A large-scale field experiment to disentangle sources of statistical discrimination in a social setting*

Carlos Gomez-Gonzalez[1*], Gwen-Jirō Clochard[2,3], Helmut Dietl[4], and Juan Cruz Duhalde[3,5]

[1]Faculty of Social and Political Sciences, University of Lausanne
[2]Institute of Social and Economic Research, Osaka University
[3]Joint Initiative for Latin American Experimental Economics
[4]Department of Business Administration, University of Zurich
[5]Universidad del CEMA
[*]Corresponding author: carlos.gomez@unil.ch

Draft version: February 12, 2025

Abstract

This paper considers two types of statistical discrimination: individual (productivity) and collective (team fit). We conduct a large-scale correspondence study in 15 Latin American countries in the context of sports to test their influence on individual behavior. We send over 10,000 applications to male amateur soccer clubs and ask them to participate in a practice session. Each club receives one application, randomly varying the applicant's origin and signals about individual and collective productivity. We find no evidence of discrimination against immigrants overall, but we observe heterogeneity that is consistent with individual statistical discrimination. Productivity signals have no significant influence.

Keywords: correspondence study, discrimination, football
JEL Classification: J15, C92, C93

*The project was approved by the Human Subjects Committee of the Faculty of Economics, Business Administration and Information Technology from the University of Zurich on February 12, 2024 (OEC IRB # 2024-013). The study was pre-registered with the AEA RCT Registry on Feb. 21, 2024 (AEARCTR-0013077). The authors acknowledge financial support from UZH Grant Office, no. FK-22-024, as well as the Grants-in-aid for scientific research (KAKENHI, grant number: 20H05631) from the Japan Society for the Promotion of Science (JSPS). The funders had no role in study design, data collection and analysis, preparation of the manuscript, or decision to publish. For their comments and suggestions on the experimental design, pilot and main study, the authors want to thank Antonio Alvarez, David Boto García, Anna Dreber, John Duffy, María Edo, Julio Elias, Guillaume Hollard, Nobuyuki Hanaki, Taisuke Imai, John List, Mari Rege and Martín Rossi, as well as seminar participants at the University of Stavanger, the University of San Andres, the University of Oviedo, ROSES, OSIPP, CAF, and internal seminars at the Chicago Experiments group and the ISER Experimental Seminar. We would also like to thank conference participants at the AFE 2024, ESA AP 2024, RCFK 2024, and ASFEE 2024. The authors also acknowledge the help of research assistants Diego Acuña Picado, Franco Albino, Juan Martin Favetto, André Ibañez, Tomas Kay, Martina Lacoma, Federico Obermann, Luisa Quesada, Lucas Santos, and Saydee Sierra in data collection. The authors declare no competing interests.

1 Introduction

The integration of immigrants and their interactions with host majority members has become one of the greatest challenges of modern societies, as the share of the foreign-born population has substantially increased in many countries. Despite the benefits for local communities of successful integration of immigrants (Kaplan et al., 2023), previous research has found evidence of substantial discrimination (see Bertrand and Duflo (2017) for a review) which has not diminished over time (Heath and Di Stasio, 2019; Quillian et al., 2017).

In the discrimination literature, a long-standing debate is whether observed differences between groups stem from statistical discrimination, where using group characteristics could be rational in the presence of uncertainty (Arrow, 1972), or taste-based discrimination, where differences arise from an innate animus against (or fondness for) a particular group (Becker, 1957). Correspondence studies, where researchers send fictitious applications, changing only the characteristics of interest (e.g., name, gender, level of education, origin), are often viewed to be the gold standard of measurement of discrimination (Verhaeghe, 2022). Because they allow for control of productivity-relevant signals, correspondence studies are often thought to identify taste-based discrimination (Kline et al., 2022). However, even in a context where the research design could control all signals sent by the applicant, some uncertainty would still remain about the quality of the applicants, their background, or their motivations, leaving room for statistical discrimination (Lippens et al., 2022).

This study proposes a distinction between two types of statistical discrimination: individual and collective. Individual statistical discrimination would occur when there exist doubts about someone's individual abilities, such as whether a particular worker would be able to perform a specific task. Collective statistical discrimination would occur when the uncertainty does not lie with the individual's abilities but rather with the individual's capacity to fit with a team. We first illustrate the distinction in a toy model.

Then, using a large-scale correspondence study, we aim to identify the relative importance of the two sources of statistical discrimination. Our study contributes to understanding what type of information about productivity drives decision-makers' choices (Martin and Marx, 2022) and the mechanisms related to particular skills that may determine discrimination (Deming, 2017; Esses, 2021; Hurst et al., 2024).

We examine discrimination against immigrants applying to join an amateur football club in Latin America.[1] This context is particularly suited to distinguish between the individual competence of a player and their ability to collaborate with other players on the team. The economic importance of amateur sports relates to the strength of weak ties (Granovetter, 1973), as contacts and personal relationships acquired in relaxed environments can create broader social networks facilitating access to employment or housing, thereby reducing inequality (Chetty et al., 2022). Additionally, football is the most popular sport in the world, and the structure of amateur leagues and clubs is identical in most countries. This allows us to standardize the design across countries, to include a large number of existing clubs, and to identify small effects. Compared to professional sports and teams in other industries, amateur clubs are not dominated by economic interests; members are not remunerated for the functions and aim to maximize wins in a friendly environment. In our sample, clubs compete in the lowest leagues organized by regional federations and have full autonomy on membership fees and admission, facing no restrictions on the number of (foreign) players.

The study involves 15 countries and contacting over 10,000 male clubs via email, phone, and social media. Each club receives one application, randomly varying the origin of the applicant[2] and signals about individual and collective productivity (football proficiency and language proficiency, respectively – see Section 4 for details). We monitor the accounts set up for applicants for four weeks (more than 90%

[1] See the list in Appendix A.

[2] for each country, we select the four most-represented foreign groups for each of the three following regions: Latin America, Europe, and Asia. The design allows us to examine heterogeneity among foreign groups.

of responses were sent within two days) and measure the share of positive responses that they receive. To alleviate the concern that correspondence studies only monitor initial callback rates (Bertrand and Duflo, 2017), we use a standardized response when the clubs ask for further information about the applicant.

Our results show no significant differences in positive response rates between local and foreign applicants. However, we observe significant heterogeneity depending on the origin of the applicant. The response rate is negatively correlated with our proxy of football proficiency (the FIFA ranking of the country), consistent with individual statistical discrimination. Regarding collective statistical discrimination, the coefficient for language proficiency is positive, but we cannot reject the null hypothesis. Lastly, we do not find that our signals significantly affect response rates.

The rest of the paper is structured as follows. Section 2 presents a brief overview of the literature. Section 3 presents a toy model to clarify ideas about individual and collective statistical discrimination. Section 4 describes the experimental design and Section 5 reports the main results, focusing on differential treatment by type of immigrant background and the effect of individual and collective productivity signals. Section 6 discusses the findings and concludes.

