Paris School of Economics
Paris, August 2020

Discrimination in the Hiring Process in France: The Role of Spatial and Within Firm Diversity

Clara Schaeper
Supervisor: Thomas Breda

Master thesis, Applied Economics
Major: Public Policy and Development

PARIS SCHOOL OF ECONOMICS

Abstract

This thesis investigates the interdependence between spatial as well as within-firm diversity and discrimination in hiring. For that purpose I make use of preliminary data from an ongoing correspondence testing in France, covering applications to 12 different occupations using French and Maghrebian sounding names. My results provide evidence of ethnic homophily and the importance of favoritism for majority-group members both within the neighbourhood of the working environment and within a firm. Discrimination is systematically higher when the local share of foreigners, or the share of births with an Arabic last name between 1966-1990, is low. Moreover recruiters with European names are more likely to call back French named applicants while there is no significant effect on discrimination when the recruiter has a non-European name. In a further step the study analyses the anticipated customer discrimination as a potential driver, taking the composition of residents in a geographical area as a proxy for the consumer composition of the firms located in that area. This exercise shows trends of applicant discrimination being systematically highest when the occupation is classified as one with a high amount of customer contact and the diversity of the neighbourhood is low.

JEL classification: C39, J15, J64, J71, J78, J82, J88

Contents

List of Figures

List of Tables

1 Introduction

"Social integration and diversity are among the salient policy issues in modern societies. The size and speed of immigration processes and the heterogeneity of European national societies pose new challenges that require a deep understanding of the cultural, social and economic dynamics that are involved.", Quote from the Syllabus of Sergio Calarrili for his PhD course on "Homophily, Segregation and the Evolution of Preferences" at Bocconi University 2018/2019

When asked if they think they have experienced unequal treatment or discrimination, 31% of French-born people who are descendants of two immigrant parents answer in the affirmative (compared to 13% of the French population) (Simon and Safi (2013)). These figures are particularly high for those whose parents were born in the Maghreb, sub-Saharan Africa and the French overseas department. In their perception, the most frequent reasons for that discrimination are origin, skin colour and nationality. Those subjective measures are also reflected in measures of discrimination. A recent meta-study by Quillian et al. (2019) compares discrimination rates in countries for which at least three different correspondence studies are available. The study suggests that France leads the way when it comes to hiring discrimination, far ahead of Germany, for example, and with even higher levels of discrimination than in the United States[1].

Existing studies provide strong evidence of discrimination[2] in hiring, but also in other structural processes such as the housing market[3], against equally qualified applicants from ethnic minority populations, compared to the ethnic majority applicants (see Bertrand and Duflo (2017) for a recent summary of all such studies). At the same time literature on the development of segregation of immigrant communities over time (and generations) find persistent levels of segregation, despite assimilation theories arguing that those patterns should decrease with time and over generations (Alba (2009)).

This thesis connects these two findings, aiming to quantify the effect of hiring discrimination in shaping the ethnic composition of a locality. I make use of a large ongoing correspondence study, which looks at response rates for applications sent to vacancies in over 400 municipalities spread over the whole of France. France as a setting is interesting, because it is both found to have large and persisting (labour market) discrimination rates as well as persistent levels of segregation.

Although levels of ethnic/racial segregation are found to be lower than segregation of blacks in the U.S., *"there are relevant reasons to think that place stratification dynamics are at play in the French society"*, McAvay and Safi (2018). The authors of a comprehensive analysis on the effects of assimilation variables such as immigrant generation, age at migration, parental age at migration, mixed ascendance, and socio-economic status

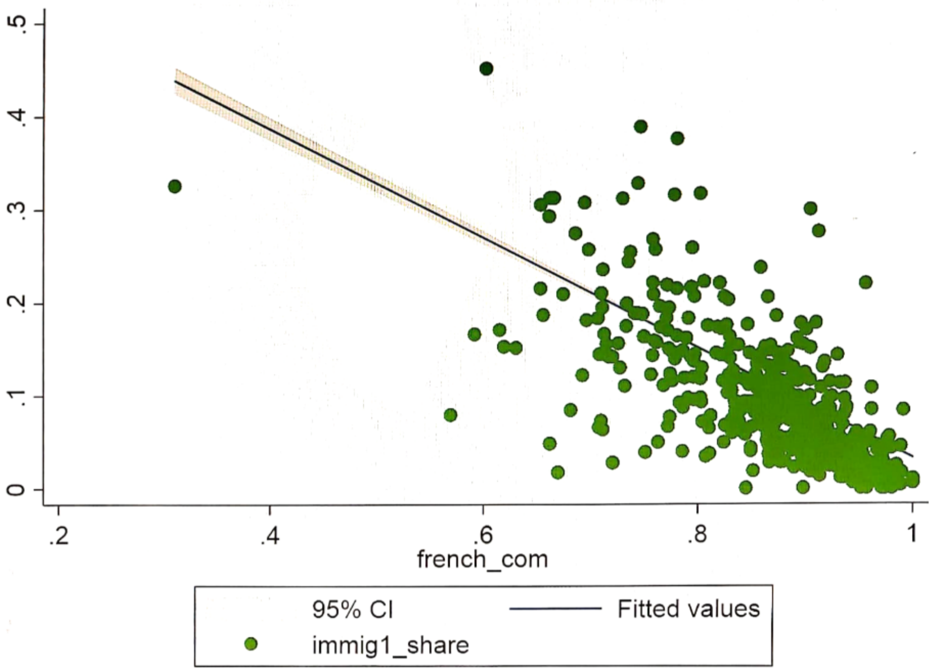
[1]Note that discrimination in hiring does not necessarily reflect ethnic inequalities. In their study, Quillian et al. (2019) note for instance that the cross-national variation in hiring discrimination does not reflect documented patterns of disparities in unemployment rates between immigrants and natives reported by the OECD (2015).

