Language Skills and Homophilous Hiring Discrimination: Evidence from Gender-and Racially-Differentiated Applications

Anthony Edo, Nicolas Jacquemet, Constantine Yannelis

Documents de Travail du
Centre d'Economie de la Sorbonne

CES Working Papers

Language Skills and Homophilous Hiring Discrimination: Evidence from Gender- and Racially-Differentiated Applications

Anthony EDO, Nicolas JACQUEMET, Constantine YANNELIS

2013.58

Maison des Sciences Économiques, 106-112 boulevard de L'Hôpital, 75647 Paris Cedex 13
http://centredeconomiesorbonne.univ-paris1.fr/bandeau-haut/documents-de-travail/
ISSN : 1955-611X

Language Skills and Homophilous Hiring Discrimination: Evidence from Gender- and Racially-Differentiated Applications*

Anthony Edo† Nicolas Jacquemet‡ Constantine Yannelis§

October 2013

Abstract

This paper investigates the importance of ethnic homophily in the hiring discrimination process, and provides a novel test for statistical discrimination. Our evidence comes from a correspondence test performed in France, in which we use three different kinds of ethnic identification: French sounding names, North African sounding names, and "foreign" sounding names with no clear ethnic association. Within both male and female groups, we show that all non-French applicants are equally discriminated against when compared to French applicants. This indicates that racial discrimination in employment is directed against members of non-majority ethnic groups, and highlights the importance of favoritism for in-group members. Moreover we find direct evidence of homophily: recruiters with European names are more likely to call back French named applicants and female recruiters are more likely to call back women. The paper also directly tests for statistical discrimination by adding a signal related to language skill ability in all resumes sent to half the job offers. Although the signal inclusion significantly impacts the discrimination experienced by non-French females, it is much weaker for male minorities.

Keywords: Correspondence testing, Gender discrimination, Racial discrimination, Ethnic homophily, Language skills.

JEL Classification: J15, J64, J71.

*We gratefully acknowledge David Neumark for sharing his data and estimation programs with us. We also thank Francis Bloch, Nick Bloom, Emmanuel Duguet, Christelle Dumas, Raquel Fernandez, Stéphane Gauthier, James Heckman, Shelly Lundberg, Muriel Niederle, Phillip Oreopoulos, Chris Taber, Marie-Anne Valfort as well as participants to various conferences and seminars for their thoughtful comments on the paper. We are grateful to the CEPREMAP for financial support. Nicolas Jacquemet acknowledges the *Institut Universitaire de France* for its support. Constantine Yannelis thanks the Alexander S. Onassis foundation for generous support.

†Paris School of Economics and University of Paris I Panthéon-Sorbonne. Centre d'Economie de la Sorbonne, 106-112 Boulevard de l'Hopital, 75647 Paris, France. `Anthony.Edo@univ-paris1.fr`.

‡Université de Lorraine (BETA) and Paris School of Economics. 3 Place Carnot, 54035 Nancy. `Nicolas.Jacquemet@univ-lorraine.fr`

§Stanford University. 579 Serra Mall, Stanford, California 94305 `yannelis@stanford.edu`.

Language Skills and Homophilous Hiring Discrimination: Evidence from Gender- and Racially-Differentiated Applications

Anthony Edo Nicolas Jacquemet Constantine Yannelis

Octobre 2013

Résumé

Cet article s'intéresse au rôle de l'homophilie ethnique dans la discrimination à l'embauche, et propose une nouvelle façon de tester l'existence d'une discrimination statistique. Notre analyse porte sur des tests de correspondance effectués en France dans lesquels nous distinguons trois types d'identification ethnique: des noms à consonance Française, à consonance maghrébine, et à consonance étrangère sans attache claire à une origine particulière. Nos résultats montrent que la discrimination à l'embauche opère indifféremment envers tous les noms à consonance étrangère, tant chez les hommes que chez les femmes. Ce résultat indique que la discrimination ethnique s'effectue à l'encontre de tous ceux qui n'appartiennent pas à la majorité ethnique, et souligne l'importance du favoritisme au sein de groupes ethniques. Nous montrons qui plus est l'existence de l'homophilie : les recruteurs au nom européen ont tendance à rappeler davantage les noms à consonance française, et les femmes ont tendance à rappeler davantage des femmes. Cet article teste également l'existence d'une discrimination statistique en ajoutant aux CV un signal portant sur les compétences linguistiques pour la moitié des offres d'emploi. Si l'inclusion du signal réduit fortement la discrimination envers les noms féminins à consonance étrangère, cet effet est bien plus faible pour les noms masculins à consonance étrangère.

Mots-clés: Evaluation par correspondance, Discrimination de genre, Discrimination raciale, Homophilie ethnique, Compétences linguistiques

1 Introduction

Homophily, the idea that people apply preferential treatment to similar individuals, has long-standing roots in both sociology (Lazarsfeld and Merton, 1954; Hamm, 2000; Mollica, Gray, and Treviño, 2003; Wimmer and Lewis, 2010) and psychology (Vigil and Venner, 2012). While a strong interest in homophilous behavior has arisen during the last decade in the analysis of network formation in economics, it is only recently that scholars interested in discriminatory practices have started to investigate the opposition between one's own racial group and *any* other ethnic origin (Stoll, Raphael, and Holzer, 2004; Giuliano, Levine, and Leonard, 2009).

This paper contributes to this literature in two ways. First, we offer an experimental investigation of racial homophily, based on correspondence testing, while this previous evidence relies on observational data. Second, we explicitly test for statistical discrimination and provide evidence regarding the underlying reason behind homophily in hiring decisions. In the same way that discrimination is traditionally rationalized as a matter of either preferences (in line with Becker (1971)'s model of taste based discrimination) or information (as suggested by the statistical discrimination models of Arrow, 1973; Phelps, 1972) individuals can associate with similar individuals because of a preference for homogeneity, through *e.g.,* conformity motives, or because they tend to share common characteristics that make communication, mutual trust and relationship formation easier (McPherson, Smith-Lovin, and Cook, 2001; Putnam, 2007). From an empirical point of view, one testable difference between the two explanations is that only the second one is sensitive to the available information on applicants. This kind of test requires focusing on specific dimensions about which a (real or perceived) lack of information can induce discrimination. We focus on language abilities, a skill raised by employers to rationalize their own discriminatory behavior (Oreopoulos, 2011).

Our measure of homophily relies on the design of Jacquemet and Yannelis (2012). They measure homophily driven discrimination in the Chicago labor market by comparing the disadvantage experienced by clearly identified minorities – African-Americans in their context – to the one faced by foreign applicants for whom no clear origin is identified by the employer. We replicate the experiment in France, using North African applicants as a benchmark, and interact race and gender by using two sets of racially differentiated applications – one male, one female. We control for the potentially confounding effect of religion (documented, *e.g.*, by Adida, Laitin, and Valfort, 2010) by selecting foreign names that are not perceived as being Muslim. To test for the effect of enhanced information about language skills ability, we alter the resumes sent to half the job postings by adding language related extra activities or grades.

Our results are three-fold. First, we find strong evidence in favor of ethnic homophily in the French labor market. Foreign applicants, whether their specific minority group is identified or not, are equally disadvantaged as compared to French applicants across all dimensions under study – for both genders, and whether or not more information is available in the application. Second, we find evidence of statistical discrimination for female applications: once a signal related to language skill abilities is included, racial discrimination is drastically reduced. Third, we find an asymmetric effect of the signal across genders, as for males enhanced information barely changes the success of applications. One interpretation is that information has nothing to do with males being discriminated against; another one, which is more consistent with the pattern observed for women, is that the lack of information driving racial homophily amongst males applicants is not related to language skills but to other individual characteristics.

The next section summarizes the existing literature on both hiring discrimination and ethnic homophily. Section 3 describes our experimental design and its implementation in Paris and its suburbs. In Section 4, we provide an overview of the outcomes of the experiment based on descriptive statistics. Section 5 turns to a statistical analysis of the data which is robust to the bias induced by racial group differences in the distribution of unobservables (Heckman and Siegelman, 1993). We discuss the possible interpretations of our results in Section 6 and Section 7 concludes.

2 Discrimination and Racial Homophily

Accumulated evidence from correspondence studies robustly supports two main conclusions: discrimination against minority applicants, with discrimination ratios generally ranging from 1.3 to 1.7.[1] This is true in many countries with very diverse immigration and colonial histories – USA (Bertrand and Mullainathan, 2004), Canada (Oreopoulos, 2011), Australia (Booth, Leigh, and Varganova, 2012), Sweden (Carlsson and Rooth, 2007), France (Duguet, Leandri, L'Horty, and Petit, 2010); see Riach and Rich (2002) for a complete survey of existing results. Recent evidence also shows that such results vary little according to the specific minority used to test for discrimination. In Canada, Oreopoulos (2011); Dechief and Oreopoulos (2012) show that Asian, Indian and Greeks all fare similarly poorly compared to native Canadians. Duguet, Du Parquet, L'Horty, and Petit (2012) study the success of male and female applications of Senegalese, Moroccan and Vietnamese origin in France. Although they focus on a high-skill occupation (computing with a Masters degree), they obtain discrimination ratios that roughly fall in the above mentioned range, and which are similar against most non-French applicants.[2] These studies rely on correspondence tests, which have the well-known drawback that no "real" person is behind experimental applications. The strength of discrimination, and that applicants from the ethnic majority group are always favored, has however been confirmed by recent evidence from an audit study based on Latino and Black applications in New-York (Pager, Western, and Bonikowski, 2009).

Such a stability across countries and across minorities within countries suggests that a more encompassing mechanism may underlie observed discrimination. The sociological notion of homophily

[1] Discrimination ratios are defined in the literature as the relative favoritism experienced by non-minority applicants as compared to minority ones; denoting C the callback rate, with index 1 for minority applicants and 0 for non minority ones, the discrimination ratio is : C_0/C_1.

