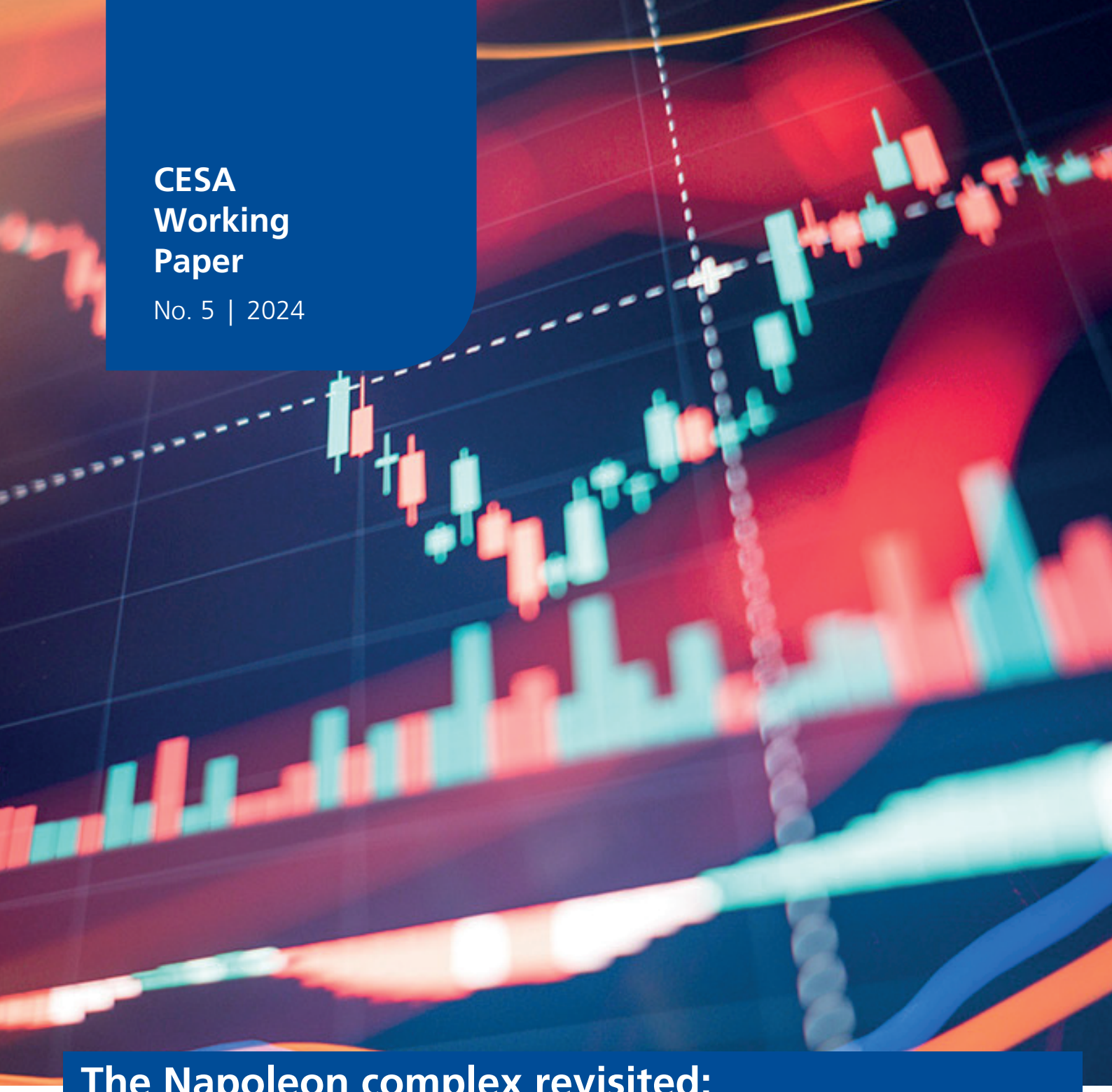


CESA
Working
Paper
No. 5 | 2024

The Napoleon complex revisited: New evidence from professional soccer

Giulio Callegaro[(a)], Mario Lackner[(b)] and Hendrik Sonnabend[(a)]

a University of Hagen (FernUniversität in Hagen)

b Johannes Kepler University Linz (JKU)

FernUniversität in Hagen

The Napoleon complex revisited: New evidence from professional soccer *

Giulio Callegaro[a], Mario Lackner[b], Hendrik Sonnabend[a]

[a]*University of Hagen*
[b]*Johannes Kepler University Linz (JKU)*

June 2024

Abstract

We analyse evaluation biases caused by physical attributes. Using data from German elite soccer, we find that referees are more inclined to sanction players when the difference in body size is sufficiently large. Moreover, we document an 'inverse Napoleon effect' in situations when the referee is confronted with smaller players, suggesting that sanctions are used as a substitute for authority gained by stature in the industry. Further analyses reveal that referees discriminate less against more talented players and teams with a higher concentration of these players. Finally, we find that the bias is reduced but still exists for the group of more experienced referees.

JEL Classification: J71, L83, Z20
Keywords: evaluation, discrimination, social dominance, referee bias, soccer

*Corresponding author: Mario Lackner, Johannes Kepler University of Linz, Department of Economics, Altenbergerstr. 69, 4040 Linz, Austria. Phone: +43 70 2468 7390, Fax: +43 70 2468 7390, email: mario.lackner@jku.at.

1. Introduction

In many situations in life, the outcome of our actions depends on the evaluation by others. Examples include the fields of education, labour market careers, and jurisdiction/crime prevention. For an optimal allocation of resources, these evaluations need to be unbiased. However, prior research has documented that this is not the case: non-performance related characteristics such as gender (e.g., Mengel et al. 2019, Card et al. 2020), ethnic background (e.g., Price & Wolfers 2010, Shayo & Zussman 2011, Parsons et al. 2011), and nationality (e.g., Spierdijk & Vellekoop 2009, Sandberg 2018) seem to influence the decisions of the evaluators even among the group of well-trained and selected experts. In the same way, premiums and penalties due to physical attributes such as attractiveness, body mass index, and height have been intensively studied as a potential source of discrimination in education and the labour market (e.g., Persico et al. 2004, Mobius & Rosenblat 2006, Case & Paxson 2008, Stinebrickner et al. 2019, Black & de New 2020, Goulão et al. 2024).