[2]In the realm of this thesis I will refer to discrimination as *"members of a minority group (women, blacks, muslims, immigrants, etc.) being treated differentially (less favorably) than members of a majority group with otherwise identical characteristics in similar circumstances"*, following Bertrand and Duflo (2017).

[3]See for instance Bonnet et al. (2016) and Laouénan and Rathelot (2020) (France), Ahmed and Hammarstedt (2008) (Sweden), Ewens et al. (2014) (US), Baldini and Federici (2011) (Italy) or Bosch et al. (2010) (Spain).

come to the conclusion, that *"overall, whether acculturative or socioeconomic, the predictive power of assimilation factors seems relatively weak compared to the persistence of considerable residential inequalities which place North and sub-Saharan Africans in the most disadvantaged positions. These findings suggest that structural mechanisms, shaped by race and ethnicity, are obstructing group desegregation dynamics in France."*, McAvay and Safi (2018).

Figure 1.1: Share of French employees, 2015 (french_com) and share of foreign population, 2017 (immig1_share) by municipality included in my sample



Data: French Census, *Tableaux détaillés - Population par situation quant à l'immigration, 2017* (Share of Foreign population, 2017); DADS, *Postes 2015* (Share of French by total no. of employees in a municipality)

Figure 1.1 above displays the correlation between the share of French employees (born in France, holding French citizenship) with the share of foreigners living in a municipality for my sample. It clearly indicates that the higher the share of foreigners, the lower the share of French employees, i.e. the higher the share of immigrants, working in that municipality. In a recent study, Combes et al. (2016) further show that the likelihood of being unemployed is significantly increasing with the share of French people living in a municipality. Their estimates predict that a one-standard-deviation increase in the proportion of French natives in an Employment Area, widens the ethnic unemployment gap by 25-30% of its standard deviation.

While discrimination, especially in the access to employment and to housing might be an important factor driving the persistent segregation in France[4], McAvay and Safi (2018) also state, that such persistent segregation effects might be a result of the "preference for

[4]See e.g. Decreuse and Schmutz (2012) who build a model of simultaneous transitions on the housing and the labor market in order to account for the residual unemployment gap between African immigrants and non-immigrants in France; or Gobillon et al. (2014). Both studies conclude that labor market factors remain the main explanation for the higher unemployment rate of Africans even when considering the housing situation.

living in proximity to co-ethnics" (see Rathelot and Safi (2014)). This might especially be relevant as it is well studied that the co-ethnic ties in immigrants' labour market outcomes play an important role[5]. There is however little evidence showing how discrimination and preference might be interconnected, as the reliance on such co-ethnic networks might be a reaction to (labour market) discrimination in the broader society in the first place[6]. McAvay and Safi (2018) further argue that "*segregation patterns are also maintained if French natives actuate preferences for non-immigrant neighborhoods, whether motivated by a desire for co-ethnic proximity or to avoid the socioeconomic disadvantage associated with immigrant areas.*". In a recent paper that focuses especially on revealing mechanisms behind the lack of assimilation and discrimination against Muslims in France, Adida et al. (2014) argue that Muslims and rooted French were locked in a sub-optimal equilibrium. Their results from altruism and trust-games conducted in lab-experiments show that *"rooted French exhibit taste-based discrimination, that is, they show lower unconditional altruism (but not lower trust), toward those they are able to identify as Muslims (due to the fact that these Muslims maintain recognizable Muslim first names) and that Muslims perceive French institutions as systematically discriminatory against them (and therefore assimilate less, although Muslims' lower assimilation may not only be due to rooted French discrimination)"*, Adida et al. (2014).

While the segregation patterns and discrimination in hiring have been studied separately from each other, there is - to the best of my knowledge - no evidence that links the composition of neighbourhoods and discrimination in hiring systematically across France. While the above cited finding of unemployment gaps increasing with the share of the rooted French population provides first evidence of discrimination being linked to the composition of the given population, their estimation method remains indirect. The analysis of such a relation in a correspondence testing set-up has the advantage of allowing better control of individual productivity characteristics and completely neutralises the effect of immigrants' preferences for working in a particular municipality, since it compares the chances of employment conditional on having applied for a specific type of job in a certain locality, unlike using data on overall employment rates.

For that purpose I make use of access to data from a large-scale correspondence study, allowing me to disentangle whether discrimination against applicants with a Maghrebian sounding name varies based on the diversity of the neighbourhood in which a firm is located[7]. Following the "preference argument", a candidate belonging to a certain ethnic group would not (or less often) apply to a municipality with a (relatively) low share of people from the same ethnic group. The correspondence study setting allows me to circumvent this as candidates apply to any vacancy fitting their randomly assigned profile, no matter where the firm is located in. In such a set-up one can consider segregation as the product of (labour) supply and demand, i.e. there are fewer immigrants in a municipality with an existing low share of ethnic diversity either because they want to go there less (supply) or because they are more rejected when they try to (demand). Because

[5]Note that the evidence on the outcomes is mixed as "*some argue they* [the co-ethnic immigrant networks]*are a valuable resource, increasing immigrants' labour force participation and wages; others find negative effects such as trapping workers in low-quality employment*",Toma (2016).

[6]Decreuse and Schmutz (2012) show for instance that *"47% of Africans "prefer" to be unemployed in a large city rather than being employed in a small town. The non-immigrant proportion is only 40%"*. As I later show large cities are also the places, where the ethnic composition of the municipalities is found to be the most diverse.

