Cultural Homophily and Collaboration in Superstar Teams

Gábor Békés* and Gianmarco I.P. Ottaviano**

* Central European University, KRTK and CEPR
** Bocconi University, Baffi-CAREFIN, IGIER, CEP and CEPR

NBER Workshop, Cambridge, MA, April 2023

Cultural Homophily and Collaboration in Superstar Teams

- Globalization - mix best global expertise in multinational teams
- Key aspect of multinationality is 'cultural diversity':
  - Benefits: talent, learning and innovation ('capabilities')
  - Costs: communication, empathy and trust ('collaboration')
- Does 'homophily' (i.e. tendency to associate with similar others) bias collaboration (i.e. ability to work for a common purpose)
  - even in superstar teams?
- Hard nut to crack:
  - Collaboration not observed directly
  - 'homophily bias' confounded

Induced vs. Choice Homophily

- Homophily = Opportunities ('induced') + Preferences ('choice')
- Opportunities vs. Preferences:
  - Distribution may mechanically determine probability of association
  - This confounds preference / choice
- Need to partial out 'induced' homophily to measure 'choice' homophily:
  - Option A: experiment with random team formation
    - Issue: Low external validity for highly skilled, lowly charged multinational workplace
  - Option B: observational data with adequate baseline
    - Issue: relevant (counterfactual) baseline

European Football as an allegory

- Teams: pro football clubs from the top-5 European leagues
  - Superstar team = global elite, top 5% of pro players
- New data: 5 countries, 11 million passes

- Collaboration: pass rate between player pairs
  - Team (squad) composition is exogenous to players
  - Collaboration is an individual choice
- Homophily = passer and receivers who share culture (nationality, history)

When I say Football, I mean Soccer

Related literature

- Cost and benefits of diversity in multicultural teams (seminal): Lazear (1999) Lang (1986)
- Cost and benefits of diversity in broader environments (cities, plants): Ottaviano and Peri (2006, 2005) Buchholz (2021)
- Cost and benefits of diversity in multicultural teams (recent developments):
  - Ethnic conflict: Hjort (2014), Laurentsyeva (2019),
  - Team formation: Calder-Wang et al. (2021)
  - Hockey: Kahane et al. (2013), Football: Nüesch and Haas (2013), Tovar (2020)
- Homophily in scientific publications: Freeman and Huang (2015), AlShebli et al. (2018)
- Homophily in friendship networks: Currarini et al. (2009, 2010)
- Literature review from psychology to management: (Lawrence and Shah, 2020; Ertug et al., 2021)

Contribution

1. Focus on everyday workplace collaboration - high skilled, lowly charged context in a developed area with no real conflicts.

2. Well defined measure of collaboration at individual level through time (not rare pair formation)

3. Careful model of baseline, both theory and empirics (go beyond randomization)

4. Rich measures of individual characteristics / large data–> FE

5. Very large, global sample - external validity

Data Collection and Definitions

Data: Scope

- 5 top leagues (France, Germany, Spain, Italy, England),
- 8 seasons (2011/12-2018/19) every teams play with every other twice
  - 20 (18) teams per league, 14,608 games in total
  - 730 passes/game
- 154 teams, each with 25-30 strong squad, regular churning (twice a year)
- 10.7 million passes ('events')
- 7,000 players from 138 countries

Raw Data: Events

- Event data – 'play by play'
  - Structured text, events with features, qualifiers:
- Separately recorded with a timestamp
  - Pass between any two players
  - Web-scraped from a `whoscored.com` website
  - Events recorded by cameras+algorithms+humans.
- Pass events separated

Raw Data: Players

- Player characteristics:
  - Nationalities (possible multiple)
  - Position in team
  - Age, height
  - Player valuations – over time
  - Web-scraped from a `transfermarkt.com` website
- Entity resolutions / coreference (accents, middle names, nicknames):
  - Matching algorithm by motifs

Measuring Cultural Homophily

- Characterize cultural background ('culture') = set of cultural traits transmitted across generations:
  - Such as language, history, norms, values and attitudes learned at home
- We measure 'culture' with four proxies:
  - Nationality, colonial legacy, federal legacy, language only
    - Alternative: linguistic similarity
    - Not alternative: Values (WVS)
- 'cultural homophily' = more intense collaboration between player pairs with same culture