[2] Vietnamese women stand as one unexplained exception, facing almost no discrimination. The comparison with other applications is difficult, though, as the resumes include gender specific information for this origin only.

points to such a mechanism, based on the principle that "birds of a feather flock together". This literature (reviewed in McPherson, Smith-Lovin, and Cook, 2001) not only shows that homogeneity is a driving force of social networks formation, but also that ethnicity is certainly the most influential factor of this process. While homophily is a well recognized phenomenon in the economics literature on network formation, the application to the effect of race on labor market outcomes only recently emerged (see Lang and Lehmann, 2012; Charles and Guryan, 2011, for up to date reviews of the theoretical and empirical literature on hiring discrimination). Based on observational data from four cities in the U.S., Stoll, Raphael, and Holzer (2004) shows that black employers tend to hire more black applicants, not only because they receive more black applications but also because they hire a greater proportion of blacks who apply. Using data from a large U.S. retail firm, the study by Giuliano, Levine, and Leonard (2009, 2011) finds an own-race bias in manager-employee relationships in regards to quits, dismissals and promotion.[3] Jacquemet and Yannelis (2012) are the first to study homophily experimentally, based on a correspondence test in Chicago including foreign names with no clear ethnic identification. They show that the discrimination rates against these applicants are the same as the one experienced by African-American applicants in all fields under study – accounting, nursing and programming.

This paper replicates the Jacquemet and Yannelis (2012) measure of homophily based discrimination, with three additions. First, the experiment is implemented in France instead of Chicago, providing a robustness check of the original results. Second, we interact racial discrimination with gender discrimination by adding a male equivalent to each female applicant. Third, we consider three different levels of job occupations (in the same sector) to assess the sensitivity of our results to the kinds of skills required. Our results generalize Jacquemet and Yannelis (2012) to the French context as well as to both males and females, and to all occupational categories under study. We take this as robust evidence that homophily has explanatory power on observed discrimination in hiring. Our main treatment of interest tries to disentangle the reasons behind this behavior.

Two main reasons have been raised to explain homophilous behavior in social relationships.[4] Currarini, Jackson, and Pin (2009) develop a model of network formation in which homophily results from an intrinsic individual preference for similar individuals. Applied to racial discrimination, this mechanism is very similar to the taste-based model introduced by Becker (1971). The laboratory

[3]Recent evidence shows that the idea that racial homophily explains discriminatory behavior extends beyond labor market outcomes. For instance, Price and Wolfers (2010); Price, Lefgren, and Tappen (2013) find evidence of a strong own-race bias in the number of fouls awarded against NBA players.

[4]These two explanations refer to what sociologist call *choice homophily* (Kossinets and Watts, 2009). A third one, formalized for instance by Jackson and Rogers (2007); Bramoullé and Rogers (2009) in the context of network formation, explains homophilous behavior by a higher probability of meeting an in-group partner – this is called *induced homophily*, because homophily results from the structural opportunities to interact. Because all applications are sent to each employer, correspondence testing is ill-equipped to provide reliable measures of such a phenomenon. It does not mean it cannot contribute to the observed differences between applicants – it will be the case, for instance, if employers' decisions are correlated to the relative composition of the overall application pool, and the pool is not exogenous to the employers' characteristics. In the empirical part, we will include covariates related to the employer location to provide some control over this channel.

experiments evidence reported in Habyarimana, Humphreys, Posner, and Weinstein (2007) tend to support a different, belief-based, explanation. The experiments take place in the neighbourhood of Kampala, the Uganda's capital, and use as a treatment variable the ethnicity of the partners of the game by making use of the high level of diversity in terms of tribes in Uganda. The evidence relies on standard social preferences games and shows that: *(i)* higher homogeneity inside groups is associated with much higher voluntary contributions to a public good, while at the same time *(ii)* there is no difference in the level of gift chosen in a dictator game. It is only when the tribe of both players is common knowledge (the receiver knows what tribe the sender belongs to) that people treat others from their own tribe more favourably. The preferred interpretation of the authors is the existence of a norm within ethnic groups, according to which cooperation among co-ethnics should be reciprocated. Such a pattern in social behavior indicates that beliefs and information on who the others are; as well as compliance to what the others are believed to expect, plays a greater role than preferences.[5] Since information, rather than the shape of preferences, drives discriminatory behavior in this case, this mechanism echoes the statistical view of racial discrimination introduced by Arrow (1973); Phelps (1972).

In the context of hiring discrimination experiments, Oreopoulos (2011); Dechief and Oreopoulos (2012) are the first to systematically manipulate the content of the applications in order to investigate the extent of statistical discrimination. These experiments use a very rich set of resume characteristics as treatment variables – language fluency, multinational firm experience, active education from highly selective schools, extracurricular activities. None of them influences the discrimination observed, in Canada, against both male and female applicants from Indian, Pakistani, Chinese or Greek origin. The authors also conducted debriefing interviews with recruiters, from which language skills appear as a major concern – although mentioning fluency on the resume does not affect any of the discrimination ratios. This concern is also very likely to be faced by French employers, as many second and third generation immigrants have distinctive ways of speaking (Trimaille, 2004).[6] In particular, they may have more difficulties in speaking, writting or communicating with their managers and/or consumers.

Building on this evidence, we seek to measure the extent of statistical discrimination associated with language skills. Although the reasons why employers discriminate cannot be observed directly, the extent of available information has observational consequences that allow the researcher to disentangle between statistical and taste based discrimination. As employers learn more about the productivity of workers, they will rely less on noisy signals that are correlated with productivity, such as race. Altonji and Pierret (2001), for instance, use the arrival of information associated to increase experience in the same firm to test for statistical discrimination in wage setting.[7] We develop

[5]This result is supported by McPherson, Smith-Lovin, and Cook (2001); Putnam (2007) who suggest that individuals may cooperate with similar others for ease of communication, mutual trust and closed cultural features that smooth the coordination of activity.

[6]This feature is supported by Vallet and Caille (1996) who show that, in 1989, foreign pupils and children of immigrants overall achieved lower scores than other pupils in a French assessment test.

[7]Knowles, Persico, and Todd (2001); Antonovics, Arcidiacono, and Walsh (2005); Levitt (2004); Pope and Sydnor

a similar strategy, which compares discrimination in cases where a prospective employer has more or less information about an individual's underlying ability. Our treatment variable is an additional resume experience (detailed below) related to the practice of French language. We treat this addition as a between job treatment variable, adding it to all applications sent to half the job offers. If perceived language skills are correlated with both ethnic background and productivity, prospective employers that statistically discriminate should rely less on ethnic background to screen applicants. Hence, the signal premium should be higher for minority applicants implying a lower level of racial discrimination. However, if employers are engaging only in taste based discrimination the signal of language abilities should have no effect.[8]

We find a strong effect of the signal on discrimination, but mainly among females, indicating that statistical discrimination is present in the French labor market. For female applicants that include a signal for French language ability, the measured level of discrimination towards minority groups drops substantially. For male applicants – who are systematically disfavored as compared to their female counterpart – the effect of the signal is much weaker, both statistically and economically.

3 Design of the Empirical Analysis

The empirical analysis relies on three main ingredients: the identity of fictitious applicants; the content of the applications and the job offers to which we apply on this basis.

3.1 Treatment Variable I: Fictitious Applicants

We rely on six fictitious identities. To test for gender discrimination, three of them are male and three are female. Each of the three correspond to a different origin. Our benchmark is French sounding names, to which we add two origins: North African sounding names and names perceived as *(i)* foreign but with *(ii)* no clear ethnic origin, so that they are unidentifiable by employers.

(2011) also exploit the arrival of new information to test for statistical discrimination in motor vehicle searches, a game show and an online peer to peer lending service, respectively.

[8]Correspondence tests have the attractive feature of being experimental, thus generating reliable causal effects. They also have well-known drawbacks, to which our study will be no exception. First, the outcome variable is the callback rate elicited by experimental applications rather than actual job propositions. Correspondence tests may under or overstate the level of discrimination in hiring since companies generally have multistage recruiting processes, which eventually require personal interviews. The estimates overstate discrimination only if *(i)* discrimination occurs at the job interview stage, and *(ii)* operates in the opposite direction – *e.g.*, minority applicants are favoured once they are called back. This configuration is generally seen as highly unlikely. Second, for obvious practical reasons, the study is restricted to job offers that circulate through public information channels, thus excluding all positions fulfilled through individual networks. This induces two potential sources of bias: first, it can be that job finding though network compensates the observed difference in success across applicant types; second, there is a potential selection effect in the pool of job offers. Existing evidence from this literature – initiated by Granovetter (1973) and surveyed *e.g.*, in Ioannides and Loury (2004) – is mixed as regards the interaction between ethnic origin and job finding through individual networks. The composition effect of the job offers pool leads to a local treatment effects – *i.e.* the study elicits discrimination for only a sub-part of such a dual labor maket. Again, there are good reasons to think that the estimate provides a lower bound of actual discrimination based on this selected sample, as empirical evidence tends to support the existence of ethnic homophily in social networks.

The choice of the names is grounded on a preliminary survey that asks respondents to indicate to what origin, if any, they associate the names. We gather further control information by also asking respondents their guess for a gender and a religion associated to the name.[9] The survey mixes a sample of 32 names with clearly identifiable French and North African names as well as a set of names intended to be unidentifiable. The sample of names has been created based on public sources of information and the study by Duguet, Leandri, L'Horty, and Petit (2010). The survey has been conducted in various area in Paris using 150 employees mainly from the public sector (nurse, nursery nurse and temps) and 150 college students *(i.e.* a total of 300 respondents). We use public employees and college students for several reasons. On one hand, public workers may *(i)* share similar perception than employers in determining the origin of names, *(ii)* without being implicated in any decision process to recruit applicants. In fact, this population of workers does not belong to human resource departments. It is therefore unlikely that respondents see the names in the survey at later stages of the experiment once resumes are sent out. On the other hand, the sample is complemented with college students since it is a much simpler population to investigate.

The results of the survey for the six names we consequently selected for the study are displayed in Table 1. The results confirm very high rates of correct guesses in terms of origins and gender for both French and North African sounding names, suggesting a very high reliability of the treatment variable. Second, the share of respondents who are unable to associate any particular origin to the last two names (leaving the field blank or putting a question mark) is always higher than two thirds. This means both that these names are largely perceived as non-French, and at the same time that no specific immigrant group is associated to these names. In spite of this high uncertainty about the names' origins, each one is associated with one particular gender for a large majority of respondents. These two names will be used in the study as the male and female treatment variable used to test ethnic homophily, *i.e.* discrimination against applicants for the only reason that they do not belong to the majority ethnic group.