[7]I measure diversity of the neighbourhood with three different proxies, specified in section 3.3.

in the correspondence study set-up supply is identical for minority and non-minority individuals, the estimated differential in hiring discrimination between an already high diverse municipality compared to a low diverse municipality can thus be considered as the level of segregation resulting from the supply side (i.e. hiring discrimination). Consequently the results help in understanding whether segregation is in fact reinforced by minority candidates being systematically excluded from getting jobs in areas, where the composition of the population in the neighbourhood and in the work-force is less diverse (low share of minority population).

Such employer behaviour is closely linked to the concept of homophily which implies the notion that *"birds of a feather flock together"*, McPherson et al. (2001). The data at hand allows to test for ethnic homophily and thus the in-group preference hypothesis in an even more direct manner. As I observe a proxy for the ethnicity of recruiters, I can detect whether discrimination against non-French applicants is driven by recruiters calling back applicants from similar ethnic backgrounds (see also Edo et al. (2019) and Jacquemet and Yannelis (2012)). My proxy for within-firm homophily will not only be the recruiter's identity but also the share of foreigners/French employed in a firm for a sub-sample of my data.

As I am thinking about the concept of homophily not only on the within-firm level but also consider the composition of the locality in which the firm is located in, I am in a next step able to reveal some additional mechanisms driving the employers behaviour. My thesis expands on the homophily effect in also considering visibility of the hired person to customers when it comes to employer decisions. In such a case the homophily effect might be even more extreme when considering the customer pool. I thereby can add to another strand of relatively new literature, which is looking beyond the discriminatory outcome of employer decisions, aiming to disentangle consumer discrimination from pure employer discrimination. Combes et al. (2016) find evidence of employment discrimination based on the employers anticipating customer discrimination. "*Such customer discrimination occurs, when a significant proportion of consumers shy away from interacting with ethnic minority workers. The value of such workers' services is thus reduced,leading many employers to reject minority applicants, even where the said employers are themselves unprejudiced. Such discriminatory practices may have major implications for the employment odds and career choices of minority workers.* Minority candidates might thus be more likely to be discriminated against, the more visible their position is to clients and the public, thereby reducing the set of employment opportunities offered to them in homogeneous municipalities. In section 5 of this thesis I discuss the theoretical predictions and how this mechanism might help understanding the nature of the above defined demand side effect, resulting in segregation.

My study's main findings are threefold and extent the existing literature in the following ways: First, I can show that discrimination is systematically higher when the local share of diversity is low. A finding, driven by employers calling a higher number of candidates that are considered similar to the ethnic local composition of the locality of the firm. This allows me to disentangle and quantify the demand-side effect (hiring discrimination) in shaping segregation outcomes from the supply side effect (the applicant's preference). Second, I find direct evidence for homophily as a mechanism. Recruiters with European

[7] I define this term following Edo et al. (2019) who refer to "*homophily as workers from one group being favored over others from employers belonging to the same group*".

names are more likely to call back French named applicants, while there is no significant effect of discrimination when the recruiter has a non-European name. While this strategy has been applied by Edo et al. (2019) analysing correspondence data in the Paris region, I am yet the first to expand such an analysis using data on the whole of France. I can further show that the within-firm composition of employees also shapes employers hiring behaviour. By using a unique feature of combining the experimental data with additional firm-level data on the composition of the within-firm employee pool I am thereby able to test a group-composition level that has not yet been studied before in a correspondence study set-up but deemed important to drive the results in the literature (see section 2). Third, I can reveal some mechanisms driving the findings by showing trends of applicant discrimination being systematically highest when the occupation is classified as one with a high amount of customer contact and the diversity of the neighbourhood is low. This adds to the existing evidence on customer discrimination in France as it is the first attempt to disentangle the employer's preference-driven effects from applicant's preferences in an experimental set-up.

The remainder of this thesis is structured as follows: In the next section, I review the literature on the relationship between the theoretical concept of homophily and discrimination in the labour market. In section 3 I present the estimation strategy and data at hand. This section is structured in three parts. In part 3.1 I present the empirical strategy applied to detect hiring discrimination. In part 3.2 I explain the experimental design of the correspondence-study who's data I am analysing. I then, in part 3.3, provide a detailed description of the creation of my data set and the sample selection. The results are presented in section 4. This section first provides some general findings before presenting the results of my main analysis in parts 4.2 and 4.3. Section 5 discusses the concept of customer discrimination as a potential mechanism reinforcing employers discriminatory behaviour and presents findings to test the established propositions. I conclude with a discussion of my results and its implications for future research and policy perspectives in France in section 6.

2 Ethnic Homophily as a source of discrimination

"A pervasive and well-documented feature of social and economic networks is that contacts tend to be more frequent among similar agents than among dissimilar ones. This pattern, usually referred to as "homophily", applies to many types of social interaction, and along many dimensions of similarity, including ethnicity, religion, gender, age, ideology, etc.. The presence of homophily, and of the implied social segregation, has important implications on how information flows along the social network and, more generally, how agents' characteristics impinge on social behaviour. It is therefore important to understand more about the main sources of homophily, under which conditions these translate into social segregation and discrimination, and what kind of affirmative action can be desirable and effective.", Extract from the lecture summary of Sergio Calarrili's Professorial Inaugural Lecture at the University of Leicester, 2014

The concept of homophily is well established in the economics's literature on network formation, for instance when analysing friendship ties (Currarini et al. (2010)). It is usually referred to as the personal preference being biased in favour of "one's own type" and which may affect both the intensity and the direction of search and opportunities.

