

Deutscher, Christian; Sahm, Marco; Schneemann, Sandra; Sonnabend, Hendrik

Article — Published Version

Strategic investment decisions in multi-stage contests with heterogeneous players

Theory and Decision

Provided in Cooperation with:

Springer Nature

Suggested Citation: Deutscher, Christian; Sahm, Marco; Schneemann, Sandra; Sonnabend, Hendrik (2021) : Strategic investment decisions in multi-stage contests with heterogeneous players, Theory and Decision, ISSN 1573-7187, Springer US, New York, NY, Vol. 93, Iss. 2, pp. 281-317,
<https://doi.org/10.1007/s11238-021-09845-w>

This Version is available at:

<https://hdl.handle.net/10419/287092>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Strategic investment decisions in multi-stage contests with heterogeneous players

Christian Deutscher¹ · Marco Sahn² · Sandra Schneemann¹ · Hendrik Sonnabend³

Accepted: 2 September 2021 / Published online: 29 September 2021
© The Author(s) 2021

Abstract

When heterogeneous players make strategic investment decisions in multi-stage contests, they might conserve resources in a current contest to spend more in a subsequent contest, if the degree of heterogeneity in the current (subsequent) contest is sufficiently large (small). We confirm these predictions using data from German professional soccer, in which players are subject to a one-match ban if they accumulate five yellow cards. Players with four yellow cards facing the risk of being suspended for the next match are (i) less likely to be fielded when the heterogeneity in the current match increases and (ii) more likely to receive a fifth yellow card in the current match when heterogeneity in the next match increases or heterogeneity in the next match but one (when they return from their ban) decreases.

Keywords Tournaments · Multi-stage contests · Heterogeneity · Anticipating behaviour · Shadow effects

✉ Hendrik Sonnabend
hendrik.sonnabend@fernuni-hagen.de

Christian Deutscher
christian.deutscher@uni-bielefeld.de

Marco Sahn
marco.sahn@uni-bamberg.de

Sandra Schneemann
sandra.schneemann@uni-bielefeld.de

¹ Department of Sports Science, Bielefeld University, Postfach 10 01 31, 33501 Bielefeld, Germany

² Otto-Friedrich-Universität Bamberg, Feldkirchenstraße 21, 96052 Bamberg, Germany

³ Department of Economics, University of Hagen, Universitätsstraße 11, 58097 Hagen, Germany

1 Introduction

Contests are common in real-life settings, such as labour markets, industrial economics, sports, and public choice. In the majority of cases, ability levels vary across contestants, and this heterogeneity has adverse effects on contestants' willingness to invest resources, often referred to as the 'discouragement effect' (e.g. Konrad, 2009).¹ In a dynamic (tournament) setting with multiple stages and asymmetric abilities, a contestant's investment in a current stage can also be affected by the strength of the opponents in future stages.

In US presidential primaries for example, a candidate may 'skip' a state, by reducing campaign expenditures in that state, because it offers a low probability of victory and instead save and reallocate those resources to subsequent efforts in states where the outcomes of the election are less certain. More generally, disadvantaged subjects tend to drop out of competitions to save effort. In team sports, coaches are more likely to spare top players in very unbalanced matches to avoid injuries and reduce (mental) fatigue for the upcoming game—in case the upcoming game is anticipated to be a close contest. Such occurrences are most recently discussed as 'load management' in the media.

We thus hypothesise that contestants self-restrict in the current stage of a dynamic battle as long as the gap in abilities at this stage is sufficiently large. This hypothesis corresponds to the discouragement effect. We also predict a spillover effect, such that contestants tend to self-restrict in the current stage when heterogeneity in the subsequent contest (upcoming stage) is sufficiently small.²

To establish empirical evidence, we use data from the German *Bundesliga*, a professional association soccer league with a double round-robin tournament structure.³ Professional soccer leagues offer a formalised structure and full information about the number of rounds and opponents at the beginning of the tournament; they also entail contests with high incentives. To test our predictions, we leverage a unique rule stating that a player who accumulates a critical number of yellow cards (also called 'bookings') will be suspended for the next match. Teams thus may self-restrict by not fielding a player threatened by a ban or by strategically provoking a booking to obtain a suspension for the subsequent match.

In line with our predictions, we find that teams systematically self-restrict in round t if the difference in ability, or heterogeneity, in t is sufficiently large. Results regarding the effect of heterogeneity in $t + 1$ on the decision to hold back resources in t are somewhat mixed, but they provide strong evidence of anticipating behaviour.

¹ We define resources as the input needed to improve a contestant's winning probability, given the opponent's ability and investment level. Several synonyms appear in tournament theory and economic analyses of conflict, such as effort, endowments, or force size.

² These hypotheses also follow from the theoretical model of a dynamic contest that we develop in Appendix 1. The model combines a discrete decision on self-restriction with a continuous effort choice in the framework of a sequential round-robin tournament.

³ Kahn (2000) and Bar-Eli et al. (2020) highlight the advantages of using sports data to test economic theories.

This article thus contributes to three strands of literature. First, it relates to work on the impact of heterogeneity in contests in general, starting with Rosen (1986), who clearly suggests that unbalanced match-ups result in lower effort levels. Theoretical work strongly supports this discouragement effect, including contributions by Baik (1994), Stein (2002), Szymanski (2003), or Konrad (2009). Intuitively, underdogs lower their willingness to invest when winning becomes too costly, whereas favourites respond by lowering their investment when the outcome appears predefined. Empirical evidence obtained in laboratory settings (e.g., Dechenaux et al., 2015; Hart et al., 2015) and from the field (e.g., Ehrenberg & Bognanno, 1990; Iqbal and Krumer, 2019; Sunde, 2009) similarly affirms these discouragement effects. We therefore expect that the investment of resources depends critically on heterogeneity in the competition.

Second, this paper contributes to a growing body of research on anticipating behaviour and inter-temporal effort provision in multi-period contests with fixed prize structures with theoretical framework provided by Szentes and Rosenthal (2003) and Konrad and Kovenock (2009). Laboratory setups show that participants in multi-stage contests exert effort in sub-optimal ways (e.g., Altmann et al., 2012; Mago and Sheremeta, 2017).

Third, and most directly, our study also relates to ‘spillover’ or ‘carryover’ effects in tournaments with heterogeneous players. While there is much evidence suggesting strategic behaviour in sport contests in general (e.g., Genakos & Pagliero, 2012; Malueg & Yates, 2010; Taylor & Trogdon, 2002), three papers are explicitly studying the effect of past and future opponents on players’ current effort levels. In their pioneer work, Harbaugh and Klumpp (2005) consider budget constraints in a tournament model and show that both underdogs and favourites have distinct incentives to reserve resources for upcoming battles; the underdogs benefit more from investing more of their resources in initial stages. Most existing field evidence comes from sport settings, where actions and performance are observable for a schedule (or tournament tree) that is known in advance. For instance, Harbaugh and Klumpp (2005) show that the introduction of a ‘rest day’ improves the performance of the favourites in NCAA basketball, indicating a shift in incentives to conserve resources. However, direct evidence of such strategic behaviour is missing. Brown and Minor (2014), using data from top-level tennis tournaments, find that the probability that the favourite wins the current stage decreases with the strength of an expected future opponent. They argue that taking the competitor’s ability in the next round into account changes the favourite’s valuation of the tournament and hence optimal effort provision.⁴Lackner et al. (2015) also find that the intensity of play (measured by personal fouls) increases when the expected relative ability of the next-stage contestant decreases in NBA and NCAA playoffs. Similar to Brown and Minor (2014), their approach builds on the idea that future opponents affect the probability of winning the tournament, or rather the continuation value, and thus the incentives to exert effort in the current stage.

⁴ Hill (2018) offers a replication study with data from NBA playoff matches.

The main novelty of our empirical approach is that we provide direct evidence of strategic action for achieving an optimal allocation of resources across the tournament. Existing articles use observable characteristics like scores (Brown & Minor, 2014) or personal fouls (Lackner et al., 2015) as discrete proxies for unobservable effort provision, which is a continuous decision. In contrast, we consider a discrete decision that requires self-restriction (in the line-up or by deliberately picking up a fifth yellow card), which is directly observable.

This article proceeds as follows: In Sect. 2, we explain how self-restriction works in the case of soccer and derive the main hypotheses. We empirically test them with field data in Sect. 3. Finally, Sect. 4 concludes.

2 Self-restriction in soccer

European soccer leagues are commonly designed as double round-robin tournaments, where all teams are matched twice in pairwise contests. For the final table, teams are ranked according to the number of points won, and prizes and rewards are distributed according to this ranking. Teams draw on one specific resource to accomplish their seasonal goals: players under contract. These players are not of unlimited availability, though, due to specific rules that sanction illegitimate behaviour.⁵ In particular, a decision to self-restrict in the German Bundesliga emerges from the ‘yellow card rule’, which mandates that a player who has received the fifth/tenth/fifteenth⁶ yellow card in a match t is suspended for match $t + 1$.^{7,8}

A team can anticipate that a player who has accumulated four (or nine) yellow cards up to match t —hereafter referred to as yellowplayers—will not be available soon⁹, so it has some leeway to decide which match $t + n$ such a player will (not) miss. Since yellowplayers typically are those fielded very frequently¹⁰ and hence are

⁵ Players may also be unavailable for other reasons such as injuries. However, since injuries cannot be used for strategic considerations, we do not take them into account in our analysis.

⁶ Only one player has accumulated 14 yellow cards (excluding yellow–red cards) in the course of a season in our data set. Overall, we observe 370 one-game suspensions due to a fifth yellow card and 26 suspensions due to a tenth yellow card in the five seasons from 2011/12 to 2015/16. On average, a team suffers 4.4 yellow card suspensions per season.

⁷ See Fédération Internationale de Football Association (FIFA) (2017) for a list of offenses that trigger a yellow card. After each season, the number of cards and penalties are reset. Furthermore, the accumulation of yellow cards counts within competition only, and a ban received in league games can be served in league games only. After serving this suspension, the player regains a ‘clean slate’ in terms of yellow cards. It is important to mention that the rules for the yellow cards ban differ considerably between leagues in terms of the exact number of cards and when to serve the suspension.

⁸ In addition to a one-game suspension due to the yellow card rule, players can be suspended for more than one game—up to four or more matches—due to serious rule violations (red card). Because red cards significantly weaken a team in the current and future matches (Bar-Eli et al., 2006), no team would deliberately risk a red card, and accordingly, we do not consider red card suspensions as strategic action.

⁹ About 72% of the players who received a fourth yellow card in match t receive the fifth yellow card within the following five matches. On average, it takes 4.5 matches for a player to receive a fifth yellow card (after a referee has shown the fourth yellow card).

¹⁰ Players who are suspended at least once in a season due to yellow cards play on average 74 min per game and 27 games per season. Non-suspended players play on average 42 min per game and 16 matches per season.

vital for their teams, we hypothesise that yellowplayers are not suspended 'randomly' but that teams strategically decide which match a yellowplayer will (not) miss, depending on current and future opponents. The league's schedule is publicly known in advance, so a team's decision to restrict its resources in match t or $t + 1$ crucially depend on the expected heterogeneity of match t or $t + 1$.