2 Literature review

This paper primarily contributes to three strands in the literature. First, it contributes to the debate between the two workhorse discrimination models: taste-based and statistical discrimination. Taste-based discrimination models were developed by Becker (1957), where employers, coworkers, and consumers might have a distaste (i.e., animus) toward minority workers, leading to hiring refusals or wage reductions. This literature has attempted to show that manager bias can lead to homophily (Giuliano et al., 2009) or that consumer bias can influence the composition of the workforce (Holzer and Ihlanfeldt, 1998; Leonard et al., 2010; Combes et al., 2016; Bar and Zussman, 2017) – or both (Ahmadi et al., 2024).

In contrast, statistical discrimination (Arrow, 1972; Phelps, 1972) argues that differences in hiring rates do not arise from an ad-hoc animus parameter, but from imperfect information about workers' productivity. Profit-maximizing employers will infer workers' productivity from available signals (including group membership). Some scholars have tried to identify whether unequal treatment between groups was due to statistical or taste-based reasons. For example, List (2004) found evidence of statistical discrimination against minorities in the market for baseball trading cards, while Bohren et al. (2019) found evidence of *inaccurate* statistical discrimination in the evaluation of posts generated by men and women.

In the statistical-taste debate, taste-based discrimination is often viewed as the residual of group differences after removing all relevant information about individual productivity. In this paper, we contribute to the debate with a large-scale field experiment on the influence of individual -football proficiency- and collective -language proficiency- skills (Guryan and Charles, 2013; Lippens et al., 2022). We argue that a distinction should be made between individual and collective statistical discrimination. In the presence of collective statistical discrimination, i.e., when there is uncertainty about the worker's ability to collaborate with other coworkers, there will still be group differences, even after accounting for all dimensions of individual productivity and in the absence of animus. Research shows that tasks requiring interaction with others regularly influence racial discrimination outcomes in the labor market (Combes et al., 2016; Deming, 2017; Hurst et al., 2024). Implicitly, in other correspondence studies, research has randomized different types of individual and collective skills and reported an influence on callback differences (Kroft et al., 2013; Pager and Karafin, 2009; Bennett, 2023).

Second, this paper contributes to the literature on methodologies for identifying discrimination. Measuring discrimination using existing data is complex, because no two individuals are exactly alike, differing only by their race, origin, or gender. Because of this limitation, the literature has turned to running experiments to causally identify discrimination (Bertrand and Duflo, 2017). The main approach used has been correspondence studies (Verhaeghe, 2022), which use fictitious applications to get around the limitation of having differences other than the dimension under study (e.g., race, gender, or nationality).

To cite a few examples, correspondence studies have shown that prejudice against minorities exists in contexts such as employment (Bertrand and Mullainathan, 2004), housing (Auspurg et al., 2019), transportation (Ge et al., 2020), and the sharing economy (Liebe and Beyer, 2021). There also exists evidence that the extent of discrimination depends on external factors such as attention level (Bartos et al., 2013), and that the phenomenon is highly persistent; for example, the lower callback rate for African-Americans has not diminished over the past two decades (Quillian et al., 2017).

We contribute to this literature with a large multi-country correspondence study in Latin America. Despite the well-documented evidence of ethnic and racial discrimination in developed regions, only a few studies examine ethnic discrimination in developing countries.[3] In Latin America, Moreno et al. (2012) and Galarza and Yamada (2014) find discrimination against indigenous applicants in the labor market of Lima, Peru. Moreno et al. (2012) also find racial differences in the Mexican labor market (although they are only significant for female applicants). In housing, Zanoni and Díaz (2024) find evidence of discrimination against Venezuelans in the rental market in Colombia. To the best of our knowledge, no study has examined discrimination against immigrants in other social settings in Latin America.

We also contribute to the correspondence study literature by analyzing club behavior beyond the initial callback level. A major criticism of correspondence studies is that they can only capture discrimination in the initial screening of applicants (Heckman and Siegelman, 1993; Bertrand and Duflo, 2017). In this paper, we contribute by investigating the extent of discrimination after the initial screening by

[3] For a review of correspondence studies in hiring, see, e.g., Neumark (2018), Lippens et al. (2023), and Zschirnt and Ruedin (2016). For a review of correspondence studies in housing, see e.g., Auspurg et al. (2019) and Flage (2018).

providing standardized information when clubs ask for follow-up information.

Third, this paper contributes to the literature on discrimination in sports. Because of the amount of easily accessible data, professional sports have long been used to test for theories (Kahn, 2000). In particular, there is ample evidence of racial and ethnic bias in sports, for instance, in the major leagues in the US (Schmidt and Coe, 2014; Price and Wolfers, 2010) or in European football (MacLeod and Newall, 2022; Principe and van Ours, 2022; Szymanski, 2000; Quansah et al., 2024). There is also evidence that local markets influence the composition of teams in the NBA and their metropolitan areas, suggesting the influence of fans on drafting decisions (Burdekin and Idson, 1991; Hoang and Rascher, 1999; Burdekin et al., 2005). Research also finds that laws increasing job-to-job mobility can decrease racial wage gaps, as shown in English football following the Bosman ruling in 1995 (Deschamps and De Sousa, 2021).

Despite their societal impact, amateur sports are underexamined in this area. This setting covers all aspects of social interaction and enables participants to expand their network with weak ties (Duarte et al., 2012; Granovetter, 1973). Even though contact in amateur sports is often used as a gateway for social cohesion and integration (Mousa, 2020), the setting is not exempt from differential treatment. Gomez-Gonzalez et al. (2021) find that immigrants receive fewer positive responses than locals when trying to join a club in European football, and that this difference is not mitigated by low-cost information campaigns (Dur et al., 2023). However, research attempting to identify the sources of discrimination in sport is scarce.[4] We contribute to this literature by examining micro-level determinants of discrimination against immigrants related to individual characteristics (Esses, 2021), which are critical to designing effective policies. Specifically, we measure how the ability to perform and communicate with team members influences the response rate to immigrants who request to join a club.

[4] Ahmadi et al. (2024), which disentangles the effects of homophily from team scouting and fan bias in professional sports, is an exception.

3 Toy model

In this section, we develop a toy model to explain why the distinction between individual and collective statistical discrimination matters, and how this distinction is linked to the use of correspondence studies.

Assume there exist two groups $\{L, F\}$ (local or foreign) in the society. Assume a recruiter (of group L for simplicity) is considering two applicants (each of whom belongs to a group $g \in \{L, F\}$) to hire for a position within her firm (or team in the case of our experiment). To do so, the recruiter maximizes her expected utility, which takes the following form:

$$U = Production + Taste \tag{1}$$

Let us assume the $Taste$ parameter is of the form of an ingroup bias in favor of L workers (Becker, 1957):

$$Taste = \theta \times \mathbb{1}(\text{Worker is of group L}) \tag{2}$$

where θ represents the intensity of the ingroup bias.

3.1 Standard framework: no collective production

Each worker belongs to group $g \in \{L, F\}$, and has an individual ability a that depends on group membership: $a \sim \mathcal{N}(\mu_g, 1/\tau_g)$.[5] Furthermore, in the standard statistical discrimination framework, production only depends on this individual ability with the form $Production = f(a)$.