We contribute to this literature by studying the phenomenon called 'Napoleon complex' or 'short man syndrome' in the context of expert evaluation in a high-stakes, real-world setting. Basically, the 'Napoleon complex' implies that height is a positive attribute associated with qualities like social dominance and that relatively small (male) individuals show compensatory behaviours. For instance, research from the fields of evolutionary biology and psychology suggests that relatively small men compete for resources in a more aggressive way (e.g., Knapen et al. 2018) and that this behaviour might be an evolutionary stable strategy (e.g., Just & Morris 2003, Morrell et al. 2005).

Specifically, we use data from top-level soccer to provide evidence that the body size difference between player and referee can explain the tendency for sanctions. Combining features of laboratory experiments and field data, the benefits of professional sports data are quality and quantity, well-defined rules, and high incentives, among others (Balafoutas et al. 2019, Palacios-Huerta 2023). For these reasons, there is a comparably large number of papers dealing with the behaviour of referees in professional sports (e.g., Sutter & Kocher 2004, Dawson & Dobson 2010, Page & Page 2010); see Dohmen & Sauermann (2016) for an overview. Generally, suppose biased decision-making can be found in a high-stakes setting with highly trained, selected, and monitored agents. In that case, it can be interpreted as a lower bound of the effect in the general population.

Prior research that links a referee's behaviour to height in professional sports gives a mixed picture. While Stulp et al. (2012) find, for instance, that the total number of fouls per match is negatively correlated with the referees' height in their sample, McCarrick et al. (2020) find no correlation for fouls but yellow cards. A reason could be that the results are estimates from rather simple empirical models, and that a 'Napoleon complex' in a conflict situation, if it exists, would not solely depend on the referee's height but also on the height of the player involved. Most closely related, Gift & Rodenberg (2014) study the nexus between referee height and personal fouls for a player in the National Basketball Association (NBA). They find a negative and significant association, suggesting that smaller referees call more personal fouls on average. Estimates from a model that interacts player and referee height support the view that this bias solely depends on the latter. As a major drawback, the authors calculate the average height of the referee trio, and it is not clear who makes the decisions.

Contrary to Gift & Rodenberg (2014), we estimate a referee fixed effects model with different kinds of sanctions and only one person in charge. Our results indicate a height bias that depends on relative height: Compared to situations where referees and players are 'at eye level', we document an up to 9.4%

higher number of fouls called against a player in a given match when he is taller than the referee (up to a 7.2% higher probability of receiving a yellow card). In line with the idea of a 'Napoleon complex', the association increases in the size of the gap and is present only in the first part of a match.

Moreover, we can also document a phenomenon that could be described as an 'inverse Napoleon complex' or 'height dominance': players are less likely to be sanctioned when they are shorter than the referee (up to a 12.3% reduction in fouls called and 16.5% for yellow cards). A natural interpretation of our findings is that sanctions are used as a substitute for authority gained by stature in the industry.

Further analyses, novel to the literature, reveal constraints in the compensatory use of sanctions: We find the Napoleon complex to be less present when the referee is confronted with better players and better teams, probably due to higher conflict costs. Finally, we observe a slight reduction in the bias for the group of more experienced referees, suggesting a correction either by learning or selection.

2. Data and background

Our dataset contains information from 2,340 matches played in the *German Bundesliga* in seasons 2014/15 to 2021/22, which gives us 51,480 observations on the player-match level. The primary data have been collected from the `Sportmonks.com` website, including the three categories of sanctions we are focused on: fouls, yellow cards, and red cards (the order corresponds to the severity of the offence). Table 1 indicates that players are regularly cautioned for fouls, whereas dismissals are rare events. Player market value data originate from the `transfermarkt.de` website, while information on referee heights was collected from the association's official website (`www.dfb.de`). Our final sample includes 50,668 observations.

The referees' height ranges from 177 to 197 centimetres (cm). With an average of 186.8cm, the elite umpires in our sample (38 individuals) are substantially taller than the average male German in the cohorts between the ages of 18 and 50 (varying between 180.2cm and 181.7cm according to the German Microcensus 2021) and even the players, see Table 1. Consequently, the average difference between players and referees in our sample is negative. We take this apparent selection as a first hint in favour of social dominance associated with height within the industry. This aligns with Stulp et al. (2012) who find that referees in their sample tend to be taller than their assistant referees.

Table 1 also presents further descriptive statistics related to team and player market values and team performances. Note that a player's relative market value is his share of his team's market value in a given match, and that we proxy a team's performance by the average number of points per season prior to the focused match. Additionally, we call the difference between the height of the player and the referee relative to the referee's height the *relative height difference*. We explain this variable in more detail in the next section. Finally, Figure 1 illustrates the distribution of height across referees and the total number of games per height, suggesting no clear pattern in both distributions.

Referees play a crucial role in association football (soccer), ensuring that the game is conducted fairly, safely, and in accordance with the Laws of the Game. They must command respect and authority on the field to effectively manage the players and the flow of the game, potentially by sanctioning players for breaking the rules. Therefore, the elite of German referees receive above-average compensation: In 2022, this group received a fixed salary between 62,000 and 82,000 Euros according to tenure and international appointments, and a bonus of 5,600 Euros per match (Kicker 2022), whereas the average yearly income in Germany was below 50,000 Euros (according to the Federal Statistical Office). According to the

League Association (DFL), appointments for *Bundesliga* matches are based on software that considers qualification, the distance between the referee location and the venue, and a balanced number of matches. The referee's performance is evaluated by German Football Association (*Deutscher Fußball-Bund*, DFB) officials on the match level using a standardised sheet. Poor average ratings can lead to relegation to the second division (see DFB 2023). Taken together, evaluators in our setting represent the elite in their profession and have high incentives to perform well. As another source of live monitoring, the *Video Assistant Referee* (VAR) was introduced to the *Bundesliga* in the 2017/18 season. According to the rules of the game, the VAR checks the referee's actions and gets active in the event of clear errors related to goals, penalty kicks, red cards, and mistaken identities.

