Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Cultural Homophily and Collaboration in Superstar Teams

Gábor Békés; , Gianmarco I. P. Ottaviano

To cite this article:

Gábor Békés; , Gianmarco I. P. Ottaviano (2025) Cultural Homophily and Collaboration in Superstar Teams. Management Science

Published online in Articles in Advance 20 Jan 2025

. https://doi.org/10.1287/mnsc.2022.01799

informs
https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–20
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Cultural Homophily and Collaboration in Superstar Teams

Gábor Békés,[a,b,c,*] **Gianmarco I. P. Ottaviano**[c,d,e]

[a] Department of Economics, Central European University, 1100 Vienna, Austria; [b] HUN-REN Centre for Economic and Regional Studies, 1097 Budapest, Hungary; [c] Centre for Economic Policy Research (CEPR), London EC1V 0DX, United Kingdom; [d] Department of Economics, Baffi-CAREFIN and IGIER Centers, Bocconi University, 20136 Milan, Italy; [e] Center for Economic Performance, London School of Economics, London WC2A 2AE, United Kingdom

*Corresponding author

Contact: bekesg@ceu.edu, https://orcid.org/0000-0002-6331-4408 (GB); gianmarco.ottaviano@unibocconi.it (GIPO)

Received: June 17, 2022
Revised: July 3, 2023; January 25, 2024
Accepted: March 2, 2024
Published Online in Articles in Advance: January 20, 2025

https://doi.org/10.1287/mnsc.2022.01799

Abstract. One may reasonably think that cultural homophily, defined as the tendency to associate with others of similar culture, affects collaboration in multinational teams in general but not in superstar teams of professionals at the top of their industry. The analysis of an exhaustive data set on the passes made by professional European football players in the top five men's leagues reveals that on the contrary, cultural homophily is persistent, pervasive, and consequential, even in superstar multinational teams of very-high-skill individuals with clear common objectives and aligned incentives who are involved in interactive tasks that are well defined and not particularly culture intensive.

History: Accepted by Olav Sorenson, organizations.

Funding: G. Békés thanks the Hungarian Academy of Sciences [Grant "Firms, Strategy and Performance" Lendület] for support.
Supplemental Material: The online appendix and data files are available at https://doi.org/10.1287/mnsc.2022.01799.

Keywords: organizations • teams • culture • homophily • diversity • language • globalization • big data • panel data • sport

1. Introduction

Cultural homophily, defined as the tendency to associate with others of similar culture, is an extensively studied feature of social interaction investigated by a vast and interdisciplinary literature beginning in sociology (Currarini and Mengel 2016). We demonstrate that it is persistent, pervasive, and consequential, even in superstar multinational teams of very-high-skill individuals with clear common objectives and aligned incentives who are involved in interactive tasks that are well defined and not particularly culture intensive. This is a context where most of the antecedents of homophily highlighted by the literature are not present or at least are at a minimum (McPherson et al. 2001, Lawrence and Shah 2020).

Recent influential surveys on homophily emphasize different aspects and methods depending on their specific discipline of interest.[1] They stress two important distinctions that are particularly relevant for our purposes. The first is between "induced" and "choice" homophily. McPherson et al. (2001) contrast the homophily effects created by the demography of the potential tie pool, which they call "baseline" homophily, with homophily measured as explicitly over and above the opportunity set, which they call "inbreeding" homophily (McPherson et al. 2001, p. 419). When it comes to telling choice from induced homophily, the most important challenge concerns the assessment of the constraints on individual choice sets. As this assessment typically represents a daunting task in observational studies, carefully designed experiments have been increasingly used across disciplines (Jackson et al. 2017, Lawrence and Shah 2020, Ertug et al. 2021). Understanding how to properly account for opportunities in observational studies represents an ongoing challenge (Sorenson and Stuart 2001, 2008; Ertug et al. 2018).

The second distinction contrasts the antecedents with the consequences of homophily. Ertug et al. (2021) state that although the antecedents of homophily are well documented and understood, the same cannot be argued about its consequences, defined as outcomes that go beyond the formation of ties and relationships. In this respect, a crucial trade-off has been highlighted and investigated, especially in management and economics, with an emphasis on performance. On the one hand, homophily facilitates coordination, communication, and trust (Castilla 2011, Opper et al. 2015), with most studies of this aspect presuming the formation of a relationship between actors based on similarity and investigating its consequences. On the other hand, homophily reduces diversity in knowledge, perspectives, and network reach (Burt 1992, Cross and Cummings 2004, Horwitz and Horwitz 2007). Which aspect is more relevant determines whether the impact of homophily on outcomes is positive, negative, or neutral, and this in turn depends not only on the specific outcomes studied but also, on the specific antecedents of homophily identified, the context considered by the analysis, and other contingency factors that can act as moderators. In this respect, the literature has been mixed on whether homophily improves or hurts performance (Jackson et al. 2003).

For our analysis, we exploit a newly assembled data set recording all passes made by professional European football (or soccer) players in the top five men's leagues (Premier League in England, Ligue 1 in France, Bundesliga in Germany, Serie A in Italy, and La Liga in Spain) over eight sporting seasons (2012–2013 to 2019–2020) together with information on key players' and teams' characteristics. The consequence of cultural homophily that we focus on is team collaboration measured as the count of passes from a passer to a receiver relative to the passer's total passes ("pass rate") when both players are fielded together during a half-season in which squad composition is stable. Passes are the essential building blocks of football. They represent how players work together for the common objective of scoring a goal, and they are positively correlated with winning more league points and achieving higher league standings.[2]

We characterize the players' cultural background (henceforth, simply "culture") in terms of a set of cultural traits (Spolaore and Wacziarg 2016, Desmet et al. 2017). These include norms, values, and attitudes that are transmitted intergenerationally, which we operationalize through nationality.[3]

Our football data set has several appealing features for the investigation of the effect of cultural homophily in multicultural teams. Multinational teams are the rule in the top five men's leagues, with players coming from over a hundred countries and squads featuring over a dozen origin countries on average. Although the "rules of the game" are codified and team composition is exogenous to players' decisions, a player is always left with a wide range of options of which teammate to pass to. All sorts of player as well as team characteristics and performance indicators are precisely measured and fastidiously recorded, which allows one to leverage unusually large, fine-grained data on collaboration in an actual workplace rather than in an artificial experimental laboratory with an extremely rich set of team and worker controls (Jackson et al. 2003).[4] Moreover, frequent passes between most teammates generate a large number of networks of maximum density, allowing one to also control for players' unobserved characteristics through individual fixed effects without running into identification and incidental parameter biases (Andrews et al. 2012).

We find strong evidence that players have a preference to pass more to players of their same culture than to other players. In a regression with passer by half-season fixed effects as well as receiver by half-season fixed effects, conditioning on pass features shows that player pairs of same nationality have a pass rate 2.42% higher than player pairs of different nationality. Computing a monetary equivalent value of the estimated homophily, we find that passing to a receiver with the same culture is as likely as passing to a receiver with a different culture but valued a remarkable 10.5% more, which corresponds to 367,500 and 809,550 euros for the median and average players. In other words, homophily is associated with higher valuation of same-culture teammates.

We also find that homophily is more pronounced when stakes are high, decreases with shared experience in the same team, and promotes inequality by helping members of larger cultural groups as these receive a disproportionate share of passes. These findings speak to studies that have tried to assess whether the estimated homophily premium is more likely due to the passer's perception of an objective cost for the team as passes are easier to coordinate within a group ("cost saving" for short) or because of the passer's preference for keeping the ball within his own group ("favoritism" for short). Although supporting cost saving more than favoritism, the evidence that they provide is not conclusive. Based on the existing literature, we consider three aspects. First, favoritism may subside when the stakes are high for the passer as pondered decisions are more likely in this case. However, high stakes may put pressure on the passer, thus causing more instinctive decisions. Second, favoritism may be more visible when the passer belongs to a minority group because of stronger group identity. Nonetheless, minority passers may be less prone to favoritism if they fear the judgment of the majority. Third, favoritism (cost saving) may promote passes to smaller (larger) groups as these are less (more) likely to keep the ball within them. Yet, minorities may be forced to yield to the favoritism of majorities by disproportionately passing to them.

Finally, we discuss some possible mechanisms behind the observed cultural homophily that have been highlighted in the literature. First, people who share a national identity may want the same nationals to do well. We show, however, that colonial legacy matters as much as same nationality, which does not support a mechanism working through national identity over and above broader cultural identity. Second, people may have a false, overly confident belief in the ability of same-culture teammates and underestimate the ability of different culture ones. Although this prejudice mechanism would be detected if homophily declined with shared experience, we find the opposite to hold. Third, being in a small group leads to salience of cultural affiliation as people are more likely to be aware of belonging to the same group when the group is small, which is not supported by our findings on group size and intergroup passes. Fourth, people may have no personal preference when they start collaborating but will build friendships over time through interaction inside and outside the workplace. Friendship may be easier to build with people from the same culture as they typically speak the same language, have the same social cues, like the same cuisine, listen to the same music, watch the same TV series or sports, and so on. Although we do not know how much time players pass together outside the workplace, we still observe that homophily increases with the time they spend in the same clubs, suggesting off-pitch familiarity as a possible mechanism through which same culture leads to homophily.

The rest of the paper is organized as follows. Section 2 describes our data set. Section 3 introduces the estimation strategy, the results of which are discussed in Section 4. Section 5 discusses the distinction between the cost saving and in-group favoritism natures of homophily as well as the possible mechanisms behind it. Section 6 concludes.

2. Data

European football or soccer (henceforth, simply "football") is a game played between two teams of 11 players each. In addition to a goalkeeper, the other 10 players in a team are arranged on the pitch in positions that belong to three broad categories; "defenders" are closer to their own team's goal, "forwards" are closer to the opponent team's goal, and "midfielders" are between them. Within these categories, there are more detailed assignments with role names that vary with the specific layout of the team as chosen by the coach.[5]

Passing the ball is a crucial component of football as it allows a team to control and maneuver the ball with the objective of putting one of its players in the best condition to score a goal. In modern football, passing has become increasingly prominent with respect to dribbling, whereby a player alone takes the ball forward past opponents. For the purposes of the present analysis, "passes" also include free kicks (following unsanctioned action), throw-ins and corner kicks (both after the ball leaves the pitch), and goal kicks as long as the ball stays with the same team. All these together represent, however, only about 5% of all passes.

To estimate how cultural homophily affects collaboration, we create a novel data set combining information on players and passes. The raw data on passes and games are web scraped from www.whoscored.com, a sports website, whereas those on players' value and career come from www.transfermarkt.com, a football player information and valuation website.[6]

2.1. Scope

The analysis covers professional European football in the top five men's leagues: Premier League in England, Ligue 1 in France, Bundesliga in Germany, Serie A in Italy, and La Liga in Spain over eight sporting seasons. These leagues were selected for their undisputed reputation as the pinnacle of national football competitions. Moreover, data availability is the most comprehensive for these leagues.