The strategic use of the yellow card rule can be exemplified by the case of two players from the Werder Bremen team who, in a March 2016 hearing, were accused of intentionally picking up yellow cards. Their team—struggling against relegation in the end of the season—has just claimed an important home victory against their direct competitor, Hannover 96. In the final phase of this match, the two midfielders were booked for 'tactical' fouls. Since it was their fifth and tenth yellow card in season, it was assumed that the players, one of whom was the team captain, provoked the bookings in the hope of avoiding the subsequent game against the league's dominant leader Bayern Munich (which Werder Bremen ultimately lost by a crushing 5–0 score). After serving the one-game suspension, the two players would have been allowed to play the following 'more important' and more winnable games. They later admitted that their plan had been arranged in advance.¹¹

This example offers three key takeaways. First, teams have incentives to let players at risk of a ban deliberately pick up a critical yellow card and take a pause when their resources are needed less. Such a scenario likely arises if an upcoming match $t + 1$ is sufficiently heterogeneous, such that the outcome is sufficiently certain.¹² Second, players at risk of picking up a one-game suspension due to a critical number of yellow cards are fielded in match t if both the teams have similar chances of winning the game—even though their probability of being unavailable in match $t + 1$ thus increases. Third, a 'strategic' yellow card also might be more likely near the end of the game, to minimise the risk that the player receives yet another yellow card within the same match. A second yellow card (or yellow–red card) results in an immediate sending-off and a suspension for match $t + 1$ without 'resetting' the number of yellow cards.¹³

In conclusion, because soccer players are central resources for accomplishing a team's goals, we expect teams to use their resources strategically. Players are not infinitely available—due to the yellow card rule—so teams must decide when to restrict themselves. According to our strategic considerations, a team's decision for or against self-restriction crucially depends on the heterogeneity of competition at time t and $t+1$. Thus, we expect two hypotheses to hold:

- (1) Teams self-restrict in t if the heterogeneity in match t is sufficiently strong.
- (2) Teams self-restrict in t if the heterogeneity in match $t + 1$ is sufficiently low.

¹¹ Picking up yellow cards intentionally is considered unsportsmanlike conduct, so the two players were fined 20,000 Euro by the German Football Association (DFB), c.f. <https://www.espn.com/soccer/german-bundesliga/story/2826982/clemens-fritz-and-zlatko-junuzovic-fined-over-bookings>.

¹² The use of strategic resting in multi-level contest in sports has been examined by Balsdon et al. (2007) and Raya (2015).

¹³ If a player picks up two yellow cards within a match, they do not count toward the number of yellow cards that accumulate and lead to a one-game suspension. Thus, if a player picks up a further yellow card—after a suspension due to a yellow–red card—he again is suspended for a match.

While these hypotheses are based on selective observations, they also result from a theoretical model of a stylised round-robin tournament, which we present in the Appendix.

3 Empirical analysis

The empirical analysis uses a rich data set from men's German top-level (Bundesliga) soccer, which comprises detailed information for all players on a game-by-game level and covers five seasons from 2011/12 to 2015/16. Each season is organised as a double round-robin tournament among 18 teams, so there are 34 games per team and season. The order of games is publicly known in advance of the season.

Our empirical approach is twofold. First, we focus on the starting line-up and examine the impact of heterogeneity of match t and $t + 1$ on the decision to field a yellowplayer or not. Second, we focus on yellow cards and investigate whether the incidence of receiving a fifth/tenth yellow card is related to the heterogeneity of match t or $t + 1$.

3.1 Measuring heterogeneity

Our analysis focuses on the impact of heterogeneity (in match t and $t + 1$) on a team's decisions to conserve resources in t and $t + 1$. We measure the heterogeneity of match t using players' market values. The data stem from the website www.transfermarkt.de. We calculate the average market value of each team at the start of each season, considering all players listed in the squad.¹⁴

Previous research has shown that market values have a high predictive power for outcomes of soccer games (e.g., Peeters, 2018). As average market values increase over the course of time, we use relative market values, that is we divide a team's average market value by the seasonal average of all teams. The heterogeneity of match t then is calculated as the (absolute) difference between the team's and the opponent's relative market values.

Betting odds are another established measure of heterogeneity¹⁵ in sport contests (e.g., Bartling et al. 2015; Deutscher et al., 2013; Sunde, 2009) and have proven to be an efficient forecasting instrument (see e.g. Forrest et al., 2005; Groll et al., 2015; Spann & Skiera, 2009). However, in our setting, betting odds have the disadvantage of being available only a few days prior to a match and thus only one game in advance. Consequently, the betting odds for match $t + 1$ may incorporate information about match t which are not available prior to match t , such as injuries

¹⁴ Note that the market value of a team may include players who leave the team in the winter break which may lead to an over- or underestimation of a team's playing quality. For this reason, we additionally calculated the average (relative) market value per team and matchday (starting line-up + substitutes). It shows that both measures are highly correlated ($r = 0.98$). We take this as an argument that mid-season transfers do not bias our results.

¹⁵ Heterogeneity based on betting odds typically is calculated by the (absolute) difference between a team's (implicit) winning and 'losing' probability in a given match (see Deutscher et al., 2013, for details).

or expulsions of certain players in match t . Therefore, closing odds for match $t + 1$ may differ from the expected heterogeneity of match $t + 1$ prior to match t . For this reason, we prefer market values as a more constant measure of heterogeneity between teams, acknowledging that there might be some inaccuracies due to time lags (Massey & Thaler, 2013). The fact that both heterogeneity measures are closely correlated ($r = 0.715$) suggests that potential biases are rather small.¹⁶ The results from regressions in which we calculate our heterogeneity measure from betting odds¹⁷ are qualitatively similar, see Appendix 3.

3.2 The starting line-up decision

In this section, we focus on a team's starting line-up and its decision to field a yellowplayer or not. We hypothesise that this decision crucially depends on the expected heterogeneity of matches t and $t + 1$. We consider a team's decision to *not* field such a player in match t as a case of self-restriction (see Fig. 1).

According to Hypothesis 1, we expect yellowplayers *not* to be fielded in t if the heterogeneity of game t is sufficiently large. In other words, teams may self-restrict and conserve restricted resources in match t if they are sufficiently sure they will win or lose this match. According to Hypothesis 2, players threatened by a suspension may be more (less) likely to be fielded in t if the subsequent match $t + 1$ is sufficiently unbalanced (balanced), because in this case, a potential suspension of such a player in match $t + 1$, caused by a fifth yellow card in t , would be less (more) harmful to the team.

We evaluate roster data for each team and match, which refer to information about a team's entire squad (i.e., starting players plus substitutes) on a particular matchday. Prior to each match, teams can nominate 18 players: 11 players for the starting line-up and up to 7 substitutes.¹⁸ A team's squad is always announced shortly before a game starts. After excluding the first five rounds¹⁹ and the final round of each season²⁰, goalkeepers²¹, players who were never a starter in the respective time period, and missings, we are left with 37,673 player-game observations (769 players and 1260 matches).

Table 1 shows that the number of yellowplayers increases over the course of a season: The share of players at risk of a ban amounts to 3.5% in the first half of the season (matchdays 6–17) and 12% in the second half (matchdays 18–33).

¹⁶ Furthermore, there is a high correlation between (relative) market values and average winning probabilities per team and season ($r = 0.924$).

¹⁷ These data stems from the website www.betexplorer.com.

¹⁸ In 97.25% of the cases, a team nominates the maximum number of players (18).

¹⁹ Players would have to accumulate five yellow cards for a suspension, so they cannot hit this mark prior to matchday five. Since we observe the first yellowplayer at matchday 6, we exclude all prior observations, i.e. matchdays 1 to 5, from our estimations.

²⁰ A suspension due to an accumulation of five yellow cards occurring in the last game of a season does not carry over to the next season.

²¹ The importance of the goalkeeper position typically leads to few or no strategic variations. Furthermore, there are only 42 observations of goalkeepers in danger of receiving a fifth yellow card within the sample.

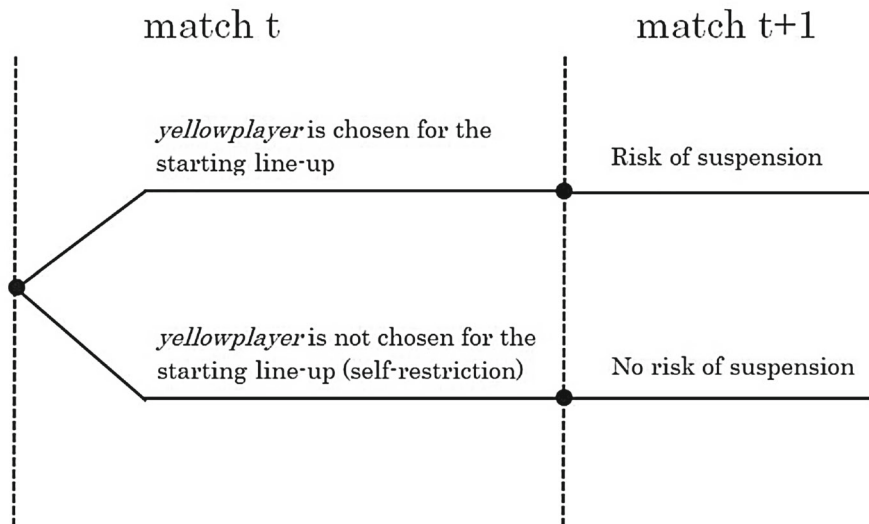


Fig. 1 Timing of events (starting line-up decision)

We estimate a logit model to investigate the impact of heterogeneity in games t and $t + 1$ on a team's decision to field a yellowplayer:

$$\begin{aligned} \text{starting11}_{i,t} = & \beta_0 + \beta_1 \text{yellowplayer}_{i,t} + \beta_2 \text{Het}_{i,t} + \beta_3 \text{Het}_{i,t+1} + \beta_4 \text{yellowplayer}_{i,t} \cdot \text{Het}_{i,t} \\ & + \beta_5 \text{yellowplayer}_{i,t} \cdot \text{Het}_{i,t+1} + \gamma' \mathbf{X} + \varepsilon_{i,t}. \end{aligned} \quad (1)$$

The dependent variable, $\text{starting11}_{i,t}$, is a binary outcome measure that takes a value of 1 if player i starts in match t and 0 otherwise, meaning this player is left on the bench.²² Our main variable of interest is the binary variable yellowplayer , such that $\text{yellowplayer} = 1$ indicates a player with a critical number of yellow cards (4, 9, or 14) prior to match t , and $\text{yellowplayer} = 0$ otherwise.

To isolate the effects of heterogeneity on the decision to choose a yellowplayer for the starting eleven (captured by β_4 and β_5), we need to control for other factors that could affect the dependent variable. Therefore, \mathbf{X} is a vector of player- and game-specific control variables, including a proxy for a player's importance to a team (% of minutes played), a player's age and position, the team size (roster), game attendance (attendance), a dummy variable signalling regional rivalry games (derby), a dummy indicating whether the match is played away (away), a variable indicating the overall quality of both teams represented by the sum of the teams' rankings prior to a matchday (quality), referee dummies, coach dummies, and data

²² We also do not include players who are completely excluded from the roster, i.e. those who are not nominated for a match, in our analysis, because we expect that the exclusion of a player from a match is not due to strategic considerations but for other reasons, such as injuries and disciplinary issues.