Interestingly, in an article on how body language can help to referee a match, published in the official DFB journal for referees, there is an explicit reference to body size as an important factor in achieving dominance in conflict situations (Altehenger 2014). While the expert emphasises that "height is not everything" and charisma and authority could also be gained via other channels, it suggests that there is already an idea of the nexus between social dominance and height in the profession.

TABLE 1 — Descriptive statistics

	Mean	**Std. dev.**
Fouls called	1.07	(1.18)
Yellow card	0.15	-
Any card	0.16	-
Dismissal[a]	0.01	-
Any card in first half.	0.06	-
Any card in second half.	0.09	-
Team market value (in millions of $)	209.02	(182.56)
Player market value (in 10-millions of $)	1.12	(1.41)
Player's relative market value[b]	9.24	(6.88)
Team past performance[c]	1.33	(0.62)
Referee experience (in years)	7.62	(4.92)
Player height (in cm)	184.03	(6.30)
Referee height (in cm)	186.78	(4.91)
Height difference (player-referee)	−2.76	(7.99)
Relative height difference[d]	−1.62	(4.41)

Notes: N = 50,668. [a] Players are suspended after two yellow or one red card. A second yellow card within a game also leads to a red card. [b] A player's relative market value is his share of his team's market value in a given match. [c] Average number of points per season prior to the focused match. [d] Difference between the height of the player and the referee relative to the referee's height.

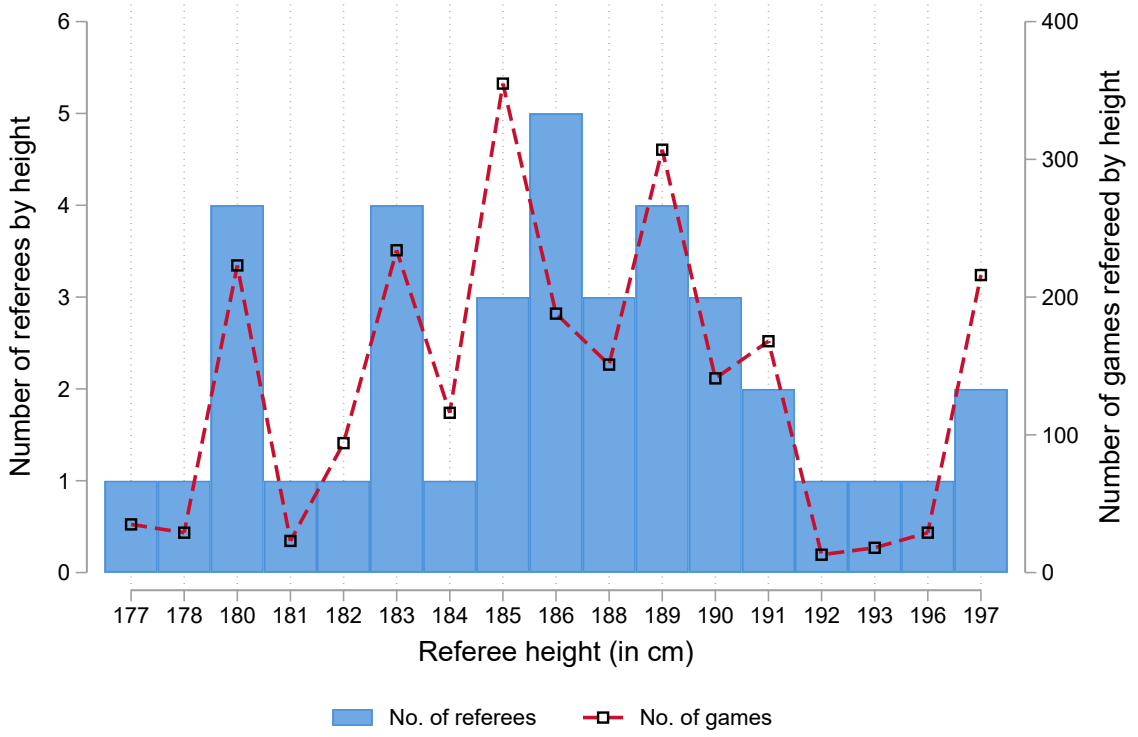
3. Empirical approach and results

We aim to identify how differences in body size affect referees' decision making in terms of sanctions. An unbiased evaluation would require that – conditional on observables such as a player's position and talent – a player's physical attributes should not affect the risk of getting sanctioned.

For the main analysis, we estimate the probability that player i of team t gets sanctioned (or the intensity of sanctions) by referee r in match m against team o in season s defined by

$$sanction_{i,r,t,s,o} = \beta_0 + \sum_{k=1}^{5} \beta_k \cdot Q_k + \xi' \mathbf{X}_{i,r,t,s,o} + \pi_r + \theta_{t,s} + \rho_{o,s} + \varepsilon_{i,r,t,s,o}\,, \tag{1}$$

FIGURE 1 — Referees and matches by referee height



where $sanction_{i,r,t,s,o}$ is a placeholder equal to the number of fouls called or equal to 1 if player i is sanctioned either by a yellow card or a dismissal (i.e. if the referee shows a red card or two yellow cards within the match).

