Birthplace diversity and team performance

Enzo Brox[1] **Tommy Krieger**[2]

February 23, 2021

Abstract

Using a hand-collected data set that includes detailed information on 7,208 matches and 3,266 players from the highest division of German male soccer, this paper examines how birthplace diversity affects team performance. The results of two different instrumental variable analyses suggest that birthplace diversity has a hump-shaped effect on team performance. To explain this finding, we argue that nationally diverse teams have a wider range of skills and face greater communication barriers.

Keywords: Birthplace diversity, globalization, multinational working teams, productivity, soccer, team composition, team performance

JEL No.: F23, F66, J01, J24, M14, M54

Acknowledgement: We are very grateful for the constructive comments from Luna Bellani, Friedrich Breyer, Sebastian Findeisen, Michael Lechner, Stephan Maurer, Guido Schwerdt, Maurizio Strazzeri, Heinrich Ursprung, Ulf Zölitz, and Nick Zubanov. We also thank seminar and conference participants for their helpful feedback. Mareike Hartlieb and Lukas Würtenberger provide excellent research assistance. We are appreciative of the financial supports from the Graduate School of Decision Science (University of Konstanz).

[1]University of St. Gallen, Chair of Econometrics, Varnbüelstraße 14, 9000 St. Gallen, Switzerland. Email: enzo.brox@unisg.ch

[2]Center for European Economic Research (ZEW), Department of Public Finance, L7 1, 68161 Mannheim, Germany. Email: tommy.krieger@zew.de

1 Introduction

The importance of team-based production increased notably over the last three decades (see Deming, 2017, Jones, 2009, Wuchty et al., 2007). Understanding the factors that determine successful collaboration is therefore at the very heart of many studies in business, labor, and organizational economics. In particular, a large body of literature examine whether teams whose members differ in their socio-demographic characteristics are more or less productive than homogeneous working groups.

Theoretical research suggests that making an unambiguous prediction about how diversity in sociodemographic worker characteristics affects the performance of a team is nearly impossible (see e.g. van Knippenberg et al., 2004, Williams and O'Reilly, 1998). A reason for this ambiguity is that diversity can have both positive and negative effects. A second reason is that various characteristics exist in which members of a team differ from each other (e.g. age, gender, tenure, religion) and that the relevance of the positive and negative aspects of diversity varies between these characteristics. Finally, the effect of diversity on team performance is likely to depend on a number of surrounding factors.

We complement the existing literature by providing causal evidence on the relationship between birthplace diversity and team performance. We choose this type of diversity for two reasons: first, the number of multinational working groups increased notably in the process of globalization, and second, compared with other socio-demographic characteristics, rather little research exists on the effects of birthplace diversity on team performance. In theory, birthplace diversity can be performance-enhancing since workers from different countries are educated under different systems and thus have different skills and experiences (Alesina et al., 2016). On the other hand, a high level of birthplace diversity might be detrimental because it increases the risk of conflicts and intergroup biases. In addition, multinational teams might have greater communication problems either due to language barriers or since people with different origins are more likely to interpret ambiguous statements differently (Lang, 1986, Lazear, 1999).

Investigating how birthplace diversity influences the performance of a working group is challenging for a number of reasons. A major challenge is to find an environment where workers from different countries collaborate with each other and where working teams are clearly identifiable. Another main difficulty is to satisfy the data requirements for such an empirical analysis. In particular, we require objective measures of performance and data on workers' abilities. Finally, we have to develop an identification strategy that allows to establish causality.

We exploit rich hand-collected data from the highest division of German male soccer (*Bundesliga*) to address the challenges that make the identification of the causal effect of birthplace diversity on team performance difficult. The *Bundesliga* is an attractive environment for our purposes because we clearly observe who collaborates with whom and have objective measures of team performance. In addition, freely accessible sources provide detailed information on soccer players. Especially notable in this regard is a time-varying measure of quality that we obtain from the video game series *FIFA*. In total, we collected data on 7,028 matches and 3,266 players, coming from almost 100 countries.

A standard approach when investigating how diversity affects team (or firm) performance is to estimate a fixed effect model (see Dezsö and Ross, 2012, Haas and Nüesch, 2012, Kahane et al., 2013, Prinz and Wicker, 2016). We illustrate that the causal effect of diversity on performance can hardly be identified with this approach. The reason is that the extent to which the composition of a team changes during the production process depends on how productive the team is. Consequently, the performance during the production process jointly affects the final output of a team and its ex-post observable composition. Econometric theory suggests that a standard fixed effect approach produces biased estimates if such a problem of reverse causality exists (see Angrist and Pischke, 2009, Wooldridge, 2010).[1] We address this and other endogeneity problems with two instrumental variable approaches. Our first approach exploits predicted line-ups published by the leading German soccer magazine *Kicker*. More specifically, we compute how diverse the players in these predicted line-ups are and use this measure as an instrument for the diversity of the fielded players. Our second approach adapts the approach by Bettinger and Long (2005) who exploit unexpected absences of university lecturers to study the importance of role models for study choice. In our case, we use unexpected replacements in the starting line-ups to produce plausibly exogenous variation in team structures. In robustness checks, we only exploit those unexpected replacements that are caused by injuries.

The results of our instrumental variable regressions provide causal evidence for a non-linear relationship between birthplace diversity and team performance. In particular, we find that the effect of birthplace diversity on team performance increases up to a certain level and becomes smaller beyond this threshold. Put differently, our findings imply that an intermediate level of birthplace diversity maximizes team performance. We also report some regression results, suggesting that the optimal level of birthplace diversity is task-specific and depends on the importance of interpersonal communication.

Our paper contributes to the literature on the relationship between birthplace diversity and team performance.[2] Broadly speaking, this literature includes two

[1] In our specific case, we show that the level of birthplace diversity increases during a match if a team does not perform well. We explain this pattern with the fact that team managers replace defensive with offensive players if their team is behind. These performance-based substitutions typically increase the level of birthplace diversity since the share of foreigners in German soccer clubs is often larger among offensive than among defensive players. The econometric consequence of these substitutions is that the standard fixed effect approach produces downward-biased estimates of the optimal level of birthplace diversity.

[2] Several studies examine the relationship between birthplace diversity and economic performance

types of studies. The first uses field or laboratory experiments. The results of these experiments are mixed. For instance, while Earley and Mosakowski (2000) conclude that the relationship between birthplace diversity and team performance is U-shaped, Lyons (2017) observes a negative effect. As an explanation, Lyons (2017) provides evidence, suggesting that nationally diverse working groups suffer from communication problems. The second type of studies applies fixed effect regressions to observable data. In many cases, this data comes from the sports industry. A remarkable exception is Freeman and Huang (2015) who examine the performance of research teams and find that birthplace diversity is performance-enhancing. By contrast, most of the studies that exploit sports data conclude that birthplace diversity has detrimental effects for team performance (see Haas and Nüesch, 2012, Kahane et al., 2013, Maderer et al., 2014).[3] We present results from instrumental variable regressions, suggesting that the relationship between birthplace diversity and team performance is hump-shaped. We also show that a fixed effect approach produces heavily biased estimates since it does not take into account that team structures change throughout the production period. To the best of our knowledge, our two instrumental variable procedures have not been used in other studies that examine how birthplace diversity affects team performance.

More generally, we contribute to the literature that analyzes how diversity in sociodemographic worker characteristics affects team performance. This literature consists of numerous studies, considering several types of diversity and applying various empirical methods (for detailed reviews, see Williams and O'Reilly, 1998, Horwitz and Horwitz, 2007, Joshi and Roh, 2009, van Knippenberg and Mell, 2016, Guillaume et al., 2017). Apart from providing causal evidence on the role of a specific type of diversity, our paper has three contributions. First, we illustrate that regression results can be severely biased if an empirical approach neglects that the composition of a working group change throughout the production process, depending on how it performs. Since such changes occur in most industries, we believe that such an illustration is important to further improve the awareness for this endogeneity problem. Second, our paper establishes two instrumental variable approaches that address the problem of reverse causality. Given that variants of our methods can be used for other types of diversity and in other institutional environments, we think that our paper can help other scholars to address their empirical problems. Third, we support the view that external factors moderate

at the country (see e.g. Alesina et al., 2016, Bove and Elia, 2017, Docquier et al., 2020), regional (see e.g. Ager and Brueckner, 2013, Ottaviano and Peri, 2005), or firm level (see e.g. Parrotta et al., 2014, Trax et al., 2015). For our purposes, these studies are of relatively little relevance since the mechanisms and empirical challenges that play a role at these levels differ to a great extent from ours.

[3] A notable exception is the study by Balsmeier et al. (2019) who show that the increase in birthplace diversity that was caused by the Bosman ruling enhanced team performance. Their result is consistent with our results since prior to this ruling birthplace diversity was much lower than the "optimal" level of birthplace diversity.

the effect of workplace diversity on team performance.

Finally, our paper contributes to the literature that exploits data from the professional sports industry to answer economic questions. Kleven et al. (2013) use information from the European soccer market to identify the effects of top tax rates on migration. Parsons et al. (2011), Price and Wolfers (2010), and Price et al. (2013) examine same-race preferences, using sports data from the United States. Arcidiacono et al. (2017), Gould and Winter (2009), and Guryan et al. (2009) exploit sports data to analyze the role of peer effects. Krumer and Lechner (2017) and Cohen-Zada et al. (2018) use data from the sports industry to study the role of scheduling in contests. Garicano et al. (2005) and Apesteguia and Palacios-Huerta (2010) show the consequences of social and psychological pressure by exploiting sports data (see also Harb-Wu and Krumer, 2019). Cohen-Zada et al. (2017) study judo tournaments to illustrate how past success and failure influence subsequent performance. Lichter et al. (2017) exploit data from the German soccer league to estimate the effect of pollution on productivity. Doerrenberg and Siegloch (2014) observe that big soccer events motivate unemployed people to search for a new job.

We proceed as follows. Section 2 reviews the existing literature. Section 3 informs about the *Bundesliga* and our data. Section 4 describes our identification strategies. Section 5 presents our estimation results. Section 6 concludes.

2 Literature review

2.1 Diversity and team performance

2.1.1 Theoretical studies

Over the last 40 years, the question of whether heterogeneous teams perform better or worse than homogeneous teams has been raised in various studies. On the one hand, diversity might increase the range of skills and thereby enhances the performance of a team (see e.g. Alesina et al., 2016, Cox et al., 1991). On the other hand, diverse working teams might have lower group cohesion and might face a higher risk of inter-group conflict since many people prefer to collaborate with people that are similar to themselves (see e.g. Jehn et al., 1999). In addition, diversity might cause misunderstandings and communication problems (see e.g. Lang, 1986, Lazear, 1999).

Because of the different channels through which diversity might influence team performance, some models predict that this effect is non-linear. For instance, the Categorization-Elaboration Model (CEM) by van Knippenberg et al. (2004) suggests that the effect of diversity on the availability of knowledge and therefore team performance is only positive up to a certain degree. If it is above this level, increasing diversity will be detrimental because the risk of misunderstandings increases and knowledge exchange becomes rather difficult. Similarly, Ashraf and Galor (2013) present a formal model that implies a hump-shaped relationship of diversity and economic performance. The CEM as well as the model by Prat (2002) suggest that the ideal level of diversity depends on the task requirements and other contextual factors.

2.1.2 Empirical studies

The literature includes numerous empirical studies that examine how diversity affects team performance (for meta analyses and comprehensive reviews, see Bell et al., 2011, Guillaume et al., 2017, Horwitz and Horwitz, 2007, Joshi and Roh, 2009, van Knippenberg and Mell, 2016, van Knippenberg and Schippers, 2007). The vast majority of them either provides evidence from field/laboratory experiments or exploits observable data. Below, we will spend more attention to those studies that apply the latter approach since we will also use observable data in our analysis.[4]

Most related studies focus on a particular type of diversity when examining how diversity affects performance (for exceptions, see Ely, 2004, Parrotta et al., 2014, Prinz and Wicker, 2016). The most common types are age (see e.g. Kunze et al., 2011, Ng and Feldman, 2008), gender (see e.g. Apesteguia et al., 2012, Adams and Ferreira, 2009, Hoogendoorn et al., 2013), ethnicity/race (see e.g. Herring, 2009, Hjort, 2014), functional background (see e.g. Ancona and Caldwell, 1992, Boone and Hendriks, 2009, Cummings, 2004), and skills (see e.g. Aggarwal and Woolley, 2019, Swaab et al., 2014, Tan and Netessine, 2019). The results of existing studies are inconclusive: while some of them suggest that the effect of diversity on team performance is positive, others provide evidence for negative, insignificant, and non-linear effects. A potential explanation for this inconsistency is that external factors moderate the relationship between diversity and performance (Guillaume et al., 2017). Another explanation might be that some studies fail to address endogeneity issues and thus report biased estimates (Hoogendoorn et al., 2013).

Compared with the huge number of studies that exploit observable data to analyze how diversity affects performance, relatively few studies apply methods that allow for a causal interpretation of the results. One of these studies is Delis et al. (2017) who use an instrumental variable approach to show that firm performance increases if the board directors come from countries with different levels of genetic diversity. Tan and Netessine (2019) report instrumental variable estimates, implying a hump-shaped relationship between skill diversity and team performance. Kesavan et al. (2014) apply a two-stage least squares approach to show that an optimal mix exists between part- and full-time workers. Using

[4] Compared to running an experiment, exploiting observable data has the advantage that we analyze the performance of teams in a real rather than an artificial environment. The main disadvantage is that addressing endogeneity issues is much more difficult (Hoogendoorn et al., 2013).

the same approach, Adams and Ferreira (2009) find that gender diversity reduces performance.

In our empirical analysis, we will provide evidence for a non-linear effect of birthplace diversity and team performance (for more details, see Section 5). Such a relationship has recently also been identified for other types of diversity and other team characteristics. For example, Staats et al. (2012) show that the effect of group size on performance is hump-shaped, whereas Tan and Netessine (2014) observe that there is an optimal level of workload. As mentioned above, other studies are Kesavan et al. (2014) and Tan and Netessine (2019). Schwab et al. (2016) and Owen and Temesvary (2018) conclude that the effect of gender diversity in boards on firm performance is non- linear.

2.2 Birthplace diversity and team performance

In this paper, we analyze whether the performance of a working group differs, depending on the number of countries in which the group members were born. Theoretically, birthplace diversity might affect team performance for three main reasons. First, people from different countries have been educated in different systems, and are thus likely to have different skills and abilities (Alesina et al., 2016). Second, birthplace diversity might create communication problems due to language barriers and different cultural habits (Lang, 1986, Lazear, 1999). Finally, the risk of conflict and inter-group bias will increase if people from different countries collaborate with each other (Earley and Mosakowski, 2000).

Inspired by the seminal models of Ashraf and Galor (2013) and Lazear (1999), Appendix A presents a simple theoretical framework on the relationship between birthplace diversity and team performance. The key prediction of our model is that the positive and negative aspects of birthplace diversity can lead (under a couple of relatively mild assumptions) to a hump-shaped net effect. The second major prediction of our simple model is that the optimal level of birthplace diversity differs for different tasks because of differences in the importance of effective communication.

Only a few empirical studies investigate how birthplace diversity affects team performance. Their results are quite mixed. Earley and Mosakowski (2000) present correlations, suggesting a U-shaped relationship between birthplace diversity and performance. Lyons (2017) shows results from a randomized field experiment in which programmers from Bangladesh, India, and Pakistan collaborate with each other.[5] She finds that nationally heterogeneous groups perform worse because of greater communication barriers. Freeman and Huang (2015) conclude that multi-national research teams are more productive. Exploiting data from the sports

[5] A weak spot of her experiment is that the working groups only consists of two people. Lyons (2017) therefore has no opportunity to examine whether the relationship between birthplace diversity and team performance is non-linear.

industry, Haas and Nüesch (2012), Kahane et al. (2013), and Maderer et al. (2014) suggest that birthplace diversity is detrimental, while Balsmeier et al. (2019) and Ingersoll et al. (2017) present estimates that suggest the opposite. To our best knowledge, none of these closely related studies addresses the issue that team composition might change endogenously during the production process.