The recruiter will hire the worker of group F over the worker of group L if

$$\mathbb{E}[f(a_F)] \geq \mathbb{E}[f(a_L)] + \theta \tag{3}$$

If there is a difference in average abilities between the two groups, then even without ingroup bias ($\theta = 0$) the recruiter will prefer recruiting the worker from the group with the highest average productivity. This is a standard model of statistical discrimination using individual characteristics (Phelps, 1972; Arrow, 1972), which we label *individual statistical discrimination*.

In the context of a correspondence study, the researchers will artificially fix the ability level of the two applicants to a common value $\bar{a}$ by providing information about individual productivity. It is then

[5] We are using the same notations as Bohren et al. (2019).

assumed that if we observe a difference in hiring rates between L and F, it must mean that the recruiter displays bias ($\theta \neq 0$).[6]

3.2 Collective production

In the previous model, we assumed that the production function only depended on individual ability a. Let us assume now that the productivity is a function of both individual ability and *collective fit*. This collective fit represents the ability of a worker to work with the team and contribute to its joint production. The production becomes $Production = g(a, c)$ where c represents the collective fit, and g is an increasing function of both its parameters. For simplicity, let us assume that $c_F < c_L$, i.e., the local worker can contribute more easily to the team's joint production than the foreign worker.[7]

With this new production function, the recruiter will hire the worker of group F over the worker of group L if

$$\mathbb{E}[g(a_F, c_F)] \geq \mathbb{E}[g(a_L, c_L)] + \theta \tag{4}$$

Importantly, if the collective fit parameter c depends on group membership, then *conditional on a common (perfect) signal* $\bar{a}$, it would still be rational for the recruiter to hire the local worker, even with no ingroup bias ($\theta = 0$). This difference, conditional on the worker's individual productivity, is what we refer to in this paper as *collective statistical discrimination*.

In a standard correspondence study, individual characteristics are fixed and group membership is randomly varied (Verhaeghe, 2022). If the production function solely depends on individual performance, as is often assumed in the discrimination literature (Schwab, 1986; Fernandez and Greenberg, 2013; Echenique and Li, 2025), it is correct to attribute any difference in callback rates to ingroup bias. However, in the presence of collective productivity and if productivity is linked -or thought to be linked- to

[6] Note that if the signal is not precise enough, i.e., if there is still uncertainty about the worker's ability a, then it would still be rational to discriminate in the recruitment process even with $\theta = 0$.

[7] If skills between groups are complementary, it would be feasible to have $c_F > c_L$.

group membership, then part of the difference in callback rates would be *falsely* attributed to ingroup bias, when the reason is in fact attributable to collective statistical discrimination.

In the experiment below, we will assume that in the presence of uncertainty about the quality of a player, the recruiters will use the proxy of the quality of the national team for a measure of μ_g. As for collective ability, we will proxy the parameter c_g by whether the country of origin of the applicant speaks the same language as the country of the club.

4 Experimental design

In this section, we present the experimental design of the correspondence study we performed in Latin America. The study was approved by the Human Subjects Committee of the Faculty of Economics, Business Administration and Information Technology from the University of Zurich on February 12, 2024 (OEC IRB # 2024-013). The study was pre-registered with the AEA RCT Registry on Feb. 21, 2024 (AEARCTR-0013077). Deviations from the pre-analysis plan are presented in Appendix B.

Setup. We ran the experiment in the context of amateur football (soccer) in Latin America, focusing on prejudice against immigrants. The study involved 15 countries (Appendix A) and contacting over 10,000 male clubs.[8] We sent each club one application, randomly varying the origin of the applicant and signals about productivity (see below).

The experiment was conducted between February and May, 2024. Clubs were contacted via email, WhatsApp, Facebook, and Instagram, which are considered common modes of communication in Latin America. Accounts were monitored for at least four weeks after the last application was sent. As more than 90% of responses were received within two days, this is a comfortable delay to receive all responses.

[8]The contact information for only approximately 500 female clubs could be identified, 300 of which were in Argentina, so we decided not to include them in the experiment.

Treatments. There are two dimensions of treatment in our experiment. The first dimension is relative to the *content of the message*. We randomized the content of the application message that clubs receive. The control treatment is a core *neutral* message (Appendix C). For the treatment groups, we added an additional message to the core of the email.

For the *Individual* treatment ($T1$), the following sentence was added: "When I was young, I played in the academy of Cremonese in Italy." Cremonese is a professional club in the second division of Italy. For the amateur clubs in our sample, the treatment would probably signal a level far superior to their average player, so we expected the treatment to have an effect on individual statistical discrimination. We decided to have every individual in treatment T1 send the same information, i.e., having been in the academy of Cremonese, rather than, say, a team from their origin country, because we wanted a uniform signal across countries of origin. In the toy model presented in Section 3, we use treatment $T1$ as a signal regarding ability ($\bar{a}$).

For the *Collective* treatment ($T2$), the added sentence was "I did part of my high-school curriculum in Spanish [Portuguese in Brazil] so I speak Spanish [Portuguese] well." The *Collective* treatment is designed to signal the player's ability to communicate with others on the team, i.e., to have an effect on collective statistical discrimination. In the toy model, this would correspond to a positive shock on the collective production parameter c. Treatment $T2$ was only implemented for applicants coming from countries where the primary language is not Spanish [Portuguese in Brazil].

We also include a treatment that combines the two treatments (individual and collective).

The second dimension of treatment is relative to the origin of the applicants. In each country, census data were used to determine the countries of origin of the applicants used. We used the four countries with the highest numbers of immigrants in the country for each of the following three regions: Latin America, Europe, and Asia.[9] The list of countries of origin for each club country in our data is presented

[9]Many Latin American countries have too small a population from individual African countries, or sometimes do not report

in Appendix A. We used census data to identify the two most common first and last names for children born in 2000.

Treatments are summarized in Figure 1. Each club was randomly placed in one of the 16 (4x4) treatment cells. All the analyses are between clubs, meaning that the clubs do not receive similar-sounding applications from two different players, and we are not able to identify individual measures of discrimination of non-consenting participants. Assignment to treatment was stratified by country.