The economic literature linking this concept to hiring discrimination is relatively new but indicates that homophily might play a relevant role in shaping the pool of employees within a firm, and hence municipalities and ultimately re-inforcing segregation on several levels. Mirroring this idea to the stated suggestions on the desire of French natives for co-ethnic proximity in section 1, one should expect that the desire for co-ethnic proximity should also be found when choosing the work-place composition (after all people spend most of their day at work). Jacquemet and Yannelis (2012) and Edo et al. (2019) are yet the first to directly link the concept of homophily to the stable and consistent findings of discrimination rates based on the ethnicity of the candidates not only in France but *"across countries and across minorities within countries"*, Jacquemet and Yannelis (2012).

Within-firm level

A literature in sociology has long shown strong connections of within-firm's social networks and homophilious behaviour in hiring (Pager and Shepherd (2008)). Mouw (2002) for instance finds that the use of employee referrals in firms with less than 10% of black employees reduces the probability of a black hire by nearly 75% relative to the use of newspaper ads. This implies that the active reliance on the current network, in this case the composition of the employees, reproduces the existing racial composition of an organization, irrespective of an employer's personal racial attitudes (see Petersen et al. (2000), Royster (2003), Waldinger et al. (2003)).

In the economics literature, Giuliano et al. (2011) find evidence of an "own-race bias" in manager-employee relationships when it comes to quits, dismissals and promotions in large U.S retail firms. Stoll et al. (2004), using large U.S. observational data, show that black recruiters hire a greater proportion of black applicants, compared to white recruiters. The paper closest to mine is Edo et al. (2019) who find that ethnic discrimination "almost vanishes" when the recruiters have a non-European origin (proxied by the origin of their name).

Location of the firm level

While one might argue that in the set-up of a correspondence study "referrals" do not play a major role as neither candidate is known to anyone in the firm that one applies to, Mouw (2002)'s findings further suggest that employee referrals *"are just as important as the geographic location of the firm in generating employment segregation: both increase the predicted level of inter-firm racial segregation among blue-collar workers in the cities studied by about 10%"*.

Already in the first, and certainly one of the most discussed, correspondence studies by Bertrand and Mullainathan (2004), the authors included *"a measure capturing the marginal effect of employer location on the racial gap in callback"*. For that purpose they use zip-code level data on the fraction of African-Americans living in a certain neighbourhood of the cities included in their study. They find a very small but significant positive effect of the share of African-Americans in the employer location on African-American call-backs.[8].

[8]Note that potential explanations for the small effect might first be the fact that they only look at two large cities (Boston and Chicago) and that in those cases the *"zip-codes may not be the best proxy to measure racial differences in areas as postal zip codes do not measure any real boundaries between community zones in Chicago"*, Jacquemet and Yannelis (2012).

Jacquemet and Yannelis (2012) also looking at hiring discrimination in Chicago proxy diversity by splitting their sample into firms located in the suburbs (where whites make up 78.1% of the population) and firms located in the city (where whites make up 37.6% of the population). They find that ethnic origin variables are insignificant in the city-center, while they are statistically significant in the suburbs. They argue that this finding is in line with *"literature in economics on suburban and central city employer preferences"*, Jacquemet and Yannelis (2012). This literature indicates that the mechanism driving the location of the firm to be a relevant factor for differential discrimination rates is not only the concept of homophily per-se. The location becomes especially relevant when the neighbourhood population makes up the establishment's pool of customers. Following this literature, employment discrimination in the U.S is shown to be consistently higher, when firms are located in suburbs (see e.g. Raphael et al. (2000), Holzer (1996), Holzer and Ihlanfeldt (1998)). This is assumed to be more relevant for small companies and occupations for which the visibility to and the direct contact with customers is important. Jacquemet and Yannelis (2012) for instance find larger discrimination gaps in hiring for the occupation of nursing (which is considered to involve high levels of contact with customers) compared to the occupations in programming (which is considered to have low customer contact levels).

I will discuss and disentangle this particular mechanism of customer discrimination in section 5 but it certainly needs to be kept in mind as one driver for homophilious behaviour of employers already for the understanding of the main findings.

Theoretical predictions

This thesis expects that homophily has a twofold effect. On the one hand it drives employer's behaviour because they themselves are seeking to employ people from their own ethnic group and on the other hand because they seek to keep the composition of the location that they are located in as it is. I will proxy the firm's "identity" by the recruiter's suggestive ethnic origin and the within-firm composition of the employees. The composition of the neighbourhood of a firm will be proxied with three different measures of neighbourhood diversity, introduced in section 3.3.

Following the concept of homophily it is expected that the context in which the employer finds oneself will be mirrored in the pool of candidates that are considered employable. This leads to the following predictions tested in my thesis: (1) Discrimination rates against Maghrebian candidates should be higher when the level of diversity in the municipality is low compared to the level of diversity being high; (2) The pool of candidates that are invited should be as diverse or homogenous as the existing pool of employees. This implies that discrimination is expected to vary both with the recruiter's identity as well as with the within-firm composition of employees.

Note that the line between homophilious behaviour working through "statistical" or "taste-based" discrimination is not clear-cut. Homophily can be explained by the "intrinsic individual preference" for similar individuals, as modeled in Currarini et al. (2009). This would be translated into a variation of the "taste-based" discrimination parameter (Becker (2010)). There are however also so-called "belief-based" explanations for homophily. These indicate 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. Since information drives discriminatory behavior in this case, this mechanism echoes the statistical view of* *discrimination"*, Edo et al. (2019). Such an effect would be translated into a variation of the statistical discrimination parameter. In both cases homophilious behaviour would however be expected to work in the same direction. Thereby, it should not matter if it works through one or the other channel, as I capture both effects in my estimation of discrimination, as can be seen in the next section.