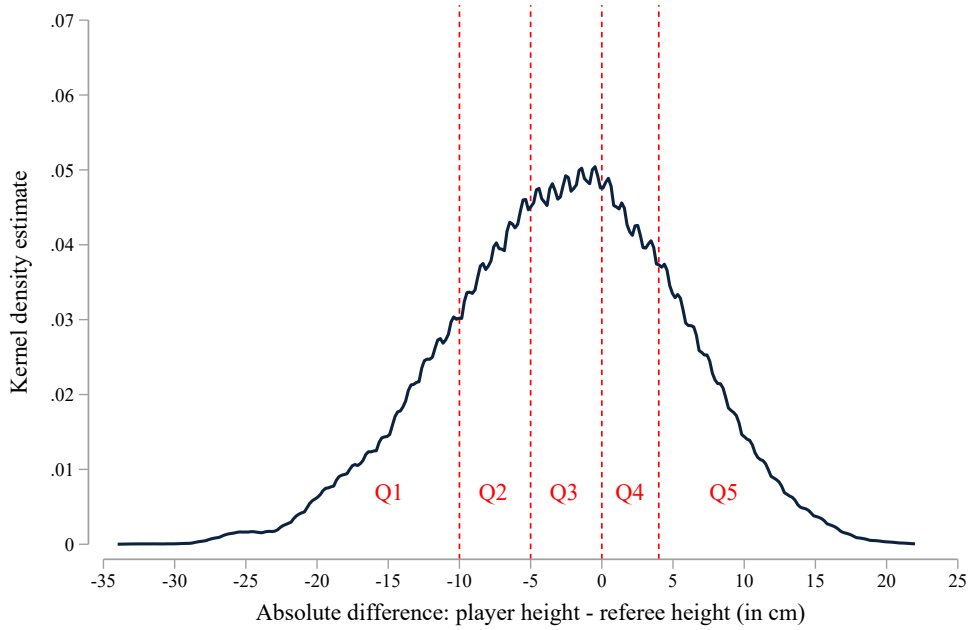
Our focus explanatory variable Q_k is an indicator for the kth-quintile of the distribution of the relative relative height difference between i and r defined by $\frac{player\ height_i - referee\ height_r}{referee\ height_r}$. To give an idea of what these differences look like, Figure 2 presents kernel density estimates of the distribution of the absolute height difference. The third quintile – which will work as our reference category – refers to a situation where the player and referee are of equal size or the referee is slightly taller.

$\mathbf{X}_{i,r,t,s,o}$ is a vector of controls, including player i's market value, the teams' average points in the season before match m in the season, and a binary variable that takes the value 1 if player i plays at home and 0 otherwise. While market values and points serve as a proxy for talent and performance, the home game indicator is used to address prior evidence on home-biased refereeing (Dohmen & Sauermann 2016). In addition, position dummies (defender, midfielder, forward) account for defensive players being more engaged in tackles and, hence, more at risk of being sanctioned.

Furthermore, π_r are referee fixed effects we add to ensure that estimates of β_k capture within-referee variations in the sense that the same referees encounter players who are taller or shorter than him.[1] Finally, to account for unobserved heterogeneity, such as differences in the styles of play and coaching, a team's talent, and local rivalries ('derbies'), we include team-season ($\theta_{t,s}$) and opponent-season (ρ_o) fixed effects.

[1]We prefer referee fixed effects over player fixed effects because of the lower loss of degrees of freedom and the substantial share of players with only a very low number of observations.

FIGURE 2 — Distribution of the difference between players' and referees' height



Notes: Kernel density estimates of the distribution of the difference between player and referee height (in cm). Positive values indicate that the player is taller than the referee.

3.1. Main results

Table 2 presents our main results. Compared to a situation when players and referees are at eye level, players are penalised less frequently for fouls and have a lower probability of receiving a yellow card when they are shorter than the referee (columns 1 and 2). Specifically, the estimates of β_1 and β_2 indicate a reduction by 0.132 and 0.059 fouls on the match level (12.3% and 5.5% at the sample mean) when the relative height difference is in the first or second quintile and 16.5% and 9.2% reductions in the likelihood of being sanctioned with a yellow card. We take this as evidence of social dominance through stature in the profession, implying that referees resort less to sanctions when they can establish authority and their role as enforcers of the rules simply by physical superiority. Moreover, it is interesting to note that $\hat{\beta}_1$ clearly exceeds $\hat{\beta}_2$, implying that social dominance is, in fact, increasing in relative body size.

On the other side of the spectrum, it shows that referees resort more to sanctions in situations of physical inferiority: players tend to be penalised more often in terms of fouls called and yellow cards when they exceed the referee by height, the classic Napoleon complex. In detail, compared to the reference category, we find an increase of up to 9.4% for fouls and 7.2% for the likelihood of receiving a yellow card when the relative height difference is in the fourth and fifth quintile.

Now, given that the Napoleon complex implies in our setting that referees tend to gain authority by higher strictness levels when confronted with taller players, we would expect them to 'send the message' at the beginning of a match rather than at its end. The estimates in columns (4) and (5) support this idea: After splitting the sample into first and second halves, we find that the Napoleon complex (in terms of a higher probability for players taller than the referee of getting a yellow card) is present only in the first half of match but more pronounced (see that the point estimate for the fifth quintile indicator translates

into a 13.8% when evaluated at the sample mean). For referees with physical superiority, however, their lower tendency to sanction does not depend on the time of the game (e.g., we document a 13.8% and 17.2% reduction for first quintile situations in the first and second half).

Finally, the findings presented in column (3) suggest that the referee's leeway in handling conflict situations is limited, as parameter estimates for our main explanatory variables are precisely zero when the outcome is a suspension. Suspension is the most drastic intervention and hence less suitable for 'sending a message', and it could also be challenged by the VAR (see Section 2).

In conclusion, we interpret our main findings as referees using sanctions to substitute authority gained by stature. However, this behaviour results in a biased evaluation since the evaluation of players' actions should not depend on their stature.

