the Second-Half Stoppage Time for
Various Score Differences
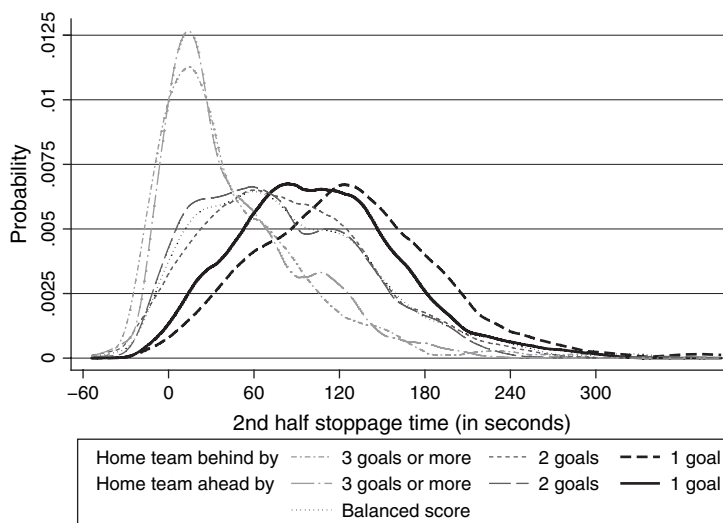
Note: The figure shows kernel density estimates of second-half injury time distributions conditional on the score line at the end of regulation time.

Estimates from ordinary least squares regressions reported in Panel A of Table 1 confirm the descriptive results that more stoppage time is granted when the home team is one goal behind than in matches in which the home team is one goal ahead at the end of second-half regulation. The home bias in awarding stoppage time, which is measured by the coefficient on the "Home Ahead" dummy that takes the value of 1 if the home team is one goal ahead and 0 if the home team is one goal behind at the end of the first half, amounts to about 22 sec of additional stoppage time.[13] This difference in stoppage time is statistically significant at the 1% level, even controlling for factors that should affect the amount of stoppage time according to the *Laws of the Game*, such as the number of

injury treatments, substitutions, and cards (Column 1). Actual injury treatments on the field have the expected positive effect on the duration of stoppage time, but the effect is small; referees add about 6 sec of extra time for each time a player has been yellow carded.[14] The home bias remains statistically significant and its size is virtually unchanged when allowing for season-specific effects and controlling for other potentially confounding factors such as the relative number of shots on goal, fouls, corner kicks, crosses, and other proxies for relative strength (Table 1, Column 2, and table notes for

---

13. On average, 0.05 goals are scored during each minute of stoppage time so that the impact of the bias might seem small. However, this can be argued since goals scored in the last seconds of a match might have enormous consequences. For example, FC Bayern München won the German championship in 2001 due to a goal they scored during injury time in the last match of the season.

14. Very similar coefficient estimates, which are reported in Table 3 of the working paper version (Dohmen 2005), are found in drawn matches. Interestingly, referees also tend to favor the home team by lengthening drawn matches in which the home team is more likely to score next. In addition, games end about 10 sec earlier on average when neither team had scored during regulation time and matches with more shots on goal, more tackles, and more crosses last longer. This indicates that referees lengthen more exciting matches, potentially being influenced by spectators who like suspense.