One potential challenge to such an interpretation of this set of names is that their treatment effects could potentially be driven by strong and consistent beliefs from a minority of employers. Inspection of the middle column of Table 1 discards such an interpretation. Among the 17% of respondents answering the origin question about ALDEGI Jatrix, and the 30% answering it for HADAV Alissa, the guesses are dispersed in terms of both geographic location and religion.[10] Interestingly, both names are associated by a majority of respondents to different religions: Christian for Jatrix, Jewish for Alissa. These are also different from the perceived religion of the North African names, which are unsurprisingly identified as Muslims, while French names are widely identified as Christian.[11]

[9] The question about religion was not included in the first wave of the survey. We ran a second wave, with again 300 respondents, including only this question. We decided not to ask the origin again to this second sample in order to avoid that respondents mechanically relate religion to origin.

[10] The most frequent origins are Eastern Europe (5%) and Southern Europe (3%) for Jatrix Aldegi; and North African (9%) and Israeli (6%) for Alissa Hadav.

[11] Among the residual respondents about the religion question, Jatrix is identified as Jewish by 16% of the sample

Table 1: Perceived Origins, Gender and Religion for the Six Identities used in the Study

Names	Origin and Sex Guesses			Other Perceived Origin				Perceived Religion	
1. French	Correct	M	F						
LECLERC Pascal	99%	97%	1%	Unknown	1%	-		Christian	96%
ROUSSET Sandrine	97%	1%	98%	Unknown	2%	-		Christian	95%
2. North African	Correct	M	F						
BENBALIT Rachid	94%	96%	3%	Unknown	2%	Israeli	2%	Muslim	94%
BENOUNIS Samira	92%	1%	99%	Israeli	3%	Unknown	3%	Muslim	77%
3. Foreign	Unknown	M	F						
ALDEGI Jatrix	83%	73%	10%	East Eur.	5%	South Eur.	3%	Christian	67%
HADAV Alissa	70%	1%	83%	North Afr.	9%	Israeli	6%	Jewish	55%

Notes: The table reports the perceived origins, gender and religion of names for the six identities used in this study. The first two variables have been collected on a sample of 300 respondents; the religion question has been asked in the a second wave with 300 other respondents. The first column shows the share of correct guesses for the two first sets of names (French and North African) and the share of respondents who do not identify the origin of Foreign names. The percentage of respondents who perceived the name as male or female is indicated in columns M (male) and F (female). In addition to the prevalent guess, the *middle* columns display the most frequent answers among residuals respondents. Finally, the *right-hand* side of the table provides the most likely perceived religion of those names.

3.2 Treatment Variable II: Content of the Applications

To maximize the job offers collected during the study, we select a dynamic sector which is less sensitive than others to economic fluctuations hence offering a large quantity of vacancies. We focus on accounting jobs, in particular vacancies advertising openings for accounting assistant, accounting secretary and accountants.[12]

We construct six resumes from actual ones accessible online and alter them to create a distinct set that would not be associated with their owners. Resumes are further arranged so that they differ in terms of content and form (layout, style and typeface) to limit the risk of detection. These small differences between applications could induce uncontrolled systematic differences in the perceived quality of resumes by the employer. To orthogonalize the differences in the content of the application and the applicant's identity, a random rotation system is implemented across names and resumes

and as Muslim by 9% of the sample; Alissa is perceived as either Christian (26%) or Muslim (14%).

[12]This choice will also ease comparison with Duguet, Leandri, L'Horty, and Petit (2010), who selected the same sector. We moreover chose jobs with varying shares of women – females amount to 66% of people employed in accounting jobs and 84% of people in assistant and secretary jobs (*source*: French labor force survey, 2010 wave).

from one job offer to the other.

To limit the noise in the relative success of applicants introduced by the randomization, we constrained (resume specific) characteristics of the applicants to be similar according to the following dimensions. Applicants were born in 1988 (aged 23) and have never repeated a year. They all are single. In line with the habits of the French labor market, the nationality of the applicant never appears on the resume – which signals they all are French. The postal addresses are also application specific (they do not change as the applicant name changes) and have been chosen within the same area of the South of Paris.[13] Applicants have a BTS (*Brevet de Technicien Supérieur* - BTS) in accounting and management obtained in 2009.[14] The educational background of the six fictitious applicants is also similar, and consistent with their educational attainment. Applicants are not employed when they apply to show their immediate availability. Their work experience ranges from 18 to 22 months composed of two or three different spells, whose job titles and corresponding job descriptions are real but slightly altered and randomly assigned to the six resumes. Moreover, a motivation letter is created for each application, whose form is consistent with the associated resume. In order to be realistic and representative, the motivation letters are also created from multiple examples accessible online. Last, we account for gender differences between our applicants by adding clear and implicit indicators of gender through spelling in resumes and/or motivation letters. For instance, we added gender links such as "assistant" for male and "assistant<u>e</u>" for female, as well as some abbreviation like "M." and "Mlle".

All of the above features are common to all applications. The content of the applications is also used as a treatment variable, through additional skill signals related to language ability. In the applications sent to half the job postings, we add one of each of the six following sentences on all resumes sent to the employer:

- Tutorial for pupils with difficulties in lecture and redaction – *Work Experience Category* ;
- French Tutoring – *Work Experience Category* ;
- Member of a reading club – *Leisure Category* ;
- Participation in Scrabble and Crossword competition – *Leisure Category* ;
- Writer of an inter-college newspaper – *Leisure Category* ;
- Reward for a well-known competition on French language skills in 2003 (rank: 54/8,500) – *Leisure Category* ;

[13] Although Duguet, Leandri, L'Horty, and Petit (2010) show that location of the applicants strongly interacts with racial discrimination, our design neutralizes this dimension.

[14] According to the International Standard Classification of Education (ISCED), this level of education corresponds to the first stage of tertiary education. This educational attainment is the most requested for accounting jobs.

The signals are neither application- nor identity- specific: we match them to applications and identity using an additional systematic rotation scheme. All signals are strongly related to language skills – two of them refer to past experience in French teaching, the others are related to past leisure or extra activities centered around use of the French language.

3.3 Implementation of the Experiment

The experiment was conducted between September 2011 and February 2012 in Paris and its suburbs. We responded daily to job advertisements provided by the major employment agency *Pôle Emploi*. Other sources of job vacancies, such as `APEC.fr` and `cadreemploi.fr`, were also used to complete our sample. No unsolicited applications were sent. Jobs listings are included in the study according to the following criteria: full or part-time job, short or long-term contracts (excluding temporary jobs) and located in the Ile-de-France (Paris and its suburbs). Offers were not included if they required a Master's degree, some specific work experience, or more than two years of experience.

We sent six applications (one for each name in Table 1) to each job offer. They are sent sequentially and in random order, usually between 10 AM and 2 PM on the same day as it appeared on the website. Resumes are addressed to employers in word format via direct email.[15] For each fictitious identity, an email address and a telephone number (including an automatic answering service) were registered at large internet providers (*i.e.* Yahoo, Laposte, Hotmail or Gmail) and at a phone company (SFR). All of this information is applicant-specific and moves across resumes along with the identity of the applicant. Our main outcome variable is the number of callbacks elicited by each applicant. In order to minimize the effect of the study on the labor market, any invitation is promptly declined by email.

Because almost all job advertisements are collected through *Pôle Emploi*, we gather several additional control variables. First, the advertisements are standardized and provide a multitude of information on the number of employees in the firm, its location, the term of job contract, as well as the gender and name to whom we send the applications. Second, the agency works almost only with employers themselves, rather than specialized intermediaries such as interim or recruitment agencies. This allows us to consider the name provided as the (likely) recruiter's one, from which we build an origin and gender variable.

4 Descriptive Statistics

Overall, 504 job vacancy postings complied with the inclusion criteria during the period of the study; to which we sent 3024 resumes. As a benchmark with existing results, Table 2 provides the overall success rate of each applicant, by gender (in row) and origin (in column). Three lessons emerge from two-by-two comparisons.

First, applications with French names always elicit far more callbacks than non-French ones, with an overall rate of 17% against 10%. The overall discrimination ratio is equal to 1.7. Second,

[15] In a few cases, applications were sent by postal mail.

Table 2: Success Rates by Origin/Gender with and without the Inclusion of the Signal

	French Names			North African Names			Foreign Names		
	Control	Signal	Overall	Control	Signal	Overall	Control	Signal	Overall
Male	14.3%	15.5%	14.9%	6.0%	8.3%	7.1%	8.3%	7.9%	8.1%
Female	20.6%	18.7%	19.5%	10.7%	14.7%	12.7%	9.1%	15.1%	12.1%
Total	17.5%	17.1%	17.3%	8.3%	11.5%	9.9%	8.7%	11.5%	10.1%

Signal Effect on Racial Discrimination

	H_1: French≠North African			H_1: French ≠ Foreign			H_1: North African ≠ Foreign		
Male	4.17***	3.48***	5.40***	2.73***	3.91***	4.63***	-1.61	0.28	-0.96
Female	4.25***	1.77*	4.29***	4.85***	1.57	4.57***	0.94	-0.23	-0.49
Total	5.35***	3.70***	6.46***	4.96***	3.58***	6.09***	-0.36	0.00	-0.24

Gender: t-tests on Mean Callback Differences (H_1: Male ≠ Female)

	-2.63***	-1.30	-2.77***	-2.47***	-2.87***	-3.79***	-0.39***	-3.48***	-2.74***

Notes: ***, **, * denote significance at the 1%, 5%, 10% level. The *upper* part of the table displays the callback rates elicited by each applicant, by gender (in row) and race (in column). We provide the overall share of elicited callbacks, as well as its decomposition according to whether resumes included the signal or not. The total number of applications is 504 for each identity, and equal to 252 for the sub-samples with and without the signal. The *middle* part provides the statistics of student t-tests of equality in mean callbacks between origins. The *lower* part of the Table provides the statistics of student t-tests of equality in mean callbacks between gender of the same origin.

a decomposition of callback rates for each non-French origin indicates that North African (9.9%) and Foreign applicants (10.1%) are equally treated, and both disfavored relative to French ones (17.3%). The fact that Foreign applicants are discriminated against while their ethnic group is not identified by employers confirms the prevalence of ethnic homophily among employers - *i.e.* a general mistrust against all members of the non-majority ethnic group. These two observations remain true conditional on the applicant's gender: the degree of racial discrimination against both Non-French sets of names is similar inside each gender group. Third, in terms of gender based discrimination, females are considerably more successful than males in getting through the initial job screening stage: the gap ranges from 19.6% vs 14.9% among French applications, to 12.7% vs 7.1% for North African sounding names and 12.1% vs 8.1% for Foreign sounding ones. We also find some evidence of intersectionality (the interrelation between several sources of discrimination, see *e.g.*, Crenshaw, 1991) as racial discrimination ratios are more severe amongst males (around 2) than amongst females (1.6). Such a difference in the strength of discrimination faced by men and women from the same origin is compatible with either of the two following interpretations: the underlying mechanisms through which discrimination operates are the same, but they are weaker against women; or the reasons underlying discrimination against men and women from the same origin are different. The decomposition of callback rates according to whether a signal is included on the resumes provides some empirical evidence supporting the second interpretation.