Same Culture Definition

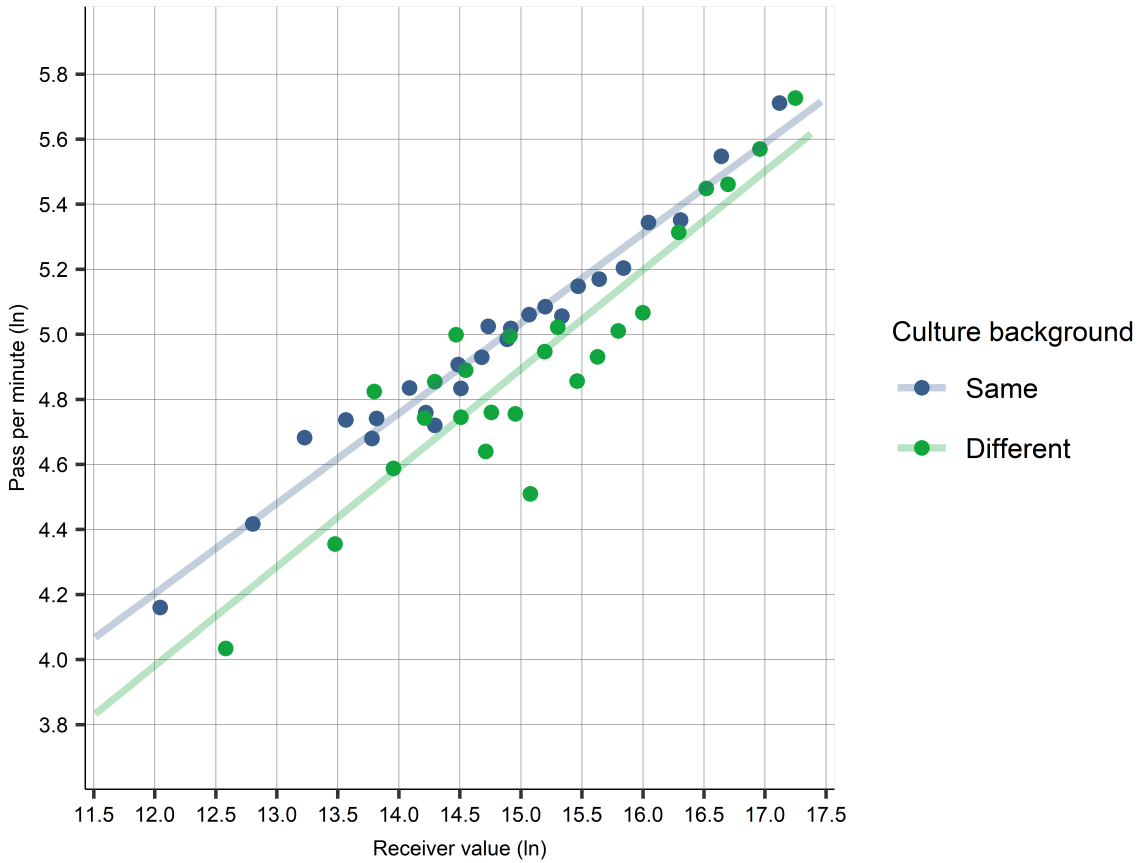
- Same nationality (citizenship)
- Same colonial legacy – different nationality
  - Argentina-Spain, England-Egypt (ruler and colony)
  - Uruguay-Argentina (colony siblings)
- Same federal legacy – different nationality
  - Russia-Georgia, Croatia-Serbia
  - Scotland, Northern Ireland, Ireland
- Same language – different nationality, colonial /federal legacy
  - Switzerland and Germany
  - DR Congo and France

Model (ideas)

Model: Setup

- Model to disentangle choice from opportunity in an internally consistent way
- Model team's utility
- Player o's passing decision is determined by the comparison of team utilities across all potential receivers $d = 1, ..., N$.
- Homophily = shifter leading to more passes between player pairs of similar culture after controlling for variables based on the model.