Table 1 Yellowplayer-observations per matchday

Matchday	Yellowplayer		Total
	0	1	
6	1335	6	1341
7	1332	11	1343
8	1327	15	1342
9	1306	24	1330
10	1293	30	1323
11	1297	43	1340
12	1297	53	1350
13	1281	63	1344
14	1280	67	1347
15	1271	79	1350
16	1265	83	1348
17	1252	94	1346
1st half	15,536	568	16,104
18	1180	95	1275
19	1232	108	1340
20	1204	132	1336
21	1213	143	1356
22	1205	155	1360
23	1213	145	1358
24	1211	149	1360
25	1206	152	1358
26	1198	158	1356
27	1193	166	1359
28	1173	184	1357
29	1159	189	1348
30	1174	194	1368
31	1152	195	1347
32	1142	204	1346
33	1150	195	1345
2nd half	19,005	2564	21,569
Total	34,541	3132	37,673

Table 2 Descriptive statistics—starting11

Variable	Mean	SD	Min	Max
Starting11	0.638		0	1
Yellowplayer	0.083		0	1
Het _{<i>t</i>}	0.888	1.040	0.001	4.265
Het _{<i>t</i>+1}	0.893	1.045	0.001	4.265
% of minutes played	0.549	0.306	0	1
Roster	33.239	3.841	25	44
Derby	0.065		0	1
Attendance	43.651	16,909	13,500	81,359
Quality	18.914	7.365	3	35
Matchday	19.536	8.078	6	33
Away	0.500		0	1
Age	25.859	3.779	17.344	37.593
Left/right defender	0.180		0	1
Left/right wing player	0.123		0	1
Left/right/central midfield	0.127		0	1
Forward	0.184		0	1
Centre back	0.210		0	1
Offensive midfield	0.062		0	1
Defensive midfield	0.113		0	1

$N = 37,673$

about the matchday (matchday).²³ We measure a player's importance by the percentage of minutes that player is on the field in matches prior to a match t . The more minutes he plays prior to match t , the more important he is for a team, and, hence, the more likely it is that he is in the starting line-up in match t . Moreover, α_i controls for unobserved player-specific effects. Finally, $\varepsilon_{i,t}$ is the error term that captures all other unobserved factors that affects starting11.

Table 2 contains the descriptive statistics for the main variables. Note that yellowplayer = 1 applies to a total of 3,277 observations for 377 different players.

Table 3 presents the estimation results from our preferred model defined in Eq. (1) (columns (1)–(3)). As a robustness check, we add player fixed effects to the regression model (columns (4)–(6)). Yet, this means that the coefficient estimates solely depend on the variation of the dependent and independent variables within players. So, given the rather small number of 'yellowplayer situations' per player

²³ The roster size reflects the number of (potential) alternatives for a suspended player and therefore could affect strategic behaviour. That is, the larger the set of alternatives, the lower the 'costs' of a suspension. A derby is a game between long-time rivals, often located in close geographical proximity. Players and coaches potentially evaluate those games higher than other games, with possible consequences for player suspensions. The matchday (6–33) is included to account for the stage of the competition. Moreover, as documented in Table 1, the number of players on the verge of being suspended increases as the season proceeds. Attendance numbers capture the atmosphere in a stadium, which can impact decision making (e.g. Dohmen, 2008). We thereby assume that teams have rational expectations about the expected number of spectators because of the high stadium utilisation in Germany and the wealth of accessible information such as advance ticket sales.

Table 3 Self-restriction and the starting line-up decision—logit regressions

	(1)	(2)	(3)	(4)	(5)	(6)
Yellowplayer	0.1947*** (0.0651)	0.1947*** (0.0651)	0.2301*** (0.0739)	0.1661** (0.0725)	0.1660** (0.0725)	0.1965** (0.0809)
Het _t	0.0378*** (0.0104)	0.0379*** (0.0104)	0.0372*** (0.0104)	0.0373** (0.0190)	0.0355* (0.0192)	0.0349* (0.0192)
Yellowplayer * Het _t	- 0.1075** (0.0513)	- 0.1075** (0.0513)	- 0.0959* (0.0506)	- 0.0994* (0.0526)	- 0.0994* (0.0526)	- 0.0903* (0.0536)
Het _{t+1}		0.0007 (0.0068)	0.0047 (0.0078)		- 0.0117 (0.0165)	- 0.0083 (0.0170)
Yellowplayer * Het _{t+1}			- 0.0545 (0.0505)			- 0.0451 (0.0527)
% of minutes played	4.0121*** (0.0580)	4.0121*** (0.0580)	4.0127*** (0.0580)	3.5540*** (0.0651)	3.5534*** (0.0651)	3.5537*** (0.0651)
Match-specific controls	Yes	Yes	Yes	Yes	Yes	Yes
Player-specific controls	Yes	Yes	Yes	No	No	No
Referee FE	Yes	Yes	Yes	Yes	Yes	Yes
Player FE	No	No	No	Yes	Yes	Yes
Coach FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	37,673	37,673	37,673	37,673	37,673	37,673
Pseudo-R2	0.2141	0.2141	0.2141	0.1203	0.1203	0.1203

This table includes player-match-level data, that is all players who belong to a team's squad (starting line-up + bench) on a particular matchday

Logit regressions, dependent variable: *starting11*

Match-specific controls: derby, attendance, quality, away, matchday, roster

Player-specific controls: age, age², position-dummies (left/right defender, left/right wing player, left/right/central midfield, forward, centre back, offensive midfield, defensive midfield)

(Robust) standard errors (clustered at the match level, column 1–3) are in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

along with variations of Het_t and Het_{t+1}, this approach should be seen as a second-best option.

Our results indicate three main conclusions. First, the percentage of minutes played prior to t increases the probability of being fielded in t . This finding is not surprising; the variable % of minutes played was designed explicitly to proxy for the player's importance to the team. Second, the estimated coefficient of the first interaction term (yellowplayer * Het_t) $\hat{\beta}_4$ is negative and differs significantly from zero at the 5% (columns (1) and (2)) and 10% significance level (columns (3)–(6)). That is, the tendency to choose a yellowplayer as a starter decreases with the heterogeneity of the current competition. In other words, the player is protected from the risk of suspension when the match appears to be decided in advance. We take this result as first (yet not strong) evidence in favour of our Hypothesis 1. To ease interpretation, Table 7 in the Appendix presents OLS estimates with standardised heterogeneity measures (z -score). A one standard deviation increase

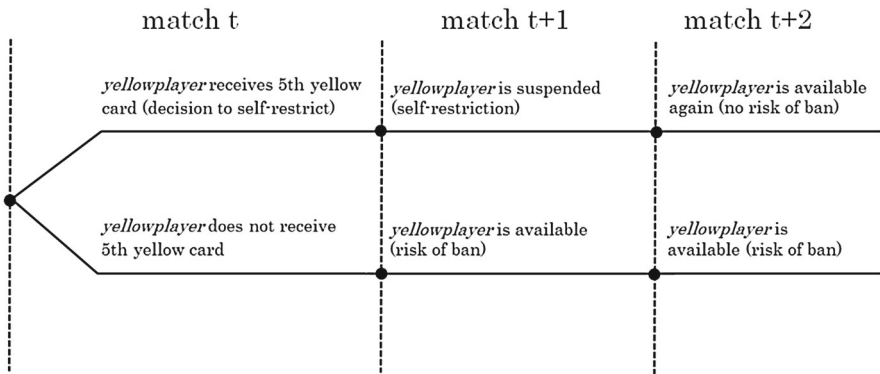


Fig. 2 Timing of events (decision to deliberately receive a fifth yellow card)

of Het_t is associated with a decrease of being in the starting eleven by 1.6% (2.5% when evaluated at the sample mean) for yellowplayers. As a complementary effect, players not at risk of a yellow-card suspension are more likely to start in match t when this match is more heterogeneous (Het_t , columns (1)–(3)). Taken together, these findings can easily be linked with the discouragement effect, as we discuss in Sect. 4.

Third, we do not find similar effects for the next round, which means that the heterogeneity of the upcoming match $t + 1$ does not affect a team's decision (rows five and six). We refrain from taking this result as a disproof of Hypothesis 2 though, because assuming that teams put a player in the starting lineup in t just to provoke a ban in $t + 1$ would probably undervalue the importance of match t . Consequently, we proceed with a more subtle approach to study anticipating behaviour and self-restriction in round $t + 1$ in the next section.

3.3 The decision to receive yellow cards

In the second part of our analysis, we seek to provide evidence of strategic self-restriction in a setting that demands anticipating behaviour. In detail, we examine whether heterogeneity in upcoming rounds influences the probability of receiving a critical yellow card followed by a suspension.²⁴ In this context, we have to separate the decision to self-restrict from being self-restricted, because they refer to different points in time: Being self-restricted in $t + 1$ results from the decision to receive a fifth yellow card in match t (see Fig. 2). Hence, Hypothesis 1 states that teams are more likely to be self-restricted in match $t + 1$ (due to a fifth/tenth yellow card in t) if the heterogeneity in match $t + 1$ is sufficiently strong. Hypothesis 2 implies that self-restriction in match $t + 1$ is less likely if the heterogeneity in match $t + 2$ is sufficiently strong. Intuitively, this means that a team tends to self-restrict in $t + 1$

²⁴ Note that injuries or disciplinary issues do not play a role in this setting.

Table 4 Descriptive statistics—yellow card

Variable	Mean	Std. Dev.	Min	Max
Yellow	0.149		0	1
Het _{<i>t</i>}	0.885	1.036	0.001	4.265
Het _{<i>t</i>+1}	0.889	1.043	0.001	4.265
Het _{<i>t</i>+2}	0.891	1.044	0.001	4.265
Yellowplayer	0.095		0	1
% of minutes played	0.626	0.287	0	1
Away	0.502		0	1
Matchday	19.007	7.796	6	32
Roster	33.248	3.837	25	44
Quality	18.931	7.405	3	35
Attendance	43,490	16,938	13,500	81,359
Derby	0.067		0	1
Age	26.009	3.666	17.361	37.580
Left/right defender	0.174		0	1
Left/right wing player	0.126		0	1
Left/right/central midfield	0.123		0	1
Forward	0.184		0	1
Centre back	0.209		0	1
Offensive midfield	0.063		0	1
Defensive midfield	0.120		0	1

$N = 29,555$

when it is important to have full resources in $t + 2$ as this match is expected to be tight.

To test our hypotheses, it is necessary to further restrict our sample. First, we exclude players without any playing time and players who were never ‘booked’. Second, as we now rely on information from the match after the next, we also have to discard those observations where this information is not available. The final sample covers 29,555 observations.

The dependent variable (yellow) is a dummy that indicates whether a player i has received an yellow card in match t . We suggest the following empirical model:

$$\begin{aligned} \text{yellow}_{i,t} = & \beta_0 + \beta_1 \text{yellowplayer}_{i,t} + \beta_2 \text{Het}_{i,t} + \beta_3 \text{Het}_{i,t+1} + \beta_4 \text{Het}_{i,t+2} \\ & + \beta_5 \text{yellowplayer}_{i,t} \cdot \text{Het}_{i,t+1} + \beta_6 \text{yellowplayer}_{i,t} \cdot \text{Het}_{i,t+2} + \gamma' \mathbf{X} + \varepsilon_{i,t}, \end{aligned} \quad (2)$$

which closely resembles Eq. (1) in Sect. 3.2 but also includes round $t + 2$.²⁵ Table 4 presents the descriptive statistics.