TABLE 2 — Main results

	(1)	(2)	(3)	(4)	(5)
	fouls called	**yellow card**	**suspension**	**yellow card first half**	**yellow card second half**
Rel. height - Quint. 1	−0.132***	−0.025***	−0.001	−0.008**	−0.016***
(1 = *yes*, 0 = *no*)	(0.017)	(0.006)	(0.001)	(0.004)	(0.004)
Rel. height - Quint. 2	−0.059***	−0.014***	0.000	−0.004	−0.010**
(1 = *yes*, 0 = *no*)	(0.015)	(0.005)	(0.001)	(0.003)	(0.004)
Rel. height - Quint. 4	0.082***	0.008	0.001	0.004	0.004
(1 = *yes*, 0 = *no*)	(0.016)	(0.005)	(0.001)	(0.003)	(0.004)
Rel. height - Quin. 5	0.101***	0.011**	0.000	0.008**	0.003
(1 = *yes*, 0 = *no*)	(0.017)	(0.005)	(0.001)	(0.004)	(0.004)
Past performance	−0.045***	−0.006	−0.001	0.000	−0.006
	(0.015)	(0.004)	(0.001)	(0.003)	(0.004)
Opponent's past performance	0.009	0.006	−0.003***	0.000	0.006
	(0.014)	(0.004)	(0.001)	(0.003)	(0.004)
Player market value (in $ Mio.)	−0.035***	−0.004***	0.000	−0.004***	−0.001
	(0.004)	(0.001)	(0.000)	(0.001)	(0.001)
Home team	−0.048***	−0.017***	−0.001**	−0.009***	−0.008***
(1 = *yes*, 0 = *no*)	(0.008)	(0.003)	(0.001)	(0.002)	(0.002)
Add. bin. controls	*yes*	*yes*	*yes*	*yes*	*yes*
Referee FE	*yes*	*yes*	*yes*	*yes*	*yes*
Team-season FE	*yes*	*yes*	*yes*	*yes*	*yes*
Opponent-season FE	*yes*	*yes*	*yes*	*yes*	*yes*
R^2	0.133	0.027	0.008	0.016	0.017
Mean dep. var.	1.075	0.152	0.006	0.058	0.093

Notes: N = 50,668; ** and *** indicate statistical significance at the 5% and 1% level. Standard errors in parentheses are clustered on the player-referee combination. The omitted base category is the third quintile of the relative height distribution, which refers to situations where the player and referee are of equal size or the referee is slightly taller (min. of −2.67%, max of 0%). Dependent variables are (1) the number of fouls called against player i in match m, (2) an indicator variable which equals 1 if player i was given a yellow card in match m (and 0 otherwise), (3) an indicator variable which equals 1 if player i was suspended in match m (and 0 otherwise), indicator variables which equal 1 if player i was given a yellow card in (4) the first or (5) second half of match m. Only players who started the game are included.

3.2. Heterogeneity analysis

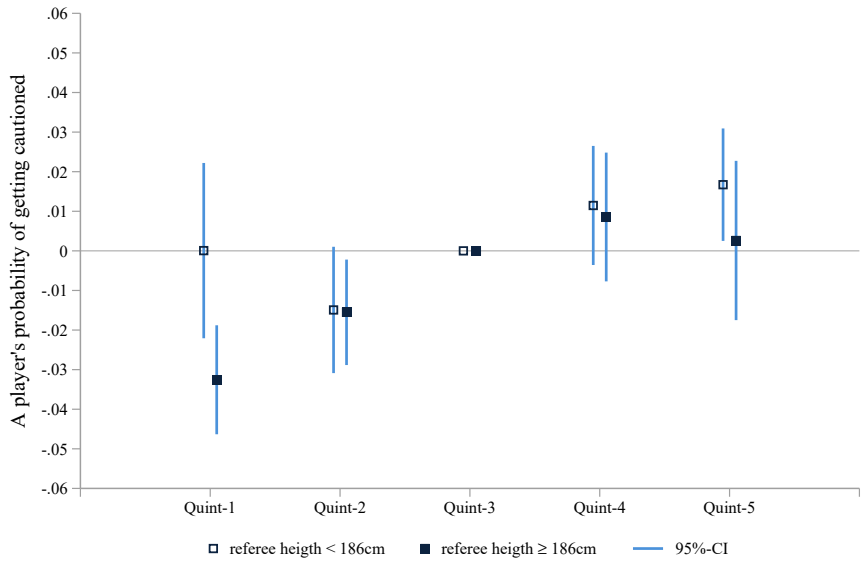
In this section, we conduct some additional analyses to uncover mechanisms explaining our previous findings. We start with the question of what role the actual height of the referee plays in our results.

We, therefore, split the sample along the median referee height of 186 cm and then re-estimate model 1. The dependent variable is player i's probability of getting cautioned in match m. Figure 3 illustrates the results. It shows that our primary finding is mainly driven by tall referees being more relaxed towards shorter players and short referees acting stricter when being confronted with significantly taller players.

Next, we turn to the unequal treatment of players and teams with regard to their talent. The idea is that more talented players have a higher reputation and status and that sanctioning these players involves higher conflict costs on the referee's side than sanctioning the 'no-names' of the league (see Lackner & Sonnabend 2023, for a further reading). Hence, if referees trade off the costs and benefits of gaining authority through sanctions when being physically inferior, we would expect them to penalise taller, high-status players less often than taller, low-status players. For the analysis, we proxy talent by a player's market value. Figure 4 then shows no significant difference between players from both groups when the relative height difference is in the first and second quintile, but that referees shy away from compensatory sanctions when facing high-status players. Although this finding reflects rational behaviour, it introduces a new facet of the evaluation bias to the literature. Figures 5 and 6 corroborate these results: Compensatory sanctions are present only in the group of low-performing teams, and a player's relative status (proxied by his share of the total market value of his team) hardly makes a difference.

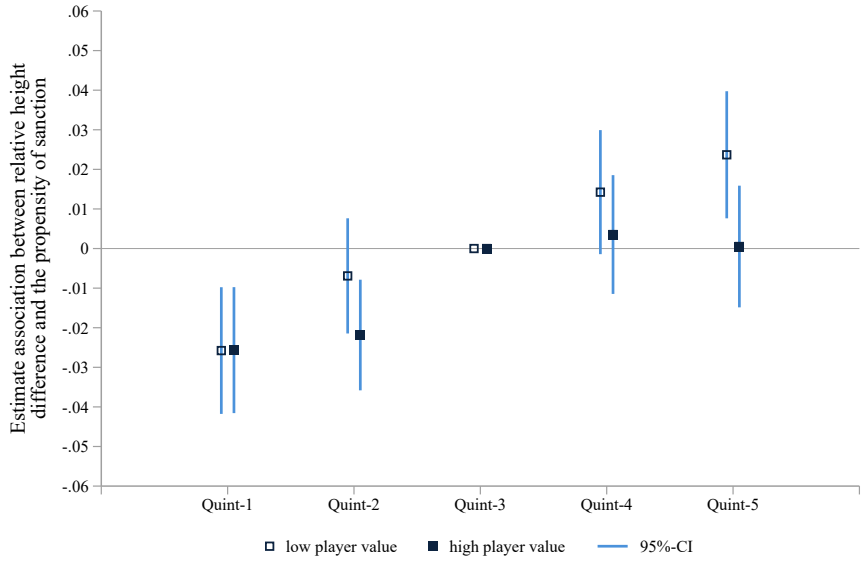
Finally, we examine whether experience can mitigate the evaluation bias in the industry. Since elite referees are constantly evaluated and can be relegated to a lower division (see Section 2), this may happen because of learning and selection. However, after splitting our sample along the median years of *Bundesliga* tenure (seven years), we only find mild evidence that experience reduces the bias. While the group of high-experienced referees tend to resort less often to sanctions in situations of physical inferiority than the low-experienced referees (acknowledging that $\hat{\beta}_5$ is not a precise zero), they act even more relaxed when confronted with shorter players. Taken together, we cannot say that experience rules out discriminatory behaviour.