Simple case = Pass rate = f(receiver value and homophily)



Passers = Spanish midfielders in La Liga, N=24,299.

Model: Passer's Decision

Player o's passing decision is determined by the comparison of team utilities across all potential receivers $d = 1, ..., N$.

$$U^o + \beta \varphi^d U^d - \widetilde{c}^{o,d} + z^d$$

- U^o = team benefit from player o with the ball
- U^d = deterministic part by player d's characteristics
- z^d realization of its random part ('shock') due to match contingencies.
- φ^d = probability of successful pass to receiver d
- β = relative importance the team attaches to passing in general ('style')
- $\widetilde{c}^{o,d}$ = 'passing cost'

Model: Forward looking dynamic model

- Dynamic model
  - Passer takes into account future state of the ball

- 'pass rate' $p^{o,d}$ as the ratio number of passes from player o to teammate d over the total number of team passes.
  - Passer characteristics including team mates fielded with him
  - Receiver characteristics including team mates fielded with him
  - Position of players and passes
  - Time spent together when passer has the ball
- Passing cost ($\widetilde{c}^{o,d}$) includes a binary same culture indicator = measure of homophily
  - Also: distance between players, forwardness

Estimation: Poisson PML + FEs

- Poisson model
  - PPML advantages w/ FEs as in gravity models
- 'pass rate' $p^{o \cdot d}$ as the ratio number of passes from player o to teammate d over the total number of team passes.
  - Passer characteristics – FE (half-season)
  - Receiver characteristics – FE (half-season)
  - Homophily measures – binary
  - Position of players and passes
  - Time spent together when passer has the ball
- Aggregate probabilities to relative frequency
- Half-season level (16-20 games)

Poisson model with double player fixed effects

$$E(pass\_count_{o,d,t}|.) = exp(\delta SameCult_{o,d} + PassF_{o,d,t} + \ln tau_{o,d,t} + \upsilon_{o,t} + \upsilon$$

- Homophily: $SameCult_{o,d}$ as the same culture indicator (0/1).
- Offset time spent together (τ)
  - Decision of the manager
- FE: passer*half-season + receiver*half-season
- $PassF$ = Football related – pass friction
  - distance, forwardness
  - cross position dummies

Estimation: role of fixed effects

- In estimation, use double player (*half-season) fixed effects
- Unobserved player characteristics
- Alternatives the passer faces in terms of receivers
  - Akin to multilateral resistance term in structural gravity

Results

Results

- Data: 5 leagues, 98 team/season, 8 seasons
  - Aggregated at half-seasons (16-20 games)
  - N=669K
- Poisson models
  - Passer*half-season and receiver*half-season FE
    - FEs soak up team*half-season dummies
  - Exposure variable is time spent together
  - Includes pass frictions (distance, forwardness index)
  - Standard errors clustered P*h-s, R*h-s