3 Data and institutional framework

The first main challenge when examining how birthplace diversity affects team performance is to find an appropriate institutional environment. Kahane et al. (2013) suggest that such an environment has to satisfy four requirements. First, individuals from different countries need to collaborate with each other. Second, the composition of the team must be observable. Third, information on workers' origins, skills, and experiences have to be available. Fourth, the output of the working group needs to be objectively measurable. This section shows that we meet these four requirements when considering the *Bundesliga*.

3.1 Institutional background

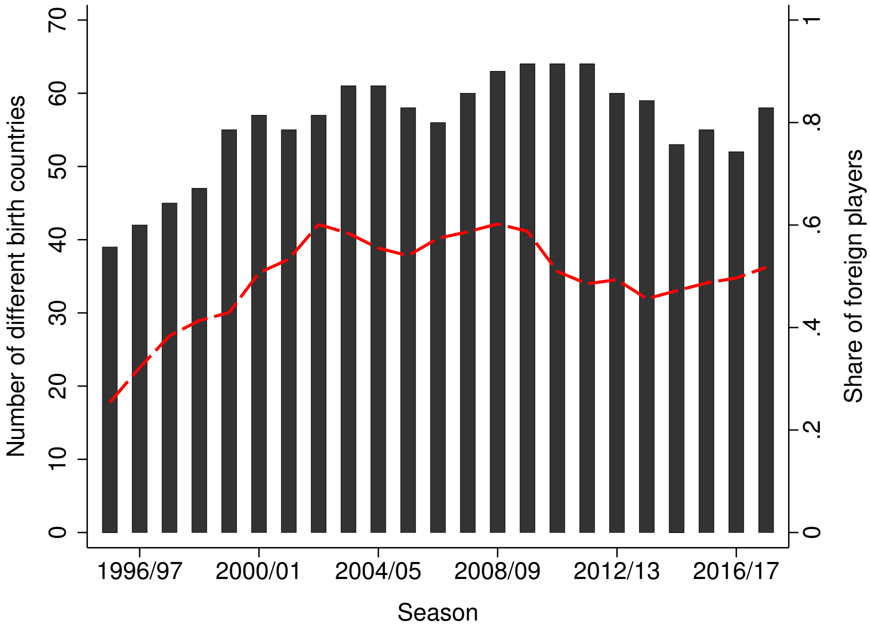
The *Bundesliga* is the highest division of German male soccer and consists of 18 clubs. This league is organized as a double round-robin system. Each clubs therefore plays 34 matches per season, 17 of them at home. A match day includes nine matches, whereby each club participates in one of them. Matches are assigned to match days prior to the start of the season and usually take place on a weekend.[6] Between the first round (August – December) and the second round (January – May) of a season, a winter break of four/five weeks without matches is made.[7] In both rounds, each club plays against all other clubs. If a club has the home field advantage in the first round, the opponent has the home field advantage in the second round.

A soccer match lasts 90 minutes and consists of two halves. Prior to the match, the team manager nominates eleven starting players (one goal keeper, ten field players) and seven substitutes. During the match, the team manager can substitute up to three players. At the end of the match, the winner obtains three ranking points, while the loser gets zero. In case of a draw, both clubs obtain one point. At the end of a season, the total number of ranking points determines the position in the table. If two clubs have the same number of points, the difference between the goals scored and goals conceded serves as a decision criterion.

[6] For organizational reasons, some match days have to take place during midweek days (Tuesday and Wednesday). For details, see Krumer and Lechner (2018).

[7] In some seasons, the first match day of the second round takes place before the winter break starts.

Figure 1 Trends in birthplace diversity (Bundesliga, 1995/96 – 2017/18).



Notes: The dashed red line illustrates how the share of foreign *Bundesliga* players developed over time. All figures are weighted by the number of matches. The bar chart shows how the number of birth countries changed over time.

The starting players and the substitutes are selected out of the squad. The squad consists of all soccer players hired by the club and is compiled by the club managers. As in most other soccer leagues, the club managers can change the squad twice per year: the first transfer period is from July to August and the second in January. There are no rules determining the size of the squad. A salary cap does not exist and budgets differ across clubs. The club budget depends on various factors, including ticket sales, transfer revenues, sponsorship contracts, TV revenues, and monetary rewards for participation in a European club tournament.

The rule governing the fielding of foreign players has only changed twice since the introduction of the *Bundesliga* in 1963. The initial regulatory scheme lasted until the middle of the 1995/96 season and allowed the fielding of three foreign players. In December 1995, the European Court of Justice declared the initial form of the *three-players rule* illegal on the grounds that it is not compatible with the treaties of the European Union.[8] Afterwards, a *three-players rule* only applied to non-European players.[9] As of the 2004/05 season, also this restriction was abolished.

[8]This landmark decision is known as *Bosman ruling* and is named after the Belgian Jean-Marc Bosman who sued his club, RFC Liège, because of contractual disputes. For details, see Dobson and Goddard (2011), Kleven et al. (2013), and Simmons (1997).

[9]This rule had, however, rather little practical relevance since most of the non-Europeans had a dual citizenship of their country of birth and an European country.

Figure 1 illustrates that the *Bundesliga* is an environment where people from various countries collaborate. The dashed red line in Figure 1 shows how the share of foreign players changed over time (right scale). In the 1995/96 season, only about one quarter of the *Bundesliga* players was not German-born. Caused by the relaxation of the *three-players rule*, the share of foreigners began to increase in the 1996/97 season and peaked at 60 percent in the 2002/03 season. Since the 2010/11 season, about half of the players are German-born.

Figure 1 also includes a bar chart, suggesting that foreign *Bundesliga* players come from various countries (left scale). The length of a bar corresponds to the number of birth countries. Between the 1995/96 season and the 2003/04 season, the number of birth countries increased from 39 to 61. Since then, the number of birth countries oscillates between 52 and 64.

3.2 Data

We created a novel database, including information on all *Bundesliga* matches from the 1995/96 season up to the 2017/18 season.[10] For each of these 7,038 matches, we know the date, the participating clubs, the final result, and the venue. We also observe the starting line-ups, have information on substitutions, and know the current ranking positions of the clubs. In addition, we identified the names of the team managers and checked whether a club participated in a match of a European club tournament or the national cup in the week just before or immediately after a *Bundesliga* match. Our key source of information was the homepage of the leading soccer magazine *Kicker* (`www.kicker.de`).

In total, 3,266 players participated in the 7,038 matches that are covered by our database. For each of them, we identified the country of birth. The total number of birth countries is 98 (for the full list, see Table E.1). Apart from Germany, the most frequent countries of birth are Brazil, France, Poland, and Denmark. For all players, we also know their date of birth, the date at which they were hired by a *Bundesliga* club, and the total number of matches that they have played prior to a particular match in (i) the *Bundesliga*, (ii) the highest soccer divisions in France, England, Italy, and Spain, (iii) an European club tournament, and (iv) European and world championships.[11] All these hand-collected information come from the website of the soccer magazine *Kicker* and the online databases *Transfermarkt* (`www.transfermarkt.com`) and *World Football* (`www.worldfootball.net`). We complement this objective information with two expert-based measures of players' quality. The first is the market value of the player as reported by *Transfermarkt*. Our second measure of quality comes from

[10]We began the data collection with the 1995/96 season for three reasons: (i) prior to this season, it was prohibited to field more than three foreigners, (ii) as of this season, the winner of a match obtains three ranking points rather than two, and (iii) data quality.

[11]We choose the leagues of France, England, Italy, and Spain since the level of play in these league is similar to the level of play in Germany.

the video game *FIFA* which is released annually by the video game company *Electronic Arts*. In this popular video game, each soccer player has a playing strength ranging from 0 to 100. We develop a PHP script to download the ratings from the *FIFA* online database (`www.fifaindex.com`). Both measures of players' quality have the shortcoming that they are only available for the 13 latest *Bundesliga* seasons (2005/06 – 2017/18).

4 Empirical framework

4.1 Fixed effect approach

We begin our analysis of the relationship between birthplace diversity and team performance with the following regression model:

$$\mathrm{Y}_{isrmd} \;=\; \mathfrak{F}\big(\mathrm{B}_{isrmd}\big) \;+\; \gamma \cdot \mathbf{X}_{isrmd} \;+\; \Lambda_{isrm} \;+\; \Theta_d \;+\; \varepsilon_{isrmd} \tag{1}$$

where i denotes the club, s the season, m the manager, and d the match day. The subscript $r \in \{1, 2\}$ indicates whether the match took place in the first or the second round of a season. $\mathbf{X}$ is the set of controls, Λ the set of club-by-season-by-round-by-manager fixed effects, and Θ the set of match day fixed effects.

A key advantage of using data from the professional sports industry is that performance (Y) can be directly and objectively measured (Kahane et al., 2013). In line with related studies, we apply two different measures. The first is the difference between the goals scored and goals allowed. Our second measure of team performance is the number of ranking points that a club obtained at the end of a match.

Our explanatory variable of interest is the level of birthplace diversity (B). Following Alesina et al. (2016), we define "birthplace diversity" as the probability that two randomly drawn team members were born in different countries. Put differently, we measure birthplace diversity with the fractionalization index:

$$\mathrm{B} \;=\; 1 \;-\; \frac{1}{n} \sum_{l=1}^{n} \sum_{k=1}^{n} b_{lk} \tag{2}$$

where b_{lk} is a dummy variable that is equal to 1 if players l and k have the same country of birth, and 0 otherwise. The parameter n denotes the total number of players that participated in a particular match.

As explained in Section 2, the relationship between birthplace diversity and team performance might be nonlinear. In our baseline analysis, we assume a quadratic functional form to test this hypothesis:

$$\mathfrak{F}\big(\mathrm{B}_{isrmd}\big) \;=\; \beta_1 \cdot \mathrm{B}_{isrmd} \;+\; \beta_2 \cdot \mathrm{B}^2_{isrmd} \tag{3}$$

The optimal level of birthplace diversity is then:

$$\mathrm{B}^{*} = -\frac{\beta_1}{2 \cdot \beta_2} \quad \text{with} \quad \beta_1 \in [0, -2 \cdot \beta_2] \quad \text{and} \quad \beta_2 < 0. \tag{4}$$

To allay concerns about our baseline approach, we use regression methods that relax the functional form assumptions in our robustness checks. For the sake of comparison, we also present results from linear regressions.

When estimating (1), we exploit variation within clubs and control for a large number of potential confounders. Our fixed effects capture all factors that are specific to a *Bundesliga* club in a round of a particular season if a particular team manager is in charge. These factors include the abilities of the manager and his staff, the composition of the squad, the budget and prestige of a club, the quality of the training facilities and youth sections, cultural differences between the manager and the players, and the fan base. Our fixed effects also make sure that our regression results are not biased due of general time trends. To control for factors that vary between single matches, we add measures to the regression model that reflect the experience and quality of the fielded players. We also control for the home field advantage and various manager and opponent characteristics. In Table E.2, we report all match-specific control variables.

4.2 Instrumental variable approaches

The estimates of β_1, β_2, and B^* might be biased for two reasons. The first is that our fixed effects and control variables might not suffice to capture all confounding factors.[12] The second potential endogeneity issue is reverse causality. Reverse causality might exist in our case because the composition of a team changes in the course of a match due to substitutions.

In Appendix B, we show that reverse causality is indeed a severe endogeneity issue. We also argue that this problem biases the OLS estimate of the optimal level of birthplace diversity towards zero. More specifically, we use within-match data to illustrate that team managers replace defensive with offensive players if their team is behind (see also Garicano and Palacios-Huerta, 2014). Usually, these performance-based substitutions do not influence the outcome of the match but increase birthplace diversity since the share of foreign offensive players is larger than the share of foreign defensive players. Put differently, our findings clearly suggest that poor (well) performing teams increase (decrease) their birthplace diversity during a soccer match. Below, we describe two instrumental variable approaches that address this issue.

[12] For instance, we lack information about the personal relationship between the team manager and the players. Our fixed effects cannot account for this relationship since it can change within a few weeks, especially if a new manager is hired or if the club has a series of victories/defeats.

4.2.1 Diversity in predicted starting line-ups

Only a few of the related studies apply an instrumental variable procedure to allay endogeneity concerns (see e.g. Delis et al., 2017, Kesavan et al., 2014, Tan and Netessine, 2014, 2019). Among these studies, the most common approach is to exploit a lagged value of the main variable of interest (see also Bloom and Van Reenen, 2007, Siebert and Zubanov, 2010). Obviously, when using this intuitive approach, the regression results cannot be biased because of reverse causality. In addition, the first-stage relationship is likely to be strong in this case because working groups often change slowly over time. A general weak spot of using lagged values as instruments is that the exclusion restriction might be violated (Angrist and Pischke, 2009, 2010, Bazzi and Clemens, 2013). More specifically, the extent to which the diversity of a team changes between two different periods depends on the performance in the first period and various other factors. The exclusion restriction only holds if one can adequately control for those factors that influence the current performance through a channel other than the level of diversity.

In this paper, we propose an instrumental variable approach that shares the strengths of the "standard" approach and for which we think that the exclusion restriction is less likely to be violated. Our basic idea is to exploit predicted line-ups that are published by the leading soccer magazine *Kicker* at the day before the first match of a match day begins. To implement our approach, we first manually digitized all 7,956 expected starting line-ups that were published between July 2005 and June 2018.[13] Afterwards, we identified the country of births of the listed players. In our analysis, we use this information to estimate the first-stage equations:

$$\mathrm{B}_{isrmd} = \rho_1 \cdot \mathrm{K}_{isrmd} + \rho_2 \cdot \mathrm{K}^2_{isrmd} + \alpha \cdot \mathbf{X}_{isrmd} + \Lambda_{isrm} + \Theta_d + \eta_{isrmd} \tag{5}$$

$$\mathrm{B}^2_{isrmd} = \delta_1 \cdot \mathrm{K}_{isrmd} + \delta_2 \cdot \mathrm{K}^2_{isrmd} + \lambda \cdot \mathbf{X}_{isrmd} + \Lambda_{isrm} + \Theta_d + \mu_{isrmd} \tag{6}$$

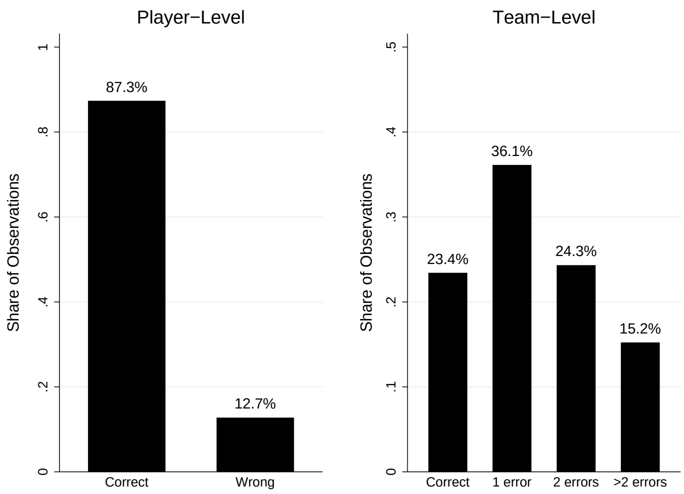
where K denotes the birthplace diversity in the starting line-up predicted by *Kicker*. All other variables have the same meaning as in (1).[14]

Figure 2 shows the relationship between the actual starting line-ups and the starting line-ups predicted by *Kicker*. The left graph indicates that 87 percent of the start players also belong to the expected starting line-up. This overlap is reassuring since it suggests strong first-stage relationships. A concern might be that *Kicker* correctly predicts the actual starting line-ups in most of the cases

[13]We cannot expand the sample to earlier seasons because of limited data availability and low data quality.

[14]A question might be why we use the starting line-up predicted by *Kicker* rather than the actual starting line-up to create our instruments. The advantage of the former is that it is made by journalists rather than the manager. Thus, our estimates are less likely to be confounded by an unobserved manager characteristic when using the diversity in the *Kicker* line-up.