FIGURE 3 — Effect heterogeneity: referee height



Notes: Estimates for β_1 to β_5 with sample split. Dependent variable: player i's probability of getting cautioned. Hollow squares: referees shorter than or equal to 186 cm; full squares: referees taller than 186 cm. Standard errors are clustered on the referee-player level.

FIGURE 4 — Effect heterogeneity: players' talent

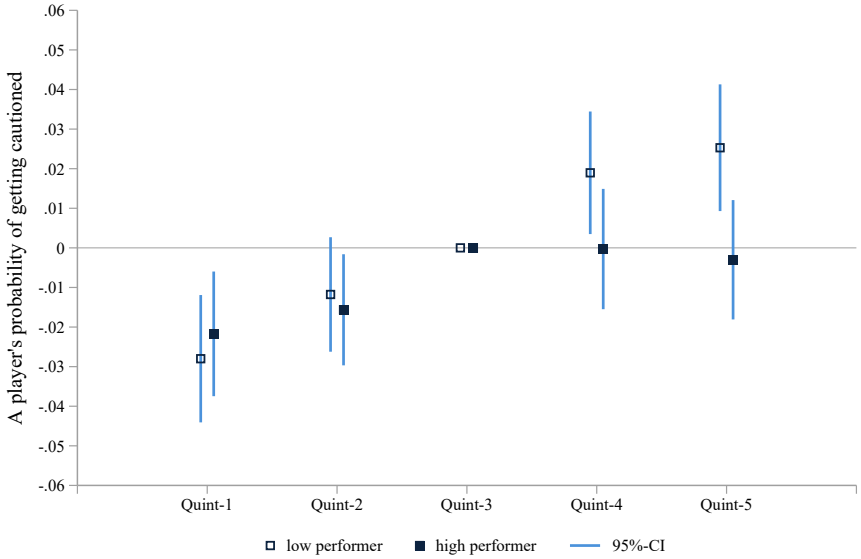


Notes: Estimates for β_1 to β_5 with sample split. Dependent variable: player i's probability of getting cautioned. Hollow squares: market value of $<$ 6.3 mio.; full squares: market value $\geq$ 6.3 mio. Standard errors are clustered on the referee-player level.

4. Robustness checks

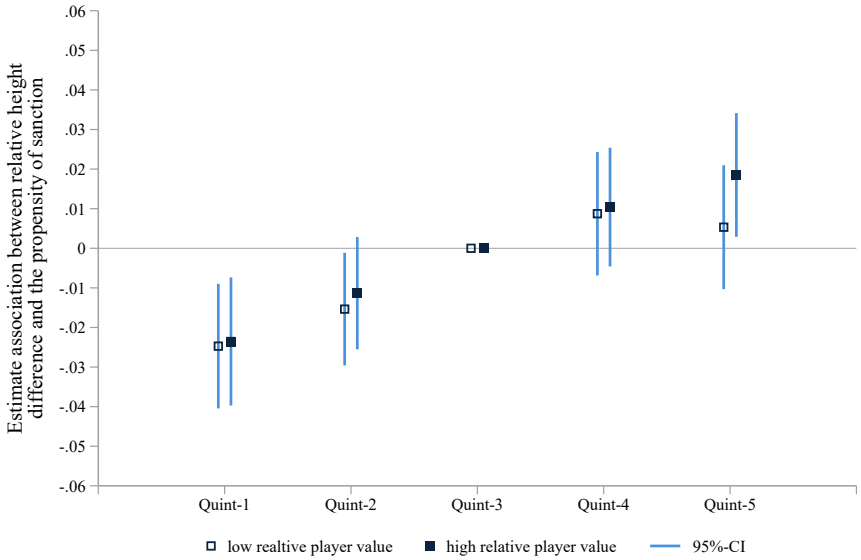
We check the robustness of our results to non-linear estimators. Specifically, we carry out logit regressions when the dependent variable is an indicator variable (yellow cards and suspensions), and Poisson

FIGURE 5 — Effect heterogeneity: teams' past performance



Notes: Estimates for β_1 to β_5 with sample split. Dependent variable: player i's probability of getting cautioned. Hollow squares: low-performing teams ($<$ 1.32 points per game); full squares: high-performing teams (point average of $\geq$ 1.32). Standard errors are clustered on the referee-player level.