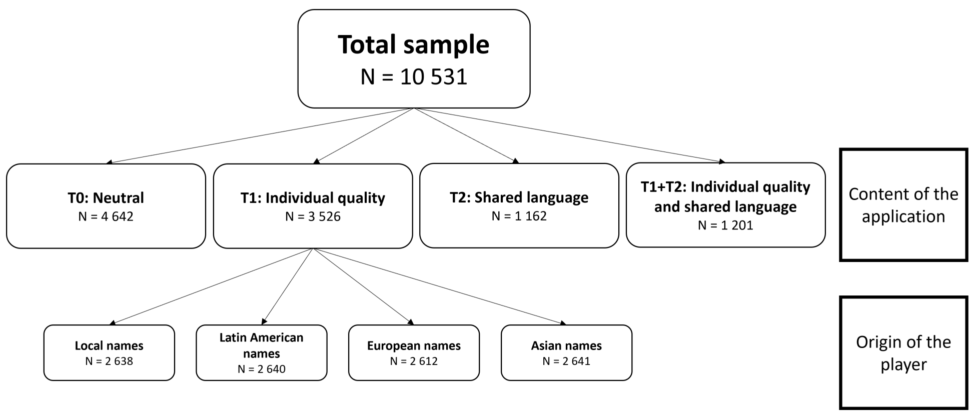


Figure 1: Treatment arms for the experiment. *Note:* The applicant's origin (Local, Latin American, European, and Asian names) are applied to each content treatment (*Neutral*, *Individual quality* and *Shared language*). Samples for the *Shared language* treatment are smaller because many countries of origin of the applicants (particularly in Latin America) already share the language with the country of the club.

Outcome variable. The outcome "Response" is coded as follows: (1) no response; (2) negative response; (3) the club asks for additional information or asks to have a call with the applicant; (4) positive response. The variable "Positive response", the main outcome used in our study, is a dummy variable equal to 1 if the "Response" is 3 or 4; and 0 otherwise. This is the only variable of interest. In Appendix G, we present the results with an alternative definition of a positive response, in this case only counting a positive response if the answer is 4 (with the above definition). The results are unchanged with this alternative definition.

In addition, if the club responded by asking for further information about the player, we sent a

the disaggregated statistics by country. We thus decided not to include African countries in the analysis.

standardized response about the player (similar in all treatments) and monitored responses. This second stage email is meant to measure discrimination in a second stage, thus partially alleviating one of the main concerns with correspondence studies, i.e., the fact that researchers can only analyze differential treatment in the first round (i.e., callbacks or invitation for an interview – see Bertrand and Duflo (2017)). The text of the second round email is presented in Appendix D.

Hypotheses. As specified in the pre-analysis plan, we test for five primary hypotheses. For this reason, we set the significance level for p-values at 0.01.

Hypothesis 1 (presence of discrimination against immigrants in the amateur football context in Latin America): Foreign applicants receive fewer positive responses than local applicants.

To test for Hypothesis 1, we estimate the following model:

$$\mathbb{P}(\text{Postitive response}_i) = \alpha_0 + \alpha_1 \text{Foreign}_i + X_i + \epsilon_i \tag{5}$$

Hypothesis 2 (presence of individual statistical discrimination): Applicants from countries with a lower FIFA ranking receive fewer positive responses.

Hypothesis 3 (presence of collective statistical discrimination): Applicants from countries that do not share a common language with the country of the club receive fewer positive responses.

To test for Hypotheses 2 and 3, we estimate the following model:

$$\mathbb{P}(\text{Postitive response}_i) = \gamma_0 + \gamma_1 \text{FIFA Ranking}_i + \gamma_2 \text{Shared Language}_i + X_i + \epsilon_i \tag{6}$$

Hypothesis 4 (reduction in individual statistical discrimination): The treatment signaling individual quality has a positive effect on response rates for applicants from countries with a lower FIFA ranking.

Hypothesis 5 (reduction in collective statistical discrimination): The treatment signaling collective ability (fluency) has a positive effect on response rates for applicants from countries who do not speak the same language.

To test for Hypotheses 4 and 5, we estimate the following model:

$$\begin{aligned}\mathbb{P}(\text{Postitive response}_i) &= \delta_0 + \delta_1\text{FIFA Ranking}_i + \delta_2\text{T1}_i + \delta_3\text{FIFA Ranking}_i \times \text{T1}_i \\ &\quad +\delta_4\text{Shared Language}_i + \delta_5\text{T2}_i + \text{T1}_i \times \text{T2}_i \\ &\quad +X_i + \epsilon_i \qquad (7)\end{aligned}$$

We have an additional hypothesis for the subsample who replied asking for more information. This hypothesis is exploratory, so we kept the conventional $p = 0.05$ level for the significance threshold.

Hypothesis 6 (second round discrimination): Foreign applicants receive fewer positive responses than local applicants in the second round of applications.

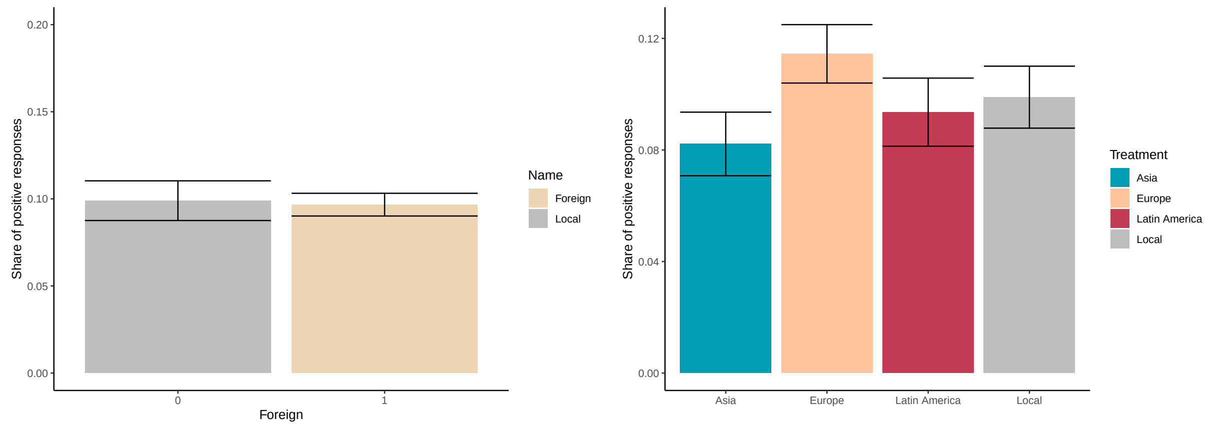
To test for Hypothesis 6, we estimate the following model only for clubs that have asked for further information in the second round:

$$\mathbb{P}(\text{Postitive response}_i) = \eta_0 + \eta_1\text{Foreign}_i + X_i + \epsilon_i \qquad (8)$$

5 Results

Foreign applicants are not less likely to receive a positive response (Figure 2a; Table F.1, Column 1). Due to the large sample size, we were able to detect effects as small as 1 pp. However, this result does not mean that there is no heterogeneity between the countries of origin of applicants. First, in a non-pre-registered result, we find that the response rate differs by the region of origin of the applicant, with Asian applicants receiving fewer positive responses than locals ($p = 0.01$) and European applicants receiving marginally significantly more ($p = 0.09$).

Second, we investigated the difference in response rates between the countries of origin, with respect to our measures of statistical discrimination (Table F.1). In Column 1, we find support for Hypothesis 2, with a negative correlation between the FIFA ranking of the country (used as a proxy for the quality of players in the country) and response rates. The correlation is statistically significant ($p = 0.007$). However, we do not find support for Hypothesis 3, in that applicants from countries that share the same



(a) All foreign names combined

(b) Disaggregating by region

Figure 2: Positive response rate by origin treatment

Note: N = 10,821. In Panel (a), the difference between the response rates for foreign and local applicants is not statistically significant ($p = 0.57$). In Panel (b), Asian applicants receive significantly less positive responses than locals ($p = 0.01$), while the European applicants receive marginally more ($p = 0.09$).

language with the country of the club do not receive significantly higher response rates (Column 2, $p = 0.74$). The results are robust if we include both sources of statistical discrimination (Column 3).