Figure 2 Accuracy of the predicted starting line-up.



Notes: The left graph shows how likely it is that a player who is part of the starting line-up predicted by *Kicker* also belongs to actual starting line-up. The right graph shows how likely it is that there are no (one, two, or more than two) discrepancies between the starting line-up predicted by *Kicker* and the actual starting line-up.

and makes serious mistakes in a few very specific cases. The right graph of Figure 2 alleviates this concern. In particular, we observe that *Kicker* correctly predicts the full starting line-up in only 23 percent of the cases and that the number of incorrect predictions per starting line-up is usually small.

Compared with an approach that exploits lagged variables as instruments, our approach is more likely to satisfy the exclusion restriction for two key reasons. First, when forming expectations about the starting line-ups, the *Kicker* editors take into account that players who are exhausted or performed poorly in the previous matches might be replaced by others. Consequently, we require fewer conditions to satisfy the exclusion restriction when using our approach. Second, the starting line-up predicted by the sport magazine *Kicker* is artificial rather than made by the team manager. The likelihood that an unobserved manager characteristic will confound our estimation results is thus lower.

In sum, in our first two-stage least squares approach, we use the birthplace diversity in the starting line-up predicted by *Kicker* as the instrument for the birthplace diversity of the fielded players. With this procedure, we can identify the causal effect of birthplace diversity on team performance if two conditions hold: first, the expected and the actual level of birthplace diversity are closely correlated, and second, conditional on our set of controls and fixed effects, the actual level of diversity is the only channel via which the predicted level of diversity is related with team performance.

4.2.2 Unexpected replacements

Even though our fixed effects and match-specific control variables block various alternative channels, we cannot be completely sure that our first instrumental variable approach meets the exclusion restriction. In this section, we therefore propose an instrumental variable approach that exploits another source of quasi-exogenous variation. Our rationale is that biased estimates are relatively unlikely if both approaches produce the same results, despite using different sources of variation.

Our second instrumental variable approach is inspired by studies in education economics, especially by the work of Bettinger and Long (2005) and Herrmann and Rockoff (2012). The common feature of these influential studies is that they exploit unexpected absences of instructors.[15] We adapt their approach by using the extent to the which the level of birthplace diversity changes because of unexpected replacements in the starting line-up as instrument for the birthplace diversity of the fielded players. As our first instrumental variable approach, our second approach addresses the problem of reverse causality since the unexpected replacements take place prior to the match.

We proceed in three steps when implementing our second instrumental variable approach. In the first step, we define three groups of players (for an example, see Figure D.1). Group $\mathcal{A}$ includes all players that belong to the starting line-up predicted by *Kicker* and are no start players. Group $\mathcal{B}$ consists of those start players that are not part of the expected starting line-up ($\mathcal{B}$). Group $\mathcal{C}$ includes all remaining start players. In the second step, we calculate the dissimilarity between the players in $\mathcal{C}$ and $\mathcal{A}$, as well as $\mathcal{C}$ and $\mathcal{B}$:

$$\Delta(\mathcal{A},\mathcal{C}) = \frac{1}{|\mathcal{A}|} \cdot \frac{1}{|\mathcal{C}|} \cdot \sum_{j \in \mathcal{A}} \sum_{k \in \mathcal{C}} (1 - s_{jk}) \tag{7}$$

$$\Delta(\mathcal{B},\mathcal{C}) = \frac{1}{|\mathcal{B}|} \cdot \frac{1}{|\mathcal{C}|} \cdot \sum_{j \in \mathcal{B}} \sum_{k \in \mathcal{C}} (1 - s_{jk}) \tag{8}$$

where s_{jk} is a dummy that is equal to 1 if the players j and k come from the same country, and 0 otherwise. In the last step, we create the instrumental variable $\mathrm{Z} \in [-1, 1]$ as the difference of two dissimilarity scores:

$$\mathrm{Z} = \Delta(\mathcal{B},\mathcal{C}) - \Delta(\mathcal{A},\mathcal{C}). \tag{9}$$

Our instrumental variable can only be different from 0 if there is an unexpected change in the starting line-up and if the players in $\mathcal{A}$ and $\mathcal{B}$ were not born in the same countries. Z is positive if the dissimilarity between the players that unexpectedly belonged to the starting line-up and the other start players ($\Delta(\mathcal{B},\mathcal{C})$)

[15] Bettinger and Long (2005) uses unexpected absences of university lecturers to investigate the importance of role models for study choice, while Herrmann and Rockoff (2012) examine how unexpected absences of teachers affect the performance of students.

is larger than the dissimilarity between the players that unexpectedly dropped out from the starting line-up and the players that belonged to both the actual and the expected starting line-up ($\Delta(\mathcal{A}, \mathcal{C})$). As intended, we find a positive correlation between our instrumental variable and the birthplace diversity of the fielded players (see Figure D.3).

As explained in Section 4.1, in our baseline analysis, we assume a quadratic functional relationship between birthplace diversity and team performance. Thus, we need instrumental variables for both the level of birthplace diversity and its squared term. Following Ashraf and Galor (2013), we use a three-stage procedure to meet this challenge.[16] First, we perform a zero-stage regression in which our measure of birthplace diversity is regressed on Z and the controls of the second-stage equation:

$$\mathrm{B}_{isrmd} \;=\; \phi \cdot \mathrm{Z}_{isrmd} \;+\; \pi \cdot \mathbf{X}_{isrmd} \;+\; \Lambda_{isrm} \;+\; \Theta_d \;+\; \zeta_{isrmd}. \tag{10}$$

Afterwards, we use the point estimates that we obtain when estimating (10) to compute predicted values of birthplace diversity ($\widehat{B}$). Finally, we exploit these predicted values and their squared terms as our instruments. Consequently, the first-stage equations are:[17]

$$\mathrm{B}_{isrmd} \;=\; \rho_1 \cdot \widehat{\mathrm{B}}_{isrmd} \;+\; \rho_2 \cdot \widehat{\mathrm{B}}^2_{isrmd} \;+\; \alpha \cdot \mathbf{X}_{isrmd} \;+\; \Lambda_{isrm} \;+\; \Theta_d \;+\; \eta_{isrmd} \tag{11}$$

$$\mathrm{B}^2_{isrmd} \;=\; \delta_1 \cdot \widehat{\mathrm{B}}_{isrmd} \;+\; \delta_2 \cdot \widehat{\mathrm{B}}^2_{isrmd} \;+\; \alpha \cdot \mathbf{X}_{isrmd} \;+\; \Lambda_{isrm} \;+\; \Theta_d \;+\; \mu_{isrmd}. \tag{12}$$

An objection against our second instrumental variable approach might be that some of the unexpected replacements are not triggered by plausibly exogenous events since they are voluntarily made by the managers. To allay this legitimate concern, we show robustness checks in which we only exploit those replacements that are caused by injuries (for details, see Section 5.2.2). In these robustness checks, we have to make some small adjustments with regard to the procedure that we use to compute Z (for details, see Appendix C).

Conceptually, the main difference between our first and second instrumental variable approach is that the first approach exploits a complete but artificial line-up, while the second approach uses the replacements of single players.[18] If

[16] For details about the econometric foundation of the approach used by Ashraf and Galor (2013), see Angrist and Pischke (2009) and Wooldridge (2010). We need a zero-stage regression since Z^2 and B^2 are only weakly correlated. Two reasons exist for this weak correlation: first, $\mathrm{Z} \in [-1, 1]$ and $B \in [0, 1]$ have different domains, and second, squaring is a non-monotonic transformation if the domain is $[-1, 1]$, while it is a monotonic transformation if the domain is $[0, 1]$.

[17] It is irrelevant for the second-stage estimates of β_1, β_2, and B^* whether we use the variable Z or the predicted value $\widehat{B}$ in the first-stage equations. We choose $\widehat{B}$ because it eases the interpretation of the reduced-form estimates.

[18] From a methodological perspective, our second instrumental variable approach has a clearer data generating process than the first approach. Therefore, for the second approach, it easier to evaluate whether the exclusion restriction holds. A key advantage of the first approach is that its implementation is much more straightforward.

both approaches satisfy their exclusion restriction, we expect them to produce similar regression results. We also think our approaches nicely complement each other since it is rather unlikely that an unobserved factor exists that causes a violation of both exclusion restrictions.

5 Results

5.1 Main results

5.1.1 Fixed effect estimates

Table 1 shows the main results of our empirical analysis. In Column 1 and 2, we report our fixed effect estimates. Columns 3 and 4 present results from instrumental variable regressions. In all specifications, we use the goal difference at the end of the match as our measure of team performance and cluster the standard errors at the club-by-season-by-round-by-manager level. Furthermore, we always assume a quadratic functional form to verify whether birthplace diversity enhances team performance only up to a certain degree.

The common way to test for the presence of a hump-shaped relationship is checking whether the estimate of the quadratic term of the variable of interest ($\hat{\beta}_2$) is negative and statistically significant. Lind and Mehlum (2010) argue that this conventional procedure creates misleading results if the optimal level is not within the lower and upper bound of the data range (see also Simonsohn, 2018). More specifically, the point estimate of β_2 indicates whether the relationship between birthplace diversity and team performance is concave, but it does not show whether the optimal level of birthplace diversity is between 0 and 1. Lind and Mehlum (2010) also present a statistical test that addresses this issue. We apply this test and thus structure our main regression table in the following manner. The upper part shows the results of the Lind-Mehlum-Test, consisting of an estimate of the optimal level of birthplace diversity (B^*) and a p-value that reveals the presence of a hump-shaped relationship within the data range. The lower part reports the regression coefficients of β_1 and β_2 and their p-values.

In Column 1, we use a sample that consists of 7,038 *Bundesliga* matches to examine whether the relationship between birthplace and team performance is hump-shaped (for summary statistics, see Table E.3). Consistent with this hypothesis, we find that the estimate of the parameter β_2 is negative. However, $\hat{\beta}_2$ is not statistically significant at conventional levels (p-value: 0.234). The Lind-Mehlum-Test suggests that the optimal level of birthplace diversity is 0.148, but also shows that $\hat{\mathrm{B}}^*$ is not statistically different from 0 (p-value: 0.387). Column 1 of Table 2 confirms this finding by showing that the linear regression model provides evidence for a negative and statistically significant correlation between birthplace diversity and team performance.

Table 1 Birthplace diversity and team performance (main results).

	(1)	(2)	(3)	(4)
	Lind-Mehlum-Test			
Optimal level of birthplace diversity ($\hat{B}^*$)	0.148	0.129	0.608*	0.605**
	(0.387)	(0.436)	(0.073)	(0.023)
	Regression coefficients			
Birthplace divers. ($\hat{\beta}_1$)	0.311	0.238	5.372	41.924**
	(0.774)	(0.873)	(0.103)	(0.044)
Birthplace divers. sq. ($\hat{\beta}_2$)	-1.048	-0.927	-4.420*	-34.670**
	(0.234)	(0.435)	(0.094)	(0.043)
Estimation technique	OLS	OLS	IV	IV
IV approach	-	-	1st	2nd
Outcome variable	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.
Observations	14,076	7,956	7,956	7,956
Seasons	23	13	13	13
Fixed effects	Yes	Yes	Yes	Yes
Basic controls	Yes	Yes	Yes	Yes
Quality controls	No	Yes	Yes	Yes
First-stage F-statistic (B)	-	-	126.3	13.7
First-stage F-statistic (B^2)	-	-	286.5	14.3
Sample mean (B):	0.669	0.699	0.699	0.699

Notes: The table reports results from fixed effect and instrumental variable regressions. The upper part of the table shows the results of the Lind-Mehlum-Test, the lower part presents regression coefficients. We cluster standard errors at the club-by-season-by-round-by-manager level and report p-values in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

A concern regarding our first fixed effect analysis might be that we do not adequately control for the quality of the teams. This problem might exist since our two measures of players' quality (market value, strength in the video game *FIFA*) are only available for the 13 latest *Bundesliga* seasons. In Column 2 of Table 1, we therefore analyze a reduced sample of 3,978 matches (for summary statistics, see Table E.4). The results of this analysis are virtually the same as before. In particular, the estimated optimal level of birthplace diversity is still small ($\hat{B}^* = 0.129$) and not statistically different from 0 (p-value: 0.436). Column 2 of Table 2 shows that the result of the linear regression model also hardly changes if we reduce our sample and control more explicitly for the quality of the fielded players.

Taken together, our fixed effect regressions provide evidence for a negative rather than a hump-shaped relationship between birthplace diversity and team performance. Our results are thus consistent with other studies that use data from the professional sports industry and fixed effect approaches (see e.g. Haas and Nüesch, 2012, Kahane et al., 2013, Maderer et al., 2014). These related studies also suggest different explanations for why nationally diverse teams perform less well than nationally homogeneous teams. The most common arguments are that heterogeneous teams suffer from cultural and linguistic barriers, have lower group coherence, and are more prone to inter-group biases. Our key objection against these explanations is that they leave open why the actual levels of birthplace

Table 2 Birthplace diversity and team performance (linear model).

	(1)	(2)	(3)	(4)
Birthplace divers. ($\hat{\beta}_1$)	-0.903***	-0.883**	0.059	0.296
	(0.002)	(0.020)	(0.956)	(0.858)
Estimation technique	OLS	OLS	IV	IV
IV approach	-	-	1st	2nd
Outcome variable	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.
Observations	14,076	7,956	7,956	7,956
Seasons	23	13	13	13
Fixed effects	Yes	Yes	Yes	Yes
Basic controls	Yes	Yes	Yes	Yes
Quality controls	No	Yes	Yes	Yes
First-stage F-statistic	-	-	640.5	272.80
Sample mean (B):	0.669	0.699	0.699	0.699

Notes: This table presents results from linear fixed effect and instrumental variable regressions. We cluster standard errors at the club-by-season-by-round-by-manager level and report p-values in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

diversity differ considerably from 0. Put differently, it remains rather unclear why managers field nationally diverse teams if birthplace diversity has a strong negative effects on team performance.[19] A potential answer to this question is that managers might heavily overestimate (underestimate) the positive (negative) aspects of birthplace diversity. However, since the team managers and their staff members are experts and frequently analyze the performance of their team, we have some doubts whether such gross misjudgment is plausible.

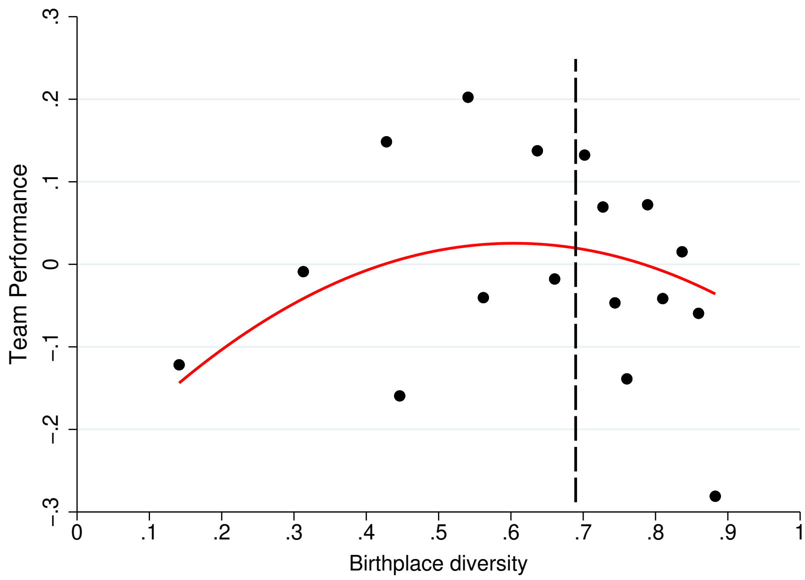
5.1.2 Instrumental variable estimates

In Appendix B, we develop an alternative explanation for why our fixed effect approach creates estimates of the optimal level of birthplace diversity that are considerably smaller than the average birthplace diversity of the fielded players. Our explanation is that a fixed effect regression produces biased estimates of the effect of birthplace diversity on team performance because this conventional approach does not address the problem of reverse causality. More specifically, in our supplementary analysis, we first show that the managers replace defensive with offensive players during the match if their team performs poorly (see also Garicano and Palacios-Huerta, 2014). Afterwards, we argue that these performance-based substitutions often have no notable effect on the outcome of the match, but increase the level of birthplace diversity. The second effect sets in because the share of foreign offensive players in the *Bundesliga* is much larger than the share of foreign defensive players.