TABLE 1
Length of Stoppage Time in Close Games

	Panel A: Second-Half Stoppage Time			Panel B: First-Half Stoppage Time		
	(1)	(2)	(3)	(4)	(5)	(6)
Home ahead	−22.142*** (4.168)	−20.601*** (4.501)	−21.660*** (4.478)	−6.11* (3.651)	−7.285* (3.828)	−7.833** (3.786)
No. of treatments	−0.187 (0.219)	1.364 (0.460)	1.262 (0.447)	0.781*** (0.156)	1.201*** (0.330)	1.317*** (0.369)
No. of substitutions	1.915 (1.168)	1.562 (1.602)	2.172 (1.421)	18.396*** (2.406)	18.131*** (2.197)	17.89*** (2.214)
No. of fouls	0.284 (0.251)	−0.819** (0.360)	−0.558* (0.311)	0.058 (0.172)	−0.066 (0.258)	0.046 (0.260)
No. of yellow cards	6.667*** (0.998)	6.041*** (0.929)	6.118*** (1.162)	5.754*** (1.160)	5.926*** (1.129)	5.818*** (1.176)
No. of 2nd yellow cards	3.694 (3.952)	2.871 (3.899)	1.056 (4.231)	15.433*** (5.816)	13.989** (5.822)	11.154* (6.459)
No. of red cards	11.238* (6.527)	12.703** (5.749)	12.519** (5.176)	27.724*** (6.943)	31.182*** (7.129)	33.767*** (7.315)
Season dummies	No	Yes	Yes	No	Yes	Yes
Controls for relative strength	No	Yes	Yes	No	Yes	Yes
Fixed effects for home team, visiting team, and referee	No	No	Yes	No	No	Yes
Constant	91.178*** (11.264)	74.724*** (14.831)	73.615*** (16.947)	34.52*** (7.020)	52.914*** (11.366)	63.567*** (16.483)
Observations	1,166	1,117	1,117	1,495	1,403	1,403
R^2	.07	.15	.31	.10	.13	.26

Notes: Ordinary least squares estimates. The dependent variable in Panel A is the length of second-half stoppage time (in seconds) in matches in which one team was leading by one goal after regular time. The dependent variable in Panel B is the length of first-half stoppage time (in seconds) in matches in which one team was leading by one goal after regular first-half playing time. The variable Home ahead takes the value 1 if the home team is one goal ahead and 0 if the home team is one goal behind. All listed independent variables except for the total number of corresponding events in the respective half of the match except for "No. of fouls," for which separate information for each half is not available. Controls for relative strength include the relative number of tackles won, shots on goal, fouls, corner kicks, and crosses as well as the absolute difference in rankings before the match and the ranking of the home team. These variables are jointly not significant in the regression models of Panel A (p value corresponding to Wald test statistic exceeds .7). The rank of the home team is significant at the 5% level in Specification (5) of Panel B. Information on relative strength is missing for the majority of matches in the season 1992/1993, which causes the number of observations to drop when relative strength is controlled for. Standard errors, given in parentheses, allow for correlation between observations of the same referee.

*Significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

details) as well as team- and referee-fixed effects (Column 3).[15]

Pressure from the crowd is arguably less intense at the end of the first half when fans might still have hope and faith that their team at least levels the score in the second half. This gives rise to the hypothesis that first-half stoppage time decisions are less biased, but that delays of the match, which should be compensated for by stoppage time according to the official rules, have the same impact on the length of stoppage time in both halves. The regression results in Panel B of Table 1 support this hypothesis: the home bias in awarding stoppage time is generally much smaller in the first half compared to the second half. Explanatory variables that capture match delays, in contrast, have similar effects in both halves.[16]

The systematic differential treatment of home and visiting teams in second-half stoppage time decisions is less pronounced in the German *Bundesliga* than in the Spanish *Primera Division*, where home teams are granted about 113 sec more of stoppage time in close matches (Garicano, Palacios-Huerta, and Prendergast 2005). Home teams might be favored less in Germany because relatively more visiting supporters accompany their team to away matches in Germany, where the opponents' home cities are typically geographically less distant than in Spain. This conjecture entails that countervailing social forces mitigate referee bias. Since the data do not contain information about crowd composition, I construct two proxy variables that

exploit the fact that supporters are more likely to accompany their team to away matches the shorter the traveling distance is. The first variable proxies the traveling distance of fans who live close to their home team's city by the linear distance between opponents' home cities and is calculated based on coordinates.[17] Fans living close to their preferred team's home city account for a substantial part of supporters, as is evident by the geographical distribution of fan clubs of German soccer teams. The second proxy variable is a measure of relative popularity and accounts for traveling distance of another group of visiting team's supporters, namely those who live close to the away match location (and potentially further away from their team's home city). Since more popular teams have more fans nationwide, the number of fans who live close to any match location is positively correlated with a team's popularity, which I measure by the attendance-to-capacity ratio in away matches.[18] The relative popularity index is calculated as the ratio of the average attendance-to-capacity ratio in the visiting team's away matches to the average attendance-to-capacity ratio in the home team's home matches. Since the linear distance and the attractiveness of the visiting team have a positive and statistically significant effect on match attendance, we can be confident that the two proxy variables capture variation in the composition of the crowd.[19]