²⁵ In addition, one might think of including the score as a further control variable. However, this would necessitate a switch from match to minute of game as unit of observation. Since our main explanatory variables are constant for each match and thus would lose their explanatory power in a minute-by-minute analysis, we do not follow this idea.

Table 5 The effect of heterogeneity on yellow card suspensions—logit regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	> 45 min	> 60 min	> 75 min	All	> 45 min	> 60 min	> 75 min
Yellowplayer	- 0.1561* (0.0869)	- 0.2097** (0.1056)	- 0.2367* (0.1231)	- 0.2651 (0.1661)	- 0.4094*** (0.0879)	- 0.4348*** (0.1099)	0.4477*** (0.1260)	- 0.4390*** (0.1665)
Het _t	- 0.0991*** (0.0261)	- 0.0999*** (0.0306)	- 0.1255*** (0.0354)	- 0.1619*** (0.0462)	- 0.0972*** (0.0254)	- 0.1000*** (0.0311)	- 0.1254*** (0.0362)	- 0.1640*** (0.0483)
Het _{t+1}	- 0.0185 (0.0199)	- 0.0176 (0.0244)	- 0.0018 (0.0274)	- 0.0004 (0.0349)	- 0.0148 (0.0223)	- 0.0134 (0.0269)	- 0.0038 (0.0302)	- 0.0039 (0.0386)
Het _{t+2}	0.0296 (0.0189)	0.0386* (0.0235)	0.0541** (0.0271)	0.0602* (0.0327)	0.0351* (0.0213)	0.0435* (0.0254)	0.0572** (0.0286)	0.0652* (0.0364)
Yellowplayer *	0.1184* (0.0615)	0.1712** (0.0755)	0.2093** (0.0825)	0.2464** (0.1135)	0.0991* (0.0567)	0.1470** (0.0681)	0.1946*** (0.0742)	0.2273** (0.0933)
Yellowplayer *	- 0.1101* (0.0631)	- 0.1903** (0.0765)	- 0.2110** (0.0830)	- 0.3083** (0.1215)	- 0.1362** (0.0636)	- 0.2195*** (0.0836)	- 0.2281** (0.0948)	- 0.3294*** (0.1323)
% of minutes played	0.2477*** (0.0610)	0.1135 (0.0731)	0.0337 (0.0841)	- 0.1795 (0.1102)	0.4977*** (0.0877)	0.4324*** (0.1056)	0.3958*** (0.1204)	0.2339 (0.1534)
Match-specific controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Player-specific controls	Yes	Yes	Yes	Yes	No	No	No	No
Referee FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Coach FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Player FE	No	No	No	No	Yes	Yes	Yes	Yes
Obs.	29,555	27,095	25,706	22,719	29,555	27,095	25,706	22,719

Table 5 continued

(1) All	(2) > 45 min	(3) > 60 min	(4) > 75 min	(5) All	(6) > 45 min	(7) > 60 min	(8) > 75 min
Pseudo-R2	0.0230	0.0205	0.0209	0.0129	0.0148	0.0169	0.0229

Player-match-level data (all fielded players)

Logit regression, dependent variable: yellow

Matchspecific controls: derby, attendance, quality, roster, matchday, away

Playerspecific controls: age, age², position-dummies (left/right defender, left/right wing player, left/right/central midfield, forward, centre back, offensive midfield, defensive midfield)

Models 2–4 and 6–8 refer to the last 45/30/15 min of a match, so observations of bookings in prior minutes are excluded

(Robust) standard errors (clustered at the match level, column 1–4) are in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The results from the logit regressions are presented in Table 5. Again, as a robustness check, player fixed effects were added to our preferred model (columns (5)–(8)). Our main finding is that the estimated β_5 differs significantly from zero, meaning that the probability of reaching the limit of five yellow cards increases with the heterogeneity of the next round's competition ($\text{yellowplayer} * \text{Het}_{t+1}$), in accordance with our Hypothesis 1. The result points to anticipating behaviour among competitors in a multi-stage tournament setting. Moreover, we find significant effects of the heterogeneity of the match after the next one on a yellowplayer's probability to get booked ($\text{yellowplayer} * \text{Het}_{t+2}$). The likelihood of being suspended for the next match decreases with increasing heterogeneity of the match after next, which is in line with our Hypothesis 2.

Again, we demonstrate economic significance using OLS regressions and standardised values of Het_t , Het_{t+1} and Het_{t+2} in Table 11 in Appendix 3. A one standard deviation increase of Het_{t+1} (Het_{t+2}) increases (decreases) the likelihood of getting a 'booking' for yellowplayers by around 1.5 percentage points (or 11% when evaluated at the sample mean).

Despite our set of controls, these findings do not establish direct proof of strategic behaviour though. Yet, if players provoke yellow card suspensions on purpose, we expect them to do so at the end of a game to minimise the risk of a second yellow card within the same game. We observe exactly this trend, such that when we split the sample according to playing time remaining (columns (2) to (4) and (6) to (8)), it becomes apparent that the estimated coefficients of the interaction term (fifth row) increases toward the end of a 90-minute game. In addition, we estimate a simple model of substitution where *substituted* is a binary variable equal to one if a player is substituted, and zero otherwise. The sample is restricted to yellowplayers only. Results suggest that the likelihood of being substituted after getting a fifth yellow card increases significantly in the heterogeneity of the next match (see Table 13 in Appendix 3).

We take both findings as evidence in favour of deliberate and strategic behaviour.

4 Conclusion and implications

This article provides evidence of strategic investment decisions in anticipation of the future need for resources. This empirical analysis of German Bundesliga soccer indicates that anticipated heterogeneity in contests decreases the willingness to invest resources. Players therefore, self-restrict in current and future competition when the heterogeneity between contestants is sufficiently large. Players at risk of a ban tend to be excluded from the starting eleven when the contest is unbalanced. Furthermore, players are more likely to receive a crucial yellow card that triggers a ban if the subsequent contest (which they will miss due to the ban) is sufficiently unbalanced. Finally, players also tend to receive a crucial yellow card if the second to next game (for which they return from their ban) is rather homogeneous.

These findings reflect a discouragement effect such that effort provision in contests falls short of what might be expected—considering the prize at stake—due to lopsided competition. We demonstrate that the decision to self-restrict is affected

not only by the strength of the current opponent but also by the anticipated strength of the future competitor in the next stage of the tournament. Thus, our study not only expands the literature on ‘spillover’ or ‘carryover’ effects in tournaments with heterogeneous players, but also provides direct evidence for strategic behaviour in the field.

Our results have important policy implications for soccer itself but also for other contests like promotion tournaments or political campaigns. In multi-stage contests with multiple contestants, heterogeneity is detrimental to mediocre contestants. The best contestants can save resources against the weakest contestants without substantially lowering their chances of winning. Correspondingly, the weakest contestants save their resources in matches when playing the strongest opponents, because their *ex ante* chances of winning are very low. Abilities, and thus winning probabilities, are more balanced against mediocre teams, such that neither the strongest nor the weakest contestants conserve resources and instead meet them at full strength. Mediocre contestants cannot afford to self-restrict in any contest without risking a negative impact on the outcome. Thus, the structure of the contest and the possibility to self-restrict for the next competition creates a disadvantage for mediocre teams and has a stabilising effect on the top-level hierarchy. The schedule itself could be disadvantageous to some competitors too, especially if they face an opponent that most recently played in a heterogeneous match-up (Krumer and Lechner 2018). For the case of soccer (or any other sports played in league system), the findings are also critically important. The prize distribution is typically top-heavy, with the very top performers receiving disproportionately large parts of the prize at stake. The findings presented in this paper clearly indicate a disadvantage to mediocre teams and result in an entry barrier to join the top level and receive the highest awards.

Tournament organisers could counteract these effects by making self-restriction advantages less pronounced. First, in the case of soccer, they could change the rules, such that a crucial fifth yellow card would lead to a randomly drawn ban some time in the next five matches. Then the strategic element of self-restriction would diminish, because contestants would not know in advance whether they would lose their valuable resources at exactly the moment they do not need it. In the same way, penalties like in Ice Hockey where the offending player is sent off the field for a set number of minutes instead of yellow cards would prevent self-restricting behaviour. Second, schedule imbalances can be lowered if the round-robin format repeats multiple times (as in the Bundesliga, with two rounds of round-robin). Rearranging the sequence of games after each round then would prevent teams from facing opponents that profit from self-restriction because their last match was unbalanced.

Appendix 1: Theoretical model

We introduce the possibility of self-restriction into the model of strategic effort choices in round-robin tournaments with three heterogeneous players. The players may restrict their resources in the first match to increase the probability of having unrestricted resources in their second match. Whether a player uses this option in

equilibrium depends on the extent of heterogeneity between that player and the opponents in the first and second match.

Assumptions

The structure of the tournament

A *round-robin tournament*, also referred to as an all-play-all tournament, matches each participant with each other participant in a pairwise contest, ranks them according to the number of matches won, and awards prizes according to this ranking. For simplicity, we focus on round-robin tournaments with three risk-neutral players and consider an exogenous sequence in which player 1 is matched with player 2 in the first match, player 1 is matched with player 3 in the second match, and player 2 is matched with player 3 in the third match (Krumer et al., 2017; Sahn, 2019).²⁶

If a player earns two victories, he is ranked first, and the player with one victory is ranked second; if there is a tie because each player has won one match, the ranks are assigned randomly with equal probabilities of 1/3 for each player and rank. The player ranked first (last) receives the first (last) prize, the value of which is identical for all players and normalised to 1 (0). The value of the second prize is assumed to be identical for all players as well and fixed at 1/2, or half of the first prize. This prize structure (is the only one that) ensures a fair tournament, in the sense that all potential differences in players' equilibrium winning probabilities and expected payoffs result from differences in their abilities, not from their position in the sequence of matches (Laica et al. 2017).

The structure of the resulting sequential game with its $2^3 = 8$ potential courses is depicted in Fig. 3. The seven nodes $k \in \{A, \dots, F\}$ represent all combinations for which the ranking of the tournament has not been fully determined when the respective match starts.

Tournament matches

Each match of the tournament is organised as a lottery contest between two potentially heterogeneous players, A and B , with linear costs of effort.²⁷ Specifically, player A 's probability of winning match $k \in \{A, \dots, F\}$ is

²⁶ Apart from renumbering players, this exogenous sequence is unique. The use of exogenous sequences is prevalent in sports tournaments. Sahn (2019), Krumer et al. (2020), and Laica et al. (2017) also discuss endogenous sequences in which the outcome of the first match determines the order of the two remaining matches.

²⁷ The non-perfectly discriminating lottery contest offers concreteness and accounts for random elements that appear in most sport competition. Sahn (2019) and Laica et al. (2017) consider round-robin tournaments with general Tullock contests, including the limit case of all-pay auctions, as studied by Krumer et al. (2017, 2020). The alternative assumption of zero effort costs but limited resources in round-robin tournaments with three players implies a quasi-simultaneous structure with multiple equilibria (Dagaev & Zubanov, 2017) and, thus, is not a fruitful modeling option in our context.