As explained in Section 4.2, we address the problem of reverese causality and other endogeneity issues with two instrumental variable approaches. In our first

[19]The fixed effect estimates reported in Columns 1 and 2 of Table 2 suggest that the goal difference decreases by almost one goal per match if a nationally homogeneous team will be replaced by a highly heterogeneous team.

Figure 3 Birthplace diversity and team performance.



Notes: This figure presents a graphical illustration of the results of our first instrumental variable approach. The red solid line shows the predicted relationship between birthplace diversity and team performance. For the sake of vividness, we use a binned scatterplot to present the underlying raw data.

approach, we use the birthplace diversity in the starting line-up predicted by the soccer magazine *Kicker* as instrumental variable. Our second approach exploits unexpected replacements in the starting line-up as source of plausibly exogenous variation.

Column 3 of Table 1 reports the regression results produced by our first instrumental variable approach (for a graphical illustration, see Figure 3). The first-stage and reduced-form estimates can be found in Table E.5 and E.6. Compared with the results of our fixed effect approach, we observe three key differences. First, the estimate of β_2 is now statistically significant at the ten percent level. Second, the implied optimal level of birthplace diversity is 0.609 and thus much larger than the fixed effect estimates of B*. Finally, the results of the Lind-Mehlum-Test provide statistical evidence for the presence of a hump-shaped relationship between birthplace diversity and team performance (p-value: 0.073). Another remarkable result is that the linear model (see Column 3 of Table 2) produces an estimate of β_1 that is not statistically different from 0 (p-value: 0.946). By contrast, our non-linear model suggests that changing the birthplace diversity by 10 percentage points in either direction at the optimal level reduces the goal difference by 0.044, which is 3.149 percent of the sample mean. Projected to the season as a whole, such a change thus amounts to a decrease in the goal difference by more than one goal.

Column 4 of Table 1 shows that our second instrumental variable approach produces similar results as our first approach. A notable difference is that the estimates of β_2 and B^* are now statistically significant at the five rather than the ten percent level. Another difference is that the strength of the first-stage relationship decreases considerably when applying the second approach. A weak-instrument problem is nonetheless unlikely because the first-stage F-statistics are well above 10.

Taken together, our two instrumental instrumental variable approaches provide empirical evidence for a causal non-linear effect of birthplace diversity on team performance. This finding is consistent with economic models, suggesting that teams benefit from diversity up to a certain degree but that diversity beyond this level has no further benefits or even negative effects due to increased communication problems (see e.g. Ashraf and Galor, 2013, van Knippenberg et al., 2004).[20] Interestingly, we also observe that our instrumental variable approaches identify an optimal level of birthplace diversity ($\approx$ 0.61) that is similar to the average birthplace diversity of the fielded players ($\approx$ 0.70).[21]

5.2 Robustness checks

5.2.1 Alternative measure of team performance

As many related studies, our baseline analysis uses the difference between goals scored and goals allowed as measure of team performance. An objection against this approach might be that maximizing the goal difference is only a secondary objective because the final ranking position is primarily determined by the total number of ranking points. The goal difference is only of relevance if two clubs have the same number of ranking points at the end of a *Bundesliga* season. Put differently, a legitimate question is why we use the goal difference instead of the number of ranking points to measure team performance. We prefer the former approach because the goal difference is a more fine-grained measure of team performance and thus has more discriminating power.[22]

To allay the concern that our results are driven by our measure of team performance, we rerun our baseline analyses, using the number of ranking points as the dependent variable. We find that the results of this robustness check do not considerably differ from our main results (see Tables E.7 and E.8). In

[20] In Section 5.3, we will present qualitative and quantitative evidence, suggesting that misunderstandings that are caused by linguistic barriers indeed play an important role.

[21] We need to leave open why the average level of birthplace diversity exceeds the predicted optimal level. A potential reason is that managers overestimate/underestimate the positive/negative effects of birthplace diversity. Another explanation might be that there are internal constraints (e.g. budget restrictions) or market frictions (e.g. limited supply of appropriate players) that prevent managers from fielding a team with the optimal level of birthplace diversity.

[22] Frequently, the number of ranking points does not indicate performance differences even if they exist. For example, if a team loses a match zero to one, it receives the same number of ranking points as it would obtain if it loses zero to five. Due to this relatively low distinctiveness, we argue that the number of ranking points is a less appropriate measure of team performance than the goal difference.

particular, we observe that both of our instrumental variable approaches provide statistical evidence for a hump-shaped effect of birthplace diversity on team performance. The estimated optimal level of birthplace diversity is 0.58 in both cases. However, compared with our baseline analyses, the instrumental variable estimates of β_2 and B^* are on average slightly less statistically significant. We believe that this change is plausible because the number of ranking points is a less sensitive measure of team performance than the goal difference.

5.2.2 Using inquiries as source of variation

A concern regarding our second instrumental variable approach might be that some of the unexpected replacements are not triggered by plausibly exogenous events. The estimates reported in Column 4 of Table 1 are thus biased if the cause of the replacements is neither captured by our fixed effects nor by our match-specific control variables. To substantiate that this is unlikely to be the case, we perform a robustness check that only exploits those replacements that are explainable with short-term injuries. The basic idea behind this robustness check is that short-term injuries occur coincidentally and can thus be used to create plausibly exogenous variation in teams' composition.

We define that an injury serves as the explanation for an unexpected replacement in either of the following cases: (i) The soccer magazine *Kicker* predicted that a player would be a starting player, but this player did not play since he was injured or did not completely recover from an injury. (ii) *Kicker* predicted that a player could not play because of an injury, but he was fit enough for attending the match. To identify the incorrect predictions that are caused by an injury, we read articles and match reports published by *Kicker* and exploited the online database *Transfermarkt*. In total, injuries explain 30 percent of the incorrect predictions.

Column 1 of Table E.13 presents the results from the instrumental variable regression that exploits short-term injuries. We observe only small changes when comparing these results with the baseline results. For example, the estimated optimal level of birthplace diversity is now 0.631 (p-value: 0.049) rather than 0.605 (p-value: 0.023) and the p-value of $\hat{\beta}_2$ increases from 0.043 to 0.093. In total, we thus conclude that this robustness check confirms the presence of a hump-shaped relationship between birthplace diversity and team performance.

5.2.3 Alternative definition of team

In our baseline analyses, a team consists of all players who participated in a match for at least one minute. This approach might be criticized since it is debatable whether players who are fielded for a few minutes actually have a notable effect on the performance of a team. As a robustness check, we thus count only those players as team members who played for at least 30 minutes. Columns 2 and 3 of Table E.13 show that our results hardly change if we use this alternative definition of "team".

5.2.4 Alternative procedure for measuring unexpected changes in birthplace diversity

Our second instrumental variable approach distinguishes three groups of players and uses a three-stage procedure to create an instrumental variable (Z) that reflects the extent to which the level of birthplace diversity changes due to an unexpected replacement in the starting line-up (see Section 4.2.2). An alternative approach is to calculate the difference of the birthplace diversity in the actual starting line-up and the birthplace diversity in the starting line-up predicted by *Kicker*. Column 4 of Table E.13 shows that this alternative approach produces almost the same results as our basic approach. We prefer the latter for two reasons: first, the first-stage F-statistics are slightly higher, and second, it is easier to restrict the analysis to those unexpected replacements that are caused by injuries.

5.2.5 Past performance

Another objection against our basic regression models might be that we do not adequately control for past performance. Controlling for the performance in the previous matches might be important because it affects the self-confidence of the players. Furthermore, the likelihood that a manager unexpectedly changes his starting line-up might depend on the past performance of his team. To allay the concern that this omitted factor might cause a violation of the exclusion restriction and thus biases our instrumental variable estimates, we augment our regression models by the first two lags of the dependent variable. For our first approach, we observe that the estimates of β_1, β_2 and B^* become statistically more significant when controlling for past performance (see Column 5 of Table E.13). The results of our second instrumental variable approach remain virtually unchanged (see Column 6 of Table E.13).

5.2.6 Migration during childhood

A standard argument for why birthplace diversity might be conducive to team performance is that people from different countries were trained under different education systems and thus have different skills (Alesina et al., 2016, Freeman and Huang, 2015). Obviously, this argument can only apply if workers receive their basic training in their country of birth. Among others, in our case, this key condition might not hold because soccer clubs might scout and hire players at very young ages.

To investigate whether migration during childhood indeed gives cause for concern, we proceed in two steps. In the first step, we use the web database *Transfermarkt* to identify for each player the country where he lived during his youth. We observe that this country differs from the country of birth for 222 players (6.8 percent). About half of them lived in Germany during their childhood. For almost 100 of the players who are not German-born but grew up in Germany, we find biographical information that imply why they move to Germany. In the vast majority of cases, the parents of the player migrated to Germany in order to work there or because they fled from their home country for political reasons. For only three players, we find evidence, suggesting that they moved to Germany due to an offer from a soccer club.[23]

To illustrate that our baseline results are not driven by those players who moved from one country to another during their childhood, we rerun our main analyses, using the country in which the players lived during their youth rather than the country of birth. Columns 7 and 8 of Table E.13 suggest that our instrumental variable estimates are robust to this methodological change.

5.2.7 Alternative measure of birthplace diversity

From a conceptual perspective, it is straightforward that two individuals either have the same country of birth or were born in two different countries. As a consequence, most studies assume a binary similarity matrix $\mathbf{B} = \{b_{lk}\}_{l,k=1,\ldots,n}$ when measuring birthplace diversity (see e.g. Alesina et al., 2016). In our main analyses, we followed these studies and treated all the country differences as equivalent. A potential objection against this classical approach for measuring birthplace diversity is that some countries are perceived to be more similar to each other than others and that neglecting these differences in the degree of similarity might create a systematic bias. To address this issue, Bossert et al. (2011) propose a fractionalization index that takes into account different degrees of similarities. The key problem when using this more general fractionalization index is to operationalize the similarity matrix. Put differently, while probably most people agree that the similarities between German and Swiss players are larger than the similarities between players from Japan and Serbia, creating an uncontroversial measure for the extent to which the German and Swiss players are more similar is nearly impossible.

In Columns 9 and 10 of Table E.13, we present results from instrumental variable regressions in which we use a more distinctive approach for measuring the similarity between players from different countries. More specifically, in this

[23]For example, the parents of the former German national team players Miroslaw Klose and Lukas Podolski were labor migrants. The father of the former German national team player Gerald Asamoah was a political refugee. The current captain of the Austrian national team, Julian Baumgartlinger, was hired by a German soccer club (*TSV 1860 München*) at the age of 13.

robustness check, we use data on linguistic distance from Spolaore and Wacziarg (2016) to proxy the degree of similarity.[24] Compared with our main results, we observe only minor changes. If anything, we obtain slightly smaller estimates of B^* (≈ 0.56). We think that the robustness of our results is hardly surprising because our alternative measure of birthplace diversity is highly correlated with our baseline measure.

5.2.8 Dynamic assessment of player's quality

Our basic regression models include two control variables that reflect the quality of the fielded players. The first is the market value as indicated by the web database *Transfermarkt*. Our second measure of quality is an expert-based rating provided by the popular video game *FIFA*. A concern regarding both measures might be that they are only updated once a year at the beginning of a season. Consequently, our two measures change if there are changes in the line-up but not because of form fluctuations. For our identification strategy, this might be problematic because form fluctuations are a common reason for changes in the line-up.

To produce an index that reflects the current form of the fielded players, we exploit individual-level data provided by the soccer magazine *Kicker*. After each match, *Kicker* grades the individual performance of all players that played at least 30 minutes. The grade scaling ranges from 1.0 (very good) to 6.0 (very bad) in increments of 0.5.

We proceed in three steps to measure player's current form. In the first step, we hand-collected the grades of all matches in our database. In total, these are 163,971 grades. In the second step, we calculate for each player (p) and each match (q) the difference between his grade (G) and his average grade during the season ($\bar{G}$):

$$\Delta_{p,q} = G_{p,q} - \bar{G}_p \tag{13}$$

In the last step, we compute a team-level measure of the extent to which the fielded players deviated from their normal form in the previous match:

$$F_{i,q} = \frac{1}{n_{i,q}} \sum_{p=1}^{n_{i,q}} \Delta_{p,q-1} \tag{14}$$

where i denotes a team and n the number of fielded players.[25] F is positive

[24] Despite the fact that the measurement of linguistic distances is a controversially discussed issue, we think for two reasons that using linguistic distance as measure for similarity has two weaknesses. The first is that we treat players whose countries of birth differ but have the same official language as equal. The key second weak point is that we put too much emphasis on one particular channel (communication) through which birthplace diversity might affect team performance.

[25] If a particular player did not attend in the previous match, we set Δ equal to 0 and thus implicitly assume that he has his average form.

(negative) if the players underperformed (overperformed) in the last match.

Columns 11 and 12 of Table E.13 illustrate how our instrumental variable estimates react when controlling for the current form of the players. Compared with our baseline regressions, we only observe minor changes. For example, the optimal level of birthplace diversity estimated by our first instrumental variable approach shifts from 0.608 (p-value: 0.073) to 0.593 (p-value: 0.066) if we add the measure F as a control variable to our regression model.

5.2.9 Alternative clustering methods

In our main regressions, we cluster the standard errors at the club-by-season-by-round-by-manager level and thus in accordance with our main panel dimension. A concern regarding this strategy might be that the number of observations per cluster is too low for producing reliable estimation results.[26] To alleviate this concern, we show in Table E.9 how our results change when using alternative clustering methods. In Columns 1 and 2, we cluster at the club-by-season level. The number of observation per cluster increases to 34 in this case. In Columns 3 and 4, we apply a two-way clustering. Following the guidelines of Cameron et al. (2012) and Cameron and Miller (2015), we cluster at the cross-sectional (club-by-season-by-round-by-manager) and the panel dimension (match day). In Columns 5 and 6, we use standard errors that are clustered at the club-by-season and at the match day level to determine how statistically significant our estimates are. For each of the three alternative clustering methods and both instrumental variable approaches, we find that the Lind-Mehlum-Test confirms the presence of a hump-shaped relationship between birthplace diversity and team performance.

5.2.10 Piecewise linear regressions

Our baseline estimates suggest that the optimal level of birthplace diversity is around 0.6. This result implies that an increase in birthplace diversity enhances (reduces) team performance if the actual level of birthplace diversity is smaller (larger) than the predicted optimal level. Put differently, if we estimate a linear regression model and restrict the sample to those observations with $\mathrm{B} < \mathrm{B}^*$ ($\mathrm{B} > \mathrm{B}^*$), we should observe that birthplace diversity has a positive (negative) effect on team performance. Using our first instrumental variable approach, we show in Columns 1 and 2 of Table E.10 that this hypothesis indeed holds. Notably, the slope parameter of the two regression lines have similar absolute values. This similarity suggests that the quadratic functional form assumption made by the Lind-Mehlum-Test is appropriate.