15. Note that the regression results indicate that a large fraction of the variation in stoppage time remains unexplained in the specification in Column 3, even though the regression includes controls for the host of variables that are explicitly mentioned in the *Laws of the Game*. This might reflect the fact that the amount of time that is wasted varies even for a particular type of match interruption, for example, yellow carding a player, so that control variables are only an imperfect measure of the actual time lost in a particular instance. But the size of the unexplained variation also provides suggestive evidence of the importance of referee discretion in stoppage time decisions.

16. Social pressure exerted by the crowd is likely to increase not only toward the end of the game but also toward the end of the season, when more is at stake so that the difference in stoppage time awarded in close matches is expected to be larger toward the end of the season. While this difference widens by about 10 sec in the last five rounds compared to earlier rounds, this increase is not statistically significant at conventional levels.

17. Linear distances are obtained using the program on http://www.koordinaten.de/online/dist_wel.shtml (last seen on July 11, 2007).

18. The constructed popularity index varies slightly for a given team because observations on matches that involve both opponents are neglected when calculating this measure. Reassuringly, this proxy for popularity is highly correlated with the number of fan clubs of the different teams throughout the country. In fact, the four highest ranked teams according to the constructed popularity index are FC Bayern München, Borussia Dortmund, Schalke 04, and Borussia Mönchengladbach, arguably the most popular teams in Germany, and certainly those with most supporters' clubs.

19. An increase of 0.1 in the popularity index of the away team raises attendance by about 4,900 visitors. The effect of distance is nonlinear: compared to the reference case in which opponents' home towns are 300–450 km apart, 4,638 more spectators are on average attracted if the distance between the opponents' home towns is less than 150 km and 865 additional spectators attend when this distance is between 150 and 300 km, while there is no statistically significant difference in attendance when the distance is even longer. Detailed results are reported in the working paper version (Dohmen 2005).

To assess whether the size of the home bias depends on the composition of the crowd, I estimate the regression model from Column 3 of Table 1 separately for matches in which both teams come from cities that are less than 150 km apart and cities that are further apart. At the same time, I augment the regression model with an interaction term between the score difference and a dummy variable that equals 1 if the visiting team's relative popularity is in the upper third of the distribution. The regression results in Table 2 reveal that the home bias is negligible and statistically not different from zero if the distance between the cities is less than 150 km, unless the visiting team is relatively popular (see Column 1). When the cities are further than 150 km apart, the home bias (about 24 sec of additional stoppage time when the home team is behind in score) is statistically significant (see Column 2).[20] These findings support the social pressure hypothesis.

To investigate next whether the intensity of social pressure, as measured by distance between the crowd and the field, affects the size of the bias, I exploit the fact that some teams play their home matches in stadiums in which an athletics track separates the stands from the field, while other teams play their home matches in stadiums without such a running track. Table 3 reports separate regressions for matches that take place in stadiums with a running track and for matches that take place in stadiums without a track. The regressions include controls for crowd size in addition to the set of explanatory variables of Specification (3) in Table 1. The variable of interest is the "Score difference," which equals 1 if the home team is one goal ahead and −1 if it is one goal behind. The results indicate that the estimated difference in stoppage time awarded in close matches amounts to almost 1 min (twice the coefficient on Score difference) of additional stoppage time when the match takes place in a stadium without a running track (Column 1). However, when the match takes place in a stadium in which a running track separates the stands from the field, this effect is much weaker and statistically not significant. This finding indicates that physical distance extenuates the bias in stoppage time decisions, suggesting that the intensity of crowd pressure is an