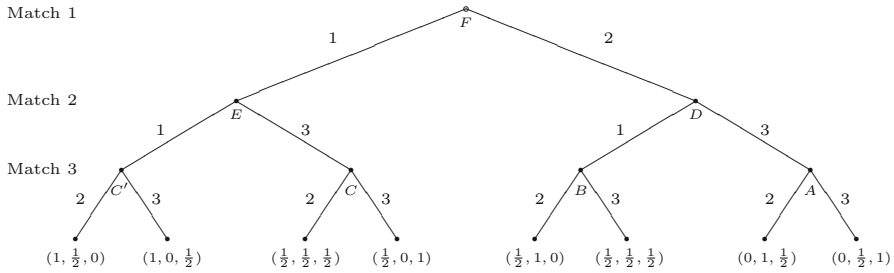


Fig. 3 Three-player round-robin tournament with exogenous sequence

$$p_A^k = \begin{cases} 1/2 & \text{if } x_A^k = x_B^k = 0, \\ \frac{\theta_A^k x_A^k}{\theta_A^k x_A^k + \theta_B^k x_B^k} & \text{else,} \end{cases}$$

where θ_i^k describes ability, and x_i^k denotes the effort of player $i \in \{A, B\}$ in match k (Leininger 1993; Baik 1994).

Player A chooses $x_A^k \geq 0$ to maximise the expected payoff

$$E_A^k = p_A^k(w_A^k - x_A^k) + (1 - p_A^k)(\ell_A^k - x_A^k), \tag{3}$$

where w_i^k denotes player i 's expected continuation payoff from winning match k , and ℓ_i^k denotes the expected continuation payoff from losing match k , with $w_i^k \geq \ell_i^k \geq 0$ for $i \in \{A, B\}$. For $w_A^k = \ell_A^k$, the optimal choice is $x_A^k = 0$ for any $x_B^k \geq 0$. If $x_A^k = 0$ and $w_B^k > \ell_B^k$, player B has no best response unless there is a smallest monetary unit $\varepsilon > 0$, in which case the best response is $x_B^k = \varepsilon$. As $\varepsilon \rightarrow 0$, in the limit, $x_B^k \rightarrow 0$, and $p_B^k \rightarrow 1$. Otherwise, there is a unique Nash equilibrium in pure strategies (e.g., Cornes & Hartley, 2005). The equilibrium effort levels can be derived from the following necessary conditions:

$$\frac{\partial E_i^k}{\partial x_i^k} = \frac{\theta_i^k \theta_j^k x_j^k}{(\theta_i^k x_i^k + \theta_j^k x_j^k)^2} (w_i^k - \ell_i^k) - 1 = 0,$$

yielding

$$x_i^k = \frac{\theta_i^k \theta_j^k (w_i^k - \ell_i^k)^2 (w_j^k - \ell_j^k)}{[\theta_i^k (w_i^k - \ell_i^k) + \theta_j^k (w_j^k - \ell_j^k)]^2} \tag{4}$$

for $i, j \in \{A, B\}$ with $i \neq j$. The resulting equilibrium winning probabilities equal

$$p_i^k = \frac{\theta_i^k (w_i^k - \ell_i^k)}{\theta_i^k (w_i^k - \ell_i^k) + \theta_j^k (w_j^k - \ell_j^k)}. \tag{5}$$

Inserting Eqs. (4) and (5) into Eq. (3) yields the expected equilibrium payoffs:

$$E_i^k = \ell_i^k + \frac{(\theta_i^k)^2 (w_i^k - \ell_i^k)^3}{[\theta_i^k (w_i^k - \ell_i^k) + \theta_j^k (w_j^k - \ell_j^k)]^2}. \quad (6)$$

The possibility of self-restriction

In practice, most tournaments offer the possibility that players restrict their current resources to increase the probability of having unrestricted resources available in future matches. In team sports, such self-restriction might involve giving particular important players a break, so they can rest for a future contest. As a new element, we now introduce this self-restriction opportunity into the above tournament model. To keep the analysis tractable, we make the following assumptions.

Players differ in their basic abilities.²⁸ Let $\theta_i > 0$ denote the basic ability of player $i \in \{1, 2, 3\}$. Before the tournament starts, players 1 and 2 make a binary choice to restrict themselves (R) or not (NR) in their first match (match 1). If player $i \in \{1, 2\}$ chooses R, his ability in the first match will be restricted to $\theta_i^F = r\theta_i$ with some $r \in (0, 1)$, but he will have unrestricted ability $\theta_i^k = \theta_i$ in the second match ($k \in \{D, E\}$ if $i = 1$, and $k \in \{A, B, C, C'\}$ if $i = 2$). In contrast, if player $i \in \{1, 2\}$ chooses NR, his ability in the first match will be unrestricted, $\theta_i^F = \theta_i$, but he faces some positive probability $\pi \in (0, 1)$ of having to compete with restricted ability in the second match. To capture the related uncertainty in reduced form, we assume that the player's (expected) ability in the second match equals $\theta_i^k = q\theta_i$, with $q = 1 - \pi(1 - r) \in (r, 1)$. For simplicity, we assume that player 3 has unrestricted ability in both matches²⁹ which is normalised such that $\theta_3^k = \theta_3 = 1$ for all $k \in \{A, \dots, E\}$.³⁰

This round-robin tournament with three players and an opportunity of self-restriction (for players 1 and 2) constitutes a sequential game Γ with four stages: In Stage 0, players 1 and 2 decide simultaneously whether to self-restrict or not in match 1; their choices are observed by all players. In Stage 1, players 1 and 2 decide simultaneously about their effort in match 1, and the outcome of match 1 is observed by all players. In Stage 2, players 1 and 3 decide simultaneously about their effort in match 2, and the outcome of match 2 is observed by all players.

²⁸ In the present framework of a Tullock contest with linear effort costs, heterogeneity in players' abilities is equivalent to heterogeneity in their (linear) costs of effort, as well as to heterogeneity in their valuations of the prizes (e.g., Cornes & Hartley, 2005; Ryvkin, 2013).

²⁹ Allowing player 3 to self-restrict considerably complicates the analysis, due to second-order strategic incentives: The self-restriction of one or both players in the first match may trigger/prevent the self-restriction of player 3 in the second match. The first-order strategic incentives of players 1 and 2 for restricting themselves remain present and do not fundamentally change though, because in the subgame perfect equilibrium of the game, players 1 and 2 correctly anticipate the (expected) ability of player 3 as their upcoming opponent.

³⁰ Fixing one of the three abilities is possible without loss of generality, because only relative abilities matter.

Finally, in Stage 3, players 2 and 3 decide simultaneously about their effort in match 3.

Results

For each feasible combination of exogenous parameters $(\theta_1, \theta_2, r, q)$, the game Γ can be solved by backward induction for its subgame perfect equilibrium (SPE) through repeated use of Eqs. (4)–(6). Appendix 2 illustrates this procedure for the example of $(\theta_1, \theta_2, r, q) = (1, 1, 1/2, 3/4)$.

We are particularly interested in identifying the conditions in which players 1 or 2 choose self-restriction in the first match as part of an equilibrium strategy. Because the comparative statics are analytically not tractable, we study the effect of variations of the exogenous parameters $(\theta_1, \theta_2, r, q)$ on equilibrium behaviour numerically. We use *Microsoft Excel* to compute the SPE of Γ on a grid with more than 200,000 grid points, increasing θ_i from 0.1 to 10 in steps of 1% (and from 0.001 to 1000 in steps of 10%, respectively) for $i \in \{1, 2\}$, together with r and q from 0.5 to 0.9 in steps of 0.1, subject to $r < q$.³¹ The calculations demonstrate that, depending on the parameters, all kinds of equilibrium behaviours by players 1 and 2 may arise in the first stage.

Proposition 1 For each of the following combinations of first-stage behaviour by players 1 and 2, there are parameters $(\theta_1, \theta_2, r, q)$, such that the respective combination is part of the players' strategies in the SPE of game Γ :

- (a) No player chooses self-restriction R.
- (b) Only the weaker player chooses self-restriction R.
- (c) Both players choose self-restriction R.
- (d) Only the stronger player chooses self-restriction R.
- (e) Both players choose self-restriction R with positive probability (equilibrium in mixed strategies).

Figure 4 illustrates Proposition 1. The players' behaviour only depends on their relative basic abilities, such that the diagrams are symmetric around the 45°-line on which $\theta_1 = \theta_2$. We thus can focus on cases in which player 1 is at least as able as player 2, with grid points on and below the 45°-line. Moreover, the left panel of Fig. 4 encompasses a range of parameters in which the maximum difference in relative abilities of players 1 and 2 is 1:1,000,000, whereas, the right panel zooms in on the practically more relevant range in which this maximum difference is 1:100.

Around the 45°-line, players 1 and 2 have similar basic abilities, and they never self-restrict. Self-restriction by one or both players occurs only if the difference between their abilities θ_1 and θ_2 is sufficiently large.³² Therefore,

³¹ The spreadsheets are available on request.

³² More precisely, *ceteris paribus*, the greater the ability of the stronger (weaker) player, the more (less) likely it is that the weaker (stronger) player self-restricts.

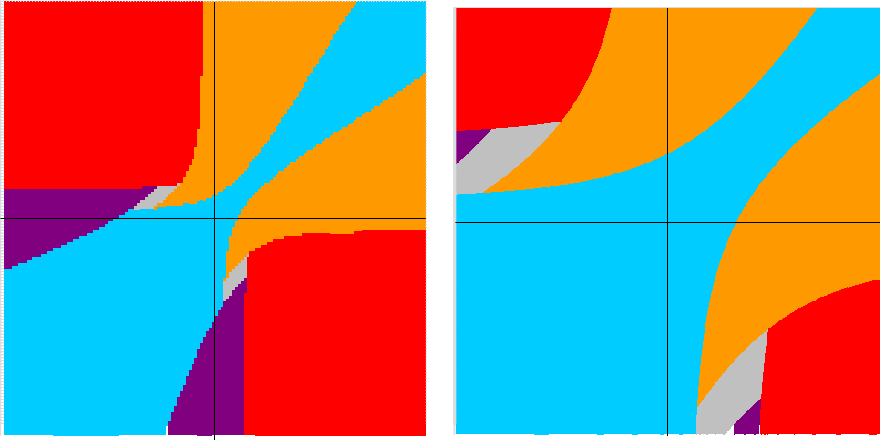


Fig. 4 Comparative statics w.r.t. θ_1 and θ_2 . *Note:* This figure illustrates the effects of variations in θ_1 and θ_2 . On the horizontal (vertical) axis of the left diagram, θ_1 (θ_2) increases from 0.001 to 1000 in steps of 10% for normalised $\theta_3 = 1$. The right diagram zooms in, such that θ_1 (θ_2) increases from 0.1 to 10 in steps of 1%, again for normalised $\theta_3 = 1$. In both diagrams, the vertical (horizontal) bold line represents $\theta_1 = 1$ ($\theta_2 = 1$). Each colour represents a certain type of equilibrium behaviour of players 1 and 2 in the first stage: ● blue $\hat{=}$ no player chooses R. ● orange $\hat{=}$ only the weaker player chooses R. ● red $\hat{=}$ both players choose R. ● violet $\hat{=}$ only the stronger player chooses R. ● grey $\hat{=}$ both players choose R with positive probability (equilibrium in mixed strategies)

Hypothesis 1 *The larger the difference in basic abilities θ_1 and θ_2 in the current match, the more likely a player chooses self-restriction.*

The underlying intuition is that players cannot afford to restrict themselves in their first match if the match will be close. Only if they are sufficiently sure they will win or lose, due to a large gap in abilities, will players conserve their strength for their second match.