FIGURE 6 — Effect heterogeneity: relative player value

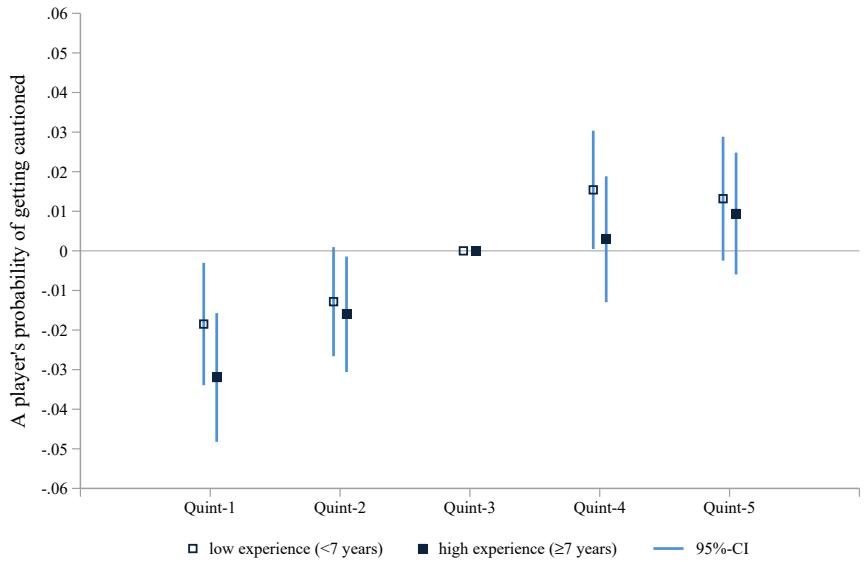


Notes: Estimates for β_1 to β_5 with sample split. Dependent variable: player i's probability of getting cautioned. Hollow squares: below-median relative market value (median at approximately 0.44%); full squares: above-median relative market value. Standard errors are clustered on the referee-player level.

regressions when the dependent variable is a count variable. Table 3 reports marginal effects for the main variable of interest. Overall, our main results are confirmed.

Another concern relates to the endogeneity of the referee appointment decision: If league officials were aware of the behavioural biases we find in our main results, they might identify potential team

FIGURE 7 — Effect heterogeneity: referee experience



Notes: Estimates for β_1 to β_5 with sample split. Dependent variable: player i's probability of getting cautioned. Referee experience is calculated based on the first match refereed in the first German league. Standard errors are clustered on the referee-player level.

characteristics that moderate the effect and take this into account when scheduling referees in a season. For example, organisers could try to avoid assigning a relatively short referee to matches of teams with tall players.[2] If that would be the case, then this scheduling policy would affect our estimates.

To address this concern, we perform a balancing test for referee assignments. That is, we estimate a model of referee height on the match level, regressed on team characteristics averaged for the home, the away or both teams. The results are presented in Table 4. It shows that neither height, talent, nor performance of one or both teams explains the referee's height in a given match. We take this as evidence that there is no referee selection in anticipation of evaluation biases related to relative height.

5. Conclusion and discussion

Our study highlights the subtle yet significant influence of stature on decision-making in expert evaluations. We find robust evidence of height-related biases in refereeing decisions within professional soccer: Players taller than the referee are exposed to a higher risk of being cautioned than players of the same height, whereas shorter players benefit from leniency. This bias is consistent with the 'Napoleon complex' concept where shorter individuals show compensatory behaviours for lacking social dominance through height. In our setting, where referees are responsible for enforcing the rules, our results suggest that sanctions are used as a substitute for authority gained by stature.

Additional findings suggest that referees use these compensatory sanctions rationally, in the sense that they refrain from using them in situations with high conflict costs associated with the player's or his team's status. On the contrary, we do not find that experience unambiguously reduces the bias.

The presence of a height bias among highly trained and monitored referees in professional sports

[2] Such a policy, however, would deviate from the official rules for appointments, see Section 2.

TABLE 3 — Robustness: Non-linear models

	(1)	(2)	(3)	(4)	(5)
	fouls called	**yellow card**	**suspension**	**yellow card first half**	**yellow card second half**
Rel. height - Quint. 1	−0.120***	−0.024***	−0.002	−0.008**	−0.016***
(1 = *yes*, 0 = *no*)	(0.017)	(0.005)	(0.002)	(0.004)	(0.004)
Rel. height - Quint. 2	−0.053***	−0.013***	−0.001	−0.004	−0.009**
(1 = *yes*, 0 = *no*)	(0.015)	(0.005)	(0.002)	(0.003)	(0.004)
Rel. height - Quint. 4	0.080***	0.009	0.002	0.005	0.004
(1 = *yes*, 0 = *no*)	(0.016)	(0.005)	(0.002)	(0.003)	(0.004)
Rel. height - Quin. 5	0.104***	0.011**	0.001	0.008**	0.003
(1 = *yes*, 0 = *no*)	(0.017)	(0.005)	(0.002)	(0.004)	(0.004)
Full set of controls	*yes*	*yes*	*yes*	*yes*	*yes*
Referee FE	*yes*	*yes*	*yes*	*yes*	*yes*
Team-season FE	*yes*	*yes*	*yes*	*yes*	*yes*
Opponent-season FE	*yes*	*yes*	*yes*	*yes*	*yes*
Mean dep. var.	1.075	0.152	0.006	0.058	0.093

Notes: N = 50,668; ** and *** indicate statistical significance at the 5% and 1% level. Standard errors in parentheses are clustered on the player-referee combination. The omitted base category is the third quintile of the relative height distribution, which refers to situations where the player and referee are of equal size or the referee is slightly taller (min. of −2.67%, max of 0%). Dependent variables are (1) the number of fouls called against player i in match m, (2) an indicator variable which equals 1 if player i was given a yellow card in match m (and 0 otherwise), (3) an indicator variable which equals 1 if player i was suspended in match m (and 0 otherwise), indicator variables which equal 1 if player i was given a yellow card in (4) the first or (5) second half of match m. Only players who started the game are included.

TABLE 4 — Association of team-game characteristics and referee height

	(1)	(2)	(3)
	pooled	**home team**	**away team**
Avg. player value	−0.032	−0.028	−0.034
	(0.033)	(0.049)	(0.044)
Avg. player height	−0.261	−0.130	−0.111
	(0.195)	(0.108)	(0.116)
Team success	0.097	0.358	−0.339
	(0.296)	(0.253)	(0.262)
Home team-season FEs	*yes*	*yes*	*yes*
Visiting team-season FEs	*yes*	*yes*	*yes*
Gameday FEs	*yes*	*yes*	*yes*
N	2,340	2,340	2,340
R-sq	0.135	0.135	0.134

Notes: ** and *** indicate statistical significance at the 5-percent level and 1-percent level, respectively. Standard errors–clustered on the pairing level–in round parentheses. Dependent variable is the height of the referee in cm (mean 186.78).

suggests that similar biases could be pervasive in other professional settings such as corporate boardrooms, hiring committees, and performance evaluations. Understanding these biases is crucial because they can impact career opportunities, promotions, and overall workplace dynamics. Hence, training programs that combine awareness with methods of verbal and non-verbal communication that supports individuals in leadership positions to emphasise authority – regardless of their physical attributes and also in the presence of high conflict costs – may help mitigate the bias.