The data set includes all games played in the sporting seasons from 2012–2013 to 2019–2020, which offer the highest data quality that we could access. A season is the time period between mid-August and mid-May, during which each team plays twice (home and away) with every other team in its league. A season is composed of two halves; the fall half-season runs from mid-August until the end of December, and the winter-spring runs until mid-May. The Premier League, La Liga, Serie A, and Ligue 1 are all composed of 20 teams (playing $20 \times 19 = 380$ games), whereas there are 18 teams ($18 \times 17 = 306$ games) in the Bundesliga. In any given season, there are 98 teams in our sample, and we have $98 \times 16 = 1{,}568$ team by half-season units in our data set. Because of relegation and promotion, we have a total of 154 teams in the sample. Overall, our data set covers a total of $8 \times (380 \times 4 + 306) = 14{,}608$ games.[7]

2.2. Player Data Set

We have 6,998 players in our sample for whom we can fully map their entire career, with a typical team relying on a squad of about 30 players. For every player, data include his country of birth, single or multiple citizenship information, country of birth, date of birth, height, and participation in a national team. These are all time invariant in our data set.

European football is truly globalized as there are players from 138 countries of citizenship in our sample. French, Spanish, and Italian players make up the largest citizenship groups followed by Germans, English, Brazilians, and Argentinians. Other countries of citizenship with several players include the Netherlands, Serbia, Senegal, and Uruguay.

We characterize the cultural background (or simply, "culture") of team members in terms of nationality, but we also look at other traits, such as native language, colonial legacy (past membership of a colonial empire), and federal legacy (past membership of a political union).[8]

To determine whether two players have the same culture, we characterize the chosen cultural traits as follows. First, nationality is defined on the basis of citizenship of a country. As some players have multiple citizenships, we define two players as "same nationality" if they share at least one of them or have the same country of birth. Second, to ascertain common colonial legacy, we use colonial links data from CEPII[9] as in Head and Mayer (2014). We define two players as sharing the same colonial legacy if their nationalities include a former colonial ruler and its subject (e.g., Spain and Argentina) or two subjects of the same colonial ruler (e.g., Argentina and Uruguay). Third, by common federal legacy, we refer to countries that formed political unions some time in the twentieth or twenty-first century. These include (i) countries of the former Soviet Union, including Russia and Ukraine; (ii) countries of the former Yugoslavia, including Croatia and Serbia; (iii) the Czech Republic and Slovakia; and (iv) Ireland, Northern Ireland, and Great Britain (itself including three constituent footballing countries: England, Wales, and Scotland). Although possible because of multiple citizenship, it is extremely rare that players share both colonial and federal legacies. For these players, same colonial legacy subsumes same federal legacy.

Fourth, for language, we rely on CEPII data as in Head and Mayer (2014) to ascertain whether two countries share one or more common languages. We assume that a player speaks (as mother tongues) the official and widely spoken languages of his country of citizenship at the beginning of his career. We consider some languages that are very close, even if not identical, as one language (see Section 2 in the Online Appendix for details). The fact that our language variable refers essentially to mother tongue implies that it should indeed be seen more as a cultural marker than as a means of communication. For many players (such as Argentinean and Spanish, Brazilian and Portuguese, or French and Senegalese players), same colonial legacy subsumes same language. For other players (such as Croatian and Serbian, Czech Slovakian, or Irish and British players), same federal legacy subsumes same language. As a result, there is a small residual group of players who share the same language but neither colonial nor federal legacies.

Based on nationality, language, colonial legacy, and federal legacy, we define the following categories: "same nationality" if two players share nationality (e.g., two Argentinian players); "same colonial legacy" if they have different nationality but same colonial legacy (e.g., Argentinian and Spanish); "same federal legacy" if they have different nationality but same federal legacy (e.g., Croatian and Serbian); "same language" if they have different nationality, different colonial legacy, and different federal legacy but same language (e.g., Belgian and French); and "no shared culture" if they have different nationality, different colonial legacy, different federal legacy, and different language (e.g., Argentinian and French).

More than a quarter of players have multiple citizenship. In such cases, if two players are citizens of the same country or of at least two different countries with common language, they are considered as speaking the same language. Analogously, if two players are citizens of the same country or of at least two different countries with common colonial (federal) legacy, they are considered as having the same colonial (federal) legacy.

In our data set, 37.9% of the players have the same nationality; 8.4% have the same colonial legacy; 1.5% have the same federal legacy; and 2.8% have the same language but different colonial legacy, different federal legacy, and different nationality. We consider all of these players as having the same culture. According to this definition, 50.6% of the players in our sample have the same culture, whereas 49.4% of them do not.

2.3. Pass Data Set

The pass data set contains aggregate information about passes between any two players at the half-season level. A pass is defined as any movement of the ball from one player to another (including free kicks, throw-ins, corner kicks, and goal kicks). There are about 365 successful passes on average per game, which for two teams, implies 730 passes per game or about 8.1 passes per minute. We have 10.73 million passes in total.

In a game, most players pass to each other but with different frequency depending on their positions. On average, the pass frequencies of outfield players between broad positions (defender, midfielder, and forward) are fairly balanced. The three highest frequencies are observed from defenders to midfielders (11.58%), from defenders to defenders (10.9%), and from midfielders to midfielders (8.86%). The lowest frequency is observed from goalkeepers to forwards (1.89%).

Aggregation is at the level of half-seasons to average out match contingencies. The partition in half-seasons is determined by the timing of the transfer windows, which are located between seasons (summer transfer window) and at the beginning of the calendar year (winter transfer window). It also splits the number of games during a season into two approximately equal parts; the number of games per team in a half-season ranges between 16 and 20 compared with the exact equal split of 17 for the German Bundesliga and 19 for the other top five leagues.

We deal with player pairs with zero pass count as follows. If two players never pass to each other during a half-season, it must be that it is impossible for them to do so because of fielding or positioning reasons (e.g., the two players are only fielded to substitute each other as forwards). We thus drop the corresponding player pairs from the data set. However, if we observe that a player passes to a given teammate but is never reciprocated, we keep the player pair. This implies that we have some zeros in the data set recording the lack of passes from a player to a teammate from whom he nonetheless receives passes. Only 7.8% of the observations give rise to such zeros.

An alternative to half-seasons would be to consider full seasons. However, half-seasons have advantages compared with seasons. The presence of the winter transfer window implies that during a season, a team's squad may change composition. A player's quality can be more reasonably considered constant in a half-season than in a season, especially as younger players may evolve. The fact that half-seasons are separated by transfer windows allows us to cleanly map players' careers as they change teams, thereby combining player and pass information in a consistent way. Finally, considering a half-season allows us to investigate the role of common experience as players who spend more time together on the pitch may learn to pass more to each other.

2.4. Combined Data Set

We combine player information with pass information to obtain a relational data set linked via player names as well as additional information.

To match player and pass data, one has to identify players in both data sets and create a unique identifier for players. This process has proved to be a difficult task, for which we have developed a matching algorithm based on player names and additional information. The procedure is detailed in Section 2 in the Online Appendix, where we also discuss a few decisions regarding data cleaning, such as dropping players who only have a single passing partner or those we could not identify. All results are robust to these decisions.

The resulting data set is a directional pass data set that when keeping track of who is the passer and who is the receiver, consists of 669,022 observations at the passer, receiver, and half-season level. In a half-season, a player makes a total of 288 passes on average (ranging between 0 and 2,166, with the median equal to 221). On average, he passes to 18.07 receivers (ranging from 2 to 35, with the median equal to 19). The average pass count from passer to receiver is 15.92 (ranging from 0 to 488, with the median equal to 8).

3. Separating Choice from Opportunities

A crucial challenge in assessing how common culture affects collaboration through passes arises from the conflation of choice and opportunity. As discussed in Section 1, individuals may collaborate more with similar others because they choose to do so ("choice homophily") or because collaboration with similar others is forced on them by circumstances ("induced homophily").

A naive model of passing behavior would explain the number of passes between two players intuitively in terms of the time they are fielded together, the number of passes made by the passer to all teammates, the number of passes received by the receiver from all teammates, and the average bilateral distance between the two players on the pitch over a half-season. The first three variables would be expected to positively affect the number of passes between the player pair, whereas the fourth would be expected to have a negative influence. Homophily would then act as a shifter, leading to more passes between player pairs of similar culture after controlling for all of the above variables as nondistance-related frictions because of cultural differences make it more difficult to pass the ball between player pairs of different culture than between those of same culture.

Although intuitively attractive, the naive approach may lead to biased estimates of homophily as it neglects the potential role of other player characteristics, in particular those that are hardly observable and cannot be taken off the shelf to enrich the model's specification. Among these, the most important source of concern is arguably that any reasonable model of passing behavior should account for the alternatives that the passer faces in terms of receivers as the decision of passing to one of them is made after assessing the benefit of giving the ball to each of them. Analogously, one should also account for the alternatives that the receiver faces in terms of passers as the passes that he receives depend on which teammates are actually fielded.

To capture players' unobserved choice sets, we use player by half-season fixed effects. The idea is that a player's opportunities may vary between half-seasons because of transfers and different squad composition. They may also vary between games in a half-season because of different team composition or because of the fact that in some games, the player is fielded in a different position than the one that he is typically assigned to. However, as these idiosyncratic events should be averaged out over a half-season, in any given half-season, a player's choice set becomes one of his fixed characteristics. Such fixity allows the player's choice set to be absorbed by his fixed effect. This clears the ground for the identification of choice homophily as the player's fixed effect also includes the average attractiveness of all available alternatives against which his choice can be benchmarked.

In the appendix, we show how this insight can be obtained as a structural implication of a forward-looking discrete choice model of passing behavior.[10]

The empirical implementation of this approach requires a generalized linear estimator with a log link function. In such a setup, there is a rich literature on the benefits of using a Poisson model rather than a log(*count*) model with a large number of fixed effects. The former is in line with best practices in the estimation of gravity equations through fixed effects Poisson pseudomaximum likelihood (FE-PPML) in international trade (see, e.g., Fally 2015 and Santos-Silva and

Tenreyro 2022).[11] Nonetheless, we will also provide robustness checks with a log(*count*) model.

We implement the Poisson model as follows. We use the number of passes from player o to player d in half-season t as dependent variable $pass\_count_{o,d,t}$, and we introduce team by half-season fixed effects to control for the fact that the number of passes between teammates may be mechanically driven by their team's total number of passes.