The scale on both axes of the diagrams in Fig. 4 also is exponential, such that the ratio of the basic abilities of players 1 and 2 is constant along any parallel to the 45°-line. Moving from the lower left to the upper right on any such parallel, the ability of player $i \in \{1, 2\}$ increases relative to the ability of player 3. Because the region below the 45°-line in which player $i \in \{1, 2\}$ self-restricts is convex in the relevant range (see the right panel of Fig. 4), there is some ratio α_i of the basic abilities of players 1 and 2 for which the corresponding parallel $\theta_1 = \alpha_i \theta_2$ is tangential to the region in which player $i \in \{1, 2\}$ self-restricts. In this sense, the relative ability $\delta_i = \theta_i / \theta_3$ that characterises the tangent point is that for which player i 's probability of self-restriction is maximal (see Fig. 5). These observations suggest:

Hypothesis 2 *The greater the difference in (weighted) basic abilities θ_i and $\delta_i \theta_3$ in the future match, the less likely player $i \in \{1, 2\}$ chooses self-restriction.*

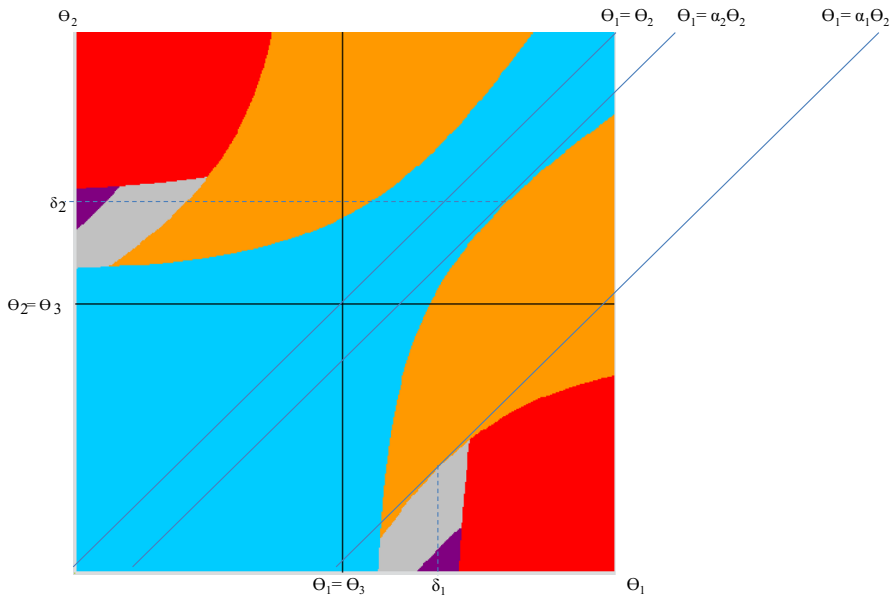


Fig. 5 Relative abilities θ_1 and θ_2 compared with $\theta_3 = 1$

Here, the intuition is that it will not pay off for players in the first match to conserve their strength for their second match if they expect the first match to be closer than the second match or, more formally, if the (weighted) heterogeneity in the second match is too pronounced.³³

Finally, if we vary the (expected) levels of restriction r and q , the number of grid points (θ_1, θ_2) for which the players choose self-restriction R in the SPE of game Γ increases in r and decreases in q . Figure 6 illustrates this finding. Again, the intuition is straightforward: An increase in r means that the restriction in the players' first match is less severe, and therefore, choosing R is less expensive. An increase in q instead means that the possible restriction in the players' second match is less severe or less likely, so the incentive to conserve their full strength by choosing R is less pronounced. Moreover, the variations of r and q in Fig. 6 make clear that our hypotheses are stable across diverse specifications of the model.

Appendix 2: Backward Induction of Game Γ

This appendix illustrates how to solve game Γ (see Fig. 3) by backward induction, considering the parameters $(\theta_1, \theta_2, r, q) = (1, 1, 1/2, 3/4)$.

³³ In our theoretical model, $\delta_i > 1$, which may be due to the assumption that player 3 has the advantage of unrestricted ability $\theta_3 = 1$ in both matches. In the baseline estimations for our empirical analysis, we set $\delta_i = 1$ and check the robustness of our results for alternative values of δ_i .

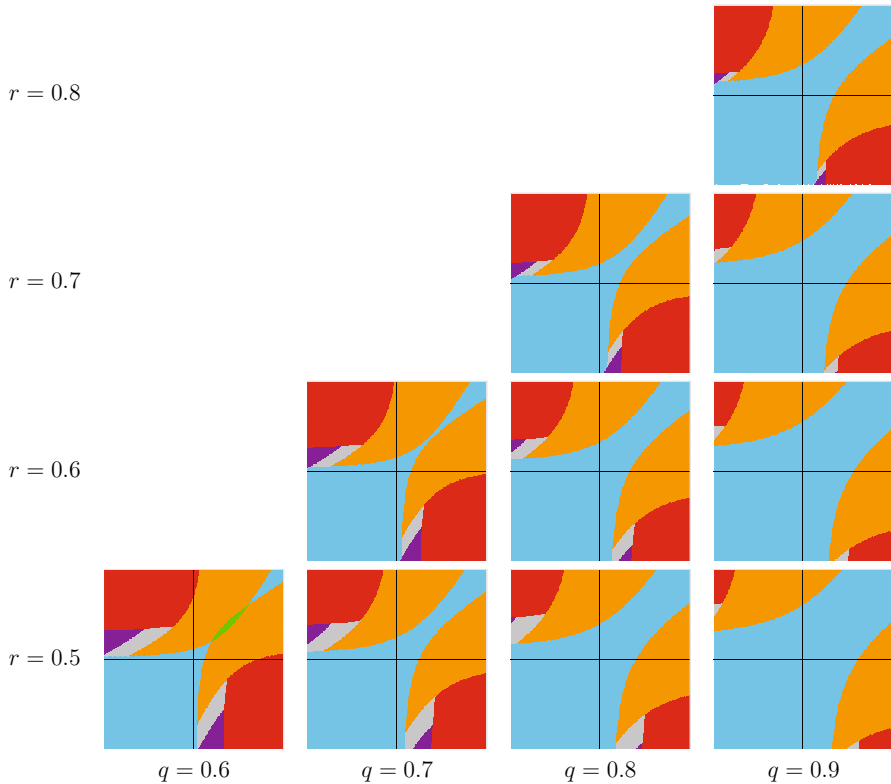


Fig. 6 Comparative Statics w.r.t. r and q . *Note:* This figure illustrates the effects of variations in r and q . On the horizontal (vertical) axis of each diagram, θ_1 (θ_2) increases from 0.1 to 10 in steps of 2% for normalised $\theta_3 = 1$. The vertical (horizontal) bold line represents $\theta_1 = 1$ ($\theta_2 = 1$). Each colour represents a certain type of equilibrium behaviour of players 1 and 2 in the first stage: ● blue $\hat{=}$ no player chooses R. ● orange $\hat{=}$ only the weaker player chooses R. ● red $\hat{=}$ both players choose R. ● violet $\hat{=}$ only the stronger player chooses R. ● grey $\hat{=}$ both players choose R with positive probability (equilibrium in mixed strategies). ● green $\hat{=}$ only the weaker player chooses R or only the stronger player chooses R or both players choose R with positive probability (multiple equilibria)

Case (R,R)

Suppose that both players 1 and 2 have chosen R in stage 1 and thus restrict themselves in the first match.

4th stage: Player 2 vs player 3

In the third match, player 2 is unrestricted, and thus $\theta_2^k = \theta_3^k = 1$ in all nodes $k \in \{A, B, C, C'\}$.

In node *A*, player 2 wins the first match, and player 3 wins the second match. Thus $w_2^A = w_3^A = 1$, $\ell_2^A = \ell_3^A = 1/2$, which yields

$$x_2^A = x_3^A = \frac{(1 - \frac{1}{2})^2 \cdot (1 - \frac{1}{2})}{[(1 - \frac{1}{2}) + (1 - \frac{1}{2})]^2} = \frac{1}{8},$$

$p_2^A = p_3^A = \frac{1}{2}$, and $E_2^A = E_3^A = \frac{1}{2} + \frac{1}{8} = \frac{5}{8}$ by Eqs. (4)–(6).

In node *B*, player 2 wins the first match, and player 1 wins the second match. Thus $w_2^B = 1$, $w_3^B = 1/2$, $\ell_2^B = 1/2$ and $\ell_3^B = 0$, which yields $x_2^B = x_3^B = 1/8$, $p_2^B = p_3^B = 1/2$, $E_2^B = 5/8$, and $E_3^B = 1/8$.

In node *C*, player 1 wins the first match, and player 3 wins the second match. Thus $w_2^C = 1/2$, $w_3^C = 1$, $\ell_2^C = 0$, and $\ell_3^C = 1/2$, which yields $x_2^C = x_3^C = 1/8$, $p_2^C = p_3^C = 1/2$, $E_2^C = 1/8$, and $E_3^C = 5/8$.

In node *C'*, player 1 wins the first and the second matches. Thus $w_2^{C'} = w_3^{C'} = 1/2$, and $\ell_2^{C'} = \ell_3^{C'} = 0$, which yields $x_2^{C'} = x_3^{C'} = 1/8$, $p_2^{C'} = p_3^{C'} = 1/2$, and $E_2^{C'} = E_3^{C'} = 1/8$.

3rd stage: Player 1 vs player 3

In the second match, player 1 is unrestricted, and thus $\theta_1^k = \theta_3^k = 1$ in both nodes $k \in \{D, E\}$.

In node *D*, player 2 wins the first match. Thus, $w_1^D = \frac{1}{2}p_2^B + \frac{1}{2}p_3^B = \frac{1}{2}$, $w_3^D = E_3^A = \frac{5}{8}$, $\ell_1^D = 0$, and $\ell_3^D = E_3^B = \frac{1}{8}$, which yields $x_1^D = x_3^D = 1/8$, $p_1^D = p_3^D = 1/2$, $E_1^D = 1/8$, and $E_3^D = 1/4$.

In node *E*, player 1 wins the first match. Thus, $w_1^E = 1$, $w_3^E = E_3^C = \frac{5}{8}$, $\ell_1^E = \frac{1}{2}p_2^C + \frac{1}{2}p_3^C = \frac{1}{2}$, and $\ell_3^E = E_3^{C'} = \frac{1}{8}$, which yields $x_1^E = x_3^E = 1/8$, $p_1^E = p_3^E = 1/2$, $E_1^E = 5/8$, and $E_3^E = 1/4$.