Table 1: Presence of individual and collective statistical discrimination

	Individual (1)	Collective (2)	Both (3)
FIFA ranking	-0.001^{***}		-0.001^{***}
	(0.001)		(0.001)
Shared Language		0.012	-0.003
		(0.036)	(0.036)
Constant	-0.870^{***}	-0.923^{***}	-0.868^{***}
	(0.049)	(0.051)	(0.055)
N	10,531	10,531	10,531
Pseudo R^2	0.073	0.072	0.073

Note: The dependent variable is a dummy variable for whether the applicant received a positive response. * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses. Controls include country and means of contact fixed effects.

Regarding support for Hypotheses 4 and 5 and the reductions in statistical discrimination from our signals of individual and collective ability, we do not find evidence that the signals had any significant effect on response rates (Table 2). If anything, the signals work in opposite directions to our pre-registered hypotheses. Potential explanations are detailed in the next section. All these results qualitatively hold if we change the definition of a positive response to include only a strict invitation to train at the club (Table G), although the significance levels change.

Table 2: Reduction in individual and collective statistical discrimination

	Probability of positive callback (1)
FIFA ranking	-0.001^{**}
	(0.001)
T1: Quality treatment	-0.048
	(0.057)
FIFA ranking x T1	-0.0002
	(0.001)
Shared Language	-0.024
	(0.042)
T2: Language treatment	-0.025
	(0.068)
T1 x T2	-0.051
	(0.089)
Constant	-0.857^{***}
	(0.058)
N	10,531
Pseudo R^2	0.074

Note: The dependent variable is a dummy variable for whether the applicant received a positive response. * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses. Controls include country-fixed effects.

The last result concerns discrimination in the second round (Figure 3, Table F.2), for clubs who asked applicants for further information about themselves. While we do observe that foreign applicants receive fewer responses, the difference is not statistically significant ($p = 0.37$). Moreover, the magnitude of the difference between local and foreign applicants falls well below the minimum detectable effect size in the pre-analysis plan (approximately 10pp in the pre-analysis plan, less than 2pp in the data).

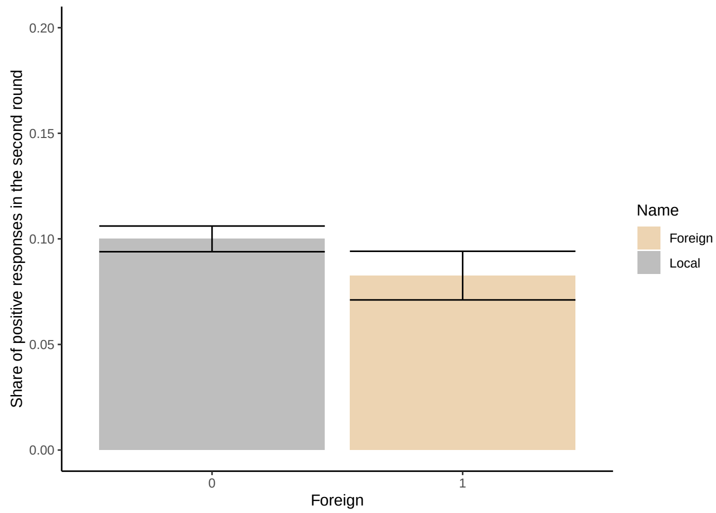


Figure 3: Positive response rate in the second round, by origin of the applicant

Note: N = 310. The difference in response rate between local and foreign applicants is not statistically significant ($p = 0.37$). The same message was sent to all clubs who requested future information about the applicant; see Section 4 for details.

6 Discussion and conclusion

The debate between statistical and taste-based discrimination lies at the heart of discussions about fairness in socioeconomic contexts. Despite concerns about the costs imposed on companies or harm to other applicants (Kessler et al., 2019), the gold standard for identifying taste-based discrimination has been the correspondence study (Bertrand and Duflo, 2017; Verhaeghe, 2022). This methodology allows researchers to hold all signals about jobseekers or applicants constant in real life scenarios, meaning that the only remaining difference must be due to taste-based discrimination. In this paper, we argue to the contrary, that in the presence of collective statistical discrimination – that is, when there exists uncertainty about whether the applicant would fit well with peers – group differences may persist even in the absence of animus on the part of decision-makers.

We test the distinction between individual and collective statistical discrimination in the context of amateur football in Latin America, focusing on prejudice against immigrants. We leverage the similarity in league structures and the large amount of publicly available contact information of clubs to run a standardized correspondence study in 15 countries covering over 10,000 clubs. As such, the study sheds light on discrimination in an understudied region.

While our results show no statistical differences in positive response rates between local and foreign applicants, we find significant heterogeneity depending on the origin of the applicant. We find that European applicants are favored over locals, but Asian applicants receive significantly fewer positive responses. The finding sheds some light on the preferences toward specific minorities or outgroup members as predicted by hierarchies in multiple group systems theory (Hagendoorn, 1995). Future research is needed to better understand these group dynamics across regions.

Additionally, the analysis shows that statistical discrimination plays a significant role. Specifically, we find a robust negative relationship between the positive response rate and the FIFA ranking of the country of origin (a proxy for player quality), thus highlighting a form of individual statistical discrimination. While the point estimate for shared language – our proxy for collective productivity – is positive, we do not find that the difference is statistically significant.

In our experiment, we also randomly sent additional information to clubs to signal individual and collective productivity. We find that the signals do not significantly affect response rates, and the point estimates are even opposite to what we had hypothesized (both coefficients being insignificantly negative when we hypothesized significant positive coefficients). Alternative explanations could be put forth to explain this result, but our design does not allow us to distinguish between them.

One potential explanation is simply that the signals were too weak or too noisy, thereby reducing the reliability of information (Neumark, 2018). For example, our signal of individual productivity implies that the individual has spent some time in a foreign country, which adds noisy. Other individual productivity signals are possible, e.g., the individual played pro-youth level in his respective country, but we disregarded them because they would create heterogeneity within the same treatment. We chose the youth level because career information is not always readily accessible on the Internet. An alternative explanation could be that clubs did not perceive the signals in the way we anticipated. Going back to the model presented in Section 3, we hypothesized that the individual productivity signal affects the perception of individual ability a. It is not inconceivable, though, that the signal of belonging to a professional team's academy could also signal information about collective ability, e.g., "If a professional team wants to recruit this player, it means that they think he can play as a team," or even as a form of taste, e.g., "This player needs to brag that he used to be in the academy of a professional club so he is not someone likeable." Other forms of bias could influence the responses, e.g., "This player is too good for our small club, so he will likely receive other positive offers and the club should not waste time with him." Similar arguments could be made for the collective productivity signal. Identifying treatments that can isolate the effect of individual and collective productivity (especially satisfying exclusion restrictions) remains a challenge and an essential next step for the study of statistical discrimination.