Combining a piecewise linear regression approach with our second instrumental variable approach is slightly more difficult since it exploits changes in birthplace

[26] The maximal number of observations per cluster is 17 when using our basic approach. The average number of observations per cluster is

diversity to create plausibly exogenous variation. Assume that the actual level of birthplace diversity exceeds the ideal level (B > B*) and that an unexpected change in birthplace diversity takes place ($Z \neq 0$). If this change is positive (negative), birthplace diversity moves away from (towards) the optimal level of birthplace diversity. We therefore expect a negative (positive) estimate when estimating a linear model and restricting the sample to the observations with $Z > 0$ ($Z < 0$). Columns 3 and 4 of Table E.10 confirm this expectation.[27]

5.2.11 Simonsohn's two-lines test

So far, we have used the Lind-Mehlum-Test to check whether the relationship between birthplace diversity and team performance is hump-shaped. A potential concern regarding this approach is that it imposes a quadratic functional form. To address this issue, Simonsohn (2018) presents an alternative procedure. The basic idea behind this method is to estimate two regression lines (one for low and another for high levels of the variable of interest) and to test whether the slope parameters of these regression lines differ in their sign. To determine the threshold value that splits the sample in two parts, Simonsohn (2018) proposes a method, called Robin-Hood-Algorithm, that consists of five steps. In the first step, this algorithm estimates a cubic spline for the relationship between the explanatory variable (x) and the dependent variable (y). Afterwards, Simonsohn's algorithm computes the most extreme internal fitted value ($\hat{y}_{max}$). In the third step, the Robin-Hood-Algorithm identifies those predicted values that differ from $\hat{y}_{max}$ by at most one standard deviation. All predicted values that satisfy this condition belong to the set of potential threshold values ($\mathcal{Y}_{flat}$). In the fourth step, the algorithm estimates an interrupted regression model, using the median value of x within $\mathcal{Y}_{flat}$. The output of this interrupted regression will be two slope-parameters (z_1, z_2). In the last step, the Robin-Hood-Algorithm computes the actual threshold value as the $\frac{|z_2|}{|z_1|+|z_2|}$th percentile of the x values that are associated with $\mathcal{Y}_{flat}$ (for further details and an illustrative example, see Simonsohn, 2018).

When applying Simonsohn's Robin-Hood-Algorithm to our data, we obtain a threshold value of 0.62. Reassuringly, this threshold is almost the same as the predicted optimal level of birthplace diversity that we find when using the Lind-Mehlum-Test. Column 5 of Table E.10 illustrates that the results of an interrupted instrumental variable regression that uses 0.62 as threshold value confirm the presence of a non-linear relationship between birthplace diversity and team performance. Importantly, we also find that the slope parameters of both regression lines are statistically significant at conventional levels (p-values: 0.069, 0.060).

[27] We cannot conduct this robustness check for the case B < B* since the sample size is too small.

5.3 Mechanisms and moderating factors

The regression results presented in the previous sections provide strong evidence for a hump-shaped effect of birthplace diversity on team performance. In this section, we discuss potential explanations for this relationship. In addition, we examine whether the optimal level of birthplace diversity depends on contextual factors.

Several existing theories suggest that diversity increases the range of skills and thus enhances the performance of a working group (see e.g. Alesina et al., 2016, Lazear, 1999). For our case, this implies that our results are only plausible if soccer players have country-specific skills. We are convinced that this key prerequisite holds since various sports science studies suggest that country-specific soccer skills exist due to different education systems. For example, Sarmento et al. (2013) and Basevitch et al. (2013) show that Italian players have particularly high tactic skills, while players from Spain have outstanding passing skills (for further evidence, see Dellal et al., 2011, Mitrotasios et al., 2019, Sarmento et al., 2014).[28]

The common argument for why an increase in birthplace diversity does not necessarily improve team performance is that highly diverse groups are likely to suffer from communication problems (see e.g. Lang, 1986, van Knippenberg et al., 2004).[29] Unfortunately, we cannot run an empirical analysis that directly confirms this assumption. The reason is that the language skills of soccer players cannot systematically be measured due to limited data availability. However, we found various interviews in which players and managers stressed the crucial role of language skills and interpersonal communication on the field (for an overview, see Table E.14). In addition, several linguistic studies suggest that communication barriers affect the performance of soccer teams. For example, Kellerman et al. (2005) and Ringbom (2012) run surveys among soccer managers and players in the Netherlands and Finland and observe that language problems are perceived as the key reason for misunderstandings if a team consists of players from different countries (see also Lavric et al., 2008).

Another potential explanation for why teams might benefit from birthplace diversity only up to a certain degree is that social categorization might lead to performance-reducing inter-group biases if the team members are highly diverse (O'Reilly et al., 1989, van Knippenberg et al., 2004). Although we lack data that allows us to precisely assess the relevance of this mechanism, we think for two reasons that severe inter-group biases are relatively unlikely in our case. First, Price et al. (2013) present empirical results from the NBA, suggesting that racial

[28] We also found some interviews in which players and managers suggest that country-specific skills exist in soccer. For example, Nuri Sahin, a famous soccer player in Germany, mentioned in an interview with the soccer magazine *Kicker* that Japanese offensive players have special tactical skills (see `https://www.kicker.de/732654/artikel`).

[29] For empirical evidence on the negative effect of linguistic diversity on team performance, see Dale-Olsen and Finseraas (2020).

differences do not cause inter-group biases in sports teams. Of course, ethnic diversity is not the same as birthplace diversity, but we believe that inter-group biases are more prevalent in racially diverse working groups than in nationally diverse working groups. Second, studies in psychology suggest the risk of inter-group biases is relatively low if a team has a common goal or competes with another team (see e.g. Gaertner et al., 1993, 2000, Lowe, 2020).

To empirically support our explanations, we exploit the facts that the overall performance of soccer teams depends on two different aspects (goal scoring, goal preventing) and that interpersonal exchange on the field is more important for the defensive performance than for the offensive performance. Put differently, if our hunch regarding the underlying channels is correct, we should find that the optimal level of birthplace diversity is larger for offensive than for defensive players.[30] Table E.11 shows that this is indeed the case. More specifically, we observe that the optimal level of birthplace diversity for offensive performance ($B^* = 0.65$) exceeds the optimal level for defensive performance ($B^* = 0.52$). However, for two reasons, we need to interpret this result with some caution. First, the two estimates of the optimal levels of birthplace diversity are only statistically significant at relatively low levels, according to the Lind-Mehlum-Test (p-values: 0.171, 0.097). Second, the Lind-Mehlum-Test does not allow to check whether the two predicted optimal levels of birthplace diversity are statistically different from each other. However, to support the view that the importance of interpersonal exchange is likely to be a moderating factor for the relationship between birthplace diversity and team performance, we can at least say that the difference of the two estimates is larger than a standard deviation.

6 Conclusion

Multinational working groups are becoming more and more common in virtually all advanced economies. For many practitioners, it is thus important to know whether a nationally diverse team performs better or worse than a homogeneous team. We address this crucial question, using rich hand-collected information on 7,208 matches and 3,266 players from the *Bundesliga*, the highest division of German soccer. In our empirical analysis, we apply two different instrumental variable approaches to identify the causal effect of birthplace diversity on team performance. Their results suggest that this effect is hump-shaped. In other words, we find that birthplace diversity positively affects team performance only up to a certain level. Beyond this threshold, birthplace diversity reduces team performance.

An objection against our work might be that professional soccer is a rather special industry and that the external validity of our results is thus low. We

[30] We also believe that such a pattern could hardly be explained with inter-group biases since this requires that the defensive players are more likely to discriminate their co-workers than the offensive players. If anything, our anecdotal evidence suggests the opposite.

argue that this is not the case since effective exchange and diversity in skills are both factors for success in many fields. Examples include consultancy, arts and music, research and development, as well as marketing. Obviously, the level of birthplace diversity that optimizes the performance of a team differs within and between these working fields. Consistent with the view that the optimal level of diversity is moderated by task-specific factors, we present regression results suggesting that the optimal team structure depends on how important interpersonal exchange is in the production process.

We believe that the contribution of our paper goes beyond the result that birthplace diversity has a hump-shaped effect on team performance since we illustrate that a frequently applied empirical method is likely to produce misleading regression results. More specifically, we find that classical fixed effect regressions produce heavily downward-biased estimates of the optimal level of birthplace diversity. The main reason for this bias is that team compositions change endogenously during the production process. In our particular case, we observe that managers (unintentionally) increase the birthplace diversity of the team during a match if their team does not perform well.[31] Basic econometric theory implies that this problem of reverse causality biases the predicted ideal level of birthplace diversity towards 0. An important question in this regard is whether we need to expect similar problems when studying other types of diversity or when using data that does not come from the professional sports industry. From our point of view, these questions can hardly be negated since endogenous changes in team compositions exist in various fields.[32] We therefore believe that conclusions should always be drawn with a lot of caution if fixed effect methods are used to provide evidence on the consequences of diversity for team performance.[33]

Our study suggests two paths for future research. First, we think that it is crucial to check whether the fixed effect estimates reported in previous studies can be confirmed when using a method that addresses the problem of reverse causality. Our paper helps to answer this question since our two instrumental variable approaches can easily be adjusted to different types of diversity and institutional environments. Second, we are convinced that future studies should provide more evidence on the factors that moderate the relationship between diversity and team performance. Without such evidence, we can hardly provide

[31] More specifically, soccer managers replace defensive with offensive players if their teams are behind. These performance-based substitutions increase the level of birthplace diversity because the share of foreign offensive players in the *Bundesliga* is larger than the share of foreign defensive players.

[32] For instance, young researchers might be more likely to ask a senior researcher to join their team if they have problems with their project. Similarly, managers might change the age or gender composition of a team depending on how it performs.

[33] Of course, if the problem of reverse causality exists, it depends on the specific research environment whether a fixed effect approach creates a downward-biased or an upward-biased estimate of the optimal level of diversity.

specific practical advices on how to structure a working group.

References

Adams, R. B. and Ferreira, D. (2009). Women in the boardroom and their impact on governance and performance. *Journal of Financial Economics*, 94(2):291–309.

Ager, P. and Brueckner, M. (2013). Cultural diversity and economic growth: Evidence from the US during the age of mass migration. *European Economic Review*, 64(1):76–97.

Aggarwal, I. and Woolley, A. W. (2019). Team creativity, cognition, and cognitive style diversity. *Management Science*, 65(4):1586–1599.

Alesina, A., Harnoss, J., and Rapoport, H. (2016). Birthplace diversity and economic prosperity. *Journal of Economic Growth*, 21(2):101–138.

Alesina, A. and La Ferrara, E. (2005). Ethnic diversity and economic performance. *Journal of Economic Literature*, 43(3):762–800.

Ancona, D. G. and Caldwell, D. F. (1992). Demography and design: Predictors of new product team performance. *Organization Science*, 3(3):321–341.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Angrist, J. D. and Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24(2):3–30.

Apesteguia, J., Azmat, G., and Iriberri, N. (2012). The impact of gender composition on team performance and decision making: Evidence from the field. *Management Science*, 58(1):78–93.

Apesteguia, J. and Palacios-Huerta, I. (2010). Psychological pressure in competitive environments: Evidence from a randomized natural experiment. *American Economic Review*, 100(5):2548–64.

Arcidiacono, P., Kinsler, J., and Price, J. (2017). Productivity spillovers in team production: Evidence from professional basketball. *Journal of Labor Economics*, 35(1):191–225.

Ashraf, Q. and Galor, O. (2013). The "Out of Africa" hypothesis, human genetic diversity, and comparative economic development. *American Economic Review*, 103(1):1–46.

Balsmeier, B., Frick, B., and Hickfang, M. (2019). The impact of skilled immigrants on their local teammates' performance. *Applied Economics Letters*, 26(2):97–103.

Basevitch, I., Yang, Y., and Tenenbaum, G. (2013). Is the best defense a good offense? Comparing the Brazilian and Italian soccer styles. *International Journal of Fundamental and Applied Kinesiology*, 45(2):213–221.

Bazzi, S. and Clemens, M. A. (2013). Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. *American Economic Journal: Macroeconomics*, 5(2):152–86.

Bell, S. T., Villado, A. J., Lukasik, M. A., Belau, L., and Briggs, A. L. (2011). Getting specific about demographic diversity variable and team performance relationships: A meta-analysis. *Journal of Management*, 37(3):709–743.

Bettinger, E. P. and Long, B. T. (2005). Do faculty serve as role models? The impact of instructor gender on female students. *American Economic Review*, 95(2):152–157.

Bloom, N. and Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *Quarterly Journal of Economics*, 122(4):1351–1408.

Boone, C. and Hendriks, W. (2009). Top management team diversity and firm performance: Moderators of functional-background and locus-of-control diversity. *Management Science*, 55(2):165–180.

Bossert, W., d'Ambrosio, C., and La Ferrara, E. (2011). A generalized index of fractionalization. *Economica*, 78(312):723–750.

Bove, V. and Elia, L. (2017). Migration, diversity, and economic growth. *World Development*, 89(1):227–239.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2012). Robust inference with multiway clustering. *Journal of Business and Economic Statistics*, 29(2):238–249.

Cameron, A. C. and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372.

Cohen-Zada, D., Krumer, A., and Shapir, O. M. (2018). Testing the effect of serve order in tennis tiebreak. *Journal of Economic Behavior & Organization*, 146:106–115.

Cohen-Zada, D., Krumer, A., and Shtudiner, Z. (2017). Psychological momentum and gender. *Journal of Economic Behavior & Organization*, 135:66–81.

Cox, T. H., Lobel, S. A., and McLeod, P. L. (1991). Effects of ethnic group cultural differences on cooperative and competitive behavior on a group task. *Academy of Management Journal*, 34(4):827–847.

Cummings, J. N. (2004). Work groups, structural diversity, and knowledge sharing in a global organization. *Management Science*, 50(3):352–364.

Dale-Olsen, H. and Finseraas, H. (2020). Linguistic diversity and workplace productivity. *Labour Economics*, 64:101813.

Delis, M. D., Gaganis, C., Hasan, I., and Pasiouras, F. (2017). The effect of board directors from countries with different genetic diversity levels on corporate performance. *Management Science*, 63(1):231–249.

Dellal, A., Chamari, K., Wong, d. P., Ahmaidi, S., Keller, D., Barros, R., Bisciotti, G. N., and Carling, C. (2011). Comparison of physical and technical performance in European soccer match-play: FA Premier League and La Liga. *European Journal of Sport Science*, 11(1):51–59.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *Quarterly Journal of Economics*, 132(4):1593–1640.

Dezsö, C. L. and Ross, D. G. (2012). Does female representation in top management improve firm performance? A panel data investigation. *Strategic Management Journal*, 33(9):1072–1089.

Dobson, S. and Goddard, J. (2011). *The Economics of Football*. Cambridge University Press.

Docquier, F., Turati, R., Valette, J., and Vasilakis, C. (2020). Birthplace diversity and economic growth: evidence from the us states in the post-world war ii period. *Journal of Economic Geography*, 20(2):321–354.

Doerrenberg, P. and Siegloch, S. (2014). Is soccer good for you? The motivational impact of big sporting events on the unemployed. *Economics Letters*, 123(1):66–69.

Earley, C. P. and Mosakowski, E. (2000). Creating hybrid team cultures: An empirical test of transnational team functioning. *Academy of Management Journal*, 43(1):26–49.

Ely, R. J. (2004). A field study of group diversity, participation in diversity education programs, and performance. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 25(6):755–780.

Freeman, R. B. and Huang, W. (2015). Collaborating with people like me: Ethnic coauthorship within the United States. *Journal of Labor Economics*, 33(S1):S289–S318.

Gaertner, S. L., Dovidio, J. F., Anastasio, P. A., Bachman, B. A., and Rust, M. C. (1993). The common ingroup identity model: Recategorization and the reduction of intergroup bias. *European Review of Social Psychology*, 4(1):1–26.

Gaertner, S. L., Dovidio, J. F., Samuel, G., et al. (2000). *Reducing intergroup bias: The common ingroup identity model*. Psychology Press.

Garicano, L. and Palacios-Huerta, I. (2014). Making the beautiful game a bit less beautiful. In *Beautiful game theory: How soccer can help economics*. Princeton University Press.

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

Gould, E. D. and Winter, E. (2009). Interactions between workers and the technology of production: Evidence from professional baseball. *Review of Economics and Statistics*, 91(1):188–200.

Guillaume, Y. R., Dawson, J. F., Otaye-Ebede, L., Woods, S. A., and West, M. A. (2017). Harnessing demographic differences in organizations: What moderates the effects of workplace diversity? *Journal of Organizational Behavior*, 38(2):276–303.