---

TABLE 2

The Effects of Crowd Composition

	Distance ≤150 km (1)	Distance >150 km (2)
Score difference	−4.199 (8.237)	−12.055*** (2.590)
Attraction × score difference	−20.335* (11.704)	3.596 (4.170)
Observations	199	918
R^2	.58	.34

Notes: Ordinary least squares estimates. The dependent variable is second-half stoppage time in close matches. The regressions include controls for the number of treatments, substitutions, fouls, and cards and controls for relative strength, season dummies, referee dummies, and team dummies. The variable "Attraction × score difference" is the product of the Score difference variable, which takes the value −1 if the home team is one goal behind and 1 if the home team is one goal ahead at the end of regulation time, and a dummy that takes the value 1 if the visiting team ranks in the top third of relative crowd composition index. This index is equal to the ratio of the average attractiveness of the visitor (i.e., the average attendance-to-capacity ratio in all away games except those that are played on the current opponent's turf) to the average attendance-to-capacity ratio in home games of the home team (calculated over all home matches except for those against the current opponent). The variable "Distance" measures the shortest distance between the home cities of opponents. Standard errors are given in parentheses.

*Significant at the 10% level; ***significant at the 1% level.

important determinant of referee favoritism.[21] This empirical result provides additional new support for the hypothesis that social forces affect referees' decisions.

B. Decisions to Award Goals and Penalty Kicks

Since goals and penalty kicks have a much more immediate impact on the ultimate match outcome than additional playing time, I examine next whether referees also favor

---

20. This bias tends to be reduced if the visiting team attracts more fans when playing away, but this effect is not statistically significant.

21. The regression results also show that neither the number of spectators nor the attendance-to-capacity ratio significantly affects the length of stoppage time. Coefficient estimates for interactions between these variables and the score difference indicator suggest that a higher attendance-to-capacity ratio and a higher absolute number of spectators tends to reduce the home bias in stadiums without a running track and to increase it in stadiums with a running track. Alternative specifications, in which controls for attendance and attendance-to-capacity ratio and their interactions with the score difference are dropped sequentially, indicate that a larger crowd and a higher attendance-to-capacity ratio is required in stadiums with a running track to induce referee bias. The difference in the size of the home bias in stadiums with and without a running track disappears when both interaction terms are excluded from the regression.

TABLE 3

The Stadium, the Crowd, and Second-Half Stoppage Time in Close Games

	Stadiums without Track (1)	Stadiums with Track (2)
Score difference	−29.297** (13.938)	−7.392 (7.569)
Attendance (in thousands)	0.986 (0.680)	−0.367 (0.612)
Attendance × score difference	0.126 (0.231)	−0.123 (0.223)
Ratio of attendance-to-capacity	−39.833 (33.191)	8.775 (34.761)
Ratio of attendance-to-capacity × score difference	18.447 (18.257)	−2.129 (14.381)
No. of treatments	1.499** (0.618)	−0.412 (0.527)
No. of substitutions	2.941 (2.390)	1.250 (2.299)
No. of yellow cards	8.004*** (1.952)	5.943*** (1.737)
No. of 2nd yellow cards	−3.135 (8.970)	1.419 (6.379)
No. of red cards	15.202 (9.831)	9.210 (7.328)
Constant	45.464 (52.650)	76.008 (46.951)
Observations	554	563
R^2	.38	.41

Notes: Ordinary least squares estimate. The dependent variable is the length of stoppage time awarded at the end of the match in matches where the home team is either one goal behind (Score difference = −1) or one goal ahead (Score difference = 1). The sample is split into matches that took place in stadiums without a running track separating the field and the stands (Column 1) and matches in stadiums with a running track (Column 2). All regressions include the same set of controls for relative strength as in Table 1, as well as fixed effects for home teams, visiting teams, and referees. The effects of controls for relative strength are never significant. The results concerning the home bias are not affected by the inclusion or exclusion of team-fixed effects. Standard errors are given in parentheses.