In terms of future research, we believe the study of collective statistical discrimination naturally raises questions about its sources. The inability to work with a team could be due to animus from coworkers who do not want to collaborate with minority workers (e.g., coworker bias in Becker (1957)). In this case, collective statistical discrimination could indeed be understood as animus, but not on the side of the manager. However, collective statistical discrimination could also result from a genuinely reduced ability to work with a team. In this paper, we argue that language is a factor influencing collective productivity, making it a legitimate factor for profit-maximizing firms to consider.

Lastly, in terms of policy, this work sheds light on the need to better understand the sources of discrimination to combat it (Ahmadi et al., 2024). Understanding the origins and mechanisms of discrimination is essential to effectively combat its pervasive effects on individuals and society as a whole. By examining the root causes of discrimination, we better understand how and why certain groups are marginalized or disadvantaged. This examination enables us to develop more targeted and effective interventions to dismantle discriminatory structures and promote equal treatment. Future research should bear in mind the particularities of our context, as it is a social setting with no explicit economic incentives.

References

Ahmadi, M., G.-J. Clochard, J. Lachman, and J. A. List (2024). Toward an understanding of discrimination when multiple channels exist. Technical report, National Bureau of Economic Research No. w33391.

Arrow, K. J. (1972). Some mathematical models of race discrimination in the labor market. In *Racial discrimination in economic life*, pp. 187–204. Lexington Books.

Auspurg, K., A. Schneck, and T. Hinz (2019). Closed doors everywhere? A meta-analysis of field experiments on ethnic discrimination in rental housing markets. *Journal of Ethnic and Migration Studies 45*(1), 95–114.

Bar, R. and A. Zussman (2017). Customer discrimination: Evidence from Israel. *Journal of Labor Economics 35*(4), 1031–1059.

Bartos, V., M. Bauer, J. Chytilová, and F. Matějka (2013). Attention discrimination: Theory and field experiments. Technical Report 499, CERGE-EI Working Paper Series.

Becker, G. S. (1957). *The economics of discrimination*. University of Chicago Press.

Bennett, C. T. (2023). Labor market returns to mbas from less-selective universities: Evidence from a field experiment during covid-19. *Journal of Policy Analysis and Management 42*(2), 525–551.

Bertrand, M. and E. Duflo (2017). Field experiments on discrimination. In *Handbook of economic field experiments*, Volume 1, pp. 309–393. Elsevier.

Bertrand, M. and S. Mullainathan (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review 94*(4), 991–1013.

Bohren, J. A., A. Imas, and M. Rosenberg (2019). The dynamics of discrimination: Theory and evidence. *American Economic Review 109*(10), 3395–3436.

Burdekin, R. C., R. T. Hossfeld, and J. K. Smith (2005). Are NBA fans becoming indifferent to race? Evidence from the 1990s. *Journal of Sports Economics 6*(2), 144–159.

Burdekin, R. C. and T. L. Idson (1991). Customer preferences, attendance and the racial structure of professional basketball teams. *Applied Economics 23*(1), 179–186.

Chetty, R., M. O. Jackson, T. Kuchler, J. Stroebel, N. Hendren, R. B. Fluegge, S. Gong, F. Gonzalez, A. Grondin, M. Jacob, et al. (2022). Social capital I: Measurement and associations with economic mobility. *Nature 608*(7921), 108–121.

Combes, P.-P., B. Decreuse, M. Laouenan, and A. Trannoy (2016). Customer discrimination and employment outcomes: Theory and evidence from the french labor market. *Journal of Labor Economics 34*(1), 107–160.

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

Deschamps, P. and J. De Sousa (2021). Labor mobility and racial discrimination. *European Economic Review 135*, 103738.

Duarte, R., D. Araújo, V. Correia, and K. Davids (2012). Sports teams as superorganisms: Implications of sociobiological models of behaviour for research and practice in team sports performance analysis. *Sports Medicine 42*, 633–642.

Dur, R., C. Gomez-Gonzalez, and C. Nesseler (2023). How to reduce discrimination? Evidence from a field experiment in amateur soccer. *Journal of Ethnic and Migration Studies 49*(1), 175–191.

Echenique, F. and A. Li (2025). Rationally inattentive statistical discrimination: Arrow meets phelps. *Journal of the European Economic Association*, jvaf002.

Esses, V. M. (2021). Prejudice and discrimination toward immigrants. *Annual Review of Psychology 72*, 503–531.

Fernandez, R. M. and J. Greenberg (2013). Race, network hiring, and statistical discrimination. In *Networks, Work and Inequality*, pp. 81–102. Emerald Group Publishing Limited.

Flage, A. (2018). Ethnic and gender discrimination in the rental housing market: Evidence from a meta-analysis of correspondence tests, 2006–2017. *Journal of Housing Economics 41*, 251–273.

Galarza, F. B. and G. Yamada (2014). Labor market discrimination in Lima, Peru: Evidence from a field experiment. *World Development 58*, 83–94.

Ge, Y., C. R. Knittel, D. MacKenzie, and S. Zoepf (2020). Racial discrimination in transportation network companies. *Journal of Public Economics 190*, 104205.

Giuliano, L., D. I. Levine, and J. Leonard (2009). Manager race and the race of new hires. *Journal of Labor Economics 27*(4), 589–631.

Gomez-Gonzalez, C., C. Nesseler, and H. M. Dietl (2021). Mapping discrimination in Europe through a field experiment in amateur sport. *Humanities and Social Sciences Communications 8*(1), 1–8.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology 78*(6), 1360–1380.

Guryan, J. and K. K. Charles (2013). Taste-based or statistical discrimination: The economics of discrimination returns to its roots. *The Economic Journal 123*(572), F417–F432.

Hagendoorn, L. (1995). Intergroup biases in multiple group systems: The perception of ethnic hierarchies. *European Review of Social Psychology 6*(1), 199–228.

Heath, A. F. and V. Di Stasio (2019). Racial discrimination in Britain, 1969–2017: A meta-analysis of field experiments on racial discrimination in the British labour market. *The British Journal of Sociology 70*(5), 1774–1798.

Heckman, J. J. and P. Siegelman (1993). The urban institute audit studies: Their methods and findings. In *Clear and convincing evidence: Measurement of discrimination in America*, pp. 187–258. Urban Institute Press, Washington, DC.

Hoang, H. and D. Rascher (1999). The nba, exit discrimination, and career earnings. *Industrial Relations: A Journal of Economy and Society 38*(1), 69–91.

Holzer, H. J. and K. R. Ihlanfeldt (1998). Customer discrimination and employment outcomes for minority workers. *The Quarterly Journal of Economics 113*(3), 835–867.

Hurst, E., Y. Rubinstein, and K. Shimizu (2024). Task-based discrimination. *American Economic Review 114*(6), 1723–1768.