Guryan, J., Kroft, K., and Notowidigdo, M. J. (2009). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics*, 1(4):34–68.

Haas, H. and Nüesch, S. (2012). Are multinational teams more successful? *International Journal of Human Resource Management*, 23(15):3105–3113.

Harb-Wu, K. and Krumer, A. (2019). Choking under pressure in front of a supportive audience: Evidence from professional biathlon. *Journal of Economic Behavior & Organization*, 166:246–262.

Herring, C. (2009). Does diversity pay?: Race, gender, and the business case for diversity. *American Sociological Review*, 74(2):208–224.

Herrmann, M. A. and Rockoff, J. E. (2012). Worker absence and productivity: Evidence from teaching. *Journal of Labor Economics*, 30(4):749–782.

Hjort, J. (2014). Ethnic divisions and production in firms. *Quarterly Journal of Economics*, 129(4):1899–1946.

Hoogendoorn, S., Oosterbeek, H., and Van Praag, M. (2013). The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management Science*, 59(7):1514–1528.

Horwitz, S. K. and Horwitz, I. B. (2007). The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of Management*, 33(6):987–1015.

Ingersoll, K., Malesky, E., and Saiegh, S. M. (2017). Heterogeneity and team performance: Evaluating the effect of cultural diversity in the world's top soccer league. *Journal of Sports Analytics*, 3(2):67–92.

Jehn, K. A., Northcraft, G. B., and Neale, M. A. (1999). Why differences make a difference: A field study of diversity, conflict and performance in workgroups. *Administrative Science Quarterly*, 44(4):741–763.

Jones, B. (2009). The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *Review of Economic Studies*, 76(1):283–317.

Joshi, A. and Roh, H. (2009). The role of context in work team diversity research: A meta-analytic review. *Academy of management journal*, 52(3):599–627.

Kahane, L., Longley, N., and Simmons, R. (2013). The effects of coworker heterogeneity on firm-level output: Aassessing the impacts of cultural and language diversity in the National Hockey League. *Review of Economics and Statistics*, 95(1):302–314.

Kellerman, E., Koonen, H., and van der Haagen, M. (2005). Feet speak louder than the tongue": A preliminary analysis of language provisions for foreign professional footballers in the Netherlands. In *Second Language Needs Analysis*, pages 200–22. Cambridge University Press.

Kesavan, S., Staats, B. R., and Gilland, W. (2014). Volume flexibility in services: The costs and benefits of flexible labor resources. *Management Science*, 60(8):1884–1906.

Kleven, H. J., Landais, C., and Saez, E. (2013). Taxation and international migration of superstars:

Evidence from the European football market. *American Economic Review*, 103(5):1892–1924.

Krumer, A. and Lechner, M. (2017). First in first win: Evidence on schedule effects in round-robin tournaments in mega-events. *European Economic Review*, 100:412–427.

Krumer, A. and Lechner, M. (2018). Midweek effect on soccer performance: Evidence from the German Bundesliga. *Economic Inquiry*, 56(1):193–207.

Kunze, F., Boehm, S. A., and Bruch, H. (2011). Age diversity, age discrimination climate and performance consequences—a cross organizational study. *Journal of Organizational Behavior*, 32(2):264–290.

Lang, K. (1986). A language theory of discrimination. *Quarterly Journal of Economics*, 101(2):363–382.

Lavric, E., Pisek, G., Skinner, A., and Stadler, W. (2008). *The linguistics of football.* Narr Francke Attempto Verlag.

Lazear, E. P. (1999). Globalisation and the market for team-mates. *Economic Journal*, 109(454):15–40.

Lichter, A., Pestel, N., and Sommer, E. (2017). Productivity effects of air pollution: Evidence from professional soccer. *Labour Economics*, 48(1):54–66.

Lind, J. T. and Mehlum, H. (2010). With or without U? The appropriate test for a U-shaped relationship. *Oxford Bulletin of Economics and Statistics*, 72(1):109–118.

Lowe, M. (2020). Types of contact: A field experiment on collaborative and adversarial caste integration.

Lyons, E. (2017). Team production in international labor markets: Experimental evidence from the field. *American Economic Journal: Applied Economics*, 9(3):70–104.

Maderer, D., Holtbrügge, D., and Schuster, T. (2014). Professional football squads as multicultural teams: Cultural diversity, intercultural experience, and team performance. *International Journal of Cross Cultural Management*, 14(2):215–238.

Mitrotasios, M., Gonzalez-Rodenas, J., Armatas, V., and Aranda, R. (2019). The creation of goal scoring opportunities in professional soccer. tactical differences between spanish la liga, english premier league, german bundesliga and italian serie a. *International Journal of Performance Analysis in Sport*, 19(3):452–465.

Ng, T. W. and Feldman, D. C. (2008). The relationship of age to ten dimensions of job performance. *Journal of Applied Psychology*, 93(2):392.

O'Reilly, C. A., Caldwell, D. F., and Barnett, W. P. (1989). Work group demography, social integration, and turnover. *Administrative Science Quarterly*, pages 21–37.

Ottaviano, G. I. and Peri, G. (2005). Cities and cultures. *Journal of Urban Economics*, 58(2):304–337.

Owen, A. L. and Temesvary, J. (2018). The performance effects of gender diversity on bank boards. *Journal of Banking & Finance*, 90:50–63.

Parrotta, P., Pozzoli, D., and Pytlikova, M. (2014). Labor diversity and firm productivity. *European Economic Review*, 66(1):144–179.

Parsons, C. A., Sulaeman, J., Yates, M. C., and Hamermesh, D. S. (2011). Strike three: Discrimination, incentives, and evaluation. *American Economic Review*, 101(4):1410–35.

Prat, A. (2002). Should a team be homogeneous? *European Economic Review*, 46(7):1187–1207.

Price, J., Lefgren, L., and Tappen, H. (2013). Interracial workplace cooperation: Evidence from the NBA. *Economic Inquiry*, 51(1):1026–1034.

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

Prinz, J. and Wicker, P. (2016). Diversity effects on team performance in the tour de france. *Team*

Performance Management.

Ringbom, H. (2012). Multilingualism in a football team: the case of ifk mariehamn. In *Cross-linguistic Influences in Multilingual Language Acquisition*, pages 185–197. Springer.

Sarmento, H., Anguera, M. T., Pereira, A., Marques, A., Campaniço, J., and Leitão, J. (2014). Patterns of play in the counterattack of elite football teams-a mixed method approach. *International Journal of Performance Analysis in Sport*, 14(2):411–427.

Sarmento, H., Pereira, A., Matos, N., Campaniço, J., Anguera, T. M., and Leitão, J. (2013). English premier league, spaińs la liga and italẙs seriés a–what's different? *International Journal of Performance Analysis in Sport*, 13(3):773–789.

Schwab, A., Werbel, J. D., Hofmann, H., and Henriques, P. L. (2016). Managerial gender diversity and firm performance: An integration of different theoretical perspectives. *Group & Organization Management*, 41(1):5–31.

Siebert, W. S. and Zubanov, N. (2010). Management economics in a large retail company. *Management Science*, 56(8):1398–1414.

Simmons, R. (1997). Implications of the Bosman ruling for football transfer markets. *Economic Affairs*, 17(3):13–18.

Simonsohn, U. (2018). Two lines: A valid alternative to the invalid testing of u-shaped relationships with quadratic regressions. *Advances in Methods and Practices in Psychological Science*, 1(4):538–555.

Spolaore, E. and Wacziarg, R. (2016). Ancestry, language and culture. In *Palgrave Handbook of Economics and Language*, pages 174–211. Springer.

Staats, B. R., Milkman, K. L., and Fox, C. R. (2012). The team scaling fallacy: Underestimating the declining efficiency of larger teams. *Organizational Behavior and Human Decision Processes*, 118(2):132–142.

Swaab, R. I., Schaerer, M., Anicich, E. M., Ronay, R., and Galinsky, A. D. (2014). The too-much-talent effect: Team interdependence determines when more talent is too much or not enough. *Psychological Science*, 25(8):1581–1591.

Tan, T. F. and Netessine, S. (2014). When does the devil make work? An empirical study of the impact of workload on worker productivity. *Management Science*, 60(6):1574–1593.

Tan, T. F. and Netessine, S. (2019). When you work with a superman, will you also fly? An empirical study of the impact of coworkers on performance. *Management Science*, 65(8):3495–3517.

Trax, M., Brunow, S., and Suedekum, J. (2015). Cultural diversity and plant-level productivity. *Regional Science and Urban Economics*, 53(1):85–96.

van Knippenberg, D., De Dreu, C. K., and Homan, A. C. (2004). Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology*, 89(6):1008.

van Knippenberg, D. and Mell, J. N. (2016). Past, present, and potential future of team diversity research: From compositional diversity to emergent diversity. *Organizational Behavior and Human Decision Processes*, 136:135–145.

van Knippenberg, D. and Schippers, M. C. (2007). Work group diversity. *Annual Review of Psychology*, 58:515–541.

Williams, K. Y. and O'Reilly, C. A. I. (1998). Demography and diversity in organizations: A review of 40 years of research. *Research in Organizational Behavior*, 20:77–140.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.

Wuchty, S., Jones, B. F., and Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827):1036–1039.

For online publication

A Theoretical framework

In this supplementary section, we develop a simple theoretical framework to illustrates how birthplace diversity may affect team performance. The basic structure of our model is inspired by the work of Ashraf and Galor (2013) and Lazear (1999).

We consider a team that consists of n workers originating from m countries. Each worker i has two traits: an origin-specific trait ($q_i \in [0,1]$) and a personality trait ($p_i \in [0,1]$) that does not depend on the country of origin. $\mathbf{A} = \{a_{ij}\}_{i,j=1,\ldots,n}$ is a similarity matrix in which $a_{ij} = 1 - |q_i - q_j| \in [0,1]$ reflects the extent to which the origin-specific traits of workers i and j resemble each other. Following Bossert et al. (2011), we define that the birthplace diversity of the team ($\delta \in [0,1]$) only depends on the similarity matrix $\mathbf{A}$ and assume that:

$$\delta = \delta(\mathbf{A}) = 0 \quad \Leftrightarrow \quad a_{ij} = 1 \quad \text{for all } i,j = 1,\ldots,n \tag{A1}$$

$$\delta = \delta(\mathbf{A}) = 1 \quad \Leftrightarrow \quad a_{ij} = 0 \quad \text{for all } i,j = 1,\ldots,n \text{ and } i \neq j \tag{A2}$$

$$\delta_{a_{i,j}} = \frac{\partial\,\delta(\mathbf{A})}{\partial\, a_{i,j}} < 0 \quad \text{for all } i,j = 1,\ldots,n \text{ with } a_{ij} \in (0,1). \tag{A3}$$

Our model presumes that the performance of teams depends on two factors. The first factor is the available knowledge:

$$H = H(\delta,\, \rho,\, \Sigma_Q,\, \Sigma_P) \geq 1 \quad \text{with} \quad \Sigma_Q = \sum_{i=1}^{n} q_i \quad \text{and} \quad \Sigma_P = \sum_{i=1}^{n} p_i \tag{A4}$$

where Σ_P and Σ_Q capture the abilities of the team members and $\rho \in [0,1]$ is the diversity of the personality traits. We assume that birthplace diversity increases the available knowledge because workers born in different countries grow up under different cultural and educational systems and are thus likely to have distinct productive skills (Alesina and La Ferrara, 2005, Alesina et al., 2016, Ashraf and Galor, 2013):[34]

$$H_\delta = \frac{\partial H}{\partial \delta} > 0 \quad \text{with} \quad \lim_{\delta \to 0} H_\delta = \infty \quad \text{and} \quad \lim_{\delta \to 1} H_\delta = 0 \quad \text{and} \quad H_{\delta\delta} \leq 0. \tag{A5}$$

The other factor of team performance is the efficiency with which the team members collaborate with each other:

$$E = E(\delta,\, \rho) \in (0,1] \quad \text{with} \quad E(0,0) = 1. \tag{A6}$$

[34] In line with several related studies, we implicitly assume that different skills complement each other.

The efficiency of the collaboration decreases in birthplace diversity since workers from different countries are likely to face cultural and linguistic barriers (Freeman and Huang, 2015, Lazear, 1999). Furthermore, inter-group biases and conflicts are more likely to exist in diverse teams (Earley and Mosakowski, 2000).

$$E_\delta = \frac{\partial E}{\partial \delta} < 0 \quad \text{with} \quad \lim_{\delta \to 0} E_\delta > -\infty \quad \text{and} \quad E_{\delta\delta} \leq 0. \tag{A7}$$

Assuming a Cobb-Douglas production function, the output of a team is then given by:

$$Y = E\big(\delta, \rho\big)^{\alpha} \cdot H\big(\delta, \rho, \Sigma_Q, \Sigma_P\big)^{1-\alpha} \quad \text{with} \quad \alpha \in (0,1), \tag{A8}$$

and the first-order condition with respect to δ can be written as:

$$0 = \alpha \cdot E_\delta \cdot \left(\frac{H\big(\delta, \rho, \Sigma_Q, \Sigma_P\big)}{E\big(\delta, \rho\big)} \right)^{1-\alpha} + (1 - \alpha) \cdot H_\delta \cdot \left(\frac{E\big(\delta, \rho\big)}{H\big(\delta, \rho, \Sigma_Q, \Sigma_P\big)} \right)^{\alpha} \tag{A9}$$

From (A9), we obtain the following results:

Proposition 1. *Conditional on all other determinants of team performance, it holds that:*

(a) Team output Y is a concave function of birthplace diversity δ.

(b) Team output Y is maximized at an intermediate level of birthplace diversity $\delta^* \in (0,1)$.

(c) The more important efficient collaboration is for team output, the lower is the optimal level of birthplace diversity:

$$\frac{\partial \delta^*}{\partial \alpha} < 0 \quad \text{with} \quad \lim_{\alpha \to 0} \delta^* = 1 \quad \text{and} \quad \lim_{\alpha \to 1} \delta^* = 0.$$

B The effect of performance-based substitutions on birthplace diversity

In this supplementary section, we illustrate why the conventional fixed effect approach produces downward-biased estimates of the optimal level of birthplace diversity. More specifically, we present regression results, showing that the level of birthplace diversity increases during a match because of performance-based substitutions.

Our analysis consists of three parts. The first part shows how team managers react during a match if their team does not perform well. To this end, we divide each match into eighteen non-overlapping five-minute periods and estimate the regression model:

$$\mathrm{O}^{e}_{isrmdp} = \alpha \cdot \mathrm{O}^{b}_{isrmdp} + \beta \cdot \mathrm{P}^{b}_{isrmdp} + \gamma \cdot \mathbf{X}^{b}_{isrmdp} + \xi_{isrmd} + \theta_p + \varepsilon_{isrmdp} \tag{A10}$$

where i is a club, s a season, m a manager, d a match day, r a round, and p a five-minute period. O^b/O^e is the number of offensive players (forwards) at the beginning/end of a period. P is a dummy variable that is equal to 1 if the team is behind at the beginning of a period. The vector $\mathbf{X}$ includes team and opponent controls. ξ is the club-by-match fixed effect, θ the period fixed effect, and ε the error term.

Column 1 of Table E.12 shows results from estimating (A10). Consistent with the results of Garicano and Palacios-Huerta (2014), we find that team managers increase the number of fielded forwards if their team is behind.

In the second part of this supplementary analysis, we study whether the birthplace diversity of the fielded players depends on the score of the match. We address this question with the regression model:

$$\mathrm{B}^e_{isrmdp} = \alpha \cdot \mathrm{B}^b_{isrmdp} + \beta \cdot \mathrm{P}^b_{isrmdp} + \gamma \cdot \mathbf{X}^b_{isrmdp} + \xi_{isrmd} + \theta_p + \varepsilon_{isrmdp} \qquad \text{(A11)}$$

where B^b/B^e is the level of birthplace diversity at the beginning/end of a five-minute period. All other components of (A11) have the same meaning as in (A10).