4.1 Differences according to Language Skills Signals

Our main treatment variable of interest is the explicit mention of language skills ability. Table 2 also provides an overview of the comparison between the control applications (no additional skill related to language ability on the resume) and the treatment group. Focusing first on the control group, we observe that all trends discussed based on the overall rate of success are reinforced: both non-French origins are discriminated against when compared to the French applicant; this is true for both males and females, while females on average elicit more positive answers than their male counterparts.

The effect of the signal on French applications is negligible: while the inclusion of a signal generates slightly more callbacks for male applicants, it induces a moderate decrease for female ones (both are statistically not significant). These two slight variations together are enough to absorb the statistical significance of the difference between control applications – the t-test statistic of the difference between French males and females on applications once the signal is included raises to -1.30 (with a corresponding p-value equal to 0.20).

Among non-French applicants, the inclusion of a signal drastically improves the success rates, with similar magnitude among both non-French origins. The callback rate raises from 8.3% to 11.5% on average for North African sounding names and from 8.7% to 11.5% for Foreign sounding ones. However, for each of these two race categories, we notice an important asymmetry between male and female applicants. The change in callback rates is much stronger on females applications than on males ones – for whom it is positive for North African applicants, and slightly negative for Foreign ones. Among females, the raise is both economically and statistically significant:[16] thanks to the inclusion of a signal, the callback rate raises from 10.7% to 14.7% for North African applications, from 9.1% to 15.1% for foreign ones. Conditional on the signal, the difference with French female applicants remains marginally significant only for females of North African origin; for foreign ones, the callback rates are statistically the same.

[16] The t-test statistic (p-value) of equality in mean callbacks within the group of the non-French female applicants is equal to -1.86 (0.06).

Table 3: Callback Rates Decomposition According to Employers' Characteristics

	Type of Job			Firm Size		Job Contract		Ethnic Diversity		Recruiter Identity			
	Assistant (208)	Secretary (116)	Account (180)	Small (259)	Large (115)	Short (166)	Long (283)	High (400)	Low (104)	Male (203)	Female (246)	European (381)	Non-Eur. (63)
Pascal	16.8%	11.2%	15.0%	12.5%	14.8%	17.0%	13.7%	13.8%	19.2%	11.3%	17.5%	15.2%	11.1%
Rachid	6.3%	6.9%	8.3%	5.5%	8.7%	9.3%	5.9%	6.5%	9.6%	4.0%	9.8%	7.1%	6.4%
Jatrix	7.7%	6.0%	10.0%	5.8%	7.8%	9.9%	7.1%	6.8%	13.5%	4.4%	10.2%	7.1%	9.5%
Sandrine	19.2%	22.4%	18.3%	20.1%	18.3%	20.3%	19.3%	18.3%	25.0%	13.3%	26.8%	21.0%	19.1%
Samira	13.0%	16.4%	10.0%	10.6%	11.3%	14.3%	11.8%	11.3%	18.3%	9.9%	15.0%	11.6%	19.1%
Alissa	12.5%	11.2%	12.2%	10.6%	11.3%	12.6%	11.8%	11.0%	16.4%	8.9%	15.9%	12.6%	12.7%

Notes. The table distinguishes callback rates elicited by the six applicants based on type of job, firm size, type of job contract, location of the firm (based on the level of ethnic diversity at the city level) and identity (gender and origin) of potential recruiters. For each column, the total number of job offers to which we responded is in brackets.

4.2 Decomposition according to Job Offer's Characteristics

Due to the design of the experiment, the randomization of applications applies at the job level on the basis of the openings during the experiment. The aforementioned discrimination patterns may thus be subject to composition effects at both the job's and the employer's characteristics levels. Beyond the type of occupation, the characteristics of the job that we observe from the offers are the type of job contract offered (either long-term or short-term), the number of employees in the firm and its location. We use the location variable to classify jobs' location according to the share of immigrants in the neighbourhood.[17] The employer's characteristics are always hard to measure in correspondence testing studies, because the experimenter does not observe who is actually in charge of screening and selecting the applications. We rely on proxies of the employers' characteristics by using the observed identity of the person to whom we send the applications. Although it is likely that this name refers to the person in charge, this can also be the secretary of the human resource department or any other intermediary inside the organization. Such data are certainly subject to measurement error, so that the effect of employer's identity on discrimination has to be interpreted with caution. Based on this identity, we first construct a measure of the gender of the employer. We also constructed a variable reflecting the perceived origin of the employer deduced from their names. We divide the origin of employers into two categories: European and non-European.[18]

Table 3 disaggregates callbacks according to these characteristics. The first row of the Table indicates the distribution of the sample according to each dimension. Composition effects are actually far from being negligible: as the experiment does not randomize these characteristics, the sample is not balanced in most dimensions – except for the gender of the employer, which reflects the overall gender composition of the employees of this sector. However, from comparison of callback rates among applicants conditional on each employer's characteristics, the patterns describe in the previous sections does not seem to be sensitive to these decompositions: racial and gender discrimination operate in the same way across the Table's columns.

More precisely, non-French applications are disfavored within both groups of male and female applicants. The degree of racial discrimination is moreover approximately the same across characteristics. Conditional upon ethnic origin, callback rates of female applicants are always higher than those of males. The intensity of this kind of discrimination does not vary according to either the job

[17] We use the 2008 wave of the French census (providing exhaustive information on the population living on the territory) to compute the share of immigrants in the population of each city. This is linked with the location variable based on the zip code. Ethnic diversity is assumed to be low (high) if the immigrant share is lower (higher) than 20%.

[18] The gender is deduced from first names and/or abbreviations before the last names like "M." and "Mme". The decomposition of callbacks by employer origin has to be interpreted with caution inasmuch the sample size of non-European sounding names of employers is low, around 15 % of all job offers. The crucial issues to assess the extent of the bias induced by the use of these measures are *(i)* the relative size of the noise captured when our correspondent is not the actual recruiter; *(ii)* whether this noise is likely to be correlated with callbacks. On the second issue, one important factor likely driving force is the size of the firm, as this may jointly affect the likelihood that collecting the application and making the final decision are separated, and the tendency to discriminate. We do not find such a pattern once the relationships between employer's characteristics and callback rates are interacted with firm size.

contract or the ethnic composition of the job location. The type of job and the employer's gender, by contrast, seem to play a role. The gap in callback rates between male and female applications is the highest in secretary occupations (and at its lowest in accounting jobs). This is in line with the fact that female workers are overrepresented in secretarial jobs. Although this certainly drives part of the observed favoritism for women, it should be noted that females remain favored (although to a lesser extent) on all other types of jobs. Last, female recruiters, as measured by our proxy deduced from our correspondent's gender, exhibit a systematically higher callback rate.[19] They also appear as almost entirely responsible for the discrimination in favor of female applicants among French-sounding ones. Among male recruiters, both French applications are treated the same way. In order to test this potential favoritism (*i.e.* gender homophily among female), we implement two statistical tests of equality in callback rates between both male and female French applicants according to the gender of the employer. They confirm our conjecture: while callback rates between French applications are not significantly different when employers are male, they are different at a 1% level when employers are female.

5 Results

We now turn to parametric models aimed at measuring the marginal effect of discrimination, conditional on the heterogeneity of employer's and job's characteristics. We first sketch a theoretical model of employer decision making, based on Neumark (2012), showing that simple Probit estimations of callback rate equations can provide biased measures of discrimination. In a recent methodological paper, Neumark shows that consistent estimates can be derived from heteroscedastic Probit models, which account for the confounding effect of varying dispersion of abilities across groups – a limitation known as the Heckman critique in the literature, in reference to Heckman and Siegelman (1993); Heckman (1998).. To the best of our knowledge, we are the first to apply this methodology to original correspondence test data. We then turn to a formal characterization of our test for statistical discrimination, based on language skills signals.

5.1 Hiring Discrimination with heteroscedastic Unobserved Heterogeneity

The data generating process of a correspondence study stems from employers treatment of the content of the applications. We denote $P_i^*(J, X, Z)$ the productivity of an application i in a given position. This productivity depends on the job's characteristics (measured mainly through firm specific variables) J. At the individual level, productivity results from two components: a set of individual characteristics that are observable to both the econometrician and the employer, X_i; and a component which is unobservable to them both, Z_i. We assume that race, denoted $R_i = 0$ for non-minorities

[19]This result is partly explained by the fact that female recruiters are overrepresented in firms which tend to exhibit higher callbacks. Female recruiters mainly work in large firms (70%) located in Paris (56%) and offering short-term contracts (60%).