I will now present my estimation strategy as well as the experimental design of the correspondence study at hand. I will also thouroughly describe the creation of my variables and especially of the way I measure diversity on the different suggested levels.

3 Estimation Strategy and Data

3.1 Detecting discrimination in the hiring process

The general methodology

There are several methods applied in current research to show the extent of discrimination[9]. Among these, correspondence studies such as the one whose data I will be using in this thesis have become by far the most used method to detect discrimination in hiring among the field experiments (Bertrand and Duflo (2017)). Applications of individuals who are matched in all observable characteristics except for the one in question (suggestive origin, gender, etc.) are sent to a job vacancy. Applying lawful behaviour should result in equal propensities for being contacted by a potential employer for both groups of candidates. Observing differences in those contacts allows to reveal the existence of average discrimination rates between the minority and majority candidates. Let's assume the productivity of an applicant can be denoted as $P_i(J, X, Z)$, where i stands for the individual applicant, J are job and firm characteristics (i.e. type of occupation, job contract and other firm-specific variables), X are observable characteristics (years of education, age, marital status, etc.) and Z are un-observable characteristics. Suggestive ethnicity being $M_i = 0$ for non-minority candidates and $M_i = 1$ for minority candidates does not enter the productivity function directly. An employer takes the call-back decision, defined by the latent variable T_i^*. It determines the treatment of an application i and takes the value one, when a candidate i is called back and zero otherwise. The outcome of T_i is assumed to be dependent on the productivity (P_i) and potentially on suggestive ethnicity (M_i) if there exists discrimination based on a pure dislike of the minority group:

$$T_i^* = T^*[P_i, M_i] = P_i^* + \gamma M_i = \delta J + \beta X_i + Z_i + \gamma M_i, \tag{3.1}$$

where δ captures job-specific effects, β captures the effect of observable applicant's characteristics and γ is the discrimination effect, which captures what some studies call "taste-based" discrimination (Becker (2010)), i.e. differential treatment purely relying on the suggestive origin of a candidate. In this setting, statistical discrimination would exist, if $E(Z|M_i) \neq E(Z_i)$. The average discrimination effect can thus be denoted by the

[9] These methods include observational studies of "unexplained" gaps such as wage differentials (Oaxaca, 1978); audit and correspondence studies, natural experiments (e.g. blind vs non blind auditions, Goldin and Rouse (2000)); or lab experiments (e.g. Implicit Association Test).

difference in the outcome of T^* by origin:

$$\begin{aligned}E(T^*|M=0)-E(T^*|M=1) =& \beta[E(X|M=0)-E(X|M=1)] \\ &+[E(Z|M=0)-E(Z|M=1)]+\gamma.\end{aligned}$$

Note that the job-specific characteristics are by construction not included in this specification. Therefore one might think of them as being *"implicitly included in the deterministic part of the model, βX"*, Edo et al. (2019). As a result, the aggregate discrimination effect, measured by the correspondence study can be denoted as:

$$\underbrace{\mu}_{\text{aggregate discrimination}} = \underbrace{[E(Z|M=0)-E(Z|M=1)]}_{\text{statistical discrimination}} + \underbrace{\gamma}_{\text{taste-based discrimination}}, \tag{3.2}$$

In this analysis I aim to estimate the interaction of discrimination against Maghrebian origin candidates with an indicator for the diversity of the municipality where the firm is located as well as the diversity of the firm's employees. To do so, the most straight forward method would be to opt for an OLS model including an interaction term of the origin and the proportion of Maghrebian people living in the firm's municipality or that are employed in the firm:

$$T_i = \alpha + \beta_1 M_i + \beta_2 D_{j/m} + \gamma\left(M_i \times D_{j/m}\right) + X_{i(j/m)k} + \epsilon_{ijk}, \tag{3.3}$$

where D captures diversity on either the firm (j) or the municipality (m) level[10]. γ denotes the heterogenous impact that diversity has on the discrimination estimate β_1 for applicants with a Maghrebian background.[11] Controls, $X_{i(j/m)k}$, include individual controls (i) but also municipality-level controls (m) in order to rule out the possibility that the diversity measure simply captures unemployment or population size effects. Standard errors will be clustered on the vacancy-level (k). I also include department dummies in the full samples. While I do estimate the OLS model with interaction terms in a first attempt, I opt for the estimation of a heterogeneous Probit model following recent developments in the literature (Neumark (2012)) explained in the following section.

The Heckman - Siegelman critique and Neumark's methodology (hiring discrimination with heteroscedastic unobserved heterogeneity)

In their "Heckman - Siegelman critique" (Heckman and Siegelman (1993), Heckman (1998)), the authors show that even in an ideal set-up of a correspondence study, with no observable average differences and no difference in the average of unobserved characteristics, the discrimination measure might still be spurious. Consequently, Neumark (2012) proposed a solution to recover an unbiased estimate of discrimination even when there are group differences in the variances of the unobservables. This method has become a common practice and is thus included as the main specification in this thesis.

[10] This measure can either be a dummy variable or measured continuously. As I will do heterogeneity analysis of subsamples using the Probit, I will opt for a measure of high vs. low diversity.

[11] If for instance β_1, i.e. the estimate of μ, was negative and significant and γ was positive and significant, this would indicate that higher levels of diversity were associated with less discrimination.