2nd stage: Player 1 vs player 2

In the first match, players 1 and 2 are restricted, and thus $\theta_1^F = \theta_3^F = \frac{1}{2}$ in node *F*. Moreover $w_1^F = E_1^E = \frac{5}{8}$, $w_2^F = p_1^D E_2^B + p_3^D E_2^A = \frac{5}{8}$, $\ell_1^F = E_1^D = \frac{1}{8}$, and $\ell_2^F = p_1^E E_2^{C'} + p_3^E E_2^C = \frac{1}{8}$, which yields $x_1^F = x_2^F = 1/8$, $p_1^F = p_2^F = 1/2$, $E_1^F = 1/4$, and $E_2^F = 1/4$. Moreover, player 3's expected payoff equals $E_3^F = p_1^F E_3^E + p_2^F E_3^D = 1/4$.

Table 6 Game matrix of stage 1

	Player 2	
	R	NR
Player 1		
R	0.25 0.25	0.18 0.31
NR	0.31 0.18	0.22 0.22

Cases (NR,R), (R,NR), and (NR,NR)

Using analogous procedures, we can calculate the players' expected payoffs in cases in which only one or none of them self-restricts in the first match.

1st stage: Decision on self-restriction

The results for the expected payoffs of players 1 and 2 are in Table 6 (rounded to two decimal places). For the parameters under consideration, NR is a dominant strategy for both players, and thus, (NR,NR) is the only Nash equilibrium in the first stage of game Γ .

Appendix 3: Additional tables

See Tables 7, 8, 9, 10, 11 and 12.

Table 7 Self-restriction and the starting line-up decision—OLS regressions

	(1)	(2)	(3)	(4)	(5)	(6)
Yellowplayer	0.0174** (0.0074)	0.0174** (0.0074)	0.0170** (0.0075)	0.0059 (0.0083)	0.0059 (0.0083)	0.0056 (0.0083)
Het _t	0.0074*** (0.0018)	0.0073*** (0.0018)	0.0072*** (0.0018)	0.0067** (0.0032)	0.0063*** (0.0032)	0.0063*** (0.0032)
Yellowplayer * Het _t	- 0.0162** (0.0081)	- 0.0162** (0.0081)	- 0.0145* (0.0080)	- 0.0131 (0.0082)	- 0.0131 (0.0082)	- 0.0121 (0.0083)
Het _{t+1}		- 0.0004 (0.0012)	0.0004 (0.0014)		- 0.0024 (0.0027)	- 0.0019 (0.0028)
Yellowplayer * Het _{t+1}			- 0.0082 (0.0078)			- 0.0055 (0.0082)
% of minutes played	0.7792*** (0.0080)	0.7792*** (0.0080)	0.7792*** (0.0080)	0.6453*** (0.0100)	0.6452*** (0.0100)	0.6453*** (0.0100)
Match-specific controls	Yes	Yes	Yes	Yes	Yes	Yes
Player-specific controls	Yes	Yes	Yes	No	No	No
Referee FE	Yes	Yes	Yes	Yes	Yes	Yes
Player FE	No	No	No	Yes	Yes	Yes
Coach FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	37,673	37,673	37,673	37,673	37,673	37,673
R ²	0.2592	0.2592	0.2593	0.1286	0.1286	0.1286
Adj. R ²	0.2571	0.2571	0.2571	0.1080	0.1080	0.1080

This table includes player-match-level data, that is all players who belong to a team's squad (starting line-up + bench) on a particular matchday.

OLS regressions, dependent variable: *starting 11*.

Het_t and Het_{t+1} are standardised (*z*-score).

Match-specific controls: derby, attendance, quality, away, matchday, roster

Player-specific controls: age, age², position-dummies (left/right defender, left/right wing player, left/right/central midfielder, forward, centre back, offensive midfielder, defensive midfielder)

(Robust) standard errors (clustered at the match level, column 1–3) are in parentheses, **p* < 0.1, ***p* < 0.05, ****p* < 0.01

Table 8 Descriptive statistics—starting11 (betting odds)

Variable	Obs	Mean	Std. Dev.	Min	Max
Het _t (odds)	37,673	0.285	0.215	0	0.864
Het _{t+1} (odds)	37,673	0.288	0.216	0	0.864

Table 9 Self-restriction and the starting line-up decision—logit regressions (betting odds)

	(1)	(2)	(3)	(4)	(5)	(6)
Yellowplayer	0.2545*** (0.0849)	0.2546*** (0.0849)	0.2667** (0.1062)	0.2212** (0.0919)	0.2213** (0.0919)	0.2411** (0.1156)
Het _t (odds)	0.1127*** (0.0423)	0.1087** (0.0422)	0.1086** (0.0422)	0.1129 (0.0792)	0.1053 (0.0799)	0.1053 (0.0799)
Yellowplayer * Het _t (odds)	- 0.5395** (0.2473)	- 0.5399** (0.2474)	- 0.5384** (0.2476)	- 0.4998* (0.2604)	- 0.5005* (0.2604)	- 0.4993* (0.2604)
Het _{t+1} (odds)		- 0.0310 (0.0289)	- 0.0280 (0.0324)		- 0.0507 (0.0703)	- 0.0460 (0.0722)
Yellowplayer * Het _{t+1}			- 0.0455 (0.2453)			- 0.0726 (0.2570)
% of minutes played	4.0111*** (0.0581)	4.0111*** (0.0581)	4.0111*** (0.0581)	3.5518*** (0.0651)	3.5517*** (0.0651)	3.5517*** (0.0651)
Match-specific controls	Yes	Yes	Yes	Yes	Yes	Yes
Player-specific controls	Yes	Yes	Yes	No	No	No
Referee FE	Yes	Yes	Yes	Yes	Yes	Yes
Player FE	No	No	No	Yes	Yes	Yes
Coach FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	37,673	37,673	37,673	37,673	37,673	37,673
Pseudo R2	0.2141	0.2141	0.2141	0.1203	0.1203	0.1203

This table includes player-match-level data, that is all players who belong to a team's squad (starting line-up + bench) on a particular matchday.

Logit regressions, dependent variable: *starting 11*

Match-specific controls: derby, attendance, quality, away, matchday, roster

Player-specific controls: age, age², position-dummies (left/right defender, left/right wing player, left/right/central midfielder, forward, centre back, offensive midfielder, defensive midfielder)

(Robust) standard errors (clustered at the match level, column 1–3) are in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10 Descriptive statistics–yellow card (betting odds)

Variable	Obs	Mean	Std. Dev.	Min	Max
Het _t (odds)	29,555	0.284	0.215	0	0.864
Het _{t+1} (odds)	29,555	0.286	0.215	0	0.864
Het _{t+2} (odds)	29,555	0.287	0.215	0	0.864

Table 11 The effect of heterogeneity on yellow card suspensions—OLS regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	> 45 min	> 60 min	> 75 min	all	> 45 min	> 60 min	> 75 min
Yellowplayer	- 0.0189** (0.0076)	- 0.0193*** (0.0067)	- 0.0157** (0.0061)	- 0.0128** (0.0052)	- 0.0602*** (0.0078)	- 0.0476*** (0.0070)	- 0.0360*** (0.0064)	- 0.0251*** (0.0056)
Het _t	- 0.0124*** (0.0031)	- 0.0091*** (0.0026)	- 0.0090*** (0.0023)	- 0.0073*** (0.0020)	- 0.0119*** (0.0031)	- 0.0090*** (0.0028)	- 0.0089*** (0.0026)	- 0.0075*** (0.0023)
Het _{t+1}	- 0.0023 (0.0026)	- 0.0016 (0.0024)	0.0001 (0.0022)	0.0002 (0.0020)	- 0.0017 (0.0029)	- 0.0011 (0.0026)	0.0000 (0.0024)	0.0000 (0.0021)
Het _{t+2}	0.0044* (0.0026)	0.0043* (0.0025)	0.0049** (0.0024)	0.0038* (0.0021)	0.0050* (0.0028)	0.0047* (0.0026)	0.0051** (0.0024)	0.0040* (0.0021)
Yellowplayer *	0.0163* (0.0086)	0.0177** (0.0079)	0.0182** (0.0077)	0.0148** (0.0075)	0.0140* (0.0076)	0.0153** (0.0068)	0.0168*** (0.0062)	0.0137** (0.0054)
Yellowplayer *	- 0.0147** (0.0074)	- 0.0168*** (0.0058)	- 0.0156*** (0.0052)	- 0.0135*** (0.0046)	- 0.0169** (0.0077)	- 0.0185*** (0.0069)	- 0.0161** (0.0063)	- 0.0141** (0.0055)
% of minutes played	0.0310*** (0.0075)	0.0106 (0.0068)	0.0026 (0.0064)	- 0.0096 (0.0059)	0.0581*** (0.0103)	0.0373*** (0.0092)	0.0274*** (0.0086)	0.0119 (0.0076)
Match-specific controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Player-specific controls	Yes	Yes	Yes	Yes	No	No	No	No
Referee FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Coach FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Player FE	No	No	No	No	Yes	Yes	Yes	Yes
Obs.	29,555	27,095	25,706	22,719	29,555	27,095	25,706	22,719

Table 11 continued

(1) All	(2) > 45 min	(3) > 60 min	(4) > 75 min	(5) all	(6) > 45 min	(7) > 60 min	(8) > 75 min
R2	0.0193	0.0137	0.0119	0.0096	0.0086	0.0082	0.0080

Player-match-level data (all fielded players)

OLS regressions, dependent variable: yellow

Het_{*t*} and Het_{*t+1*} are standardised (*z*-score)

Matchspecific controls: derby, attendance, quality, roster, matchday, away

Playerspecific controls: age, age², position-dummies (left/right defender, left/right wing player, left/right/central midfielder, forward, centre back, offensive midfielder, defensive midfielder)

Models 2–4 and 6–8 refer to the last 45/30/15 minutes of a match, so observations of bookings in prior minutes are excluded

(Robust) standard errors (clustered at the match level, column 1–4) are in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 12 The effect of heterogeneity on yellow card suspensions—logit regression (betting odds)

	(1) All	(2) > 45 min	(3) > 60 min	(4) > 75 min	(5) all	(6) > 45 min	(7) > 60 min	(8) > 75 min
Yellowplayer	- 0.1786 (0.1255)	- 0.2438 (0.1517)	- 0.2974* (0.1807)	- 0.4050 (0.2513)	- 0.4104*** (0.1256)	- 0.4361*** (0.1567)	- 0.4855*** (0.1797)	- 0.5829** (0.2361)
Het _t (odds)	- 0.3817*** (0.1006)	- 0.2194* (0.1196)	- 0.3741*** (0.1354)	- 0.0768 (0.1738)	- 0.3649*** (0.1015)	- 0.1986 (0.1231)	- 0.3499** (0.1407)	- 0.6554*** (0.1821)
Het _{t+1} (odds)	- 0.1133 (0.0827)	- 0.0874 (0.1010)	- 0.0149 (0.1154)	- 0.6743*** (0.1461)	- 0.0941 (0.0951)	- 0.0557 (0.1146)	- 0.0068 (0.1295)	- 0.0773 (0.1656)
Het _{t+2} (odds)	0.2040** (0.0833)	0.2095** (0.1026)	0.2711** (0.1217)	0.2273 (0.1586)	0.2479*** (0.0927)	0.2588** (0.1115)	0.3088** (0.1262)	0.2543 (0.1627)
yellowplayer *	0.7385** (0.3073)	0.7759* (0.3976)	0.8924* (0.4611)	1.1271* (0.6465)	0.6647** (0.2759)	0.6654** (0.3385)	0.8222** (0.3809)	1.1144** (0.4902)
Yellowplayer *	- 0.6276** (0.2964)	- 0.6586* (0.3591)	- 0.6026 (0.3977)	- 0.6610 (0.5099)	- 0.7654*** (0.2938)	- 0.8228** (0.3660)	- 0.7067* (0.4128)	- 0.7517 (0.5427)
% of minutes played	0.2511*** (0.0610)	0.1179 (0.0731)	0.0383 (0.0841)	- 0.1753 (0.1102)	0.5014*** (0.0877)	0.4368*** (0.1055)	0.4009*** (0.1204)	0.2374 (0.1533)
Match-specific controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Player-specific controls	Yes	Yes	Yes	Yes	No	No	No	No
Referee FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Coach FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Player FE	No	No	No	No	Yes	Yes	Yes	Yes
Obs.	29,555	27,095	25,706	22,719	29,555	27,095	25,706	22,719