($e.g.$, whites) and $R_i = 1$ for minorities ($e.g.$, North Africans), do not enter productivity directly.[20] This allows focus on discrimination, which in this framework might occur for two reasons. First, let the employer callback decision be described by the latent variable T_i^*, which determines the treatment applied to a given application (it will drive a dichotomous choice variable below, but it can be though of as any continuous outcome, such as $e.g.$, wage offers). Taste-based discrimination implies that such treatment depends not only on the productivity of the applicant but also on unproductive observable characteristics such as race:[21]

$$T_i^* = T^*[P_i^*, R_i] = P_i^* + \gamma R_i = \delta J + \beta X_i + Z_i + \gamma R_i \,.$$

Second, the invitation decision relies on employer's perceived productivity of applicant i, to whom Z_i is not observed while R_i is:

$$\mathbb{E}[P_i^*(J, X, Z)|X, J, R_i] = \delta J + \beta X_i + \mathbb{E}(Z|R_i) \,,$$

which implies statistical discrimination as long as $\mathbb{E}(Z|R_i) \neq \mathbb{E}(Z_i)$ – the unobserved productivity of an applicant is thought of as being dependent on race. When each employer receives two applications, one from a nonminority applicant and another from a minority one, discrimination shows up in the average differential treatment reserved to applications,

$$\mathbb{E}[T^*|R = 0] - \mathbb{E}[T^*|R = 1] = \beta[\mathbb{E}(X|R = 0) - \mathbb{E}(X|R = 1)] + \mathbb{E}(Z|R = 0) - \mathbb{E}(Z|R = 1) + \gamma \,.$$

By construction, this difference in callback cannot be driven by employer's characteristic. We thus omit the term δJ below, implicitly including it in the deterministic part of the model, βX.

5.1.1 The Content of Correspondence Test Data

By way of construction, a correspondence test controls the observables X in such a way that they do not systematically differ across sub-groups: the difference in observed characteristics must balance over job applications, hence $\mathbb{E}(X|R = 0) = \mathbb{E}(X|R = 1)$.[22] The difference in treatment over all

[20] To ease exposition, we focus on the two races case with $R = 1$ for minority applicants, although the framework easily generalizes to more ethnic origins or to other specifications of the characteristics describing discriminated sub-populations, such as gender. It is also worth mentioning that we gather all sources of unobserved heterogeneity in Z. One can add an $i.i.d.$ error term to the functionals without changing the main conclusions.

[21] In the following, we will impose the linearity assumptions that will be used to estimate the model. Note that Z is not only unobserved to the econometrician but also to the employer in a correspondence test (as opposed, $e.g.$ to an audit study, where employers might gather additional information by meeting experimental applicants). As a result, employer's decisions should be deterministic conditional on the observables. To ease exposition and save on notation, we assume that employers take their callback decisions based on a random draw in the relevant distribution of unobservables – which generates random variations in decisions across firms. As discussed in (Neumark, 2012, p.1135), an alternative way to arrive at a statistical model would be to assume random productivity differences across firms that are multiplicative in the observed productivity of a worker.

[22] This is the case for two reasons. First, the resumes are calibrated in such a way that all observed productive characteristics are equally likely. There still remains differences from one resume to another, but systematic differences across race are ruled out through randomization of the names-resumes matching.

job applications thus arises due to two parameters: taste based discrimination, γ, and statistical discrimination which induces a gap in the (perceived or actual) means of the unobserved productive characteristics $\mathbb{E}(Z|R=0) - \mathbb{E}(Z|R=1)$. Note that the two parameters can only be identified together unless the study provides enough control to guarantee that the distribution means are exactly equal, so that only taste based discrimination occurs.[23] However, all components of this aggregate effect arise as the result of an unequal treatment based on unproductive observable characteristics. As such, they correspond to both the legal and the economic definition of discriminatory behavior: in the remaining, we thus focus on the identification of the gross handicap experienced by minority applicants, μ – standing for the sum of these two parameters: $\mu = \mathbb{E}(Z|R=0) - \mathbb{E}(Z|R=1) + \gamma$.

As noted by Heckman and Siegelman (1993); Heckman (1998) the observational context of a correspondence study may lead to biased estimates if the variance of the error term is race-specific.[24] This is the case because the observed outcome is non-linearly related to the underlying discriminatory decisions. To see this explicitly, denote c the perceived quality threshold an applicant has to reach to be invited for an interview: the outcome variable of the experiment is $T = \mathbf{1}[T^* > c]$. Further assume that the unobserved productivity Z is $i.i.d$ normally distributed in each race sub-group, but with a (perceived or actual) variance that varies across groups. To ease exposition, we denote $\sigma_0^2 = Var(Z|R=0)$ and $\sigma_1^2 = Var(Z|R=1)$. The statistical model generating the observed outcome thus gives the following specifications for the probability of obtaining a callback:

$$\begin{aligned}\mathbb{P}[T=1|R=1,X] &= 1 - \Phi[(c - \mathbb{E}(Z|R=1) - \beta X + \gamma)/\sigma_1] = \Phi[(\beta X + \mathbb{E}(Z|R=1) + \gamma - c)/\sigma_1] \\ \mathbb{P}[T=1|R=0,X] &= 1 - \Phi[(c - \mathbb{E}(Z|R=0) - \beta X)/\sigma_0] = \Phi[(\beta X + \mathbb{E}(Z|R=0) - c)/\sigma_0]\,,\end{aligned}$$

where Φ denotes the standard normal distribution. The difference between these two expressions is the empirical source of identification for discrimination. However, even in the extreme case with neither statistical nor taste-based discrimination – *i.e.*, $\gamma = 0$ and $\mathbb{E}(Z|R) = \mathbb{E}(Z)$, so that $\mu = 0$ – the difference in probabilities still depends on the comparison between σ_1 and σ_0. For instance, if the X have been chosen at the lower tail of the skills distribution, so that $\beta X < c$, and $\sigma_0 < \sigma_1$ then it has to be that $\Phi[(\beta X - c)/\sigma_1] < \Phi[(\beta X - c)/\sigma_0]$, $\forall X$: the experiment produces (spurious) evidence of discrimination against people from $R=1$ origin. In the more general case with both discrimination and differences in the variance of unobservables, this framework highlights the identification problem faced by correspondence test data.

[23] This condition is more likely to be met as the set of observable characteristics is wider. Still, the relative share of taste-based and statistical discrimination in the aggregate effect of race remains a matter of interpretation and cannot be empirically tested even in this kind of framework.

[24] The general principle that heteroscedasticity causes the coefficient estimates in discrete choice models to be inconsistent draws back to Yatchew and Griliches (1985).

5.1.2 Unbiased Measures of Discrimination

The identification problem arises because the average treatment effect $\Delta = \overline{T}_{|R=0} - \overline{T}_{|R=1}$ provides an estimate of $\frac{\mathbb{E}(Z|R=0)-c}{\sigma_0} - \frac{\mathbb{E}(Z|R=1)+\gamma-c}{\sigma_1}$ (up to the functional form of the distribution), while what one seeks to measure is $\mu = \mathbb{E}(Z|R=0) - \mathbb{E}(Z|R=1) + \gamma$: the systematic disadvantage experienced by minority applicants due to their belonging to an observable sub-population. As a result, the confounding effect of the dispersion in the observables shows up in the comparison of callback probabilities, thus affecting the mean comparisons between outcomes from the study.

As noted by Neumark (2012), one can however use restrictions on β to restore identification. If the study provides enough variation on relevant applications characteristics, grouped in X in the model above, one can estimate β/σ_0 and β/σ_1 from the (heteroscedastic) Probit model derived in the previous section. Under the assumption that the effect of X on the callback is homogeneous across race (*i.e.* the true value of β is the same), the ratio of the two point estimations identifies σ_0/σ_1, which in turn allows us to estimate μ after the usual normalization setting one of the variance terms equal to 1. Since only one such identifying regressor is needed to achieve identification, any additional productivity control provides the usual specification tests of an over-identified model.[25]

5.1.3 Statistical and Taste Based Discrimination

Any empirical test of statistical discrimination is conditional on the choice of (at least) one dimension of unobserved heterogeneity in productivity. Let η be such a dimension, standing for a measure of French language ability in our experiment, and S be the treatment variable affecting the content of the applications: $S_i = 0$ for baseline applications with no additional signal, $S = 1$ for treated applications with an explicit mention of language skills ability.

In baseline applications, η is part of the unobserved heterogeneity affecting the success of applications: $Z = z + \eta$, so that:

$$\begin{aligned}
\mathbb{P}[T=1|R=1,S=0,X] &= \Phi[(\beta X + \mathbb{E}(z+\eta|R=1) + \gamma - c)/\sigma_1] \\
\mathbb{P}[T=1|R=0,S=0,X] &= \Phi[(\beta X + \mathbb{E}(z+\eta|R=0) - c)/\sigma_0]\,,
\end{aligned}$$

This productivity characteristic is moreover a source of statistical discrimination if $\mathbb{E}(\eta|R) \neq \mathbb{E}(\eta)$, generating observations of unequal treatment for baseline applications that mix tasted-based discrimination and all sources of statistical discrimination, including language skills. Otherwise, it leaves discrimination unchanged as $\mathbb{E}(z+\eta|R) = \mathbb{E}(z|R) + \mathbb{E}(\eta)$. In treated applications, by contrast, the resume provides observable information about language skills in such a way that η enters the observed

[25] When there is more than one race category, one needs as many identification restrictions as the number of variance parameters.

productivity term rather than unobserved heterogeneity: [26]

$$\begin{aligned} \mathbb{P}[T=1|R=1,S=1,X] &= \Phi[(\beta X+\alpha\eta+\mathbb{E}(z|R=1)+\gamma-c)/\sigma_1] \\ \mathbb{P}[T=1|R=0,S=1,X] &= \Phi[(\beta X+\alpha\eta+\mathbb{E}(z|R=0)-c)/\sigma_0]\,, \end{aligned}$$

Since the signal is added to all applications sent to a given employer under $S = 1$, the treatment effect measured on treated applications includes only taste based discrimination and the remaining sources of statistical discrimination included in z, beyond the ones measured by η. The difference in difference estimate between control and treated applications thus identifies the extent of statistical discrimination related to language skill ability:

$$\mathbb{E}(\Delta^{DD}) = \mathbb{E}[(\overline{T}_{|R=0,S=1} - \overline{T}_{|R=1,S=1}) - (\overline{T}_{|R=0,S=0} - \overline{T}_{|R=1,S=0})] = -[\mathbb{E}(\eta|R=0) - \mathbb{E}(\eta|R=1)]$$

The variation in discrimination between the two groups of applications thus provides an empirical measure of the unequal treatment driven by different believes as regards language skills ability.