**Significant at the 5% level; ***significant at the 1% level.

home teams in decisions to award a goal or a penalty kick. I begin by investigating the correctness of goal and penalty kick decisions when a goal or a penalty kick was awarded. I then also take into account a second type of misjudgment, namely not granting a goal or a penalty kick when it should be awarded, by also considering critical situations that could have lead to a goal or a penalty kick.

The raw data on ratings of correctness of referee judgments, summarized in Table 4A, indicate that decisions to grant a goal are more likely to be wrong or disputable when the goal is awarded for the home team: 95.99% of goals scored by the visiting team, but only 95.05% of goals scored by the home team are rightly awarded. This 0.94 percentage point difference is statistically significant at the 5% level and considerable in magnitude since it implies that home teams were awarded 57 more goals based on wrong or disputable decisions. Remarkably, granted goals are less likely to be correctly awarded when a team is behind in score, especially when the home team is behind. The home team is particularly likely to be granted a goal based on a wrong or disputable decision if it is behind by one or two goals.

Referees also seem to favor home teams in penalty kick decisions (see also Sutter and Kocher 2004). The raw data reveal that a smaller fraction of penalties kicks for the home team is rightly awarded (65.20% vs. 72.57%, see Table 4B).[22] Observed differences in the frequencies of wrong, correct, and disputable decisions are statistically significant.[23] Again, the fraction of wrong or disputable decisions in favor of the home team is largest when the home team is behind in score. However, it must be noted that referees also make more disputable decisions in favor of the visitor, when the visitor is just one goal behind.

---

22. Home teams were awarded roughly twice as many penalty kicks as visitors (569 vs. 288) during the seasons from 1993/1994 until the first half of the season 2003/2004. This difference largely results from different strategies: there is simply more play in the visiting team's penalty area. Consequently, home teams are also more often involved in critical situations in which no penalty kick was awarded.

23. The null hypothesis that observed differences are purely due to random chance is rejected by a chi-square test at the 10% significance level. The $\chi^2(2)$ statistic equals 4.75. Grouping wrong and disputable decisions into one category and testing for the significance of differences in the observed frequencies of correct and not correct decisions for the home and visiting team yields a $\chi^2(1)$ test statistic of 4.41 rejecting the null hypothesis that referees decide in the same way for both teams at the 5% significance level.

TABLE 4
Correctness of Decisions to Award Penalties and Goals

(A) Decisions to Award a Goal

	Goal for Home Team Decision			Goal for Visiting Team Decision		
Score Difference	Wrong	Correct	Disputable	Wrong	Correct	Disputable
≤ -2	1.68	92.26	6.06	0.68	97.62	1.70
-1	1.43	93.69	4.88	0.15	97.51	2.34
0	1.56	95.04	3.39	0.88	96.08	3.04
1	1.38	95.48	3.13	1.62	93.73	4.65
≥ 2	1.01	96.64	2.35	0.95	96.76	2.29
Total	1.43	95.05	3.52	0.92	95.99	3.09

(B) Decisions to Award a Penalty Kick

	Penalty Kick for Home Team Decision			Penalty Kick for Visiting Team Decision		
Score Difference	Wrong	Correct	Disputable	Wrong	Correct	Disputable
≤ -2	8.57	45.71	45.71	11.76	70.59	17.65
-1	14.29	62.86	22.86	0.00	87.10	12.90
0	6.38	62.55	31.06	8.04	68.75	23.21
1	5.00	70.00	25.00	6.25	67.50	26.25
≥ 2	4.05	78.38	17.57	2.08	81.25	16.67
Total	7.38	65.20	27.42	5.90	72.57	21.53

Notes: The upper panel of the table shows the percentages of goals for the home team and the visiting team that were wrongly, correctly, or disputably awarded by score difference. The information is based on all 10,163 goals (6,025 for the home team and 4,138 for the visitor team) that were awarded from the start of the season 1992/1993 until the end of the first half of the season 2003/2004 and for which the correctness indicator is available. The lower panel of the table shows the percentages of awarded penalties for the home team and the visiting team that were wrongly, correctly, or disputably awarded. The information is based on all penalties that were awarded from the start of the season 1993/1994 until the end of the first half of the season 2003/2004 and covers 857 penalty kick decisions, 569 for the home team and 288 for the visitor team. The score difference is calculated as the number of goals that had been scored by the home team minus the number of goals that had been scored by the visitor team at the time a decision was made.