There are two types of frictions that can make it difficult for player o to pass to player d. First of all, there are distance-related frictions, for which we construct the following measure:

$$PassFric_{o,d,t} = \gamma_1 PassDist_{o,d,t} + \gamma_2 Forwardness_{o,d,t} + \eta Position_o Position_d,$$

where on average in a half-season, $PassDist_{o,d,t}$ is the distance of passes between the two players, $Forwardness_{o,d,t}$ is the share of passes between the two players with a forward direction, and $Position_o Position_d$ is a dummy variable capturing the two players' broad positions (such as defender, midfielder, and forward). As we acknowledge that $PassDist_{o,d,t}$ and $Forwardness_{o,d,t}$ may actually be a mechanism rather than a confounder, we will show results with and without them.

Second and foremost, there may be nondistance-related frictions because of cultural differences that make it difficult to pass the ball from o to d independently of their positions. In this respect, we measure cultural similarity through the time-invariant variable $SameCult_{o,d}$, which combines the categories described in Section 2.2; it takes value 1 if two players share nationality, colonial legacy, federal legacy, or just language and value 0 otherwise.

In light of the previous discussion, we estimate two different passing models, all by Poisson regressions with standard errors clustered at the passer by half-season and receiver by half-season levels. The first is a mechanical model explaining the number of passes between two players in terms of the number of passes made by the passer to all teammates ($P_{o,t}$), the number of passes received by the receiver from all teammates ($R_{d,t}$), bilateral distance-related frictions ($PassFric_{o,d,t}$), and cultural similarity ($SameCult_{o,d}$):

$$E(pass\_count_{o,d,t} | \ldots) = \exp(\delta SameCult_{o,d} + PassFric_{o,d,t} + \beta_1 \ln P_{o,t} + \beta_2 \ln R_{d,t} + \mathbf{1} \ln \tau_{o,d,t} + \phi_{c,t}. \quad (1)$$

The second passing model that we estimate uses player by the half-season fixed effects equation to also control for players' choice sets. For the reasons detailed above, this is our preferred specification:

$$E(pass\_count_{o,d,t} | \ldots) = \exp(\delta SameCult_{o,d} + PassFric_{o,d,t} + \mathbf{1} \ln \tau_{o,d,t} + \upsilon_{o,t} + \upsilon_{d,t}), \quad (2)$$

where $\upsilon_{o,t}$ and $\upsilon_{d,t}$ are player by half-season fixed effects.[12]

A positive estimate for the coefficient δ of the same culture indicator $SameCult_{o,d}$ would reveal the presence of choice homophily as it would imply that after partialing out distance-related frictions, relative to overall passing patterns, players with same culture pass more to each other than to players of different culture. Accordingly, we call the estimated δ the "homophily premium."

In both estimated equations, teams and their half-seasons are pooled. However, teams will vary in terms of style, especially in terms of overall passes per game. To partial out team characteristics, in Model (1), we include team by half-season dummies, $\phi_{c,t}$, whereas in Model (2), team by half-season dummies are absorbed by player fixed effects.[13]

For both models, we need to account for time spent together when passing opportunities arise. This exposure variable, τ_{o,d_t}, is calculated as the ratio of the number of passes made by player o when player d is on the pitch to player o's total passes.[14] Empirically in the Poisson model, we added an offset variable that is simply the log of the exposure variable ($\ln \tau_{o,d_t}$). The coefficient of the offset variable is forced to be equal to one.

Time spent together on the pitch as captured by the exposure variable is set by the coach. Thus, comparing estimates for the homophily premium obtained with or without the inclusion of the exposure variable allows us to tell how much of the premium is mediated by coach decisions.

4. Homophily in Collaboration

We are now ready to report and discuss our empirical findings based on the estimation of regressions (1) and (2).

4.1. Results

Table 1 presents the results from regressions (1) and (2) in columns (1) and (2), respectively. In these columns, by forcing the coefficient of $P_{o,d}$ and $\tau_{o,d}$ to equal to one, the effect of culture is estimated for the number of passes from the passer to the receiver relative to the former's total number of passes when both players are fielded together. In both models, team by half-season fixed effects absorb the total number of team passes. They also absorb team and team by half-season characteristics, such as club history or current management. Crossposition dummies capture the relative roles of players.

Column (1) in Table 1 supports the naive model's prediction that the pass count between two players is positively related with their total passes made or received and negatively related with pass distance. Passing pairs who move the ball forward tend to pass

Table 1. Baseline Results

	(1)	(2)
Same culture (any) (0/1)	0.0132***	0.0242***
	(0.0020)	(0.0025)
Passer total passes (ln)	0.9144***	
	(0.0011)	
Receiver total pass received (ln)	0.2948***	
	(0.0035)	
Average length of passes (ln)	−0.6587***	−0.7788***
	(0.0052)	(0.0053)
Average forwardness Ind (0-1)	0.0764***	0.0813***
	(0.0049)	(0.0061)
Observations	668,105	668,105
Pseudo-R^2	0.70881	0.76077
Team by half-season dummies	✓	
Crossposition dummies	✓	✓
Passer by half-season dummies		✓
Receiver by half-season dummies		✓

Notes. Poisson regression. The dependent variable is pass count. The offset variable is included. Standard errors, clustered at the passer by the half-season level and at the receiver by the half-season level, are in parentheses.
***$p < 0.01$.

more as well. As for the homophily premium, relative to overall passing patterns, players with same culture pass 1.32% more to each other than to players of different culture. Column (2) in Table 1 replaces total passes made and total passes received by a player pair with time-varying passer by half-season and receiver by half-season fixed effects. This still allows player characteristics to change over time and implies that the estimated coefficient of same culture is close to what would be the average of coefficients if estimated one by one for teams and time periods. The estimated homophily premium is 2.42%. To interpret this coefficient, consider the passes made by a player in a half-season, conditioning on constant and time-varying receiver characteristics, passer-receiver position pair, and other pass features. This player is expected to pass 2.42% more to teammates of same culture than to teammates of different culture. This is our preferred estimate of the "homophily premium."

To put the estimated choice homophily into context, we estimate how homophily affects passes without distinguishing between its choice and induced aspects. We do so by estimating the effect of $SameCult_{o,d}$ as an unconditional average difference within team and half-season in the following Poisson model:

$$E(pass\_count_{o,d,t}|\dots) = \exp(\delta SameCult_{o,d} + \nu_{team,t}), \quad (3)$$

where $\nu_{team,t}$ is a team by half-season fixed effect.

The corresponding results are reported in column (1) in Table 2, where for ease of comparison, column (3) in Table 2 recalls the baseline results from column (2) in Table 1. Column (1) in Table 2 offers clear unconditional

Table 2. From Total Homophily to Choice Homophily

	(1)	(2)	(3)
Same culture (any) (0/1)	0.0619***	0.0380***	0.0242***
	(0.0056)	(0.0033)	(0.0025)
Observations	668,105	668,105	668,105
Pseudo-R^2	0.07822	0.67154	0.76077
Exposure (Time shared together)			✓
Pass features		✓	✓
Crossposition dummies	✓	✓	✓
Team by half-season fixed effects	✓		
Passer by half-season fixed effects		✓	✓
Receiver by half-season fixed effects		✓	✓

Notes. Poisson regression. The dependent variable is pass count. Standard errors, clustered at the passer level, are in parentheses. Columns (2) and (3) include ln(*average pass distance*) and forwardness index.
***$p < 0.01$.

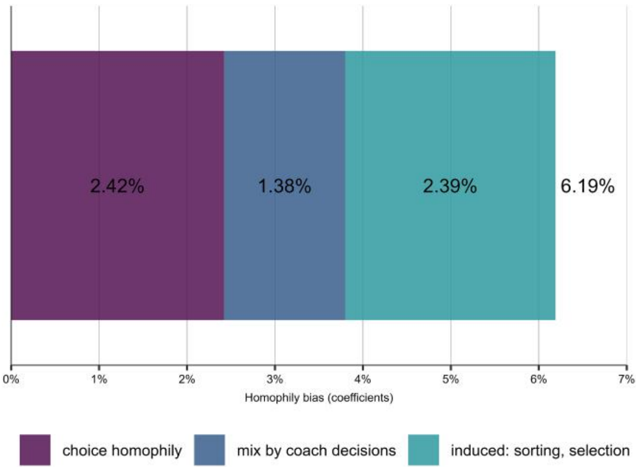
evidence of homophily; players with same culture tend to pass 6.19% more to each other than to players with different culture. Columns (2) and (3) in Table 2 introduce the full set of player fixed effects and passes controls with an exception. Specifically, although column (3) in Table 2 controls also for the time that a player pair spends together on the pitch in a half-season, column (2) in Table 2 does not.

Comparing the three columns of Table 2 suggests that 2.36% of the overall homophily premium of 6.19% vanishes when controlling for player and pass characteristics, and a further 1.38% disappears when one considers players' time spent together on pitch. This further reduction reveals the role of managers' decisions in allowing the team to internalize the effects of homophily. Endogenous team formation may mitigate the effects of homophily on collaboration, which is captured by the time that a player pair spends together on the pitch. For example, as the manager observes his players in training, he may decide to field same culture players in a game because he sees them collaborating more. In this case, the manager acts as a mediator, allowing the team to internalize the effects of homophily. Hence, the effects of homophily on collaboration that we estimate can be seen as a lower-bound estimate with respect to what would be found in a randomly composed team.

Figure 1 summarizes how the overall homophily premium of 6.19% can be decomposed into a choice homophily premium of 2.42%, a mitigation premium of 1.38% because of endogenous managerial decisions, and an induced homophily premium of 2.36% because of player and pass characteristics.

We now discuss the extent to which this finding might be affected by omitted variables or the operationalization

Figure 1. (Color online) Dissecting Total Homophily Bias



Note. Coefficient values from fixed effects regression are from Table 2.

of the notion of culture. We then comment on the importance of the premium for footballing outcomes.

4.2. Model Extensions and Potential Confounders

According to our estimated "homophily premium," conditioning on constant and time-varying receiver characteristics, passer-receiver position pair, and other pass features, a player is expected to pass 2.42% more to teammates of same culture. In Section 3 in the Online Appendix, we show that these results are robust (i) to OLS instead of Poisson, (ii) excluding special cases like goalkeepers, and (iii) to allowing for shared minutes to directly affect passes.

Identification of choice homophily according to our definition of culture assumes orthogonality with respect to other possible dimensions of passing assortativity, but some of these dimensions might actually overlap with culture. To check whether this is indeed the case, we extend the fixed effects Poisson pseudomaximum likelihood specification (2) by including a battery of player characteristics other than culture that could arguably foster reciprocal passes between players: pitch position, shared experience, quality, physical attributes, regulations, national styles, and common values. Each of them could be correlated with nationality, thereby inducing collaboration between same nationals and thus, confounding our estimated homophily. The corresponding results are reported in Table 3, where column (1) recalls the baseline homophily premium estimate for ease of comparison.