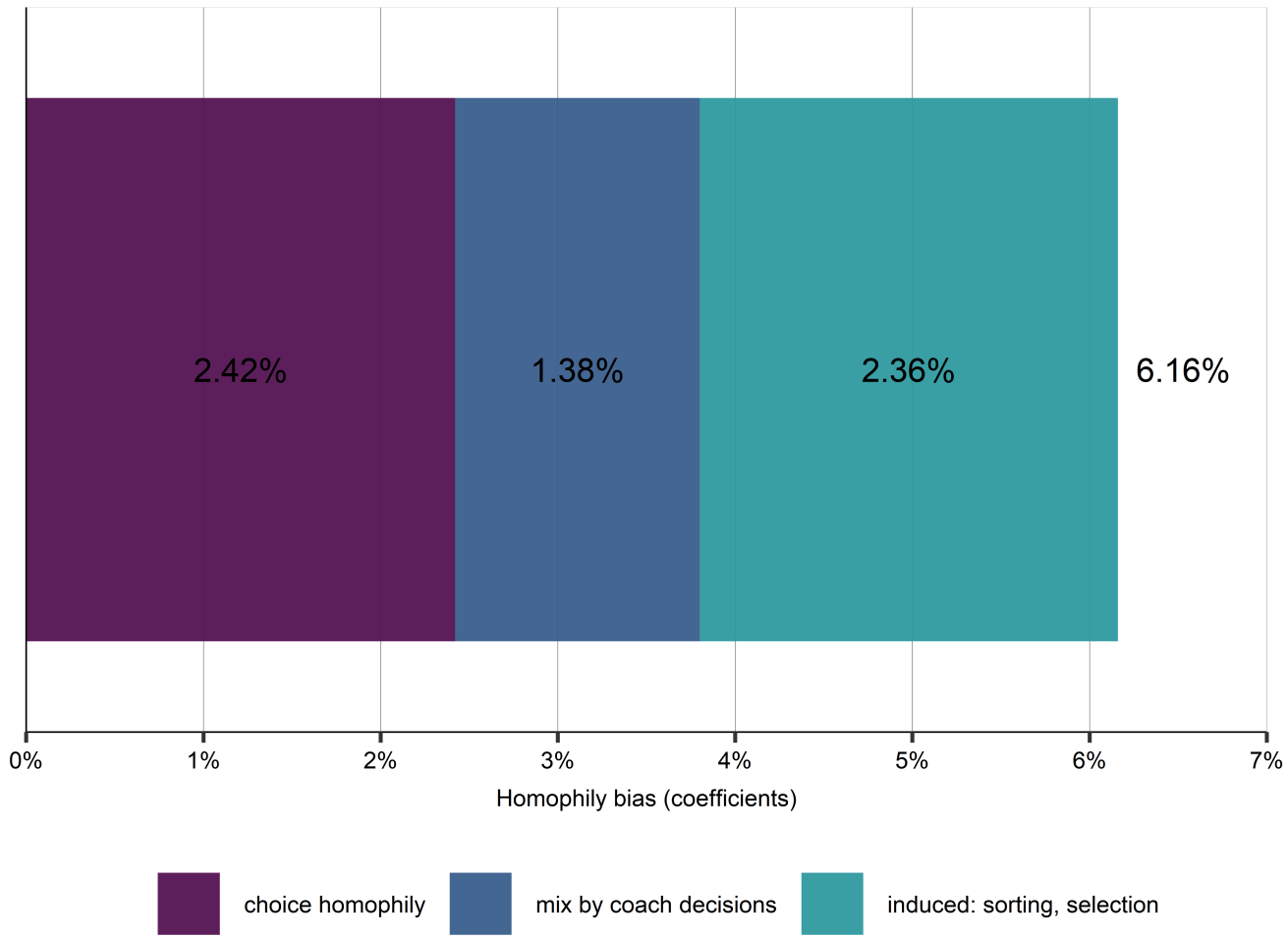
Result discussion

- Core result: *choice* homophily premium: 2.4%
  - Consider a team in half-season. Partialling out pass frictions and receiver characteristics, a player will pass 2.4% more to a same culture peer.

Result discussion

- Core result: *choice* homophily premium: 2.4%
  - Consider a team in half-season. Partialling out pass frictions and receiver characteristics, a player will pass 2.4% more to a same culture peer.
- passing to a same culture receiver is equally likely as passing to a different culture player valued a 10.5% more.
  - using transfer price estimations

Dissecting total homophily



Core results + robustness

- Core result: *choice* homophily premium: 2.4%
- Taking into account managers decision to field players: 3.8%
- (Unconditional) Same culture players tend to pass 6.2% more compared to different culture players