Probit estimates (not reported here) confirm that a correct goal decision is significantly less likely when the home team has scored. Interestingly, referees' decisions to award a goal are significantly less likely to be correct when the match takes place in a stadium without a running track that separates the stands from the field. Likewise, a probit model for the probability that a penalty kick was correctly awarded reveals that a correct judgment is 10% less likely when the game takes place in a stadium without a track.[24]

24. This effect is statistically significant at the 1% level. Estimates from a multinomial logit model, in which the categories of the dependent variable are a correct, disputable, or wrong decision, indicate that home teams are more likely to be granted disputable goals and that decisions in stadiums with a track are significantly less likely to be wrong. The results also indicate that decisions that favor the home team when it is behind by one or two goals are more likely to be disputable than plainly wrong, which squares with the fact that obvious misjudgment might have strong negative repercussions for referees.

The evidence presented so far ignores situations in which a goal or a penalty kick was not awarded. Information on the second type of decisions is available since the start of the season 1993/1994 for goal decisions and only since the start of the season 1998/1999 for penalty kick decisions. Including these situations, we can classify goals and penalty kicks as legitimate, disputable, or illegitimate in the following way: goal (penalty kick) that should be awarded according to the rules is a legitimate goal (penalty kick); illegitimate goals (penalty kicks) should not be awarded; and disputable goals (penalty kicks) comprise goals (penalty kicks) for which there is no consensus regarding their legitimacy.

Table 5 summarizes referees' goal and penalty kick decisions for home and visiting teams conditional on legitimacy. The first row of the upper panel shows that there is no indication that referees award the visitor fewer legitimate goals. There is tentative but no statistically

TABLE 5
Decisions to Award Penalties and Goals Conditional on Legitimacy

	Home Team		Visiting Team		W-M-W Test	
	Awarded	**Denied**	**Awarded**	**Denied**	**z Value**	**p Value**
(A) Decisions to award a goal						
Legitimate	5,199	68	3,637	46	−0.175	.569
Disputable	194	57	120	43	0.852	.197
Illegitimate	79	125	37	84	1.480	.070
(B) Decisions to award a penalty kick						
Legitimate	207	87	115	64	1.393	.082
Disputable	86	214	30	118	1.906	.028
Illegitimate	23	192	13	87	−0.597	.725

Notes: The upper panel of the table shows the numbers of legitimate, disputable, and illegitimate goals that were either granted or denied for home and visiting teams during all seasons from 1993/1994 until the end of the first half of the season 2003/2004. The indicator for the correctness of the referee's decision is missing for two goals that have been awarded in this period and for one goal that has not been awarded. These observations are not included when calculating the statistics. The lower panel of the table shows the numbers of legitimate, disputable, and illegitimate penalty kicks that were either granted or denied for home and visiting teams from the start of the season 1998/1999 until the end of the first half of the season 2003/2004. Statistics are based on ratings of all situations that led to penalty kicks and all critical actions that could have potentially lead to a penalty kick but were not penalized. The last column of each panel shows *p* values for the hypothesis that referees do not favor the home team against the one-sided alternative that they are more likely to award a goal (upper panel) or a penalty (lower panel) for the home team. W-M-W, Wilcoxon-Mann-Whitney.

significant evidence that referees grant the visiting team fewer disputable goals, but the home side is granted significantly ($p = .07$) more illegitimate goals than the visiting side.[25] The lower panel of the table reveals that visiting teams are more likely to be denied a legitimate or a disputable penalty kick: the visiting team was wrongly denied a legitimate penalty kick in 35.75% of cases but the home team only in 29.59% of cases. This preferential treatment of the home team is statistically significant at the 10% level as a one-sided nonparametric test reveals (see last column of the first row in Table 5B). In case of disputable penalty kick decisions, the evidence for home team favoritism is even more pronounced: home teams are awarded 28.67% of disputable penalties but visiting teams only 20.27%. Estimation results based on these data (reported in Dohmen 2005) indicate that home teams are significantly more likely to receive a penalty kick when it should objectively be awarded and when a penalty call is disputable. The estimates also show that referees tend to award fewer disputable and unjustified penalty kicks when the crowd is separated from the field by an athletics track.