4.2.1. Pitch Position as the Mechanism.

It is possible that players of same culture position themselves strategically at close (or distant) proximity to each other, leading to shorter and sideways (or longer and forward)

Table 3. Extensions: Potential Confounders

	(1)	(2)	(3)	(4)	(5)
Same culture (any) (0/1)	0.0242***	0.0238***	0.0247***	0.0213***	0.0247***
	(0.0025)	(0.0028)	(0.0026)	(0.0026)	(0.0025)
Average length of passes (ln)	−0.7788***		−0.7943***	−0.7823***	−0.7631***
	(0.0053)		(0.0052)	(0.0052)	(0.0052)
Average forwardness Ind (0-1)	0.0813***		0.0142**	0.0113*	−0.1145***
	(0.0061)		(0.0063)	(0.0062)	(0.0062)
Shared experience, half-season+ (0/1)			0.0117**	0.0110*	
			(0.0056)	(0.0056)	
Experience at any prior team (0/1)			−0.0153***	−0.0165***	
			(0.0044)	(0.0043)	
Height difference (abs, cm)				−0.0126***	
				(0.0002)	
Players value difference (dlog)				−0.0008***	
				(0.0002)	
Both treated as EU player (0/1)				0.0107	
				(0.0080)	
Observations	668,105	668,105	668,105	668,105	668,096
Pseudo-R^2/R^2	0.76077	0.74289	0.75930	0.76074	0.75731
Passer by half-season fixed effects	✓	✓	✓	✓	✓
Receiver by half-season fixed effects	✓	✓	✓	✓	✓
Crossposition dummies	✓	✓	✓	✓	✓
Crossposition dummies by nationality					✓

Notes. Poisson regression. The dependent variable is pass count. The offset variable is included. Standard errors, clustered at the passer by the half-season level and at the receiver by the half-season level, are in parentheses. "*Both treated as EU player (0/1)*" reflects the corresponding national regulations (see the Online Appendix for details).

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

passes. In this case, positioning is a mechanism, and thus, it should not be included as a confounding variable. Column (2) in Table 3 suggests that the exclusion of positions does not change the result.

4.2.2. Shared Experience. A player pair may collaborate more in the present because of previous interactions. If, for instance, these interactions were more likely between players from the same country because they grew up there, then our results would not reflect an actual relationship between same culture and collaboration. To check whether this matters, column (3) in Table 3 includes shared experience through two binary variables. Two players may pass more in the current team because they have been playing there together for at least a half-season or because they have played together somewhere else in the past.[15] For the latter variable, we collect team history for all of the players in our sample starting from youth games around the age of 14. Any shared past experience is actually rare; only 3% of player pairs have ever played together. Although the point estimates for both variables are different from zero, neither influences the estimate of homophily. Interestingly, although the coefficient for shared experience at the current club is positive, the coefficient for shared past experience is negative.

4.2.3. Quality, Physical Attributes, and Regulations. Two players may collaborate more because of similar quality or physical attributes. That would happen, for instance, if players of similar quality (or height) passed more to each other and if players of given quality (or height) came from the same country, thus confounding our results. Moreover, assortativity may be induced by regulatory constraints on players' countries of origin, making it more likely for better or taller players to hail from the same country. To check whether this is the case, column (4) in Table 3 considers assortativity in terms of quality (log player value) and physical attributes (height difference). Finally, national regulations might restrict the fielding of non-European Union (EU) nationals (with a variety of third country-specific exemptions). We coded these rules and created a "both treated as EU player" indicator.[16] Column (4) in Table 3 shows that all of these variables have only minor effects on the point estimate of the homophily premium.

4.2.4. National Style. Another confounding factor could be shared football style in a country. Countries differ in terms of their style of football, and this style is transmitted to local players at an early career stage. For instance, some countries may traditionally favor a fast "vertical" style that tries to score goals by quickly moving the ball into scoring range through long forward passes, whereas others may prefer a slow "horizontal" style, playing less quickly with short passes to find a weakness in the opposing team's tactics. Shared national style would then make it more likely for players from the same country to pass more to one another without necessarily reflecting an actual relationship between same culture and collaboration.

To partial out the features of national style, column (5) in Table 3 interacts the position-by-position set of dummies with the (first) nationality of players, which allows, say, Spanish defenders and midfielders to have a different pass intensity than Croatian defenders and midfielders. This interaction, however, does not affect our results as the point estimate for the coefficient of interest hardly changes (2.41% versus 2.42%).

4.3. Operationalization of Culture

In our regressions, two players belong to the same cultural group if they share nationality, colonial legacy, federal legacy, or language. However, treating (as we do) all cultural groups as equally different from one another may be somewhat too coarse. In the literature on culture, it has been indeed argued that there are dimensions along which some cultural groups can be considered more closely related to one another than to other cultural groups (Desmet et al. 2017).

We address this source of concern in four ways. First, we disentangle the different components of culture (as described in Section 2.2) to allow for asymmetries among cultural groups depending on whether they share (i) nationality, (ii) colonial legacy without nationality, (iii) federal legacy without nationality, or (iv) language without nationality or any historical legacy. The benchmark consists of player pairs with different nationality, different colonial legacy, different federal legacy, and different language. We then re-estimate the fixed effects Model (2), replacing $\delta SameCult_{o,d}$ with $\delta_1 SameNat_{o,d} + \delta_2\, SameCol_{o,d} + \delta_3 SameFed_{o,d} + \delta_4 SameLan_{o,d}$.

Second, we look at linguistic proximity, relying on the CEPII data set to create a similar language indicator variable based on the common language index developed by Melitz and Toubal (2014). This indicator equals one if language similarity is above the median value 0.5 (as, e.g., for Italy and Spain, Denmark and Sweden, and Croatia and Bulgaria) and zero otherwise; the indicator's mean value is 0.13. According to the indicator, in our data, 6.7% of player pair by half-season observations can be classified as speaking similar languages, despite their different nationality, different colonial legacy, different federal legacy, and different language.

Third, we consider geographical proximity as measured by another indicator variable, which equals one if two countries share a land border and zero otherwise. In our data, 7.5% of observations have a shared border, despite their different nationality, different colonial legacy, different federal legacy, and different or not even similar language.

Finally, another reason why some cultural groups might be considered more closely related is that their values are more aligned. We look into this issue by using the combined Wave7 of the World Value Survey (WVS) + European Values Survey (EVS) exercise to create a simple distance metric between countries. Specifically, following Spolaore and Wacziarg (2016) and Desmet et al. (2017), we assess the similarity of countries' answers to 72 selected questions on a variety of values, such as the role of family or political trust.

Distance between countries is computed as the average Euclidean distance between the shares of nationals giving the various admissible answers. For details, see Section 2 in the Online Appendix. The distance ranges between 0 when two countries are perfectly aligned and $\sqrt{2}$ when they are fully misaligned. In our data set, the mean and median distances are 0.25 and 0.17, respectively, whereas the minimum and maximum are 0.08 and 0.56, respectively. To check the importance of values, we create a binary variable for similar and dissimilar countries defined as those below and above the median distance, respectively.

Results are reported in Table 4. Column (1) in Table 4 shows that different cultural components have different effects on homophily. Compared with pairs not sharing any aspect of culture, we find a homophily premium of 2.85% for same nationality and 2.88% for same colonial legacy without same nationality. Same colonial legacy is, therefore, as consequential as same nationality. We find no homophily premium for player pairs from formerly federated countries (such as countries of the former Soviet Union or Yugoslavia). We also do not find any correlation for the relatively few cases of having just shared language without same colonial legacy, federal legacy, or nationality.

In column (2) in Table 4, we add linguistic proximity, partitioning players into four categories: same nationality, same language but different nationality, similar language, and dissimilar language (base). Compared with pairs who speak different and dissimilar languages (such as Russian and English), players of shared nationality (and language) pass 3.02% more, those with same language but different nationality (such as Spanish and Argentinian, Austrian and German, or Irish and English) pass 1.56% more, and pairs with similar rather than same language (such as Dutch and German) pass 1.11% more. In column (3) in Table 4, we consider geographical proximity (being neighbors) on top of language variables. We

Table 4. Dissecting Culture

	(1)	(2)	(3)	(4)
Same nationality (0/1)	0.0285*** (0.0030)	0.0302*** (0.0031)	0.0315*** (0.0031)	0.0183*** (0.0031)
Same colonial legacy (0/1)	0.0288*** (0.0042)			
Same federal legacy (0/1)	−0.0017 (0.0103)			
Just shared language (0/1)	−0.0046 (0.0070)			
CLE: diff country, same language (0/1)		0.0156*** (0.0039)	0.0140*** (0.0040)	
CLE: diff country, similar language (0/1)		0.0111** (0.0044)	0.0094* (0.0045)	
Geographical proximity (neighbors) (0/1)			0.0064* (0.0031)	
WVS: similar values (0/1)				−0.0077*** (0.0025)
Observations	668,105	668,105	668,105	668,105
Pseudo-R^2	0.76078	0.76077	0.76077	0.76076
Passer by half-season fixed effects	✓	✓	✓	✓
Receiver by half-season fixed effects	✓	✓	✓	✓
Crossposition dummies	✓	✓	✓	✓

Notes. Poisson regression. The dependent variable is pass count. The offset variable is included. Standard errors, clustered at the passer by the half-season level and at the receiver by the half-season level, are in parentheses. Column (1) indicates the same culture separated into traits top coded in this order: nationality, federal legacy, colonial legacy, and language. Column (2) keeps the same nationality and adds linguistic similarity categories as per the Common Language Index "CLE" of Melitz and Toubal (2014) (the base is a different country without a similar language). Column (3) adds geographic proximity. Column (4) has value similarity based on the 2021 WVS/EVS, cut at the median (the base is different values). All columns include pass distance and the forwardness index.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

find the results with respect to language proximity to be broadly in line with those in column (2) in Table 4, only with slightly smaller point estimates. However, for countries with the same (or similar) language that are also neighbors, coefficients should be added up. Geographical proximity alone contributes a little to passing.

Lastly, we compare player pairs depending on whether they exhibit same nationality, not the same nationality but similar values, and dissimilar values (base). Column (4) in Table 4 shows a point estimate for similar values that is actually negative; player pairs from countries with similar values seem to pass less.

To summarize, beyond nationality, we find choice homophily to be about similar language (interpreted as a cultural marker rather than a means of communication) and shared history (especially colonial ties) and unlikely to be driven by shared values.

4.4. Importance of Homophily

Beyond statistical significance, another interesting question to ask about our findings is whether the estimated homophily premium makes a difference in terms of footballing outcomes.