In Column 2 of Table E.12, we report results from estimating (A10). We find that the level of birthplace diversity increases if a team is behind. This result implies that reverse causality is non-negligible problem for our main analysis.

The last part of our analysis provides an explanation for why birthplace diversity increase if teams do not perform well. There are two reasons. First, team managers dislike losses and thus replace defensive with offensive players if their team is behind (see Column 1 of Table E.12). Second, the share of foreigners in German soccer clubs is much larger among offensive players than among defensive players.[35] Put differently, we argue that the increase in the number of forwards is the channel through which team performance affects birthplace diversity. The results shown in Column 3 of Table E.12 support this argument because we find that the estimate of the effect of team performance on birthplace diversity becomes statistically insignificant if we augment regression model (A11) by the number of forwards at the end of a five-minute period.

C Using injuries as a source of variation

To use unexpected replacements that are caused by short-term injuries as our source of plausibly exogenous variation, we have to make some small adjustments in the three-step procedure described in Section 4.2.2. The change concerns the first step in which we divide the players into different groups. Group $\mathcal{A}$ now includes those players that belonged to the starting line-up predicted by *Kicker*,

[35]Our raw data suggests that the share of foreign defenders and midfielders is around 45 percent, whereas the share of foreign forwards is around 65 percent.

but did not participate in the match because of an injury. Group $\mathcal{B}$ includes starting players that were not part of the expected starting line-up and replaced an injured player. Group $\mathcal{B}$ includes all other starting players (for an example, see Figure D.2).

The other steps of the procedure are the same as in Section 4.2.2. In the second step, we calculate the average dissimilarity between the players in $\mathcal{C}$ and $\mathcal{A}$ ($\mathcal{B}$):

$$\Delta(\mathcal{A}, \mathcal{C}) = \frac{1}{|\mathcal{A}|} \cdot \frac{1}{|\mathcal{C}|} \cdot \sum_{j \in \mathcal{A}} \sum_{k \in \mathcal{C}} (1 - s_{jk}) \tag{A12}$$

$$\Delta(\mathcal{B}, \mathcal{C}) = \frac{1}{|\mathcal{B}|} \cdot \frac{1}{|\mathcal{C}|} \cdot \sum_{j \in \mathcal{B}} \sum_{k \in \mathcal{C}} (1 - s_{jk}) \tag{A13}$$

where s_{jk} is a dummy that is equal to 1 if the players j and k come from the same country, and 0 otherwise. In the last step, we create the instrumental variable $\mathrm{Z} \in [-1, 1]$ as the difference of two dissimilarity scores:

$$\mathrm{Z} = \Delta(\mathcal{B}, \mathcal{C}) - \Delta(\mathcal{A}, \mathcal{C}). \tag{A14}$$

As expected, we observe a positive correlation between our instrumental variable and the birthplace diversity of the fielded players (see Figure D.4).

D Additional figures

Figure D.1 Classification of players (example)

Predicted Line-up	Starting Line-up
Ulreich	Ulreich
Kimmich Süle Hummels Alaba	Kimmich Süle Hummels Rafinha
Martinez Thiago James	Tolisso Thiago James
Müller Lewandowski Ribery	Müller Lewandowski Ribery

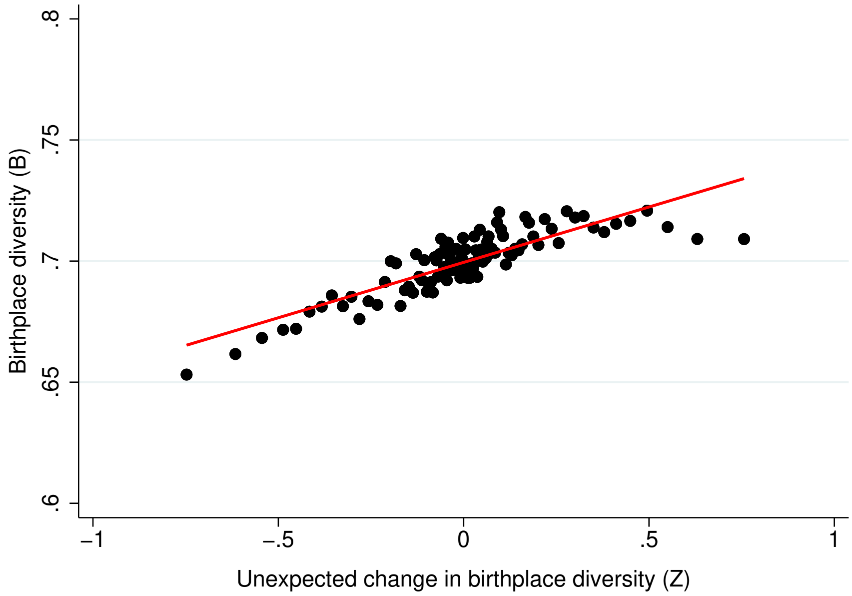
Notes: The left panel shows the expected starting line-up of *Bayern München* for the match on Saturday, 12th May 2018. *Kicker* published this line-up on Friday, 11th May 2018. The right panel shows the actual starting line-up. We use *blue* characters to indicate players (Alaba, Martinez) that belong to group $\mathcal{A}$, *red* characters to indicate players (Rafinha, Tolisso) that belong to group $\mathcal{B}$, and *black* characters to indicate players that belong to group $\mathcal{C}$.

Figure D.2 Classification of players (example)

Predicted Line-up	Starting Line-up
Ulreich	Ulreich
Kimmich Süle Hummels Alaba	Kimmich Süle Hummels Rafinha
Martinez Thiago James	Tolisso Thiago James
Müller Lewandowski Ribery	Müller Lewandowski Ribery

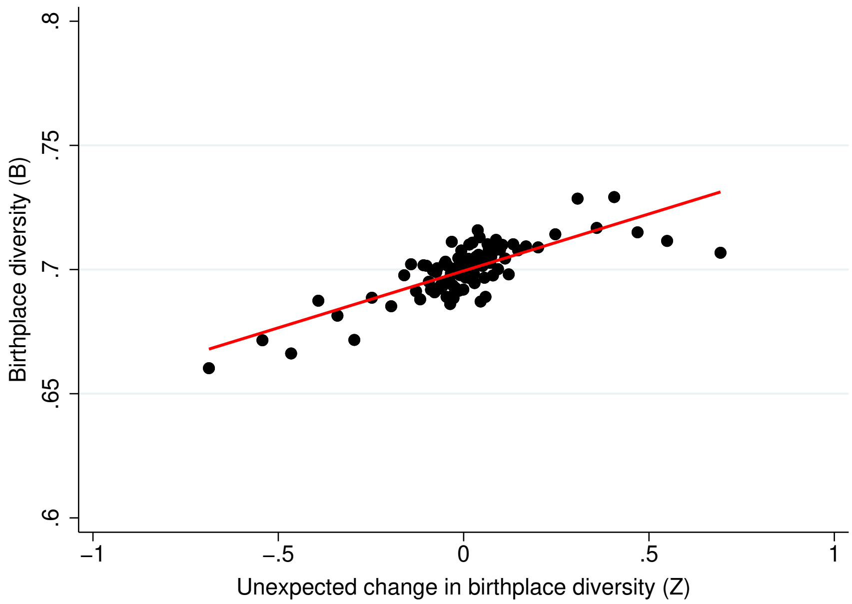
Notes: The left panel shows the expected starting line-up of *Bayern München* for the match on Saturday, 12th May 2018. *Kicker* published this line-up on Friday, 11th May 2018. The right panel shows the actual starting line-up. We use *blue* characters to indicate players (Alaba) that belong to group $\widetilde{\mathcal{A}}$, *red* characters to indicate players (Rafinha) that belong to group $\widetilde{\mathcal{B}}$, and *black* characters to indicate players that belong to group $\widetilde{\mathcal{C}}$. Players (Martinez) colored in *violet* belong to neither of the three groups. David Alaba is a member of group $\widetilde{\mathcal{A}}$ because he unexpectedly missed the match because of back problems.

Figure D.3 Zero-stage relationship (all unexpected changes).



Notes: This figure depicts the positive correlation between the birthplace diversity of the fielded player (B) and our instrumental variable Z as defined in Equation (9). For the sake of vividness, we use a binned scatterplot to present the underlying raw data.

Figure D.4 Zero-stage relationship (injuries).



Notes: This figure depicts the positive correlation between the birthplace diversity of the fielded player (B) and our instrumental variable Z as defined in Equation (A14). For the sake of vividness, we use a binned scatterplot to present the underlying raw data.

E Additional tables

Table E.1 List of countries.

Country	Country	Country
Albania (13)	Gambia (3)	Norway (36)
Algeria (1)	Georgia (9)	Paraguay (5)
Argentina (36)	Germany (1534)	Peru (6)
Armenia (1)	Ghana (26)	Poland (89)
Australia (16)	Greece (19)	Portugal (16)
Austria (71)	Guinea (3)	Romania (28)
Azerbaijan (1)	Guinea-Bissau (2)	Russia (16)
Belarus (3)	Hungary (27)	Senegal (11)
Belgium (38)	Iceland (6)	Serbia (47)
Benin (2)	Iran (10)	Sierra Leone (3)
Bolivia (1)	Israel (5)	Slovakia (23)
Bosnia and Herzegovina (35)	Italy (13)	Slovenia (14)
Brazil (141)	Ivory Coast (12)	South Africa (11)
Bulgaria (24)	Jamaica (2)	Spain (38)
Burkina Faso (3)	Japan (28)	Suriname (4)
Cameroon (25)	Kazakhstan (5)	Sweden (39)
Canada (9)	Korea (11)	Switzerland (79)
Cape Verde (1)	Kosovo (17)	Syria (1)
Chile (8)	Latvia (2)	Tajikistan (3)
China (3)	Lebanon (2)	Togo (4)
Colombia (8)	Liechtenstein (1)	Trinidad and Tobago (1)
Congo Demo. Rep. (12)	Lithuania (1)	Tunisia (10)
Cong Rep. (4)	Luxembourg (2)	Turkey (17)
Costa Rica (1)	Macedonia (16)	Ukraine (20)
Croatia (71)	Mali (2)	United Kingdom (15)
Cyprus (1)	Malta (1)	United States (32)
Czech Republic (70)	Mexico (8)	Uruguay (7)
Denmark (75)	Moldova (2)	Venezuela (5)
Ecuador (3)	Montenegro (9)	Zambia (3)
Egypt (7)	Morocco (8)	Zimbabwe (1)
Equatorial Guinea (3)	Mozambique (1)	
Estonia (1)	Namibia (2)	
Finland (12)	Netherlands (72)	
France (92)	Nigeria (26)	

Notes: This table lists the countries from which the players in our data set originate. In parentheses, we report the number of players born in a particular country.

Table E.2 List of control variables.

Variable	**Description**
Home	Dummy variable that is equal to 1 if a club had the home field advantage.
Cup$^{(a)}$	Dummy variable that is equal to 1 if a team played in the national cup in the week before (or the week after) a *Bundesliga* match.
ECT$^{(a)}$	Dummy variable that is equal to 1 if a team played in an European club tournament in the week before (or the week after) a *Bundesliga* match.
Age$^{(a,b)}$	Average age (in years).
Tenure$^{(a,b)}$	Average duration of club membership (in years).
Bundesliga matches$^{(a)}$	Average number of *Bundesliga* matches.
European matches$^{(a)}$	Average number of matches in the European club tournaments.
Championship matches$^{(a)}$	Average number of matches in European and World Championship matches.
Top-League matches$^{(a)}$	Average number of matches in highest soccer divisions in England, France, Italy, and Spain.
Market value$^{(a,c)}$	Average of logged market values.
FIFA rating$^{(a,c)}$	Average playing strength in video game *FIFA*.
Age (manager)$^{(a)}$	Age of the manager.
Tenure (manager)$^{(a)}$	Number of years in which the manager was in charge of the team.
Bundesliga matches (manager)$^{(a)}$	Total number of Bundesliga matches that a manager was in charge of a *Bundesliga* team.
Opponent-Season fixed effects	Set of season-specific dummy variables that indicate the opponent of the team.
Ranking fixed effects	Set of dummy variables that capture the current ranking positions of the opposing teams.

Notes: This table lists all variables that are included in the vector $\mathbf{X}$. (a) indicates that $\mathbf{X}$ includes this variable for both the team and its opponent. (b) indicates that $\mathbf{X}$ includes the squared term of the variable. (c) indicates that $\mathbf{X}$ does not include this variable if we use the full sample.

Table E.3 Summary statistics (full sample).

Variable	Mean	Std. Dev.	Min	Max
Goal difference	0	1.872	-8	8
Ranking points	1.372	1.313	0	3
Main explanatory variable				
Birthplace diversity	0.669	0.153	0	0.918
Control variables				
Home	0.500	0.500	0	1
Cup	0.101	0.301	0	1
ECT	0.168	0.374	0	1
Age	26.824	1.332	22.321	31.029
Tenure	2.247	0.779	0.321	6.25
Bundesliga matches	86.826	34.924	0.5	223.857
European matches	12.710	10.117	0	72.429
Championship matches	1.228	1.629	0	15.286
Top-league matches	8.492	11.073	0	90.923
Age (manager)	47.330	6.684	28.6	73.7
Tenure (manager)	2.007	2.348	0	14
Bundesliga matches (manager)	131.599	140.297	0	829

Notes: This table presents summary statistics for our reduced sample. This sample consists of 14,076 observations. In Table E.15, we indicate how strongly correlated the team characteristics are.

Table E.4 Summary statistics (reduced sample).

Variable	Mean	Std. Dev.	Min	Max
Dependent variables				
Goal difference	0	1.872	-8	8
Ranking points	1.373	1.315	0	3
Main explanatory variable				
Birthplace diversity	0.699	0.1354	0	0.918
Control variables				
Home	0.5	0.5	0	1
Cup	0.101	0.301	0	1
ECT	0.186	0.389	0	1
Age	26.261	1.237	22.321	30.6
Tenure	2.117	0.705	0.321	5.357
Bundesliga matches	81.529	30.862	6.071	193
European matches	14.010	10.873	0	72.429
Championship matches	1.276	1.840	0	15.286
Top-league matches	9.397	11.849	0	80.923
Market value (log)	14.869	0.721	13.007	17.288
FIFA rating	73.956	3.501	63.071	86.714
Age (manager)	47.660	6.964	28.6	73.7
Tenure (manager)	1.719	2.099	0	14
Bundesliga matches (manager)	129.759	140.863	0	829

Notes: This table presents summary statistics for our reduced sample. This sample consists of 7,956 observations. In Table E.16, we indicate how strongly correlated the team characteristics are.

Table E.5 First-stage estimates (non-linear model).

	(1)	(2)
	First inst. var. approach	
Birthplace divers. in Kicker line-up	0.364***	0.035
	(0.000)	(0.517)
Birthplace divers. in Kicker line-up sq.	-0.032	0.308***
	(0.528)	(0.000)
	Second inst. var. approach	
Predicted level of birthplace divers.	0.996***	1.020***
	(0.000)	(0.000)
Predicted level of birthplace divers. sq.	0.008	0.137
	(0.930)	(0.207)
Outcome variable	B	B^2
Observations	7,956	7,956
Seasons	13	13
Fixed effects	Yes	Yes
Basic controls	Yes	Yes
Quality controls	Yes	Yes

Notes: This table reports first-stage estimates. For the related second-stage estimates, see Columns 3 and 4 of Table 1. Standard errors are clustered at the club-by-season-by-round-by-manager level and p-values are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table E.6 Reduced-form estimates (non-linear model).

	(1)	(2)
	Lind-Mehlum-Test	
Optimal level of birthplace diversity	0.588*	0.723**
	(0.068)	(0.052)
	Regression coefficients	
Birthplace divers. in Kicker line-up	1.803	
	(0.103)	
Birthplace divers. in Kicker line-up sq.	-1.533*	
	(0.092)	
Predicted level of birthplace divers.		6.375**
		(0.046)
Predicted level of birthplace divers. sq.		-4.408**
		(0.026)
Outcome variable	Goal Diff.	Goal Diff.
Observations	7,956	7,956
Seasons	13	13
Fixed effects	Yes	Yes
Basic controls	Yes	Yes
Quality controls	Yes	Yes
Sample mean (B)	0.699	0.699

Notes: This table reports reduced-form estimates. For the related second-stage estimates, see Columns 3 and 4 of Table 1. Standard errors are clustered at the club-by-season-by-round-by-manager level and p-values are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table E.7 Birthplace diversity and team performance (non-linear model, alternative outcome).