Core results + robustness

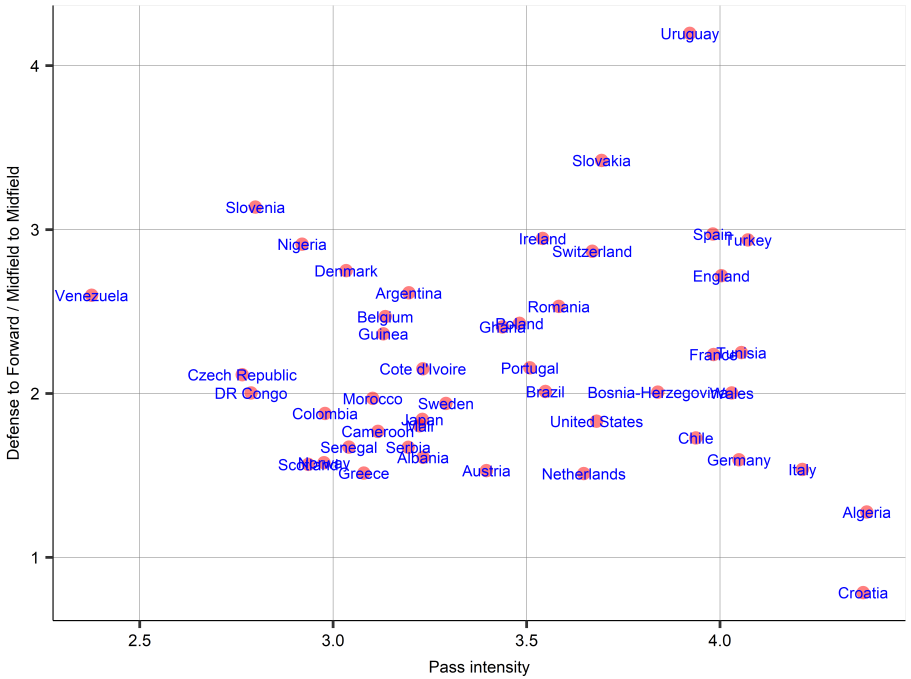
- Core result: *choice* homophily premium: 2.4%
- Taking into account managers decision to field players: 3.8%
- (Unconditional) Same culture players tend to pass 6.2% more compared to different culture players

- Robust to a variety of specifications
  - Physical differences
  - Assortative matching
  - Experience with club
  - Prior experience in youth club, other teams

  - Functional form specifications, such as ln(count)

Homophily is not common knowledge

- Players from different countries do pass differently
- May be French players learn a French "national football style"
- No



Additional facts

- Homophily is more important for complex collaboration
  - Look at pass sequences only, homophily premium is 4.8% vs 2% for single passes.
- Homophily is present for shared nationality as well as colonial links
  - It is negative for federal legacy (ie USSR, Yugoslavia)
- Alternative measure of culture: shared language, similar language works but weaker
- Shared values (World Value Survey) no correlation at all

Dissecting culture

Dep. var: pass count	(1)	(2)	(3)	(4)
Same nationality (0/1)	0.0284***	0.0302***	0.0315***	0.0186***
	(0.0030)	(0.0031)	(0.0031)	(0.0035)
Same colonial legacy (0/1)	0.0284***			
	(0.0041)			
Same federal legacy (0/1)	-0.0223**			
	(0.0106)			
Just shared language (0/1)	-0.0046			
	(0.0070)			
LC: diff country, same language (0/1)		0.0156***	0.0140***	
		(0.0039)	(0.0040)	
LC: diff country, similar language (0/1)		0.0111**	0.0094*	
		(0.0044)	(0.0045)	
Geographical proximity (neighbors) (0/1)			0.0064*	
			(0.0031)	
WVS: similar values (0/1)				-0.0064**
				(0.0029)
Observations	668,105	668,105	668,105	668,105
Pseudo R^2	0.76078	0.76077	0.76077	0.76076
passer-half_season fixed effects	✓	✓	✓	✓
receiver-half_season fixed effects	✓	✓	✓	✓
Cross position dummies	✓	✓	✓	✓

Heterogeneity

- We see more of a homophily premium
  - Young players
  - Passers in larger culture groups
- No difference
  - Receiver quality