---

25. Home teams are most likely to be awarded a disputable or illegitimate goal when being behind in score (see table 9 in Dohmen 2005 for estimation results).

IV. CONCLUDING REMARKS

This paper has provided evidence that referees in German professional soccer, who are appointed to be impartial, tend to make decisions regarding stoppage time, penalty kicks, and goals that favor the home team. Home-biased refereeing is more pronounced when the home crowd has a stronger interest in a decision that favors their team, for example, when the score margin is close and their team is behind in score. There is also evidence that the home bias is mitigated when the fraction of supporters of the visiting team rises, which indicates that conflicting social forces have countervailing effects on individuals. Strikingly, referees are more biased in stoppage time decisions and make fewer correct penalty kick and goal decisions if the match is played in a stadium without a running track that separates the stands from the field. This indicates that refereeing quality is impaired when the crowd is closer and when social pressure is arguably experienced as more intense by the referee.

A likely channel that leads to the observed systematic differences in referee decisions is that social pressure from the crowd directly affects the referee, who then departs from the decision that maximizes his expected material payoff. An alternative explanation for the

pattern that close games last longest is that referees lengthen exciting games because spectators like suspense (see Chan, Courty, and Li 2006 for a theoretical model). However, if this was the mechanism it is not clear why there is a systematic difference between close matches in which the home team is behind and those in which the home team is ahead, unless home teams are generally more likely to level the score when being behind by one goal. In any event, this alternative transmission channel presupposes that referees submit to the preferences of the crowd and would therefore not undermine the evidence that social forces affect individuals' decisions. The finding that refereeing quality is lower when no running track separates the stands from the field adds important new evidence that social pressure is the mechanism at play. It also helps to invalidate the alternative hypothesis that the governing body, the DFB, condones home team favoritism.

Another explanation emphasizes a more indirect channel that leads to biased refereeing; it entails that the crowd pressure affects the players on the field who then feel encouraged to complain about decisions that do not favor their team. It is interesting to recall in this context that the official observer of the DFB referee committee, who completes the evaluation form after the match, rates the referee's decisiveness and his competence in dealing with complaining players. As a result, referees are expected not to entertain the players complaints, which renders this indirect channel less plausible. But even if the true transmission mechanism worked through players influencing the referees, the social pressure hypothesis would not be invalidated; it would in fact be at the heart of the explanation. Regardless of the particular channel through which referees' decisions are affected, the fact that the social atmosphere in the stadium is related to refereeing quality is indicative for the role of social forces. As a result, we can conclude that the evidence is consistent with the hypothesis that social forces can affect the decisions of an individual who values not only material payoffs but also social payoffs.

These results are important for the literature on endogenous preference formation for they provide support for theories that emphasize the relevance of the social environment (e.g., Akerlof 1980; Becker and Murphy 2000; Bernheim 1994). Bernheim's model implies that we should expect a stronger influence of the social environment when social pressure rises relative to the individual's material payoff. This is likely the case in lower leagues, where monetary rewards for referees are lower. In fact, a DFB official has confirmed that referees in lower level leagues are regularly relegated or dismissed, but that dismissal or relegation is less common for *Bundesliga* referees. The DFB official attributed this to incentives effects arguing that ''much more is at stake for *Bundesliga* referees'' and to selection effects saying that *Bundesliga* referees are ''simply better referees as they have been promoted from lower leagues in several rounds upon very positive judgements of DFB observers.'' The latter statement implies that there is considerable heterogeneity in the extent to which individuals are affected by social pressure. There is evidence in the data that more experienced referees tend to be less biased, which suggests that individuals can learn to resist social pressure.