5.2 Empirical Results

The econometric model requires identifying variables to recover estimates of discrimination that are not biased by the difference in the variances of unobservables. These variables must affect the probability of hiring, regardless of race. We exploit three identifying variables. We follow Neumark (2012) and use academic honors[27] received by applicants. We also use the type of jobs (assistant, secretary or account) and the type of job contract (short or long-term) as they tend to affect hiring (see Table 3) and their effects do not vary with race.[28]

We first apply the model to the estimation of the effect of ethnic background, pooling the data from both gender. Table 4 presents the estimation results from three specifications of the model, which gradually adds in the control variables (used as identifying variables in the heteroscedastic Probit estimation) to assess the sensitivity of our results to this choice. As a benchmark, the top panel presents estimates derived from the homoscedastic Probit model, whereas the bottom panel presents results from a heteroscedastic Probit model.[29] In each model, the effect of ethnic background is captured by two dummies. The first row presents the coefficient on a dummy variable which is equal

[26] For simplicity, we assume that the signal provides perfect information about the unobserved skills to the employer. The framework could be easily generalized to an imperfect signal.

[27] At the end of high-school, students have to take a national exam called "baccalauréat". According to the test scores, students can have academic honors (no honors, relatively good, good and very good honors).

[28] To examine whether the effect of the identifying variables vary with race, we include in the Probit models interactions between these variables with the non-French origin. In all cases, the Wald tests for joint significance of interactions indicate that our identifying variables are not correlated with race.

[29] Heteroscedastic Probit models can be difficult to maximize as identification is fragile (Keele and Park, 2006). To ensure our estimates are obtained at a global maximum, we complement the simple estimation with a screening procedure on the value of the likelihood function. We embed the Probit estimation inside a loop on the value of σ_0/σ_1, changed from 0.2 to 2.5 through increments of 0.1. This procedure confirms that our estimation reaches a global maximum within this range.

Table 4: Conditional Estimates of Homophily

	Specification 1			Specification 2			Specification 3		
	Control (1)	Signal (2)	Overall (3)	Control (4)	Signal (5)	Overall (6)	Control (7)	Signal (8)	Overall (9)
				A. Basic Probit Model					
North Afr.	-0.084***	-0.053***	-0.069***	-0.084***	-0.053***	-0.069***	-0.084***	-0.053***	-0.069***
	(-5.46)	(-3.73)	(-6.58)	(-5.45)	(-3.74)	(-6.58)	(-5.48)	(-3.74)	(-6.59)
Foreign	-0.078***	-0.054***	-0.067***	-0.078***	-0.054***	-0.067***	-0.078***	-0.054***	-0.067***
	(-4.95)	(-3.66)	(-6.14)	(-4.95)	(-3.66)	(-6.13)	(-4.99)	(-3.65)	(-6.15)
Male	-0.038***	-0.057***	-0.048***	-0.038***	-0.057***	-0.048***	-0.038***	-0.057***	-0.048***
	(-2.59)	(-3.52)	(-4.35)	(-2.61)	(-3.51)	(-4.36)	(-2.60)	(-3.50)	(-4.34)
Log-lik.	-518.4	-519.0	-1111.8	-518.0	-577.7	-1110.0	-516.8	-577.7	-1109.2
				B. heteroscedastic Probit Model					
North Afr.	-0.093**	-0.056***	-0.080***	-0.093**	-0.054***	-0.082***	-0.094***	-0.054***	-0.083***
	(-2.52)	(-2.81)	(-4.53)	(-2.39)	(-2.77)	(-4.70)	(-2.90)	(-3.43)	(-4.56)
Foreign	-0.099***	-0.069***	-0.074***	-0.098***	-0.066***	-0.074***	-0.097***	-0.066***	-0.076***
	(-4.63)	(-3.38)	(-4.32)	(-4.49)	(-3.44)	(-4.47)	(-4.81)	(-3.43)	(-4.40)
Male	-0.030*	-0.056***	-0.047***	-0.030	-0.057***	-0.047***	-0.032**	-0.057***	-0.046***
	(-1.66)	(-3.30)	(-3.80)	(-1.61)	(-3.43)	(-3.85)	(-2.08)	(-3.43)	(-3.77)
Log-lik.	-577.2	-578.1	-1111.3	-516.6	-576.7	-1109.3	-515.5	-576.7	-1108.4
Identifying Variables	Type of Job			Type of Job, Job Contract			Type of Job, Job Contract, Honors		
N	1,512	1,512	3,024	1,512	1,512	3,024	1,512	1,512	3,024

Notes. ***, **, * denote statistical significance from zero at the 1%, 5%, 10% significance level. The dependent variable in each specification is an indicator of whether or not an application received a callback from an employer. T-statistics are indicated in parentheses below the point estimate. The table reports the marginal effects of race and gender on the likelihood of eliciting a callback from both homoscedastic and heteroscedastic Probit models. Panel A presents the homoscedastic Probit model while Panel B presents the heteroscedastic Probit model. Identifying variables included are noted in the bottom panel. For each regression, we provide the value of the log-likelihood. Standard errors are clustered at the firm level.

to one if an individual has a North African name. The second row presents the coefficient on a dummy variable which is equal to one if an individual has a foreign-sounding name that is unidentifiable to French employers. The results are split between applications which include a signal of ability in the French language, and those which do not. Overall, the sign and magnitudes of our results are robust to the inclusion of different sets of identifying variables. The marginal effects from the heteroscedastic specification are always very close to those from the Probit estimates, indicating that differences in the distribution of the unobservables across the French and the non-French applicants do not play a major role in our experiment. The marginal effects are also less significant, as expected given the higher flexibility of the model.

Reassuringly, both models reaffirm the experimental results found in the descriptive statistics. Applicants with both North African and unidentifiable Foreign names face a significantly reduced probability of receiving a callback in comparison to applicants with French names. Moreover, the coefficient on having a North African name is not significantly different from the coefficient on a Foreign name, suggesting that all individuals with non-French names face discrimination in the French labor market.[30] This result indicates that discrimination is directed against all non-majority ethnic groups, and highlights the importance of homophily in hiring discrimination.

Moreover, Table 4 provides evidence of statistical discrimination. In the full sample, the effect of a non-French name on callbacks pools the effects of statistical and taste based discrimination. Statistical discrimination arrives from a (real or perceived) gap in the mean or variance of characteristics of ethnic groups, as employers use ethnicity as a proxy for unobservable characteristics. The extent of this phenomenon is measured by contrasting the estimates from columns 1, 4, and 7 – that restrict the sample to resumes that did not include a signal of proficiency in the French language – while columns 2, 5 and 8 – that restrict the sample to resumes that include a signal of strong abilities in the French language. Both the ordinary Probit and the heteroscedastic Probit models indicate that the inclusion of the language signal reduces discrimination faced by applicants with non-French names, indicating the employers do engage in statistical discrimination.[31]

We now turn to conditional estimation of gender-specific discrimination, pooling the two non-French origins as a treatment dummy variable. Table 5 disaggregates the estimates of the effect of having an ethnic name on the probability of hiring for male, female and all applications. We also disaggregate the sample according to whether the signal was included or not into resumes – the homoscedastic Probit model appears in the top panel; the heteroscedastic Probit, identified thanks to the three above mentioned variables, appear in the bottom panel. The sign and significance of the

[30] The tests of equality in mean callbacks between the North African and Foreign applicants always indicate that they are equally discriminated against. For instance, the t-statistic derived from the test of equal treatment between the Non-French applicants (column 9) are equal to 0.26 and 0.59 for the homoscedastic and heteroscedastic Probit model, respectively.

[31] However, the effect of the signal on callbacks for the non-French applicants is not significant with a t-test statistic (p-value) equal to -1.50 (0.13). As explained below, the effect of the signal on callbacks is mainly driven by the sub-sample of female identities.

Table 5: Conditional Estimates of Statistical Discrimination, by Gender

	Male Applicants			Female Applicants			All Applicants		
	Control	Signal	Overall	Control	Signal	Overall	Control	Signal	Overall
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
					A. Basic Probit Model				
Non-French	-0.065***	-0.067***	-0.066***	-0.098***	-0.037*	-0.069***	-0.082***	-0.054***	-0.068***
	(-3.95)	(-4.21)	(-5.76)	(-5.21)	(-1.86)	(-5.04)	(-5.76)	(-4.16)	(-7.07)
Male	-	-	-	-	-	-	-0.038***	-0.057***	-0.048***
							(-2.59)	(-3.50)	(-4.34)
Log-lik.	-230.4	-243.8	-480.9	-283.8	-330.7	-625.2	-516.9	-577.7	-1109.2
					B. heteroscedastic Probit Model				
Non-French	-0.071***	-0.074***	-0.071***	-0.106***	-0.036	-0.073***	-0.085***	-0.056***	-0.073***
	(-3.69)	(-3.64)	(-4.75)	(-4.12)	(-1.20)	(-4.56)	(-4.19)	(-3.43)	(-6.54)
Male	-	-	-	-	-	-	-0.038***	-0.058***	-0.049***
							(-2.56)	(-3.63)	(-4.35)
Log-lik.	-230.4	-243.7	-480.9	-283.7	-330.6	-625.1	-516.8	-576.9	-1108.8
N	756	756	1,512	756	756	1,512	1,512	1,512	3,024

Notes. ***, **, * denote statistical significance from zero at the 1%, 5%, 10% significance level. T-statistics are indicated in parentheses below the point estimate. The dependent variable in each specification is an indicator of whether or not an application received a callback from an employer. The table reports the marginal effects of race and gender on the likelihood of eliciting a callback from both homoscedastic and heteroscedastic Probit models. Panel A presents the homoscedastic Probit model while Panel B presents the heteroscedastic Probit model. All specifications include type of job, type of job contract and academic honors. For each regression, we provide the value of the log-likelihood. Standard errors are clustered at the firm level.

coefficients of interest are consistent with the descriptive statistics and the results of Table 4. The results indicate evidence of a strong discrimination against minorities.