Table 12 continued

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All	> 45 min	> 60 min	> 75 min	all	> 45 min	> 60 min	> 75 min
R2	0.0232	0.0200	0.0250	0.0131	0.0142	0.0161	0.0222

Player-match-level data (all fielded players)

Logit regression, dependent variable: yellow

Matchspecific controls: derby, attendance, quality, roster, matchday, away

Playerspecific controls: age, age², position-dummies (left/right defender, left/right wing player, left/right/central midfielder, forward, centre back, offensive midfielder, defensive midfielder)

Models 2–4 and 6–8 refer to the last 45/30/15 minutes of a match, so observations of bookings in prior minutes are excluded

(Robust) standard errors (clustered at the match level, column 1–4) are in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 13 Self-restriction and the substitution of yellowplayers–logit regressions

	(1)	(2)	(3)	(4)
yellow	0.1654 (0.1592)	– 0.1317 (0.1991)	0.1474 (0.1743)	– 0.1188 (0.2200)
yellow * Het _{t+1}		0.3667*** (0.1387)		0.3392** (0.1677)
Het _t	0.0919 (0.0715)	0.0926 (0.0713)	0.1295 (0.0849)	0.1263 (0.0850)
Het _{t+1}	0.0641 (0.0627)	0.0018 (0.0696)	0.0877 (0.0697)	0.0334 (0.0754)
% of minutes played	– 1.3901*** (0.3713)	– 1.4199*** (0.3710)	– 0.8050 (0.6691)	– 0.8267 (0.6704)
Match-specific controls	Yes	Yes	Yes	Yes
Player-specific controls	Yes	Yes	No	No
Referee FE	Yes	Yes	Yes	Yes
Player FE	No	No	Yes	Yes
Coach FE	Yes	Yes	Yes	Yes
Obs.	2000	2000	2028	2028
Pseudo-R2	0.0921	0.0946	0.0733	0.0759

Player-match-level data (restricted to yellowplayers in the starting line-up)

Logit regressions, dependent variable: substitution (substituted equal to one if the player got substituted, zero otherwise). Mean of *substituted* = 0.28

Match-specific controls: derby, attendance, quality, away, matchday, roster

Player-specific controls: age, age², position-dummies (left/right defender, left/right wing player, left/right/central midfield, forward, centre back, offensive midfield, defensive midfield)

(Robust) standard errors (clustered at the match level, columns 1 and 2) are in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Acknowledgements We thank participants of the 1st Reading Football Economics Workshop, the 4th Annual Conference “Contests: Theory and Evidence” in Norwich, the 2019 CESifo Area Conference on Applied Microeconomics, and a seminar in Hagen for their valuable comments. We are also grateful to Alex Krumer, Boram Lee, and three anonymous referees for helpful comments, as well as Christoph Laica for research assistance.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Altmann, S., Falk, A., & Wibral, M. (2012). Promotions and incentives: The case of multistage elimination tournaments. *Journal of Labor Economics*, *30*(1), 149–174.
- Baik, K. H. (1994). Effort levels in contests with two asymmetric players. *Southern Economic Journal*, *61*(2), 367–378.
- Balsdon, E., Fong, L., & Thayer, M. A. (2007). Corruption in college basketball? Evidence of tanking in postseason conference tournaments. *Journal of Sports Economics*, *8*(1), 19–38.
- Bar-Eli, M., Krumer, A., & Morgulev, E. (2020). Ask not what economics can do for sports—ask what sports can do for economics. *Journal of Behavioral and Experimental Economics*, *89*, 101597.
- Bar-Eli, M., Tenenbaum, G., & Geister, S. (2006). Consequences of players' dismissal in professional soccer: A crisis-related analysis of group-size effects. *Journal of Sports Sciences*, *24*(10), 1083–1094.
- Bartling, B., Brandes, L., & Schunk, D. (2015). Expectations as reference points: Field evidence from professional soccer. *Management Science*, *61*(11), 2646–2661.
- Brown, J., & Minor, D. B. (2014). Selecting the best? Spillover and shadows in elimination tournaments. *Management Science*, *60*(12), 3087–3102.
- Cornes, R., & Hartley, R. (2005). Asymmetric contests with general technologies. *Economic Theory*, *26*(4), 923–946.
- Dagaev, D., & Zubanov, A. (2017). Round-robin tournaments with limited resources. National Research University Higher School of Economics, Working Paper 171/EC/2017.
- Dechenaux, E., Kovenock, D., & Sheremeta, R. M. (2015). A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics*, *18*(4), 609–669.
- Deutscher, C., Frick, B., Gürtler, O., & Prinz, J. (2013). Sabotage in tournaments with heterogeneous contestants: Empirical evidence from the soccer pitch. *Scandinavian Journal of Economics*, *115*(4), 1138–1157.
- Dohmen, T. J. (2008). The influence of social forces: Evidence from the behavior of football referees. *Economic Inquiry*, *46*(3), 411–424.
- Ehrenberg, R. G., & Bognanno, M. L. (1990). The incentive effects of tournaments revisited: Evidence from the European PGA tour. *Industrial & Labor Relations Review*, *43*(3), 74–88.
- Fédération Internationale de Football Association (FIFA). (2017). *Laws of the Game 2017/18*. Fédération Internationale de Football Association.
- Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: The case of English football. *International Journal of Forecasting*, *21*(3), 551–564.
- Genakos, C., & Pagliero, M. (2012). Interim rank, risk taking, and performance in dynamic tournaments. *Journal of Political Economy*, *120*(4), 782–813.
- Groll, A., Schaubberger, G., & Tutz, G. (2015). Prediction of major international soccer tournaments based on team-specific regularized poisson regression: An application to the fifa world cup 2014. *Journal of Quantitative Analysis in Sports*, *11*(2), 97–115.
- Harbaugh, R., & Klumpp, T. (2005). Early round upsets and championship blowouts. *Economic Inquiry*, *43*(2), 316–329.
- Hart, E., Avrahami, J., Kareev, Y., & Todd, P. M. (2015). Investing even in uneven contests: Effects of asymmetry on investment in contests. *Journal of Behavioral Decision Making*, *28*(4), 395–409.
- Hill, B. (2018). Shadow and spillover effects of competition in NBA playoffs. *Journal of Sports Economics*, *19*(8), 1067–1092.
- Iqbal, H., & Krumer, A. (2019). Discouragement effect and intermediate prizes in multi-stage contests: Evidence from Davis Cup. *European Economic Review*, *118*(3), 364–381.
- Kahn, L. M. (2000). The sports business as a labor market laboratory. *Journal of Economic Perspectives*, *14*(3), 75–94.
- Konrad, K. A. (2009). *Strategy and dynamics in contests*. Oxford University Press.
- Konrad, K. A., & Kovenock, D. (2009). Multi-battle contests. *Games and Economic Behavior*, *66*(1), 256–274.
- Krumer, A., & Lechner, M. (2018). Midweek effect on soccer performance: Evidence from the German Bundesliga. *Economic Inquiry*, *56*(1), 193–207.
- Krumer, A., Megidish, R., & Sela, A. (2017). First-mover advantage in round-robin tournaments. *Social Choice and Welfare*, *48*(3), 633–658.

- Krumer, A., Megidish, R., & Sela, A. (2020). The optimal design of round-robin tournaments with three players. *Journal of Scheduling*, 23(3), 379–396.
- Lackner, M., Stracke, R., Sunde, U., & Winter-Ebmer, R. (2015). *Are competitors forward looking in strategic interactions? Evidence from the field*. IZA Discussion Paper No. 9564
- Laica, C., Lauber, A., & Sahn, M. (2017). Sequential round-robin tournaments with multiple prizes. CESifo Working Paper 6685.
- Leininger, W. (1993). More efficient rent-seeking—a Mönchhausen solution. *Public Choice*, 75(1), 43–62.
- Mago, S. D., & Sheremeta, R. M. (2017). Multi-battle contests: An experimental study. *Southern Economic Journal*, 84(2), 407–425.
- Malueg, D. A., & Yates, A. J. (2010). Testing contest theory: Evidence from best-of-three tennis matches. *The Review of Economics and Statistics*, 92(3), 689–692.
- Massey, C., & Thaler, R. H. (2013). The loser's curse: Decision making and market efficiency in the National Football League draft. *Management Science*, 59(7), 1479–1495.
- Peeters, T. (2018). Testing the wisdom of crowds in the field: Transfermarket valuations and international soccer results. *International Journal of Forecasting*, 34(1), 17–29.
- Raya, J. M. (2015). The effect of strategic resting in professional cycling: Evidence from the Tour de France and the Vuelta a España. *European Sport Management Quarterly*, 15(3), 323–342.
- Rosen, S. (1986). Prizes and incentives in elimination tournaments. *American Economic Review*, 76(4), 701–715.
- Ryvkin, D. (2013). Heterogeneity of players and aggregate effort in contests. *Journal of Economics & Management Strategy*, 22(4), 728–743.
- Sahn, M. (2019). Are sequential round-robin tournaments discriminatory? *Journal of Public Economic Theory*, 21, 44–61.
- Spann, M., & Skiera, B. (2009). Sports forecasting: A comparison of the forecast accuracy of prediction markets, betting odds and tipsters. *Journal of Forecasting*, 28(1), 55–72.
- Stein, W. E. (2002). Asymmetric rent-seeking with more than two contestants. *Public Choice*, 113(3), 325–336.
- Sunde, U. (2009). Heterogeneity and performance in tournaments: A test for incentive effects using professional tennis data. *Applied Economics*, 41(25), 3199–3208.
- Szentesi, B., & Rosenthal, R. W. (2003). Three-object two-bidder simultaneous auctions: Chopsticks and tetrahedra. *Games and Economic Behavior*, 44(1), 114–133.
- Szymanski, S. (2003). The economic design of sporting contests. *Journal of Economic Literature*, 41(4), 1137–1187.
- Taylor, B. A., & Trogdon, J. G. (2002). Losing to win: Tournament incentives in the national basketball association. *Journal of Labor Economics*, 20(1), 23–41.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.