To understand how and to what extent individuals can become immune to social influences, it is important to ascertain whether referees consciously submit to social pressure or whether they are affected subconsciously. The latter conjecture appears to be more plausible. It is likely that referees' objective judgment capabilities are impaired by the emotional atmosphere in the stadium. Consistent with that explanation, Nevill, Balmer, and Williams (2002) have provided experimental evidence that soccer referees are affected by the crowd's noise. They showed videotaped tackles from an English Premier League soccer game to qualified referees who had to decide whether or not to award a foul. One group watched the videotape without the noise of the crowd being played, while the other group heard the noise. The latter group called 15.5% fewer fouls against the home team. Since referees have to judge a situation in a split of a second having very little time for deliberation, their decision-making process might be heavily influenced by cues in the environment and the atmosphere in the stadium.

REFERENCES

Akerlof, G. A. "A Theory of Social Custom, of Which Unemployment May Be One Consequence." *Quarterly Journal of Economics*, 94, 1980, 749–75.

Asch, S. E. "Effects of Group Pressure Upon the Modification and Distortion of Judgments," in *Groups,*

Leadership and Men, edited by H. Guetzkow. Pittsburgh, PA: Carnegie Press, 1951, 177–90.

Austen-Smith, D., and R. G. Fryer. "The Economics of 'Acting White'." *Quarterly Journal of Economics*, 120, 2005, 551–83.

Becker, G. S., and K. M. Murphy. *Social Economics. Market Behavior in a Social Environment*. Cambridge, MA: Harvard University Press, 2000.

Bernheim, D. B. "A Theory of Conformity." *Journal of Political Economy*, 102, 1994, 841–77.

Brock, W. A., and S. N. Durlauf. "Interactions-Based Models," in *Handbook of Econometrics*, Vol. 5, edited by J. J. Heckman and E. Leamer. Amsterdam, The Netherlands: Elsevier Science, 1999, 3297–380.

Chan, W., P. Courty, and H. Li. *Suspense*. Working Papers tecipa-223. University of Toronto, Department of Economics. 2006.

Coleman, J. S. *Foundations of Social Theory*. Cambridge, MA, Cambridge, Massachusetts: The Belknap Press of Harvard University Press, 1990.

Dohmen, T. J. *"Social Pressure Influences Decisions of Individuals: Evidence from the Behavior of Football Referees."* IZA Discussion Paper No. 1595, 2005.

Dufwenberg, M., and M. Lundholm. "Social Norms and Moral Hazard." *Economic Journal*, 111, 2001, 506–25.

Duggan, M., and S. D. Levitt. "Winning Isn't Everything: Corruption in Sumo Wrestling." *American Economic Review*, 92, 2002, 1594–605.

Fédération Internationale de Football Association. *Laws of the Game*. Zurich, Switzerland: Fédération Internationale de Football Association, 2002.

Festinger, L., S. Schachter, and K. Back. "The Spatial Ecology of Group Formation," in *Social Pressure in Informal Groups*, L. Festinger, S. Schachter and K. Back. Stanford, CA: Stanford University Press, 1950, 146–61.

Garicano, L., I. Palacios-Huerta, and C. Prendergast. "Favoritism Under Social Pressure." *Review of Economics and Statistics*, 87, 2005, 208–16.

Goldin, C., and C. Rouse. "Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians." *American Economic Review*, 90, 2000, 715–41.

Knowles, J., N. Persico, and P. Todd. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *Journal of Political Economy*, 109, 2001, 203–29.

Linn, G. "Ziel und Ablauf von Schiedsrichterbeobachtungen." *DFB Schiedsrichter-Zeitung*, 1, 2003.

Nevill, A., N. Balmer, and A. Williams. "The Influence of Crowd Noise and Experience upon Refereeing Decisions in Football." *Psychology of Sport and Exercise*, 3, 2002, 261–72.

Prendergast, C., and R. Topel. "Favoritism in Organizations." *Journal of Political Economy*, 104, 1996, 958–78.

Sutter, M., and M. G. Kocher. "Favoritism of Agents—The Case of Referees' Home Bias." *Journal of Economic Psychology* 25(4), 2004, 461–9.