This is a hard question to answer directly if one has in mind the overall impact of homophily on a team's performance as measured, for instance, by its end-of-season position in the league standings. The reason is that to answer the question, one would have to estimate not only by how much choice homophily fosters bilateral passes as we do ("pass creation") but also, by how much it diverts passes from alternative options ("pass diversion") and whether the resulting reallocation of passes fosters or hampers team performance. Only by netting out the opposite effects of pass creation and pass diversion would it be then possible to assess by how much a team's performance is affected by homophily through passes. The quantification of these effects would require a "partial equilibrium" model of team performance. Moreover, to map differential passes into a team's rank in the league standings, one would also have to deal with the fact that points gained or lost by a team are lost or gained by its competitors. The quantification of these aggregate effects would call for a "general equilibrium" model of league competition. Although this is a very interesting direction of research, it clearly goes beyond the scope of the present paper.

That said, it could still be useful to provide some benchmark against which to assess, at least indirectly, the quantitative relevance of the estimated homophily premium for team performance. To do that, we investigate the importance of the homophily premium based on the estimation of an alternative but closely related specification to the fixed effects Model (2), in which we replace player by half-season fixed effects with player values obtained from www.transfermarkt.com. We recover a player's estimated transfer market value, defined as the "expected value of a player in a free market" as determined by a group of experts. This estimate is based on how much a player may contribute to the team's success, how well he plays, and how valuable he may be to another team. A player's transfer value is considered a consensus summary measure of the quality of his footballing skills accounting for all observable (to the experts but not necessarily to us) circumstances. The most important of such circumstances include development potential, experience level, future prospects, injury susceptibility, league-specific features, marketing value, performance at the club and national team, performance potential, number and reputation of interested clubs, and prestige. Accordingly, a player's value is a more general proxy of footballing skills than his fixed effect estimated from passes. For the same reason, however, it is also less tightly linked to passing performance than his fixed effect estimate. Yet, what is particularly interesting for our purposes is that a player's value also takes into account many of the player characteristics that being unobservable to us, we soak up with player by half-season fixed effects. These characteristics importantly include the composition of the player's team in a half-season and thus, both his passing options and his coach's decisions ($\tau_{o,d,t}$).

Replacing player fixed effects in (2) with player values ($value_{o,t}$, $value_{d,t}$) and other observable player characteristics ($playerchar_{o,t}$, $playerchar_{d,t}$), we estimate

$$
\begin{aligned}
&E(pass\_count_{o,d,t}|\ldots)\\
&\quad = \exp(\delta SameCult_{o,d} + PassFric_{o,d,t} + \eta_1 value_{o,t}\\
&\qquad + \eta_2 value_{d,t} + \theta_1 playerchar_{o,t} + \theta_2 playerchar_{d,t} + \phi_{c,t}),
\end{aligned}
\tag{4}
$$

where other player characteristics include age, height, time (in days) elapsed since joining the team, and a binary indicator for being on loan.[17] We also include position by half-season dummies and nationality by half-season dummies. This specification does not feature the share of passes made by player o when player d is on the pitch because by construction, passer's and receiver's values also reflect complementary footballing skills that are intertwined with their chances of being fielded together.

Table 5 reports the result from estimating (4). It confirms that the number of passes made or received by a player depends on his value. It also reveals a statistically significant homophily premium. More to the point here, it allows us to put a monetary equivalent value on the homophily premium. In particular, contrasting the estimated 4.35% premium with the estimated 34.57% coefficient of log receiver valuation implies that passing to a receiver with same culture is equally likely as passing to a receiver with different culture but valued a remarkable 10.5% more.[18] This corresponds to 368,550 and 811,863 euros for the median and average players,

Table 5. Benchmarking Homophily

	(1)
Same culture (any) (0/1)	0.0435*** (0.0047)
Passer player valuation (ln)	0.3398*** (0.0042)
Receiver player valuation (ln)	0.3457*** (0.0041)
Observations	668,105
Pseudo-R^2	0.35204
Team by half-season dummies	✓
Passer: Position by league and half-season dummies	✓
Receiver: Position by league and half-season dummies	✓
Passer: Nationality by league and half-season dummies	✓
Receiver: Nationality by league and half-season dummies	✓
Crossposition dummies	✓

Notes. Poisson regression. The dependent variable is pass count. Standard errors, clustered at the passer by the half-season level and at the receiver by the half-season level, are in parentheses. Player values (million euros in ln) are measured at the start of the half-season. For both players, the individual controls are height, age, time since with club (in days), and binary if on loan. It includes ln(*average pass distance*) and the forwardness index.

***$p < 0.01$.

respectively, as these are valued 3.50 and 7.71 million euros, respectively. In other words, homophily is associated with a remarkably higher valuation of same-culture teammates.

5. Cost Saving vs. Favoritism

So far, we have been interested in assessing whether choice homophily actually exists while remaining agnostic about whether its existence reveals the passer's desire to promote team performance or rather, a manifestation of otherwise in-group bias detrimental to the team. We now discuss whether the estimated homophily premium is more likely due to the passer's perception of an objective cost for the team as passes are easier to coordinate within a group ("cost saving") or to the passer's preference for keeping the ball within his own group ("favoritism").

To answer the foregoing question, we rely on four working hypotheses borrowed from the literature. First, favoritism should subside when the stakes are high for the passer as pondered decisions are more likely in this case. We should then observe less homophily than when stakes are low. Second, favoritism should be more visible when the passer belongs to a minority group, in which case we should observe more homophily if the passer belongs to a small group rather than to a large group. Third, focusing on passes between players of different culture, favoritism should promote passes to smaller groups as these are less likely to keep the ball within them, whereas cost saving should promote passes to larger groups as these are more likely to keep the ball within them and thus, reduce a team's passing frictions. Fourth, because of prejudice, favoritism should decline with players shared experience in a team, in which case, we should observe that homophily weakens as players of different culture spend more time in the same team.

5.1. High Stakes

Unintentional bias is more likely when one makes fast decisions or acts on the spur of the moment (Price and Wolfers 2010). In contrast, there are circumstances in a game in which a passer might take a step back, thus reducing bias. In this respect, as high stakes raise awareness and foster reasoning, one may expect favoritism to subside when stakes are high for the passer. We consider several such circumstances determined in terms of the type of passes. We look at long passes (that are riskier; also, a wrong pass could be costly), passes before a shot on goal (that are under higher pressure), and complex pass sequences moving the ball forward (that are a crucial part of building an attack). In Table 6, we collected our key results.

First, columns (1) and (2) in Table 6 look at long passes directly. Long passes are identified by an indicator variable valued at one when their length is above the median. The indicator is then interacted with the homophily premium. Column (1) in Table 6 reveals that homophily is stronger for long passes (1.80% versus 3.29%). Column (2) in Table 6 shows that including average pass length changes the estimated coefficients only marginally.

Second, we study close-to-goal passes that either create a big chance or lead to a shot on goal, which is 2.5% of all passes in our data. Column (3) in Table 6 reports the results for the number of close-to-goal passes instead of the number of passes. The point estimate of the homophily premium for the former (2.45%) is the same as for the latter (2.42%).

Third, we distinguish between simple pass sequences (in which player o passes to player d and the ball does not come back) and complex pass sequences (in which the ball goes back and forth between the two players at least once). On average, as already mentioned, player pairs make 15.98 passes per half-season. A vast majority of them (87%) consists of simple pass sequences, but 13% are complex pass sequences featuring 3.54 passes on average. Half of the player pairs are involved in at least one complex pass sequence in our sample.

A pass sequence is truly complex only if it is part of a forward movement as it involves a coordinated movement of the first passer to receive the ball back.[19] As we know the pitch coordinates of each pass, we can filter out sequences with backward passes and focus only on

Table 6. High Stakes in Passes

	Pass count			Pass sequence count	
Dependent variable	All (1)	All (2)	Near goal (3)	Complex (4)	Simple (5)
Same culture (any) (0/1)	0.0180***	0.0191***	0.0245***	0.0225***	0.0565***
	(0.0031)	(0.0029)	(0.0073)	(0.0031)	(0.0073)
Long pass (0/1)	−0.3219***	−0.0389***			
	(0.0036)	(0.0035)			
Same culture (any) (0/1) × Long-pass (0/1)	0.0149***	0.0120***			
	(0.0042)	(0.0039)			
Average forwardness Ind (0-1)	0.0594***	0.0815***	−0.3923***	1.968***	1.554***
	(0.0067)	(0.0061)	(0.0150)	(0.0065)	(0.0138)
Average length of passes (ln)		−0.7464***	0.5896***	−0.7862***	−1.499***
		(0.0061)	(0.0107)	(0.0055)	(0.0111)
Observations	668,105	668,105	479,987	649,210	542,172
Pseudo-R^2	0.75180	0.76082	0.40895	0.74971	0.39504
Passer by half-season fixed effects	✓	✓	✓	✓	✓
Receiver by half-season fixed effects	✓	✓	✓	✓	✓
Crossposition dummies	✓	✓	✓	✓	✓

Notes. Poisson regression. The dependent variable is pass count. The offset variable is included. Standard errors, clustered at different levels, are in parentheses. "*Long pass*" refers to passes with length above the median. "*High-stakes passes*" are those creating a chance for or preceding a shot on goal.

***$p < 0.01$.

the remaining ones. Columns (4) and (5) in Table 6 reveal a large difference in homophily premium; it is more than twice as large for complex pass sequences: 2.25% versus 5.65%. This confirms that homophily is especially important for more complex collaboration.

These results are complemented by our findings about players. More pressure may be felt by younger passers and passers targeting receivers of higher quality as in both cases, a passer might feel that he is more likely to be assessed critically for possible mishaps. As we show in Section 3 in the Online Appendix, the homophily premium is higher for young passers than for experienced ones and is unaffected by receiver quality.

Taken together, these results suggest that the homophily premium is unlikely because of favoritism because homophily is not less (and in some cases, is actually more) pronounced when stakes are high.

5.2. Minorities and Within-Group Passes

Relative group size may affect the homophily of a group's members (Blau 1977, Jackson et al. 2017). Being in a relatively small group may increase in-group favoritism (Porter and Washington 1993). In our case, favoritism should be more visible when the passer belongs to a small group.

To check whether this happens, we measure group size as the number of same culture receivers who a player faces each time he passes (taking the average over a half-season). This measure ranges between 1 (when, on average, no receiver has the same culture as the passer) and 11 (when, on average all receivers share his same culture). We then define an indicator variable $Het\_grsize_{o,t}$ that records whether the passer belongs to a small or large culture group. His group is "large" ($Het\_grsize_{o,t} = 1$) if its average size is larger than four and "small" ($Het\_grsize_{o,t} = 0$) otherwise. (Results are robust to alternative cutoffs.) In our sample, 66.6% of passes are started by passers who are part of large culture groups (i.e., on average, at least four players of their group are fielded).[20] We estimate our core regression (2), adding $Het\_grsize_{o,t}$ and an interaction term with the same culture variable.