	(1)	(2)	(3)	(4)
	Lind-Mehlum-Test			
Optimal level of birthplace diversity ($\hat{B}^*$)	0.236	0.269	0.578*	0.586*
	(0.222)	(0.252)	(0.062)	(0.072)
	Regression coefficients			
Birthplace divers. ($\hat{\beta}_1$)	0.558	0.725	3.799	19.162
	(0.444)	(0.504)	(0.124)	(0.143)
Birthplace divers. sq. ($\hat{\beta}_2$)	-1.183**	-1.347	-3.284*	-16.351
	(0.049)	(0.118)	(0.100)	(0.132)
Estimation technique	OLS	OLS	IV	IV
IV approach	-	-	1st	2nd
Outcome variable	Points	Points	Points	Points
Observations	14,076	7,956	7,956	7,956
Seasons	23	13	13	13
Fixed effects	Yes	Yes	Yes	Yes
Basic controls	Yes	Yes	Yes	Yes
Quality controls	No	Yes	Yes	Yes
First-stage F-statistic (B)	-	-	126.3	13.7
First-stage F-statistic (B^2)	-	-	286.5	14.3
Sample mean (B):	0.669	0.699	0.699	0.699

Notes: The table reports results from fixed effect and instrumental variable regressions. The upper part of the table shows the results of the Lind-Mehlum-Test, the lower part presents regression coefficients. We cluster standard errors at the club-by-season-by-round-by-manager level and report p-values in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table E.8 Birthplace diversity and team performance (linear model, alternative outcome).

	(1)	(2)	(3)	(4)
Birthplace divers. ($\hat{\beta}_1$)	-0.815***	-0.913***	-0.164	-0.575
	(0.000)	(0.001)	(0.798)	(0.603)
Estimation technique	OLS	OLS	IV	IV
IV approach	-	-	1st	2nd
Outcome variable	Points	Points	Points	Points
Observations	14,076	7,956	7,956	7,956
Seasons	23	13	13	13
Fixed effects	Yes	Yes	Yes	Yes
Basic controls	Yes	Yes	Yes	Yes
Quality controls	No	Yes	Yes	Yes
First-stage F-statistic	-	-	640.5	272.80
Sample mean (B):	0.669	0.699	0.699	0.699

Notes: This table presents results from linear fixed effect and instrumental variable regressions. We cluster standard errors at the club-by-season-by-round-by-manager level and report p-values in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table E.9 Alternative clustering (instrumental variable estimates, non-linear model)

	(1)	(2)	(3)	(4)	(5)	(6)
	Lind-Mehlum-Test					
Optimal level of birthplace diversity ($\hat{B}^*$)	0.608*	0.604**	0.608*	0.604**	0.608*	0.604*
	(0.071)	(0.032)	(0.083)	(0.042)	(0.081)	(0.052)
	Regression coefficients					
Birthplace divers. ($\hat{\beta}_1$)	5.373*	41.924*	5.373*	41.924*	5.373**	41.924
	(0.068)	(0.065)	(0.069)	(0.083)	(0.036)	(0.105)
Birthplace divers. sq. ($\hat{\beta}_2$)	-4.420*	-34.670*	-4.420*	-34.670*	-4.420*	-34.670*
	(0.070)	(0.062)	(0.080)	(0.077)	(0.056)	(0.098)
Estimation technique	IV	IV	IV	IV	IV	IV
IV approach	1st	2nd	1st	2nd	1st	2nd
Outcome variable	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.
Observations	7,956	7,956	7,956	7,956	7,956	7,956
Seasons	13	13	13	13	13	13
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Basic controls	Yes	Yes	Yes	Yes	Yes	Yes
Quality controls	Yes	Yes	Yes	Yes	Yes	Yes
First-stage F-statistic (B)	136.9	9.91	127.8	13.6	137.2	9.9
First-stage F-statistic (B^2)	307.5	10.61	222.5	14.4	241.3	10.7
Sample mean (B):	0.669	0.699	0.699	0.699	0.699	0.699

Notes: The table reports results from instrumental variable regressions. The upper part of the table shows the results of the Lind-Mehlum-Test, the lower part presents regression coefficients. In Columns 1 and 2, we cluster standard errors at the club-by-season level. In Columns 3 and 4, we apply a two-way clustering procedure (club-by-season-by-round-by-manager and match day level). In Columns 3 and 4, we cluster at the club-by-season and the match day level. We report p-values in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table E.10 Piecewise and interrupted linear regression (instrumental variable estimates).

	(1)	(2)	(3)	(4)	(5)
	Piecewise linear regression				
Birthplace diversity	3.735**	-3.073*	-17.755*	18.519***	
	(0.042)	(0.072)	(0.093)	(0.003)	
	Interrupted regression (Simonsohn, 2018)				
Birthplace diversity (low)					3.328*
					(0.069)
Birthplace diversity (high)					-3.006*
					(0.060)
Estimation technique	IV	IV	IV	IV	IV
IV approach	1st	1st	2nd	2nd	1st
Observations	2,083	5,784	1,633	1,217	7,956
Seasons	13	13	13	13	13
Outcome	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.
Fixed effects	Yes	Yes	Yes	Yes	Yes
Basic controls	Yes	Yes	Yes	Yes	Yes
Quality controls	Yes	Yes	Yes	Yes	Yes
First-stage F-statistic	89.8	370.1	10.2	30.1	345.41
Diversity level ($BDiv$)	< 0.6	> 0.6	> 0.6	> 0.6	-
Change in Diversity (Z)	-	-	> 0.0	< 0.0	-

Notes: The table reports results from instrumental variable regressions. We cluster standard errors at the club-by-season-by-round-by-manager level and report p-values in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table E.11 Heterogeneity in optimal level of birthplace diversity.

	Performance in offense	**Performance in defense**
	(1)	(2)
	Lind-Mehlum-Test	
Optimal level of birthplace diversity ($\hat{B}^*$)	0.649	0.516*
	(0.171)	(0.097)
	Regression coefficients	
Birthplace divers. ($\hat{\beta}_1$)	2.778*	-3.037
	(0.062)	(0.194)
Birthplace divers. sq. ($\hat{\beta}_2$)	-2.139	2.943
	(0.114)	(0.130)
Estimation technique	IV	IV
IV approach	1st	1st
Observations	7,956	7,956
Seasons	13	13
Outcome variable	Goals scored	Goals allowed
Fixed effects	Yes	Yes
Basic controls	Yes	Yes
Quality controls	Yes	Yes
First-stage F-statistic (B)	148.98	185.85
First-stage F-statistic (B^2)	284.78	359.96
Team	Midfielder + Forwards	Defender + Midfielder

Notes: The table reports results from instrumental variable regressions. The upper part of the table shows the results of the Lind-Mehlum-Test, the lower part presents regression coefficients. We cluster standard errors at the club-by-season-by-round-by-manager level and report p-values in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table E.12 Changes in the team composition during a match

	(1)	(2)	(3)
Behind (begin of period)	0.068***	0.001***	-0.000
	(0.000)	(0.002)	(0.884)
Forwards (end of period)			0.016***
			(0.000)
Outcome variable	Forwards	Birthplace Div.	Birthplace Div.
Observations	143,208	143,208	143,208
Seasons	13	13	13
Fixed effects	Yes	Yes	Yes
Basic controls	Yes	Yes	Yes
Quality controls	Yes	Yes	Yes
R^2	0.94	0.98	

Notes: The table reports results from within-match fixed effect regressions. We cluster standard errors at the club-by-match level and report p-values in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table E.13 Robustness checks (instrumental variable estimates, non-linear model)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Lind-Mehlum-Test											
Optimal level of birthplace	0.631**	0.549**	0.566**	0.593**	0.576**	0.602**	0.657*	0.537**	0.566**	0.569**	0.593*	0.609**
diversity ($\hat{B}^*$)	(0.049)	(0.044)	(0.031)	(0.020)	(0.018)	(0.029)	(0.085)	(0.038)	(0.028)	(0.040)	(0.066)	(0.022)
	Regression coefficients											
Birthplace divers. ($\hat{\beta}_1$)	41.991*	4.434*	33.855*	45.081**	8.163**	40.010*	7.092**	24.838*	7.041**	52.442*	5.265	43.108**
	(0.098)	(0.089)	(0.061)	(0.040)	(0.035)	(0.058)	(0.013)	(0.076)	(0.039)	(0.073)	(0.116)	(0.039)
Birthplace divers. sq. ($\hat{\beta}_2$)	-33.289*	-4.038*	-29.875*	-37.979**	-7.082**	-33.244*	-5.397**	-23.112*	-6.216**	-46.104*	-4.437*	-35.419**
	(0.093)	(0.056)	(0.057)	(0.035)	(0.022)	(0.055)	(0.031)	(0.063)	(0.034)	(0.075)	(0.098)	(0.039)
Estimation technique	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV
IV Approach	2nd	1st	2nd	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
Outcome variable	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.
Observations	7,956	7,956	7,956	7,956	7,488	7,488	7,956	7,956	7,956	7,956	7,956	7,956
Seasons	13	13	13	13	13	13	13	13	13	13	13	13
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Basic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quality controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First-stage F-statistic (B)	9.8	165.8	9.4	12.9	89.6	11.7	132.4	14.9	153.7	7.4	126.6	13.7
First-stage F-statistic (B^2)	10.2	337.9	9.8	13.3	215.7	12.6	265.3	15.8	303.4	7.60	286.7	14.4
Sample mean (B):	0.699	0.684	0.684	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699

Notes: The table reports results from instrumental variable regressions. The upper part of the table shows the results of the Lind-Mehlum-Test, the lower part presents regression coefficients. We cluster standard errors at the club-by-season-by-round-by-manager level and report p-values in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. In Column 1, we use injuries as source of plausible exogenous variation. In Columns 2 and 3, we only take into account those players who participated in a match for at least 30 minutes. In Column 4, we use an alternative procedure for measuring the extent to which birthplace diversity changes due to an unexpected replacement in the starting line-up. In Column 5 and 6, we control for the performance of the club in the last two matches. In Columns 7 and 8, we take into account that some players migrated during their childhood. In Columns 9 and 10, we use an alternative measure of birthplace diversity. In Columns 11 and 12, we control for form fluctuations.

Table E.14 Newspaper articles about the importance of communication in soccer.

Interviewed person	Interviewer	Sources
Alassane Pléa (player)	Express (newspaper)	https://www.express.de/sport/fussball/borussia-moenchengladbach/gladbach-stuermer-pl%C3%A9a-spricht-nach-rekord-transfer-ueber-favre--dante-und-seine-nummer-30957612
Max Meyer (player)	Sport1 (TV-channel)	https://www.sport1.de/internationaler-fussball/premier-league/2020/02/max-meyer-exklusiv-so-lief-das-mit-schalke-04-sein-leben-in-london
Uli Hoeneß (manager)	Sport-Bild (newspaper)	https://sportbild.bild.de/bundesliga/vereine/bundesliga/hoeness-fordert-deutsch-pflicht-in-bayern-kabine-49570212.sport.html
Rene Paasch (psychologist)	Blog entry	https://www.die-sportpsychologen.de/2017/01/dr-rene-paasch-sprachbarriere-im-fussball/
Franck Ribery (player)	FC Bayern (online-interview)	https://fcbayern.com/de/news/2017/07/interview-mit-franck-ribery-teil-1/
Ciro Immobile (player)	Focus (newspaper)	https://www.focus.de/sport/fussball/bundesliga1/verlorenes-jahr-in-dortmund-groesster-heil-profi-der-bundesliga-immobile-klagt-ueber-bvb-klopp-und-deutschlnad_id_4873285.html/
Giovanne Elber (player)	Hamburger Morgenpost (newspaper)	https://www.mopo.de/sport/hsv/santos-und-walace-aussen-vor-giovane-elber-erklaert-die-brasilianer-krise-beim-hsv-26873000/
Arne Friedrich (player)	Zeit (newspaper)	/https://www.zeit.de/sport/2011-10/arne-friedrich-sprachen-lernen-deutsch-loew/
Birgit Kirschke (teacher)	Zeit (newspaper)	https://www.faz.net/aktuell/sport/fussball/bundesliga/im-interview-die-deutschlehrerin-der-auslaendischen-cottbus-spieler-1464254.html/
Pep Guardiola (manager)	Zeit (newspaper)	https://www.welt.de/sport/fussball/article182738950/Sprachprobleme-Guardiola-wollte-bei-Bayern-nach-zwei-Monaten-hinschmeissen.html/
Robert Peters (reporter)	Rheinische Post (newspaper)	https://rp-online.de/sport/fussball/international/spanien/sprache-ist-der-schluessel_aid-12594963/
Joachim Born (linguist)	scientific work	http://geb.uni-giessen.de/geb/volltexte/2015/11823/pdf/LU_8_Born_Gloning.pdf/
Ellis Palmer (reporter)	online interview	https://www.bbc.com/news/world-europe-44624066/
Anonymous (reporter)	tweet	https://translation-blog.trustedtranslations.com/language-sports-communication-european-football-2017-06-21.html/
J. Steiner und E. Lavric (linguist)	scientific work	https://www.uibk.ac.at/romanistik/personal/lavric/sprache_fussball/publications/b115_merspr_i_fussball_fallstudie_span_legionaer.pdf/
RB Leipzig	homestory	https://rblive.de/ueber-rb-leipzig/dolmetscherin-bei-rb-leipzig-hat-hausaufgaben-fuer-keita-burke-und-bernardo-1523/

Table E.15 Correlations between team characteristics (full sample).

	Birthplace diversity	Home	Cup	ECT	Age	Tenure	Bundesliga Matches	European Matches	Championship matches	Top-league matches
Birthplace diversity	1.000									
Home	0.010	1.000								
Cup	0.014	0.001	1.000							
ECT	0.026	0.003	0.107	1.000						
Age	-0.019	0.011	-0.002	-0.072	1.000					
Tenure	-0.147	0.008	-0.002	0.140	0.247	1.000				
Bundesliga matches	-0.058	0.005	0.042	0.282	0.315	0.591	1.000			
European matches	0.185	0.006	0.074	0.369	0.069	0.389	0.682	1.000		
Championship matches	0.102	0.005	0.071	0.329	0.135	0.323	0.584	0.807	1.00	
Top-league matches	0.175	0.005	0.063	0.233	0.064	0.094	0.318	0.572	0.605	1.000

Table E.16 Correlations between team characteristics (reduced sample).

	Birthplace diversity	Home	Cup	ECT	Age	Tenure	Bundesliga Matches	European Matches	Championship matches	Top-league matches	Market value (log)	FIFA rating
Birthplace diversity	1.000											
Home	0.008	1.000										
Cup	0.018	-0.001	1.000									
ECT	0.017	0.003	0.096	1.000								
Age	0.207	0.008	-0.009	-0.063	1.000							
Tenure	-0.050	0.011	0.020	0.172	0.225	1.000						
Bundesliga matches	0.090	0.005	0.065	0.306	0.252	0.628	1.000					
European matches	0.190	0.006	0.085	0.383	0.166	0.480	0.748	1.000				
Championship matches	0.144	0.003	0.082	0.347	0.180	0.402	0.581	0.830	1.00			
Top-league matches	0.240	0.002	0.062	0.236	0.156	0.134	0.310	0.570	0.600	1.000		
Market value (log)	0.124	0.005	0.103	0.460	-0.167	0.427	0.705	0.783	0.668	0.456	1.000	
FIFA rating	0.148	0.007	0.099	0.443	-0.024	0.457	0.763	0.790	0.697	0.472	0.938	1.000