Heterogeneity by age, group size, receiver quality

Dep.var: Pass count	(1)	(2)	(3)
Same culture (any) (0/1)	0.0319***	0.0174***	0.0236***
	(0.0045)	(0.0043)	(0.0027)
Same culture (any) (0/1) × Passer age (0/1, 1=Experienced)	-0.0096**		
	(0.0048)		
Same culture (any) (0/1) × Passer group size (1/1, 1 when N>=4)		0.0146***	
		(0.0059)	
Same culture (any) (0/1) × Receiver quality (0/1, 1= top 2)			0.0044
			(0.0057)
Passer group size (1/1, 1 when N>=4)		-0.0444***	
		(0.0075)	
Receiver quality (0/1, 1= top 2)			0.0129
			(0.0081)
Observations	668,105	668,105	668,105
Pseudo R^2	0.75930	0.74510	0.76077
passer-half_season fixed effects	✓	✓	✓
receiver-half_season fixed effects	✓	✓	✓
passer * receiver position dummies	✓	✓	✓

Investigating the mechanism

Mechanisms 1 – Cost vs bias

- Till now: agnostic if the existence of choice homophily represents
  - an efficient outcome promoting team performance,
  - or rather a manifestation of inefficient in-group favoritism detrimental to the team.
- No silver bullet but two arguments to support efficiency
  - Performance vs diversity shows a noisy zero, and has many problems
- Two suggestive evidence against favoritism
  - Beyond homophily, players tend pass more to large same culture groups
  - No show of lower homophily premium when under pressure

Mechanisms 1 – Cost vs bias 1

- Let's focus on passes to different culture players
- Divide receivers into small (<3) or large (>3) groups

- Group size premium (different culture passes)
  - Homophily premium here is 3.6%

	to small	to large
from small	0	2.8%*
from large	-0.6%	1.8%*

- Beyond homophily, players tend pass more to large same culture groups
  - Account for future benefits
  - Supports efficiency argument (not favoritism)

Mechanisms 1 – Cost vs bias 2

- Do players exhibit less homophily under pressure?
- Consider key passes – 2-3 passes before shot on goal
  - Really important passes
  - Under pressure from defenders
  - Sample is different = forwards and midfielders
- Homophily is unchanged

Mechanisms 2 – Motivation of players

- **What** makes same culture players find it easier to work together?
- Players of same culture being able to
  - co-operate better
  - understand each other better,
  - see each other better on the pitch
- If so, does it go away once they get to know each other?

Mechanisms 2 – Motivation of players

- Look at the evolution of homophily premium **over time**
  - Take a subset of players with some experience
  - Divide receivers into newbee vs experienced groups
  - Cutoff: median time of 7 months
  - Compare homophily premium across groups

Mechanisms 2 – Motivation of players

- Look at the evolution of homophily premium **over time**
  - Take a subset of players with some experience
  - Divide receivers into newbee vs experienced groups
  - Cutoff: median time of 7 months
  - Compare homophily premium across groups

- Homophily premium: 1.7% among newbee receivers and 2.8% among experienced
- = ... **higher** after being at the same team together
- Same culture players bond outside work – help collaborate better

Homophily over time: shared experience

	pass_count				
	(1)	(2)	(3)	(4)	(5)
Same culture (any) (0/1)	0.0166***	**0.0163*** **	0.2325	0.0131*	0.0206***
	(0.0053)	(0.0053)	(0.2156)	(0.0078)	(0.0050)
Same culture (any) (0/1) × Experience	**0.0117****	0.0127**	-0.1372	0.0191**	
	(0.0059)	(0.0060)	(0.1924)	(0.0088)	
Same culture (any) (0/1) × Experience long					0.0073
					(0.0059)
Observations	457,838	443,641	13,530	219,178	384,818
Pseudo R^2	0.76317	0.76431	0.83248	0.76578	0.76699
Early experience w other team	Include	Exclude	Only	Include	Include
Time with team capped	No	No	No	Yes	No
passer-half_season fixed effects	✓	✓	✓	✓	✓
receiver-half_season fixed effects	✓	✓	✓	✓	✓
Cross position D	✓	✓	✓	✓	✓

Summary

- Isolated choice homophily to learn about any possible premium of cultural homophily for workplace collaboration
- Even in superstar teams, workers show homophily
  - Especially when complex tasks
  - Especially when shared nationality or colonial history
  - Higher when worked together a bit
  - Favoritism as source is less likely
- Homophily is pervasive even in teams of
  - very high skill individuals
  - with clear common objectives and aligned incentives
  - and involved in well defined tasks
  - activities not particularly language-intensive.