The corresponding results are reported in column (1) in Table 7, which shows that against the minority hypothesis, the homophily premium is higher (by 1.52%) for players in large groups (3.25%) than for those in small groups (1.74%). Hence, the homophily premium is again unlikely because of favoritism.

5.3. Minorities and Between-Group Passes

Focusing on passes to receivers of different culture than the passer's, favoritism should promote passes to smaller groups, whereas cost saving should promote passes to larger groups (Currarini and Mengel 2016, Jackson et al. 2017). To investigate whether this is the case, we look at passes going from passer o to receiver d, recording whether o and d belong to small or large groups as defined before. We then rerun the fixed effects regression (2) controlling whether passes entail passing from a small group to a large group, from a large group to another large group, from a large group

Table 7. Pass Group Size and Large Group Dynamics

	(1)	(2)
Same culture (any) (0/1)	0.0173***	0.0339***
	(0.0043)	(0.0043)
Passer group: Large (0/1)	−0.0402***	
	(0.0069)	
Same culture (any) (0/1) × Passer group: Large (0/1)	0.0152**	
	(0.0059)	
Different culture pass: Small to large (0/1)		0.0239***
		(0.0065)
Different culture pass: Large to small (0/1)		−0.0053
		(0.0065)
Different culture pass: Large to large (0/1)		0.0123**
		(0.0054)
Different culture pass: Small to small (base)		0
		0
Observations	656,772	656,772
Pseudo-R^2	0.75838	0.75691
Passer by half-season fixed effects	✓	✓
Receiver by half-season fixed effects	✓	✓
Crossposition dummies	✓	✓

Notes. Poisson regression. The dependent variable is pass count. The offset variable is included. Standard errors, clustered at the passer by the half-season level and at the receiver by the half-season level, are in parentheses. In column (1), "*passer group: Large*" is one for groups of four players or more. In column (2), group size is based on the same culture groups: large with at least three players. The average length of passes and the average forward index are included.

**$p < 0.05$; ***$p < 0.01$.

to a small group, or from a small group to another small group. The last type of passes is taken as base so that all other variables (including the same culture dummy) are measured against it.

Column (2) in Table 7 reports the corresponding results, which show that passes are more likely from small to large groups (2.39%) than from large to large groups (1.23%). In turn, passes from large to large groups are more likely than those to small groups, no matter whether these originate from large (−0.52%) or small groups (0). (As passes from a small group to another small group are the base, its coefficient is zero.)

Hence, also according to the evidence on out-group passes, the homophily premium is unlikely because of favoritism. In addition, the higher frequency of passes to large culture groups is consistent with the idea that keeping the ball in those groups minimizes a team's passing frictions.

5.4. Shared Experience

The contact hypothesis in psychology suggests that prejudice and conflict between groups can be reduced if members of the groups interact with each other (Pettigrew and Tropp 2006). Common features of prejudice include having negative feelings and holding stereotyped beliefs about members of the group as well as a tendency to discriminate against them. Frequent instances involve prejudices based on characteristics, like race, sex, religion, and culture. Prejudice often supports in-group favoritism and out-group discrimination, which nonetheless, should be mitigated through intergroup contact.

To check whether this happens in our data, we study the evolution of the homophily premium over time as players repeatedly interact. In particular, let us consider a player new to a team; all teammates are initially new to him. Over time, new players will arrive, and there will be some teammates with whom he will have shared experiences and others who will be new to him. To measure shared experience, we use a threshold of 215 days (7 months), which is roughly equivalent to a half-season including summer (May to December/June to January) and very close to the median of the days spent together on a team. We define a passer and a receiver as having shared experience if they have spent at least 215 days together on their current team. We then consider only passers who have been on the team for more than 215 days and compare their passes with receivers with and without shared experience. This gives a 68% subsample of the original data, featuring 29.5% passing pairs without shared experience and 70.5% passing pairs with shared experience.

The corresponding results are shown in column (1) in Table 8, which reveals that homophily is more pronounced for pairs with shared experience. This goes against the contact hypothesis and suggests that favoritism based on prejudice is not what drives homophily.

Table 8. Homophily over Time: Shared Experience

	(1)	(2)	(3)	(4)	(5)
Same culture (any) (0/1)	0.0166*** (0.0053)	0.0163*** (0.0053)	0.2325 (0.2156)	0.0131* (0.0078)	0.0206*** (0.0050)
Same culture (any) (0/1) × Shared experience (0/1)	0.0117** (0.0059)	0.0127** (0.0060)	−0.1372 (0.1924)	0.0192** (0.0088)	
Same culture (any) (0/1) × Experience long (0/1)					0.0073 (0.0059)
Observations	457,825	443,626	13,532	219,169	384,807
Pseudo-R^2	0.76317	0.76431	0.83248	0.76578	0.76699
Early shared experience with other team	Include	Exclude	Only	Include	Include
Time with team capped	No	No	No	Yes	No
Passer by half-season fixed effects	✓	✓	✓	✓	✓
Receiver by half-season fixed effects	✓	✓	✓	✓	✓
Crossposition dummies	✓	✓	✓	✓	✓

Notes. Poisson regression. The dependent variable is pass count. The offset variable is included. Standard errors, clustered at the passer by the half-season level and at the receiver by the half-season level, are in parentheses. "*Shared experience*" is binary: one if the passer and the receiver have spent at least 215 days at the current team together. "*Shared experience long*" is 364 days or more. "*Early shared experience with other club*" is at past clubs, including youth teams. In all specifications, the average length of passes and the average forward index are included.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

We carry out two robustness tests. First, it is possible that some pairs have prior experience, having played together before in other teams, including youth teams. In columns (2) and (3) in Table 8, we then repeat the exercise for pairs without any previous shared experience and only previous shared experience, respectively. In the latter case, the sample is rather small, and having only previous shared experience has no significant effect. Second, in column (4) in Table 8, we constrain the data set to players in the first two years with their current team. This is a much smaller data set, with a higher share of new partnerships (40.5% versus 29.5%). Here, we find an even stronger relative role of shared experience (1.12% versus 1.91%).

Finally, we show that the time horizon for shared experience is fairly short. In column (5) in Table 8, we replace the 7-months threshold with 12 months, only to find that the (imprecisely) estimated effect of shared experience becomes smaller. This suggests that the impact of shared experience materializes fairly fast, although the coefficients in columns (1) and (5) in Table 8 cannot be told apart statistically.

Overall, we find evidence that homophily is higher for player pairs with shared experience on the team. This finding supports the idea that homophily is not driven by prejudice.

Taking stock, all of these findings lend more support to "cost saving" than "favoritism." The evidence is, however, not conclusive as one may argue against the hypothesis that high stakes put pressure on the passer, thus causing more instinctive decisions rather than less instinctive decisions. Minority passers may be less prone rather than more prone to favoritism if they fear the judgment of the majority, and they may be forced to yield to the favoritism of majorities by disproportionately passing to them.

5.5. Discussion

We conclude our analysis by reviewing some potential mechanisms behind the estimated homophily premium that have been highlighted in the literature in light of our findings.

5.5.1. Identity. People with strong national identity may want their compatriot to do well. This would be mostly a national issue and not a cultural one; that is, it would explain Brazilian but not Portuguese-Brazilian homophily. Our results showing how postcolonial links or other forms of cultural proximity matter do not support the national identity mechanism. Nevertheless, it may support a mechanism working through broader cultural identity.

5.5.2. Prejudice. People may have a false, overly confident belief in the ability of same culture peers and underestimate the ability of different culture ones. According to the contact hypothesis, the prejudice mechanism can be detected if homophily declines with shared experience. This not what we have found in the previous section as we actually see the opposite happening.

5.5.3. Salience. Being in a small group leads to salience of cultural affiliation as people are more likely to be aware of belonging to the same group when the group is small. This is not what we find when we look at a passer's group size. It is also inconsistent with the finding that intergroup passes favor receivers in large groups, even when the passer belongs to a small group.

5.5.4. Friendship. People may have no personal preference when they start collaborating but will build friendships over time. Friends outside work will then help each other in the workplace (i.e., on the pitch). Stronger collaboration may then ensue (on purpose or even involuntarily). Friendship may be easier to build with people from the same culture as they typically speak the same language, have the same social cues, like the same cuisine, listen to the same music, watch the same TV series or sports, and so on.[21] Friendship as a mechanism may then be detected if homophily increases with time spent together with teammates. This is consistent with our finding that shared experience amplifies the homophily premium.

6. Conclusions

We have investigated how homophily based on cultural traits affects collaboration in superstar multinational teams. In doing so, we have exploited a newly assembled exhaustive data set recording all passes by professional European football players in all teams competing in the top five men's leagues over eight sporting seasons together with full information on players' and teams' characteristics. The outcome that we have chosen as our measure of collaboration is the "pass rate" as passes represent how players work together for the common objective of scoring a goal and are positively correlated with team performance. Same culture is defined in terms of four traits: nationality, colonial legacy, federal legacy, and language.

We have shown that player pairs of same culture collaborate more, exhibiting an overall pass rate that is 6.19% higher than that for players of different culture. Separately identifying collaboration because of cultural preferences ("choice homophily") from collaboration because of opportunities ("induced homophily"), we have found strong evidence that choice homophily accounts for 2.42% of that 6.19%. Among the alternative traits, only same colonial legacy is as relevant as same nationality. To interpret this "homophily premium," consider the passes made by a player in a half-season, conditioning on constant and time-varying receiver characteristics, passer-receiver position pair, and other pass features. By choice, this player is expected to pass 2.42% more to teammates of same culture than to teammates of different culture.

The estimated homophily is consequential; passing to a receiver with same culture is as likely as passing to a receiver with different culture but valued a conspicuous 10.5% more. In other words, the passer sees the former as significantly more valuable than the latter.

We have found that homophily is more pronounced when stakes are high and players need to have more awareness of each other's movements as in the case of pass sequences that advance the ball toward the rival team's goal. Homophily also promotes inequality by favoring members of larger cultural groups as these receive a disproportionate share of passes. The estimated homophily could be because of either an objective cost for the team as passes are easier to coordinate within a group ("cost saving") or a subjective cost for the passer who prefers to keep the ball within his own group ("favoritism"). Our findings, although not exhaustive, point toward the cost-saving aspect. The most intriguing result is that homophily is stronger between players with more shared experience. Possibly spending time outside of games may build greater knowledge of each other.