Heckman and Siegelman (1993) show that *"a troubling result emerges in audit or correspondence studies because the outcome of interest is not linear in productivity (as it might be for a wage offer), but instead is non-linear"*, Bertrand and Duflo (2017). In this non-linear case the employer decides to invite a candidate - or not - depending on a "perceived quality threshold". Let's denote this threshold, c, i.e. $T = 1[T^* > c]$. The rationale behind this is that in the hiring process firms evaluate a job applicant's productivity relative to a standard, and only applicants who meet that standard ($T_i* > c$) are contacted. This non-linear relationship can make the inference of the results problematic if employers hold different believes on the variance of the unobserved productivity for each group of origin, i.e. ${\sigma^2_{M=0}}^Z \neq {\sigma^2_{M=1}}^Z$. This can be viewed as a difference in uncertainty about the un-observable value. Depending on the overall quality of applications, this difference in the variance might lead to a bias of the estimate of the aggregate discrimination. Let's assume for instance an employer believes that regardless of the same degree, the quality of school (which is unobserved) is lower for minority candidates than for majority candidates but he or she knows that with a higher uncertainty, i.e. ${\sigma^2_{M=0}}^Z < {\sigma^2_{M=1}}^Z$. If the quality of the pool of applications that an employer received was high and both candidates would thus be placed below c, a higher variance in the un-observable for the minority group might then lead to a downward biased estimation of the true discrimination effect. The estimated model hence measures the following probabilities of eliciting a callback for each group:

$$\mathbb{P}[T = 1|M = 0, X] = 1 - \Phi[\frac{(c - \mathbb{E}(Z|M = 0) - \beta X + \gamma)}{\sigma_0}] = \Phi[\frac{\beta X + \mathbb{E}(Z|M = 0) - c)}{\sigma_0}] \tag{3.4}$$

$$\mathbb{P}[T = 1|M = 1, X] = 1 - \Phi[\frac{(c - \mathbb{E}(Z|M = 1) - \beta X + \gamma)}{\sigma_1}] = \Phi[\frac{\beta X + \mathbb{E}(Z|M = 1) + \gamma - c)}{\sigma_1}], \tag{3.5}$$

where Φ denotes the assumed standard normal distribution. One can see that even with neither taste - nor statistical discrimination being present, a higher variance for one group implies a smaller effect of observed characteristics on the probability of being called back and one would therefore still capture a difference in the probabilities of a call-back of:

$$\Phi\frac{(\beta X - c)}{\sigma_0} - \Phi\frac{(\beta X - c)}{\sigma_1}. \tag{3.6}$$

When estimating a simple OLS or Probit model, where the coefficients can only be identified relative to the standard deviation of the unobservable, one would thus not capture the effect in equation 3.2, but:

$$\Phi(\frac{(c - \mathbb{E}(Z|M = 0)}{\sigma_0}) - \Phi(\frac{(c - \mathbb{E}(Z|M = 1) + \gamma)}{\sigma_1}). \tag{3.7}$$

Neumark (2012) addresses this identification problem. He shows that information on how variation in observables is related to the call-back rate can be informative about the variance of un-observables. The method requires to have variation in applicant characteristics that (1) affect hiring (2) are homogeneous accross the candidates. One can for instance consider all job-specific characteristics such as the occupation type or the type of job. In such a case, one can impose the normalization that $\sigma_0 = 1$ so that σ_1 captures the variance of the unobserved variables for the minority relative to the majority group applicants. Equation 3.5 would then be rewritten as:

$$\Phi(\beta X - c) - \Phi(\frac{(\beta X - c)}{\sigma_1}). \tag{3.8}$$

The estimate of σ_1 is then used to obtain the combined effect of discrimination. This is exactly what a heteroscedastic Probit model (Williams (2009)) does when allowing the variance of the unobservable to vary with minority/majority status. The estimated model will thus be $T_i = \alpha + \beta_1 M_i + X_{i(j/m)k} + \epsilon_{ijk}$ instead of 3.3 for the Probit estimations.

As in Edo et al. I refrain from using interaction terms in my Probit estimations[12]. My main specification will therefore be a heteroscedastic Probit model as developed above. To assess the heterogeneous impact of high and low diversity both on the local and within the firm level I will run this model on the respective sub-samples of the data which are detailed in part 3.3. Note that when the sub-samples become too small, Probit models become non-convergent when including fixed effects such as department fixed effects in the OLS above. Therefore, I opt for including dummies indicating whether a department was located at a land border and a dummy for the Paris inner circle[13] when estimating these sub-samples. The former is a way to include a proxy for the special role a certain group of Europeans from neighbouring countries might play in these departments both in the composition of the population as well as of the employees. These locations and especially Paris may further differ from other parts of France in that the internationality of firms might be differently promoted by the department administration as well as out-and inward mobility may play a bigger role[14].

3.2 The experimental design

The experiment is based on a correspondence study involving a matched-pair (quartet) design. The results of the testing are then systematically matched at the municipality and firm level with administrative data. The official purpose of the study whose data I am using is to test the interaction between gender discrimination and a plethora of different signals (including ethnicity). Although the experimental protocol of that study is mainly focused on detecting gender discrimination, it uses a multitude of occupation and individual charateristics, which also makes it fitting for my study.

3.2.1 Selection of occupations

The occupations selected for the experiment are listed in Table 3.1 below. They were chosen in order to differ on the following three dimensions: First, the level of qualification required for the occupation[15]; second, the degree of gender domination within the occupation[16] and;

[12]For a discussion on the problematic of estimating and interpreting interaction terms in Probit models see Ai and Norton (2003).

[13]Defined by the INSEE code 75056.

[14]Decreuse and Schmutz (2012) in this regard highlight the particular role of the Paris region as they find African immigrants *"to prefer living outside the Paris region, all else being equal, whereas they in general prefer living in a large city"*.