References

Altehenger, T. (2014), 'Balancieren auf der „Statuswippe"', *Schiedsrichter-Zeitung* **6**, 4–8.

Balafoutas, L., Chowdhury, S. M. & Plessner, H. (2019), 'Applications of sports data to study decision making', *Journal of Economic Psychology* **75**, 102153.

Black, N. & de New, S. C. (2020), 'Short, heavy and underrated? Teacher assessment biases by children's body size', *Oxford Bulletin of Economics and Statistics* **82**(5), 961–987.

Card, D., DellaVigna, S., Funk, P. & Iriberri, N. (2020), 'Are referees and editors in economics gender neutral?', *The Quarterly Journal of Economics* **135**(1), 269–327.

Case, A. & Paxson, C. (2008), 'Stature and status: Height, ability, and labor market outcomes', *Journal of Political Economy* **116**(3), 499–532.

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.

DFB (2023), 'Schiedsrichter-Bewertung'.
URL: *https://www.dfb.de/schiedsrichter/funktionaerin/schiedsrichter-bewertung/*

Dohmen, T. & Sauermann, J. (2016), 'Referee bias', *Journal of Economic Surveys* **30**(4), 679–695.

Gift, P. & Rodenberg, R. M. (2014), 'Napoleon complex: height bias among national basketball association referees', *Journal of Sports Economics* **15**(5), 541–558.

Goulão, C., Lacomba, J. A., Lagos, F. & Rooth, D.-O. (2024), 'Weight, attractiveness, and gender when hiring: A field experiment in Spain', *Journal of Economic Behavior & Organization* **218**, 132–145.

Just, W. & Morris, M. R. (2003), 'The Napoleon complex: why smaller males pick fights', *Evolutionary Ecology* **17**(5-6), 509.

Kicker (2022), 'Stegemann Und Zwayer: Die Top-Verdiener - Die Schiedsrichter-Honorare 2022/23'. Accessed: 2024-06-07.
URL: *https://www.kicker.de/stegemann-und-zwayer-die-top-verdiener-die-schiedsrichter-honorare-2022-23-930245/slideshow*

Knapen, J. E., Blaker, N. M. & Van Vugt, M. (2018), 'The Napoleon complex: when shorter men take more', *Psychological Science* **29**(7), 1134–1144.

Lackner, M. & Sonnabend, H. (2023), 'The older the wiser? Determinants of misbehavior in team contests', *Social Science Quarterly* **104**(5), 1170–1183.

McCarrick, D., Brewer, G., Lyons, M., Pollet, T. V. & Neave, N. (2020), 'Referee height influences decision making in British football leagues', *BMC Psychology* **8**, 1–10.

Mengel, F., Sauermann, J. & Zölitz, U. (2019), 'Gender bias in teaching evaluations', *Journal of the European Economic Association* **17**(2), 535–566.

Mobius, M. M. & Rosenblat, T. S. (2006), 'Why beauty matters', *American Economic Review* **96**(1), 222–235.

Morrell, L. J., Lindström, J. & Ruxton, G. D. (2005), 'Why are small males aggressive?', *Proceedings of the Royal Society B: Biological Sciences* **272**(1569), 1235–1241.

Page, K. & Page, L. (2010), 'Alone against the crowd: Individual differences in referees' ability to cope under pressure', *Journal of Economic Psychology* **31**(2), 192–199.

Palacios-Huerta, I. (2023), 'The beautiful dataset', *SSRN Working Paper No. 4665889* .

Parsons, C. A., Sulaeman, J., Yates, M. C. & Hamermesh, D. S. (2011), 'Strike three: Discrimination, incentives, and evaluation', *American Economic Review* **101**(4), 1410–1435.

Persico, N., Postlewaite, A. & Silverman, D. (2004), 'The effect of adolescent experience on labor market outcomes: The case of height', *Journal of Political Economy* **112**(5), 1019–1053.

Price, J. & Wolfers, J. (2010), 'Racial discrimination among NBA referees', *Quarterly Journal of Economics* **125**(4), 1859–1887.

Sandberg, A. (2018), 'Competing identities: A field study of in-group bias among professional evaluators', *The Economic Journal* **128**(613), 2131–2159.

Shayo, M. & Zussman, A. (2011), 'Judicial ingroup bias in the shadow of terrorism', *The Quarterly Journal of Economics* **126**(3), 1447–1484.

Spierdijk, L. & Vellekoop, M. (2009), 'The structure of bias in peer voting systems: Lessons from the Eurovision Song Contest', *Empirical Economics* **36**, 403–425.

Stinebrickner, R., Stinebrickner, T. & Sullivan, P. (2019), 'Beauty, job tasks, and wages: A new conclusion about employer taste-based discrimination', *Review of Economics and Statistics* **101**(4), 602–615.

Stulp, G., Buunk, A. P., Verhulst, S. & Pollet, T. V. (2012), 'High and mighty: Height increases authority in professional refereeing', *Evolutionary Psychology* **10**(3), 147470491201000314.

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

CESA Working Paper Series

CESA Working Paper No. 1 | 2023
Andreas Hauptmann, Benjamin Schwanebeck and Hans-Jörg Schmerer
Plant-level adjustments to imports and exports at the extensive margin

CESA Working Paper No. 2 | 2023
Joscha Beckmann, Marco Kerkemeier and Robinson Kruse-Becher
Regime-specific exchange rate predictability

CESA Working Paper No. 3 | 2023
Joscha Beckmann, Timo Heinrich and Jennifer Rogmann
Inflation expectations and cognitive uncertainty

CESA Working Paper No. 4 | 2024
Daniel Monsees and Matthias Westphal
Disruptions in Primary Care: Can Resigning GPs Cause Persistently Negative Health Effects?

CESA Working Paper No. 5 | 2024
Giulio Callegaro, Mario Lackner and Hendrik Sonnabend
The Napoleon complex revisited: New evidence from professional soccer

FernUniversität in Hagen