We see at least two promising directions of future research to overcome some of the limitations that we have highlighted in our analysis. First, our assessment of the importance of the estimated homophily has relied on the calculation of a monetary equivalent value for the estimated homophily premium. Hence, the question of whether the estimated homophily premium makes a difference in terms of footballing outcomes has been only indirectly answered. A direct answer would require a full-fledged model of team performance and league competition. Although such a model is beyond the scope of this paper, we view our assessment of the importance of homophily as a first step in that direction.

Second, we found that shared experience on the team amplifies the homophily premium. We interpret it as suggestive evidence of off-pitch familiarity and socializing being mechanisms through which same culture leads to homophily. In future work, this conjecture could be tested by matching passing data with the players' social media profiles to gauge their reciprocal friendship ties.

Acknowledgments

The authors thank the editor, the associate editor, and three referees as well as Paola Conconi, Marc Kaufmann, Gábor Kézdi, Mats Koster, Miklós Koren, Balázs Kovács, Alice Kugler, Balázs Lengyel, Mike Luca, Glenn Magerman, Marco Molitor, Balázs Muraközy, Dennis Novy, Eric Snowberg, Ádám Szeidl, and participants at several seminars, workshops, and conferences for useful comments and suggestions. The authors are grateful to Endre Borza and Bence Szabó for outstanding and extensive research assistance.

Appendix. A Discrete Choice Model of Passing Behavior

In this appendix, we develop a discrete choice model that provides the theoretical foundation to the empirical approach used in the main text to disentangle choice from opportunity in an internally consistent way.

Here, we offer a streamlined presentation of the model, whereas Section 5 in the Online Appendix offers reports detailed derivations.

Consider a football team of $N = 11$ players, indexed from 1 to N, who are engaged in a half-season consisting of P team passes. During the half-season, each player is assigned to a particular position on the pitch, which implies that a player's index identifies both his name and his position. Focusing on two players, labeled o (a mnemonic for "origin") and d (a mnemonic for "destination"), a "pass" from o to d is defined as a movement of the ball determined by a decision made by player o ("passer") to kick or throw the ball to teammate d ("receiver"). For $d = o$, the passer keeps possession of the ball.

In the half-season, let P^o be the number of passes made by player o to his teammates, P^d be the number of passes received by player d from his teammates, and $P^{o,d}$ be the number of passes made by player o to player d such that we have $P^o = \sum_{d=1}^N P^{o,d}$, $P^d = \sum_{o=1}^N P^{o,d}$, and $P = \sum_{o=1}^N \sum_{d=1}^N P^{o,d}$. Moreover, let $T^{o,d}$ be the number of passes made by player o when player d is on the pitch (i.e., player d is in the passer's choice set) and $\tau^{o,d} \in [0,1]$ be the share of those passes made to player d such that $T^{o,d} = P^o \tau^{o,d}$ holds. Finally, let $\pi^{o,d} \in [0,1]$ be the share of passes to player d in the total number of passes that player o makes when teammate d is on the pitch such that $P^{o,d} = T^{o,d}\pi^{o,d}$ holds. Based on these definitions, we can express the number of passes made by o to d as

$$P^{o,d} = P^o \tau^{o,d} \pi^{o,d}. \tag{A.1}$$

We are interested in characterizing $\pi^{o,d}$ in terms of the probability that player o passes to player d rather than to any of the other teammates when player d is a viable option. We assume that player o wants to maximize team payoff and understands that the benefit for the team of one of its players controlling the ball is determined by the characteristics of that player and by some randomness because of the vagaries of the game. A player's characteristics may include, for example, quality and experience as these affect what he can do with the ball. A game's vagaries may include, for instance, the performance of the opposing team, the referee's decisions, or the weather conditions. We use U^d to denote the deterministic part of the team's benefit as determined by player d's characteristics and z^d to denote the realization of its random part ("shock") because of match contingencies.

Player o also understands the challenges that he faces in passing the ball to receiver d, and we use $\tilde{c}^{o,d}$ to denote the "passing friction" capturing the cost associated with tackling those challenges. This cost may be objectively associated with the physical effort of passing or the mental effort of anticipating the receiver's moves. In this case, if the mental effort of passing to a teammate with similar cultural traits was lower, any choice homophily would be efficient for both the passer and the team as it would be because of objective constraints. Alternatively, the cost may subjectively derive from the passer's in-group favoritism. In this case, the passer's choice homophily would be efficient for him but inefficient for the team as his passes would deviate from what is objectively good for the latter.

Lastly, player o is aware of the difficulty that receiver d may face in taking control of the ball, which depends on the receiver's circumstances. We use $\varphi^d \in [0,1]$ to denote the probability that receiver d takes control of the ball, and we call it the probability of a successful pass.

Apart from passing or keeping the ball, player o may decide to do something else with the ball generating team benefit U^o. For example, he may try to score a goal or decide to kick the ball out of play to allow his team to reorganize.

We use $\beta \in [0,1]$ to denote the relative importance that the team attaches to passing in general, independently of the specific passing episode. This is an important characteristic of the team's style of play. For example, low β would be associated with teams that try to score goals by quickly moving the ball into scoring range by long passes (through long balls or long air balls), whereas high β would refer to teams that prefer to play less quickly, using many short passes (also sideways or backward) to find a weakness in the opposing team's tactics. Clearly, the weight given to U^o affects the passer's decision between passing or keeping the ball and doing something else with the ball, but it is immaterial for his choice among alternative receivers.

As a result, player o's passing decision is determined by the comparison of team utilities $U^o + \beta\varphi^d U^d - \tilde{c}^{o,d} + z^d$ across all potential receivers $d = 1, \ldots, N$. However, given that z^d is a random shock, the outcome of this decision is an array of probabilities $\pi^{o,d}$ of passing to each potential receiver (including the passer himself). These probabilities can be readily characterized under appropriate assumptions on the probability distribution of the shocks. In particular, we make a customary assumption in the discrete choice literature that z^d is the realization of a random variable Z following a Gumbel distribution with zero mode and concentration around the mode positively related to $\kappa > 0$.[22] Zero mode implies that there is no systematic deviation from the deterministic part of the team's benefit across players' assessments of match contingencies. As all players share the same κ, this is a team characteristic; players are trained to assess match contingencies in a common way. Larger κ can then interpreted as resulting from more intense training to reduce variation in their individual assessments.

Under the chosen distributional assumption, the model predicts that the probability that player o passes to player d when the latter is on the pitch evaluates to

$$\pi^{o,d} = \frac{(c^{o,d})^{-\kappa} P^d}{(\Lambda^o)^\kappa (\Lambda^d)^\kappa} \tag{A.2}$$

so that by (A.1), the number of passes made by player o to player d in a half-season is

$$P^{o,d} = \frac{P^o}{(\Lambda^o)^\kappa} (c^{o,d})^{-\kappa} \tau^{o,d} \frac{P^d}{(\Lambda^d)^\kappa} \tag{A.3}$$

with definitions

$$\Lambda^o = \left[\sum_{d=1}^N \frac{(c^{o,d})^{-\kappa} P^d}{(\Lambda^d)^\kappa}\right]^{\frac{1}{\kappa}} \text{ and } \Lambda^d = \left[\sum_{o=1}^N \frac{(c^{o,d})^{-\kappa} P^o}{(\Lambda^o)^\kappa}\right]^{\frac{1}{\kappa}}$$

for $c^{o,d} = \exp \tilde{c}^{o,d}$.

Expression (A.3) has clear implications once Λ^o and Λ^d are given intuitive interpretations. The former is a weighted geometric average of the passes received by all players in the team, with weights determined by their bilateral passing frictions with respect to passer o. It is, therefore, a compact index that measures the passer's multilateral access to teammates considering both the overall number of passes that they attract and the difficulty to reach them. Given that in this optimizing framework, the number of passes that a player attracts depends on the benefit that he also generates for the team through his own future passes, Λ^o captures the average team benefit of passer o's forward-looking options. Analogously, Λ^d is a weighted geometric average of the passes made by all players in the team with weights determined by their bilateral passing frictions with respect to receiver d. It is thus a compact index measuring the receiver's multilateral accessibility by teammates considering both their characteristics and the difficulty to be reached by them. As the number of passes that a player makes depends on the benefit that he generates for the team, Λ^d captures the forward-looking average team benefit of all passes that player d may receive.

The first implication from (A.3) is that the number of passes $P^{o,d}$ from player o to teammate d increases with the number of passes $T^{o,d} = P^o \tau^{o,d}$ made by player o when player d is on the pitch. This is beyond players' control as it reflects their coach's selection decisions. The second implication is that the number of passes $P^{o,d}$ from player o to teammate d increases with the overall number of passes P^o made by the former and the overall number of passes P^d received by the latter, where the number of passes that a player makes or receives depends on the benefit that he generates for the team. The third implication is that the number of passes from player o to teammate d decreases with the passing friction $c^{o,d}$ between them. Although these three implications are in common with the naive approach, a fourth crucial implication is unique to the structural approach; the number of passes from player o to teammate d is a decreasing function of the average team benefit of passer o's options (Λ^o) and the average team benefit of passes that player d receives (Λ^d). In other words, there are more passes between any two players when these have less attractive alternatives to pass or receive, which depends on the characteristics (including the positions) of all teammates on the pitch. Neglecting these multilateral variables, as the naive approach does, would lead to biased homophily estimation.[23]

The distinction between distance and culture-related challenges in passing from player o to player d embedded in $c^{o,d}$ can be made explicit by specifying the bilateral passing friction multiplicatively as

$$c^{o,d} = e^{\varepsilon^{o,d}} (g^{o,d})^{\gamma} (l^{o,d})^{\lambda}, \tag{A.4}$$

where $\varepsilon^{o,d}$ is an error term allowing for imprecise measurement of the bilateral friction. In this expression, $g^{o,d}$ is the physical distance between the two players' positions so that $(g^{o,d})^{\gamma}$ captures all distance-related frictions that make it hard to pass the ball from passer o to receiver d independently of their identities. The term $(l^{o,d})^{\lambda}$ captures, instead, all nondistance-related frictions that make it hard to pass the ball from passer o to receiver d independently of their positions. These may include, for instance, limited experience in playing together but crucially, also different cultural traits.