Kahn, L. M. (2000). The sports business as a labor market laboratory. *Journal of Economic Perspectives 14*(3), 75–94.

Kaplan, D. S., A. Rivera, and C. van der Werf (2023). A fair chance for migrants in the labor market. Technical report, Inter-American Development Bank Technical Note 2825.

Kessler, J. B., C. Low, and C. D. Sullivan (2019). Incentivized resume rating: Eliciting employer preferences without deception. *American Economic Review 109*(11), 3713–3744.

Kline, P., E. K. Rose, and C. R. Walters (2022). Systemic discrimination among large us employers. *The Quarterly Journal of Economics 137*(4), 1963–2036.

Kroft, K., F. Lange, and M. J. Notowidigdo (2013). Duration dependence and labor market conditions: Evidence from a field experiment. *The Quarterly journal of economics 128*(3), 1123–1167.

Leonard, J. S., D. I. Levine, and L. Giuliano (2010). Customer discrimination. *The Review of Economics and Statistics 92*(3), 670–678.

Liebe, U. and H. Beyer (2021). Examining discrimination in everyday life: A stated choice experiment on racism in the sharing economy. *Journal of Ethnic and Migration Studies 47*(9), 2065–2088.

Lippens, L., S. Baert, A. Ghekiere, P.-P. Verhaeghe, and E. Derous (2022). Is labour market discrimination against ethnic minorities better explained by taste or statistics? A systematic review of the empirical evidence. *Journal of Ethnic and Migration Studies 48*(17), 4243–4276.

Lippens, L., S. Vermeiren, and S. Baert (2023). The state of hiring discrimination: A meta-analysis of (almost) all recent correspondence experiments. *European Economic Review 151*, 104315.

List, J. A. (2004). The nature and extent of discrimination in the marketplace: Evidence from the field. *The Quarterly Journal of Economics 119*(1), 49–89.

MacLeod, S. A. and P. W. Newall (2022). Investigating racial bias within australian rules football commentary. *PLoS one 17*(7), e0272005.

Martin, D. and P. Marx (2022). A robust test of prejudice for discrimination experiments. *Management Science 68*(6), 4527–4536.

Moreno, M., H. Ñopo, J. Saavedra, and M. Torero (2012). Detecting gender and racial discrimination in hiring through monitoring intermediation services: The case of selected occupations in Metropolitan Lima, Peru. *World Development 40*(2), 315–328.

Mousa, S. (2020). Building social cohesion between christians and muslims through soccer in post-isis iraq. *Science 369*(6505), 866–870.

Neumark, D. (2018). Experimental research on labor market discrimination. *Journal of Economic Literature 56*(3), 799–866.

Pager, D. and D. Karafin (2009). Bayesian bigot? statistical discrimination, stereotypes, and employer decision making. *The Annals of the American Academy of Political and Social Science 621*(1), 70–93.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *American Economic Review 62*(4), 659–661.

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

Principe, F. and J. C. van Ours (2022). Racial bias in newspaper ratings of professional football players. *European Economic Review 141*, 103980.

Quansah, T. K., M. Lang, and B. Frick (2024). Racial bias in the stands? investigating customer-based discrimination in european soccer. *Applied Economics 56*(54), 7050–7064.

Quillian, L., D. Pager, O. Hexel, and A. H. Midtbøen (2017). Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. *Proceedings of the National Academy of Sciences 114*(41), 10870–10875.

Schmidt, A. and K. Coe (2014). Old and new forms of racial bias in mediated sports commentary: The case of the national football league draft. *Journal of Broadcasting & Electronic Media 58*(4), 655–670.

Schwab, S. (1986). Is statistical discrimination efficient? *The American Economic Review 76*(1), 228–234.

Szymanski, S. (2000). A market test for discrimination in the english professional soccer leagues. *Journal of Political Economy 108*(3), 590–603.

Verhaeghe, P.-P. (2022). Correspondence studies. In *Handbook of labor, human resources and population economics*, pp. 1–19. Springer.

Zanoni, W. and L. Díaz (2024). Discrimination against migrants and its determinants: Evidence from a multi-purpose field experiment in the housing rental market. *Journal of Development Economics 167*, 103227.

Zschirnt, E. and D. Ruedin (2016). Ethnic discrimination in hiring decisions: A meta-analysis of correspondence tests 1990–2015. *Journal of Ethnic and Migration Studies 42*(7), 1115–1134.

Appendices

A Countries included in the experiment

Figure A.1: Countries included in the experiment

Table A.1: List of countries and sample size for the main experiment

Country	**N**
Argentina[1]	1 028
Bolivia	570
Brazil	2 165
Chile	2 116
Colombia	1 719
Costa Rica	123
Ecuador	766
El Salvador	305
Honduras	125
Mexico	359
Panama	152
Paraguay	76
Peru	344
Uruguay	266
Venezuela	417
Total	**10 531**

Note: [1]Argentina was used as a pilot, with only the origin treatment and without the content treatments.

B Deviations from the pre-analysis plan

The experiment and analysis differed from the pre-analysis plan (PAP) in two dimensions. The PAP is accessible here.

The first deviation from the PAP relates to the sample size. We had anticipated a sample size of approximately 15,000 clubs to be contacted. However, only approximately 13,000 clubs could be identified, and of those, approximately 3,000 ($\sim 23\%$) returned an error. The errors can be divided into three broad categories: 1) the contact information that we found was incorrect (returned an error message), 2) the contact information existed but did not belong to a football club (another sport, or not a sports club at all), and 3) the club only had a youth team, so the application was irrelevant. For any of these errors, we removed the club from the analysis. The smaller sample size slightly reduces our statistical power, but we are still be able to identify very small effects in magnitude. Related to sample size, we also anticipated, based on our pilot study in Argentina, that the number of clubs asking for further information about the players would be approximately 600, but the actual number was much lower, thus significantly reducing our statistical power. In any case, the point estimate for the difference between foreign and local names in the second round falls well below the minimum detectable effect that we had initially planned.

The second deviation from the PAP relates to the treatment allocation. In the PAP (Figure 1), we had planned for an equal distribution of the sample between the four treatment groups (Control, Quality, Language, Quality + Language treatments). However, because many of the countries of origin of the applicants shared the same language as the country of the clubs (particularly in Latin America), the language treatment did not make any sense for these countries. For this reason, we decided to reallocate the clubs that were randomly assigned to T2 (resp. T1+T2) to be assigned to the T0 (T1) group. Therefore, the sample sizes for treatments T2 and T1+T2 are lower than expected, and the sample size for T0 and T1 are larger than expected (Figure 1).

C Text of the application

Hello,

I have just arrived in the city. I come from city Y (capital country) and I am looking for a club to meet people and enjoy playing football. I have already played at a similar level.

When I was young, I played in the academy of Cremonese. ($T1$: Individual quality treatment)

I spent my high-school in Spain/Portugal so I speak Spanish/Portuguese well. ($T2$: Shared language treatment)

Could I come for a trial training session with the senior team?

Thank you very much

Name

D Text of the second application

Note: This message was sent to clubs who requested further information about the player. The same text was sent to clubs in all treatments.