[15]Classified in three categories: (1) low skilled, (2) medium skilled (without management function) and (3) high skilled (with management function).

[16]Classified again in three categories: (1) predominantly male, (2) mixed (3) predominantly female.

third, the degree of labour market tightness (hereafter called tension) of the occupation[17].

Table 3.1: Occupations by categories

Qualification	Gender composition	Tension	Occupation	Visibility Socio-prof. category (1)	Visibility occupation French def. (2)	Visibility occupation US def. (3)
low	Feminized	Low	Administrative Employee	High	Low	Low
	Masculinized	Low	Electrical and electronic equipment assembler	Low	Low	Low
	Mixed	Low	Stock clerk, sales floor	High	High	High
	Masculinized	High	Cook	High	High	High
middle	Feminized	Low	Human resources specialist	Low	Low	Low
	Masculinized	Low	Commercial sales representative	Low	Low	Low/High
	Mixed	Low	Controlling	Low	High	Low
	Masculinized	High	Software and applications developer	Low	Low	Low
high	Feminized	Low	Human resource managers	Low	Low	Low
	Masculinized	Low	Industrial production managers	Low	High	Low
	Mixed	Low	General and operations managers	Low	Low	High
	Masculinized	High	Restaurant managers	Low	High	High

Data: DADS, *Postes 2015* were used to determine the gendered composition of the workforce. The identification of occupations under labour market tension was based on the data from *l'enquête Besoins en Main d'oeuuvre (BMO)*, which measures employers' recruitment intentions for the upcoming year by both new job creations and replacements (high tension = few job-creations foreseen; low tension = many job-creations foreseen). Finally, the level of qualification of the occupations was defined using the classification of professions and socio-professional categories (PCS) from 2009. See Table 3.6 for more details on the definition and data source for the definition of the three visibility measures.

Based on these categories, the study covers a total of 12 occupations of which four are low-skilled occupations, four are medium-skilled occupations, and four are high-skilled occupations (with managerial functions). For each skill level, one of the four occupations shows a high labor market tightness. The nine occupations with low labor market tightness were chosen according to their degree of feminisation, by retaining in each qualification level a predominantly male occupation, a balanced one and a third predominantly female one. As for the occupations with high labor market tightness, they were chosen preferably from predominantly male and/or balanced occupations.

While the above explained categories were already given when I joined the project, I added a third categorization, namely *visibility*, for the purpose of this thesis. It classifies the already-chosen occupations in two categories (low/high) based on the visibility of an occupation, measured as the importance of contact with costumers and the public. Please see the next section on variable creation for a more detailed explanation on how these are defined. It is however already worth noting that they fit the experimental design quite well. This can be seen by looking at Table 3.1: both the low and the high skilled occupations each include two highly visible and two less visible occupations and the middle skilled occupations both include one highly visible one. It is only when measuring visibility on the *"socio-professional category-level"* (2 digit classification instead of 3), that the picture does not look as balanced anymore. Here, only the socio-professional categories including low skilled professions represent high visibility. In the realm of this thesis, the two measures on the occupation level will hence be the preferred ones and the results for the socio-professional category measure will have to be looked at with caution, as it is representing only low-skilled occupations.

3.2.2 Applicant's characteristics

Main profiles

The applicant's profiles send in response to job offers of the above 12 occupations are

[17] Classified in two categories: (1) low tension, and (2) high tension.

distinguished by ethnic origin (French/Maghrebian), age group[18], gender (male/female) and social background (disadvantaged/advantaged). Based on these variables applications to each job offer are send with 4 distinguished identities which combine gender and origin in a factorial way (1: female/french, 2: male/french, 3: female/maghrebian, 4: male/maghrebian). The four applications all represent a fixed age range (all candidates are young in response to the first offer, older in response to the second offer, etc.), including only two age groups per occupation (young and middle for low skilled and middle and older for medium/high-skilled). For each CV, the exact first name that appears on the CV is randomly drawn with a probability of 0.5 of representing a higher/lower social background (see a list of the names used and their social background in Table 3.2).

Table 3.2: Selection of names by gender, origin, social background and age group

First name	Gender	Origine	Social background	Age group
RAJA	female	Maghrebian	disadvantaged	23 - 30 years
SELMA	female	Maghrebian	advantaged	23 - 30 years
SAMIA	female	Maghrebian	disadvantaged	33 - 40 years
FIONA	female	Maghrebian	advantaged	33 - 40 years
AMEL	female	Maghrebian	advantaged	48 - 55 years
NORA	female	Maghrebian	disadvantaged	48 - 55 years
ANAIS	female	French	disadvantaged	23 - 30 years
CLAIRE	female	French	advantaged	23 - 30 years
ELODIE	female	French	disadvantaged	33 - 40 years
MELANIE	female	French	advantaged	33 - 40 years
SANDRINE	female	French	disadvantaged	48 - 55 years
LAURENCE	female	French	advantaged	48 - 55 years
KARIM	male	Maghrebian	disadvantaged	23 - 30 years
OUALID	male	Maghrebian	advantaged	23 - 30 years
HASSAN	male	Maghrebian	disadvantaged	33 - 40 years
YACINE	male	Maghrebian	advantaged	33 - 40 years
MOHAMED	male	Maghrebian	disadvantaged	48 - 55 years
HEDI	male	Maghrebian	advantaged	48 - 55 years
ANTHONY	male	French	disadvantaged	23 - 30 years
CLEMENT	male	French	advantaged	23 - 30 years
SEBASTIEN	male	French	disadvantaged	33 - 40 years
GUILLAUME	male	French	advantaged	33 - 40 years
PHILIPPE	male	French	disadvantaged	48 - 55 years
JEAN	male	French	advantaged	48 - 55 years
Surnames included				
Maghrebian	**Maghrebian**	**French**	**French**	
Yousfi	Sadi	Morel	Besson	

Selection of names

In order to detect average hiring discrimination between a minority and majority group one needs to ensure that the main dependent variable - the name suggesting the origin of the candidate - is representative for existing names in the society. For that purpose the researches pre-drew a list of the 109 most common names occuring in the 2008 survey "Trajectoires et Origines" (TeO, 2008)[19]. Those names were then tested in an online

[18] Age is determined by work experience in three groups: 4 to 6 years of work experience (23 - 30 years); 14 to 16 years of work experience (33 - 40 years), 29 to 31 years of work experience (48 - 55 years).