As well as differential levels of discrimination by gender, we also find differential reactions to the language signal by gender. Including the language signal drastically reduces discrimination for women. For men, there is some evidence of a reduction in discrimination however the results are more ambiguous – both the ordinary probit models and the heteroscedastic probit models indicate that the language signal has a much larger effect for female applicants than for male applicants. For women, Table 5 indicates that when the language signal is included the marginal effect of non-French names reduces by 0.061 points. The difference between the two estimates is significant at the 1% level. However for men the inclusion of a language signal does not result in any significant change in the observed level of discrimination.

5.3 Heterogeneity

Table 6 assesses the robustness of our results to composition effects in the pool of job offers. Results are broken down according to two measures of neighbourhood diversity – the diversity of the geographic location of the employer, and the urban area – as well as the gender and ethnic origin of the employer.[32] Focusing first on diversity, Columns 1 and 2 break down the results by the percentage of foreign-born individuals in the locale. The sample is restricted to locales where more than 20% of the population is foreign-born in Column 1, while Column 2 restricts the sample to locales where less than 20% of the population is foreign-born. This split does not give rise to any noticeable difference: significant evidence of discrimination against applicants with non-French names is found in both areas, with similar magnitude. This either suggests that discrimination does not vary with the share of immigrants around the job offer location, or that the share of immigrants is not a good proxy to capture the level of diversity in cities. To further investigate this dimension, Columns 3 and 4 present the results broken down by whether or not the employer is located in the city of Paris or the surrounding suburbs – as it is a well-known phenomenon that in Paris ethnic minorities tend to live outside the city proper in the *banlieue*, or suburbs. The results indicate that applicants with non-French names are discriminated against significantly in both the city of Paris and the suburbs, however the coefficient on non-French names indicates a greater disadvantage faced by non-French applicants in the city of Paris as opposed to the suburbs. This results is consistent with the findings of Jacquemet and Yannelis (2012), who find that in Chicago, where ethnic minorities tend to live in the city center, firms in the suburbs discriminate much more against African-American applicants than firms in the suburbs. In both cases, discrimination is greater in areas where a greater portion of the population hails from the majority group. In both Chicago and Paris, employers thus favor individuals from their own ethnic group according to location specific data, which is consistent with homophily.

The right-hand side of Table 6 also provides direct evidence of both ethnic and gender based homophily. Columns 5 and 6 present results broken down by the ethnic origin of the recruiter – whether or not the recruiter has a European or non-European name. We only find significant discrimination against non-French applicants occurring for the set of recruiters with European names. While applicants with non-French names are also less likely to receive callbacks when the recruiter has a non-European name, the point estimate is much smaller than in the case of European recruiters

[32] We also performed separate estimations according to firm size and whether or not the contract is permanent or temporary. In both cases, the discrimination results are nearly identical across specifications.

Table 6: Conditional Estimates of Heterogeneity of Discrimination

	Firm Location				Recruiter Identity			
	Diverse	Homogenous	Paris	Suburb	European	Non-Eur.	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
					A. Basic Probit Model			
Non-French	-0.066***	-0.073***	-0.084***	-0.059***	-0.079***	-0.031	-0.057***	-0.095***
	(-6.12)	(-3.62)	(-4.75)	(-5.19)	(-6.97)	(-1.85)	(-3.27)	(-5.75)
Male	-0.045***	-0.060**	-0.039**	-0.054***	-0.053***	-0.031***	-0.021	-0.109***
	(-3.75)	(-2.18)	(-2.02)	(-4.06)	(-4.13)	(-1.45)	(-0.58)	(-2.83)
Log-likelihood	-820.7	-274.6	-423.4	-640.3	-826.9	-270.8	-336.7	-607.7
					B. heteroscedastic Probit Model			
Non-French	-0.076***	-0.066***	-0.093***	-0.066***	-0.087***	-0.034	-0.062***	-0.095***
	(-3.39)	(-3.75)	(-4.28)	(-4.81)	(-6.48)	(-1.02)	(-3.51)	(-5.58)
Male	-0.054***	-0.042**	-0.041**	-0.055***	-0.539***	-0.032	-0.032	-0.108***
	(-3.55)	(-2.11)	(-2.06)	(-3.97)	(-4.04)	(-1.51)	(-0.43)	(-2.70)
Log-likelihood	-820.6	-274.6	-423.1	-639.9	-826.5	-270.8	-147.0	-256.1
N	912	2,112	1,932	1,092	2,286	738	1,205	1,459

Notes. ***, **, * denote statistical significance from zero at the 1%, 5%, 10% significance level. T-statistics are indicated in parentheses below the point estimate. The dependent variable in each specification is an indicator of whether or not an application received a callback from an employer. The criteria by which each specification is restricted is given in the first row of the table. The table reports the marginal effects of race and gender on the likelihood of eliciting a callback from both homoscedastic and heteroscedastic Probit models. Panel A presents the homoscedastic Probit model while Panel B presents the heteroscedastic Probit model. All specifications include type of job, type of job contract and academic honors. Columns (7) and (8) include dummy variables for the origin of a recruiter and an interaction between the origin of the recruiter and the gender of the applicant. For each regression, we provide the value of the log-likelihood. Standard errors are clustered at the firm level.

and moreover is insignificant. We interpret this as direct evidence of homophilous discrimination – the results are driven by recruiters being more likely to call back applicants of a similar ethnic origin. Furthermore, Columns 7 and 8 also suggest that homophilous discrimination operates on the basis of gender as well as race. Once the sample is split according to the gender of the recruiter, we find that female recruiters are significantly less likely to call back French males, indicating that female recruiters prefer French female applicants.[33] This result is strongly driven by French female recruiters calling back more female applicants with French names. In particular, results from additional regressions show that once the sample is further split according to the origin of the applicant, this gender homophily does not operate between French recruiters and non-French applicants: non-French males are always significantly discriminated against whatever the recruiter's gender and the results are driven by. This confirms descriptive statistics evidence that ethnic homophily tends to overcome gender homophily.

6 Discussion

The correspondence test provides robust evidence of homophilic discrimination behavior, which is preserved both across gender and across various contents of the experimental applications – in all instances, the relative success of North African and Foreign sounding applicants is similar. In particular, the observed discrimination against applicants with non-French names remains significant when the results are broken down by geographic location and the characteristics of the recruiter. These results also provide more direct evidence of homophilous hiring discrimination– we observe that recruiters are more likely to call back similar individuals. Furthermore, employers in the city of Paris tend to discriminate against applicants with non-French names more than employers in the suburbs, which is consistent with homophily as in the Paris metro area minorities tend to live in the suburbs.

In terms of gender discrimination, consistent with previous studies such as Bertrand and Mullainathan (2004); Oreopoulos (2011), we also observe that men receive fewer callbacks than women. This effect is much more pronounced amongst minority applicants. For applicants with French names, women receive 30% more callbacks than men, while for applicants with unidentified foreign names women receive 50% more callbacks. The gender gap is largest for applicants with North African names, with women with North African names receiving 80% more callbacks than men with North African names. While it is possible that employers generally favor female applicants, this result may be due to the fact that we chose jobs with larger shares of women – females amount to 66% of people employed in accounting jobs and 84% of people in assistant and secretary jobs. Booth and Leigh (2010) find that employers in female dominated occupations tend to favor female applicants.

Our third result sheds new light on statistical discrimination, as we find that the inclusion of the language signal has an economically significant effect on ethnic discrimination against women. The effect is also asymmetric across gender, as the effect of the signal is much greater effect for female applicants than male applicants. This result is consistent with the findings of Arai, Bursell, and Nekby (2011) in Sweden using Arabic names. When applicants with Arabic names have resumes enhanced with better work experiences, the gap in callbacks drops for women but not for men. Laboratory experiments have also found that gender plays a role in the nature of discrimination.

[33] We note that all estimates include job type dummies, so the result is not driven by sectoral differences. This result is also robust to the inclusion of firm size dummies as additional controls.

Antonovics, Arcidiacono, and Walsh (2005) find that discriminatory behavior varies depending on the age and gender of participants. Moreover the effects of gender discrimination disappear when the stakes of the game are very large. The result also echoes with prior studies which have found differential effects of informational signals for men and women. Holzer, Raphael, and Stoll (2006) find that employers who use criminal background checks are more likely to hire African-Americans, and that this effect is much stronger for men than for women. The authors interpret this result as being consistent with statistical discrimination, and employer being concern with potential criminality affecting productivity for men more than for women.

The literature thus suggests a number of possible explanations for the differential effect of the language signal for men and women. First, employer discrimination towards men may be primarily taste based, while women may suffer statistical discrimination. On the other hand, employers may be discriminating on characteristics other than language for men – for instance, computing skills or heterogeneity related to social behavior in working groups. In this sense, the informational distance (Lundberg and Startz (2004)) between the majority group and minorities may differ for male and female minorities. Alternatively, the informational content of the language skill signal can go beyond language abilities *per se*. For instance, employers may view language ability as a signal of assimilation. If, at the same time, employers are concerned with women becoming pregnant – since immigrant women in France have higher birth rates (Westhoff and Frejka, 2007) – then the signal could be interpreted in terms of a lower "risk" of fertility. Last, the signal, while interpreted the same way in terms of unobservable productivity, might be seen as more credible for women if, for instance, the activities added to the resumes are perceived as more feminine. In all instances, the variation in discrimination induced by the signal is related to the extent of information available about the applicant, hence identifying a statistical reasoning behind unequal treatment. More research is needed to determine which of these hypothesis is true and whether other kinds of signals may have larger effects for men.