We are now ready to translate (A.3) into an estimating equation. Specifically, we define the half-season "pass rate" $p^{o.d}$ as the ratio $P^{o,d}/P$ of the number of passes from player o to teammate d over the total number of team passes. Then, substituting (A.4) into (A.3) and taking logs give

$$\log p^{o,d} = \log \tau^{o,d} + \log P^o (\Lambda^o)^{-\kappa} + \ln P^d (\Lambda^d)^{-\kappa} - \kappa\gamma \log g^{o,d} - \kappa\lambda \log l^{o,d} - \log P + \varepsilon^{o,d}, \tag{A.5}$$

which we will use in the Online Appendix as the theoretical basis to empirically investigate the relation between the pass rate $p^{o,d}$ and the cultural dimensions of $l^{o,d}$. Before proceeding, three remarks are in order. First, Equation (A.5) distinguishes the role of homophilous preferences ("passing to teammates"), which work through $\log l^{o,d}$, from the implications of homophilous meeting rates ("passing to teammates through teammates"), which work through the forward-looking terms Λ^o and Λ^d. Second, such distinction allows us to argue that in (A.5), the cultural dimensions of $l^{o,d}$ determine choice homophily, whereas induced homophily is determined by all other terms on the right-hand side of (A.5). Third, the pass rate in Equation (A.5) is conditional on players being together on the pitch, and homophily may play a role in the selection of fielded players. As long as this affects induced homophily, accounting for $\tau^{o,d}$ allows us to net it out.

Endnotes

[1] For example, McPherson et al. (2001) look at homophily in a sociological perspective; Lawrence and Shah (2020) and Ertug et al. (2021) emphasize the management viewpoint, whereas Jackson et al. (2017) discuss homophily from the perspective of economics with a focus on social network. See Békés and Ottaviano (2022) for more discussion of the related literature.

[2] Section 4 in the Online Appendix shows that on an aggregated data set at the level of teams and half-seasons, teams that pass 10% more will get 3%–9% more points on average (depending on specification).

[3] We also test alternative operationalizations based on colonial legacy (past membership of a colonial empire), federal legacy (past membership of a political union), native language, linguistic and geographical proximity, and shared values.

[4] Similar considerations on the appeal of team sports data can be found in, for example, Nüesch and Haas (2013), Ingersoll et al. (2017), and Tovar (2020) for European football or Kahane et al. (2013) for North American ice hockey. What distinguishes our analysis from these and related works is that we zoom in on homophilous collaboration, which we can measure accurately through the pass data.

[5] For additional details, see Section 1 in the Online Appendix.

[6] Replication codes and data for the empirical analysis are available at https://doi.org/10.1287/mnsc.2022.01799.

[7] Data quality and coverage are both very high in our data sets. Nevertheless, a few small data cleaning steps were needed, and we discuss them in Section 2 in the Online Appendix.

[8] Nationals of the same country are more likely to share a common heritage, covering not only inherited traditions, monuments, memories, and objects but also, contemporary institutions, activities, meanings, and behaviors drawn from them. Common language and shared history because of bygone political ties may allow individuals of different nationality to share at least some aspects of such a common heritage (Bobowik et al. 2018).

[9] Centre d'études prospectives et d'informations internationales (CEPII), https://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele.asp.

[10] This interpretation of individual fixed effects is analogous to that of country fixed effects in the gravity equations used to model

bilateral flows within international trade networks, where each country has several potential trade partners (Head and Mayer 2014).

[11] Specifically, we follow the procedure described in Berge (2018). As discussed by Hinz et al. (2021), a drawback of fixed effect models in general is the incidental parameter bias; having several nuisance parameters to estimate, the estimated coefficient of the variable of interest may be biased. FE-PPML estimates can deal with this type of bias better than nonlinear OLS (Santos-Silva and Tenreyro 2022). Although Weidner and Zylkin (2021) show that the Poisson model still leaves some room for potential bias, with no double-player fixed effects and a large number of observations, the bias should be small in our case.

[12] We cannot have passer by receiver fixed effects because our variable of interest is time-invariant.

[13] Alternatively, one could run team by half-season regressions. However, as we have passer by half-season fixed effects together with receiver by half-season fixed effects, our estimated impact of homophily would be very close to the mean estimate over the (1,484) team by half-season regressions. Without additional controls, using OLS and weighting by pass count, the two estimates would be exactly the same. The advantage of relying on a single regression is that we can better estimate the role played by the additional controls and have a single standard error.

[14] The exact form is based on the model described in the appendix.

[15] Results are robust to using a longer period or the log number of days spent together instead.

[16] For details, see Section 2 in the Online Appendix. In our sample, 20% of player pairs have at least one restricted player.

[17] For additional details on loans, see Section 1 in the Online Appendix.

[18] The monetary equivalent value of the homophily premium is calculated by equating $E(pass\_count^{H}_{o,d,t}|\ldots)$ to $E(pass\_count^{V}_{o,d,t}|\ldots)$. The former is the predicted $E(pass\_count_{o,d,t}|\ldots)$ using the receiver's actual value ($value_{d,t}$) and the estimated homophily premium ($\delta > 0$) from regression (4). The latter is the counterfactual $E(pass\_count_{o,d,t}|\ldots)$ computed after setting the homophily premium to zero ($\delta = 0$) and the receiver's value to $value'_{d,t}$ such that $E(pass\_count^{H}_{o,d,t}|\ldots)/E(pass\_count^{V}_{o,d,t}|\ldots) = \exp(\delta + \eta_2(value_{d,t} - value'_{d,t})) = 1$. Solving this equation gives the monetary equivalent value of the homophily premium as $value'_{d,t} - value_{d,t} = \delta/\eta_2$, with δ and η_2 estimated from regression (4). Here $0.0435/(\exp(0.3457) - 1) = 10.53$.

[19] We thank an anonymous referee for making this point. Results are robust to including all pass sequences, see Section 3 in the Online Appendix.

[20] For details, see Section 3 in the Online Appendix.

[21] We are grateful to an anonymous referee for flagging this mechanism. Interestingly, Kovacs and Kleinbaum (2020) find that even similar linguistic style will lead to friendship formation in college students.

[22] The cumulative density function of this Gumbel (or type I extreme value) distribution is $G(z) = \exp(-\exp(-\kappa z))$ for $z \in (-\infty, +\infty)$.

[23] Another unique implication of the structural approach is that the rate at which the number of passes from player o to teammate d decreases with the bilateral passing friction is higher for larger κ; when the variation in players' assessments of match contingencies is smaller, differences in bilateral frictions become more salient. We do not explore this interesting implication as it goes beyond the scope of the present paper.

References

Andrews M, Gill L, Schank T, Upward R (2012) High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias. *Econom. Lett.* 117(3):824–827.

Békés G, Ottaviano G (2022) Cultural homophily and collaboration in superstar teams. CEPR Discussion Paper No. 17618, CEPR, London, UK.

Berge L (2018) Efficient estimation of maximum likelihood models with multiple fixed-effects: The R package FENmlm. CREA Discussion Paper 2018-13, University of Luxembourg, Luxembourg.

Blau PM (1977) *Inequality and Heterogeneity: A Primitive Theory of Social Structure* (Free Press, New York).

Bobowik M, Valentim JP, Licata L (2018) Introduction to the special issue: Colonial past and intercultural relations. *Internat. J. Intercultural Relations* 62:1–12.

Burt R (1992) *Structural Holes: The Social Structure of Competition* (Harvard University Press, Cambridge, MA).

Castilla EJ (2011) Bringing managers back in: Managerial influences on workplace inequality. *Amer. Sociol. Rev.* 76(5):667–694.

Cross R, Cummings J (2004) Tie and networks correlates of individual performance in knowledge-intensive work. *Acad. Management J.* 47(6):928–937.

Currarini S, Mengel F (2016) Identity, homophily and in-group bias. *Eur. Econom. Rev.* 90:40–55.

Desmet K, Ortuño-Ortín I, Wacziarg R (2017) Culture, ethnicity, and diversity. *Amer. Econom. Rev.* 107(9):2479–2513.

Ertug G, Brennecke J, Kovacs B, Zou T (2021) What does homophily do? A review of the consequences of homophily. *Acad. Management Ann.* 16(1):38–69.

Ertug G, Gargiulo M, Galunic C, Zou T (2018) Homophily and individual performance. *Organ. Sci.* 29(5):912–930.

Fally T (2015) Structural gravity and fixed effects. *J. Internat. Econom.* 97(1):76–85.

Head K, Mayer T (2014) Gravity equations: Workhorse, toolkit, and cookbook. Gopinath G, Helpman E, Rogoff K, eds. *Handbook of International Economics* (Elsevier, Amsterdam), 131–195.

Hinz J, Stammann A, Wanner J (2021) State dependence and unobserved heterogeneity in the extensive margin of trade. Preprint, submitted July 27, https://arxiv.org/abs/2004.12655v2.

Horwitz SK, Horwitz IB (2007) The effects of team diversity on team outcomes: A meta-analytic review of team demography. *J. Management* 33(6):987–1015.

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

Jackson SE, Joshi A, Erhardt NL (2003) Recent research on team and organizational diversity: SWOT analysis and implications. *J. Management* 29(6):801–830.

Jackson MO, Rogers BW, Zenou Y (2017) The economic consequences of social-network structure. *J. Econom. Literature* 55(1): 49–95.

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

Kovacs B, Kleinbaum AM (2020) Language-style similarity and social networks. *Psych. Sci.* 31(2):202–213.

Lawrence BS, Shah NP (2020) Homophily: Measures and meaning. *Acad. Management Ann.* 14(2):513–597.

McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annual Rev. Sociol.* 27(1): 415–444.

Melitz J, Toubal F (2014) Native language, spoken language, translation and trade. *J. Internat. Econom.* 93(2):351–363.

Nüesch S, Haas H (2013) Are multinational teams more successful? *Internat. J. Human Resource Management* 23(15):3105–3115.

Opper S, Nee V, Brehm S (2015) Homophily in the career mobility of China's political elite. *Soc. Sci. Res.* 54:332–352.

Pettigrew TF, Tropp LR (2006) A meta-analytic test of intergroup contact theory. *J. Personality Soc. Psych.* 90(5):751–783.

Porter J, Washington R (1993) Minority identity and self-esteem. *Annual Rev. Sociol.* 19:139–161.

Price J, Wolfers J (2010) Racial discrimination among NBA referees. *Quart. J. Econom.* 125(4):1859–1887.

Santos-Silva J, Tenreyro S (2022) The log of gravity at 15. *Portuguese Econom. J.* 21(3):423–437.

Sorenson O, Stuart T (2008) Bringing the context back in: Settings and the search for syndicate partners in venture capital investment networks. *Admin. Sci. Quart.* 53(2):266–294.

Sorenson O, Stuart TE (2001) Syndication networks and the spatial distribution of venture capital investments. *Amer. J. Sociol.* 106(6): 1546–1588.

Spolaore E, Wacziarg R (2016) Ancestry, language and culture. Ginsburgh V, Weber S, eds. *The Palgrave Handbook of Economics and Language* (Palgrave Macmillan, London), 174–211.

Tovar J (2020) Performance, diversity and national identity evidence from association football. *Econom. Inquiry* 58(2): 897–916.

Weidner M, Zylkin T (2021) Bias and consistency in three-way gravity models. *J. Internat. Econom.* 132:103513.