Hello,

Thank you for your reply. I recently moved to Name city and I found the name of your club on the Internet. I am 26 years old and in recent years I have played in various teams at the amateur level. I usually play in the midfield, in particular as an attacking midfielder. I am now looking for a team and decided to contact you to see if I could come to a trial session. Do you think I could come to a training session with the team?

Thank you very much

Name

E List of countries of origin for applicants

Table E.1: List of countries of origin per country

Country of clubs	Latin America	Europe	Asia
Argentina	Bolivia	Italy	China
	Peru	Spain	South Korea
	Chile	France	Japan
	Paraguay	Germany	Taiwan
Bolivia	Argentina	Spain	Japan
	Brazil	Germany	South Korea
	Peru	Italy	
	Mexico	France	
Brazil	Venezuela	Portugal	Japan
	Paraguay	Italy	China
	Bolivia	Spain	Lebanon
	Argentina	Germany	South Korea
Chile	France	Bolivia	China
	Germany	Colombia	
	Italy	Peru	
	Spain	Venezuela	
Colombia	Venezuela	Spain	China
	Ecuador	Italy	Japan
	Argentina	France	
	Peru	Germany	
Costa Rica	Nicaragua	Spain	Hong Kong
	Venezuela	Italy	
	Colombia	Germany	
	El Salvador	France	
Ecuador	Colombia	Spain	China
	Peru	Germany	South Korea
	Chile	Italy	Japan
	Venezuela	France	Israel
El Salvador	Honduras	Spain	China
	Guatemala	Italy	South Korea
	Nicaragua	Germany	Japan
	Mexico	France	

Note: The table presents the countries of origins of the largest number of immigrants in each country, with a maximum of four by continent.

Table E.2: List of countries of origin per country, continued

Country of clubs	**Latin America**	**Europe**	**Asia**
Honduras	Guatemala	Spain	China
	El Salvador	Germany	Palestine
	Nicaragua	Italy	Japan
	Mexico	France	Israel
Mexico	Guatemala	Spain	Japan
	Venezuela	France	China
	Colombia	Germany	India
	Honduras	Italy	Israel
Panama	Venezuela	Spain	China
	Colombia	United Kingdom	India
	Nicaragua	Italy	Philippines
	Dominican Republic	Germany	Israel
Paraguay	Brazil	Germany	South Korea
	Argentina	Spain	Japan
	Uruguay	Italy	China
	Chile	Poland	
Peru	Venezuela	Spain	China
	Bolivia	Germany	Japan
	Argentina	Italy	South Korea
	Colombia	France	India
Uruguay	Argentina	Spain	China
	Venezuela	Italy	
	Brazil	Germany	
	Chile	Poland	
Venezuela	Colombia	Spain	Syria
	Peru	Italy	China
	Ecuador	Germany	Lebanon
	Chile	France	Saudi Arabia

F Results: Additional figures and tables

Table F.1: Differences by origin of applicant

	Foreign	Regions
	(1)	(2)
Foreign	−0.023	
	(0.040)	
Asia		-0.124^{**}
		(0.051)
Europe		0.082^{*}
		(0.048)
Latin America		−0.038
		(0.050)
Constant	-0.898^{***}	-0.896^{***}
	(0.055)	(0.055)
N	10,821	10,821
Pseudo R^2	0.072	0.075

Note: The dependent variable is a dummy variable for whether the applicant received a positive response. * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses. Controls include country and means of contact fixed effects.

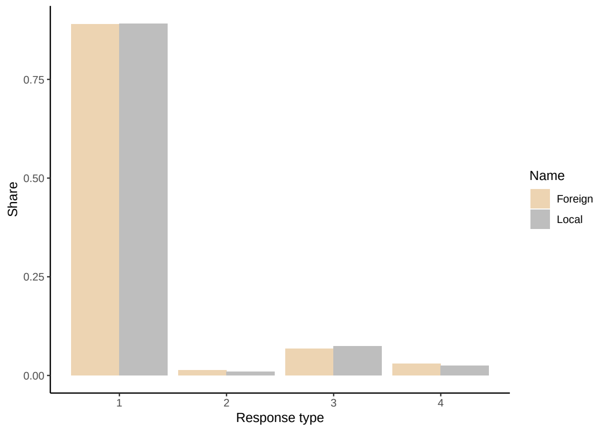
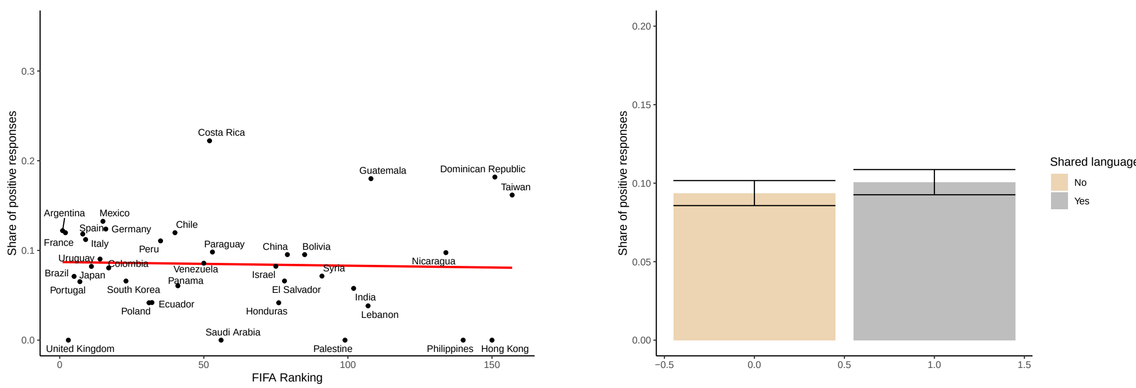


Figure F.1: Types of responses, by treatment



(a) Individual statistical discrimination: FIFA ranking. $p = 0.007$

(b) Collective statistical discrimination: shared language. $p = 0.74$

Figure F.2: Presence of individual and collective statistical discrimination

Table F.2: Discrimination in the second round

	Foreign (1)
Foreign	-0.223
	(0.253)
Constant	-2.146^{***}
	(0.639)
N	310
Pseudo R^2	0.166

Note: The dependent variable is a dummy variable for whether the applicant received a positive response in the second round. * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses. Controls include country and contact means fixed effects.

G Results: Alternative definition for positive response

Table G.1: Statistical discrimination, alternative definition

	Probability of positive callback, alternative definition (1)
FIFA ranking	−0.001
	(0.001)
T1: Quality treatment	−0.103
	(0.091)
FIFA ranking x T1	0.001
	(0.002)
Shared Language	−0.066
	(0.062)
T2: Language treatment	0.065
	(0.098)
T1 x T2	−0.055
	(0.133)
Constant	−1.412***
	(0.078)
N	10,531
Pseudo R^2	0.073

Note: The dependent variable is a dummy variable for whether the applicant received a strictly positive response (i.e., not asking for further information). * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses. Controls include country-fixed effects.