7 Conclusion

This study makes two contributions. First, for both gender groups, our experimental results indicate that non-French applications are equally treated, while they are significantly disfavored relative to the French ones. At the initial job-screening stage, racial discrimination thus operates against all non-majority members. Moreover this discrimination is driven by recruiters calling back applicants from similar ethnic backgrounds. This highlights the important role played by ethnic homophily in shaping hiring discrimination. The results thus indicate that hiring discrimination may be faced by a large number of minority groups in a society, rather than being limited to some specific targeted groups.

Second, the study is also devoted to understanding the underlying mechanisms behind homophily.

We test for and identify statistical discrimination due to language skill ability by including a signal in half the applications. The inclusion of a signal drastically undermines racial discrimination within the group of females, but it has a more ambiguous effect on the discrimination against male minorities. The overall effect confirms the role played by statistical discrimination, although such an asymmetric impact also indicates that the causes of discrimination are different across gender.

The fact that language skills are identified as an important source of variation in callbacks is not only interesting from a theoretical perspective in identifying statistical discrimination, but can also have policy implications. The fact that employers are using language skills to statistically discriminate amongst candidates, should turn attention to the implementation of learning and certification programs designed to improve language abilities. Such programs should be devoted to provide and certify sufficient language abilities for all applicants. A label could therefore be transcribed in resumes so as to modify employer beliefs or perception on applicant language skills.

References

Adida, C., D. Laitin, and M. Valfort (2010): "Identifying barriers to Muslim integration in France," *Proceedings of the National Academy of Sciences*, 107(52), 22384–22390.

Altonji, J., and C. Pierret (2001): "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics*, 116(1), 313–350.

Antonovics, K., P. Arcidiacono, and R. Walsh (2005): "Games and Discrimination: Lessons from "The Weakest Link"," *Journal of Human Resources*, 40(4), 918–947.

Arai, M., M. Bursell, and L. Nekby (2011): "The Reverse Gender Gap in Ethnic Discrimination: Employer Priors against Men and Women with Arabic Names," *CEB Working Paper*, 11(027).

Arrow, K. (1973): "The Theory of Discrimination," in *Discrimination in Labor Markets*, ed. by O. Ashenfelter, and A. Rees. Princeton University Press, Princeton (NJ).

Becker, G. S. (1971): *The Economics of Discrimination*. University of Chicago Press, Chicago (IL).

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.

Booth, A., and A. Leigh (2010): "Do Employers Discriminate by Gender? A Field Experiment in Female-Documented Occupations," *Economics Letters*, 107(2), 236–238.

Booth, A., A. Leigh, and E. Varganova (2012): "Does Ethnic Discrimination Vary Across Minority Groups? Evidence from a Field Experiment," *Oxford Bulletin of Economics and Statistics*, 74(4), 547–573.

Bramoullé, Y., and B. W. Rogers (2009): "Diversity and Popularity in Social Networks," *CIRPEE Discussion Paper*, (09-03).

Carlsson, M., and D.-O. Rooth (2007): "Evidence of ethnic discrimination in the Swedish labor market using experimental data," *Labour Economics*, 14(4), 716–729.

Charles, K. K., and J. Guryan (2011): "Studying Discrimination: Fundamental Challenges and Recent Progress," *NBER WP*, (17156).

Crenshaw, K. (1991): "Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color," *Stanford Law Review*, 43(6), 1241–1299.

Currarini, S., M. O. Jackson, and P. Pin (2009): "An Economic Model of Friendship: Homophily, Minorities, and Segregation," *Econometrica*, 77(4), 1003–1045.

Dechief, D., and P. Oreopoulos (2012): "Why do some employers prefer to interview Matthew but not Samir? New evidence from Toronto, Montreal and Vancouver," *CLSRN Working Papers*, (2012-8).

Duguet, E., L. Du Parquet, Y. L'Horty, and P. Petit (2012): "First Order Stochastic Dominance and the Measurement of Hiring Discrimination: A Ranking Extension of Correspondence Testings with an Application to Gender and Origin," *Working Paper*.

Duguet, E., N. Leandri, Y. L'Horty, and P. Petit (2010): "Are young French jobseekers of ethnic immigrant origin discriminated against? A controlled experiment in the Paris area," *Annals of Economics and Statistics*, pp. 187–215.

Giuliano, L., D. Levine, and J. Leonard (2009): "Manager Race and the Race of New Hires," *Journal of Labor Economics*, 27(4), 589–631.

Giuliano, L., D. I. Levine, and J. Leonard (2011): "Racial Bias in the Manager-Employee Relationship: An Analysis of Quits, Dismissals, and Promotions at a Large Retail Firm," *Journal of Human Resources*, 46(1), 26–52.

Granovetter, M. (1973): "The Strength of Weak Ties," *American Journal of Sociology*, 78, 1360–1380.

Habyarimana, J. P., M. Humphreys, D. N. Posner, and J. Weinstein (2007): "Why Does Ethnic Diversity Undermine Public Goods Provision?," *American Political Science Review*, pp. 709–725.

Hamm, J. V. (2000): "Do birds of a feather flock together? The variable bases for African American, Asian American, and European American adolescents' selection of similar friends," *Developmental Psychology*, 36(2), 209–219.

Heckman, J., and P. Siegelman (1993): *Clear and Convincing Evidence: Measurement of Discrimination in America*chap. The Urban Institute Audit Studies: Their Methods and Findings, pp. 187–258. The Urban Institute Press, Washington, D.C.

Heckman, J. J. (1998): "Detecting Discrimination," *Journal of Economic Perspectives*, 12(2), 101–116.

Holzer, H. J., S. Raphael, and M. A. Stoll (2006): "Perceived Criminality, Criminal Background Checks, and the Racial Hiring Practices of Employers," *Journal of Law & Economics*, 49(2), 451–480.

Ioannides, Y. M., and L. D. Loury (2004): "Job Information Networks, Neighborhood Effects, and Inequality," *Journal of Economic Literature*, 42(4), 1056–1093.

Jackson, M. O., and B. W. Rogers (2007): "Meeting Strangers and Friends of Friends: How Random Are Social Networks?," *American Economic Review*, 97(3), 890–915.

Jacquemet, N., and C. Yannelis (2012): "Indiscriminate Discrimination: A Correspondence Test for Ethnic Homophily in the Chicago Labor Market," *Labour Economics*, 19(6), 824–832.

Keele, L., and D. K. Park (2006): "Difficult choices: an evaluation of heterogeneous choice models," *2004 Meeting of the American Political Science Association*.

Knowles, J., N. Persico, and P. Todd (2001): "Racial Bias in Motor Vehicle Searches: Theory and Evidence," *Journal of Political Economy*, 109(1), 203–229.

Kossinets, G., and D. J. Watts (2009): "Origins of Homophily in an Evolving Social Network," *American Journal of Sociology*, 115(2), 405–450.

Lang, K., and J.-Y. K. Lehmann (2012): "Racial Discrimination in the Labor Market: Theory and Empirics," *Journal of Economic Literature*, 50(4), 959–1006.

Lazarsfeld, P., and R. K. Merton (1954): “Friendship as a Social Process: A Substantive and Methodological Analysis,” in *Freedom and Control in Modern Society*, ed. by M. Berger, T. Abel, and C. Page, pp. 18–66. Van Nostrand, New-York (NJ).

Levitt, S. (2004): “Testing Theories of Discrimination: Evidence from the Weakest Link,” *Journal of Law and Economics*, 47(2), 313–350.

Lundberg, S., and R. Startz (2004): “Information and Racial Exclusion,” *Journal of Population Economics*, 20(3), 621–642.

McPherson, M., L. Smith-Lovin, and J. Cook (2001): “Birds of a Feather: Homophily in Social Networks,” *Annual Review of Sociology*, 27, 415–444.

Mollica, K. A., B. Gray, and L. K. Treviño (2003): “Racial Homophily and Its Persistence in Newcomers’ Social Networks,” *Organization Science*, 14(2), 123–136.

Neumark, D. (2012): “Detecting Discrimination in Audit and Correspondence Studies,” *Journal of Human Resources*, 47(4), 1128–1157.

Oreopoulos, P. (2011): “Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Six Thousand Resumes,” *American Economic Journal: Economic Policy*, 3(4), 148–171.

Pager, D., B. Western, and B. Bonikowski (2009): “Discrimination in a Low-Wage Labor Market: A Field Experiment,” *American Sociological Review*, 74(5), 777–799.

Phelps, E. S. (1972): “The Statistical Theory of Racism and Sexism,” *American Economic Review*, 62(4), 659–661.

Pope, D., and J. Sydnor (2011): “What’s in a Picture? Evidence from Proster.com,” *Journal of Human Resources*, 46(1), 53–92.

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

Price, J., and J. Wolfers (2010): “Racial Discrimination Among NBA Referees,” *Quarterly Journal of Economics*, 125(4), 1859–1887.

Putnam, R. D. (2007): “E Pluribus Unum: Diversity and Community in the Twenty-first Century,” *Scandinavian Political Studies*, 30(2), 137–174.

Riach, P. A., and J. Rich (2002): “Field Experiments of Discrimination in the Market Place,” *Economic Journal*, 112(483), F480–F518.

Stoll, M. A., S. Raphael, and H. J. Holzer (2004): “Black job applicants and the hiring officer’s race,” *Industrial and Labor Relations Review*, 57(2), 267–287.

Trimaille, C. (2004): “Pratiques langagiéres chez des adolescents d’origine maghrébine,” *HOMMES ET MIGRATIONS*, pp. 66–73.

Vallet, L.-A., and J.-P. Caille (1996): “Niveau en français et en mathématiques des élèves étrangers ou issus de l’immigration,” *Économie et statistique*, 293(1), 137–153.

Vigil, J. M., and K. Venner (2012): “Prejudicial behavior: More closely linked to homophilic peer preferences than to trait bigotry.,” *Behavioral and Brain Sciences*, 35(6), 448–449.

Westhoff, C., and T. Frejka (2007): “Religiousness and Fertility among European Muslims,” *Population and Development Review*, 33(4), 785–809.

Wimmer, A., and K. Lewis (2010): “Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook,” *American Journal of Sociology*, 116(2), 583–642.

Yatchew, A., and Z. Griliches (1985): “Specification Error in Probit Models,” *Review of Economics and Statistics*, 67(1), 134–139.