[19] TeO is a large, cross-sectional French survey conducted in 2008 among 21,761 individuals aged 18 to 60 residing in Metropolitan France.

survey in which participants were asekd to classify them by gender, origin and social background. The final list of names includes only names that were both, by the objective definition in the TeO database, as well as by the perception of the survey participants, listed under a given category. The final list of 24 first names as well as the four surnames included in the testing are listed in Table 3.2 below.

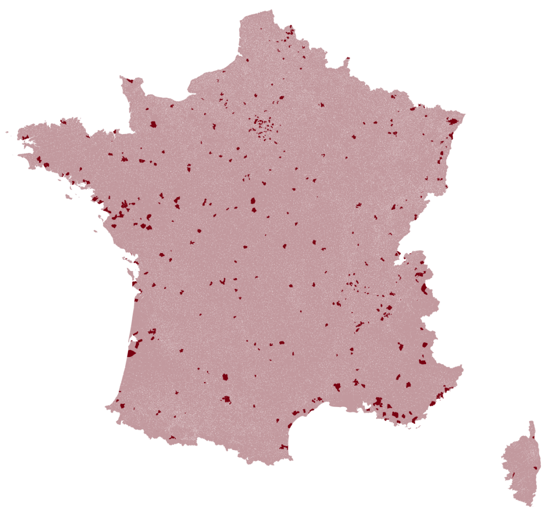
The surnames were proposed by the external company, hired to send out the CV's and collect the responses, who had already been working on such correspondence testings. The selection criteria was to choose four surnames (two French, two Maghrebian) with very different tones, in order to facilitate the identification of the candidates when receiving call-backs over voice-messages on the phone. Surnames were further chosen under the criteria that they should be associated with a wide variety of names. A particular condition for the two "French" surnames was, that they had to be equally distributed over France. The Maghrebian surnames, were conditioned on being homogeneously distributed within North African countries.

Other characteristics included as signals in the CV's
In order to take into account the possible differentiated effect of the content of CV's according to the gender of the candidates (and in the case of this thesis their origin), the researchers added signals to the applications. These signals are included in the form of signal blocks, combining inactivity and presence of children; marital status and presence of children and one block with no such signal at all[20].

3.2.3 Experiment implementation

The testing started on the 11th of December and the sending of applications continued without interruption until March 13, 2020. During that time a total of 2916 applications were sent out, responding to 729 vacancies by firms located in 464 different municipalities across the whole of France (see Figure 3.1 for their geographical distribution).

Figure 3.1: Geographical distribution of municipalities included in the sample



[20]The final signals included are thereby: a: no signal, b: in a couple and 2 children, c: period of inactivity, d: in a couple, 2 children and period of inactivity, e: in a couple, f: single, g: in a couple and 2 children, h: single and 2 children

Job offers are selected on first served basis, conditional on the inclusion criteria. The sending out of the applications is done by an external company who has already worked on a number of correspondence studies. They receive a list of web-scraped offers from the most common French job-agency "Pôle Emploi", which matched the listed occupations. The company then filters each offer again "by hand" in order to assign it to a CV template (there are four templates for each occupation*age group). Given the limited number of those CV profiles it is ensured to never re-apply to a firm more than once with the same occupation[21].

Address assignment:
The allocation of addresses to candidates is carried out by a geolocation program according to the location of a company. These locations are subject to the following restrictions: In the Paris region and in all urban units with a population of more than 100,000 inhabitants, the places of residence of the four candidates are located within a radius of 5 to 10 kilometers around the location of the station. In all other cases (offers located in urban units with less than 100,000 inhabitants), the four applicant's addresses are located within a radius of 20 kilometres of the location of the firm for unskilled occupations and more than 100 kilometres from the location of the firm for skilled occupations.

Randomization method and units
Resumes are randomly matched to applications and sent in response to a job offer. Randomisation is further taking place to create the 4 CV's with which we apply to each job offer. For each of the 12 occupations and their two age groups an application will be randomly characterised by an identity (gender, first name, surname, social class), a CV template (experiences and training), a signal block and a signal from that block.

3.3 Variable creation

3.3.1 Main Outcome: call-back measure

A callback is defined as a positive personalised phone or e-mail contact by a potential employer. This is usually a request for an interview, but employers also contact applicants asking for additional documents/information or for a call-back by the applicant. An invitation is defined as a personalised phone or e-mail contact in which the potential employer expresses interest in conducting an interview.

Potential outcomes, coded in the raw data:

1. "Neutral" response (an automatic message of receipt)
2. Invitation to an interview
3. Application rejected
4. Call without leaving a message
5. "Neutral-positive" response (request for a call-back or for additional documents)
6. Request for additional information or to respond to an online questionnaire
7. Other case

[21] Note that we indeed apply more than once to a municipality or even a firm but then for a different occupation.

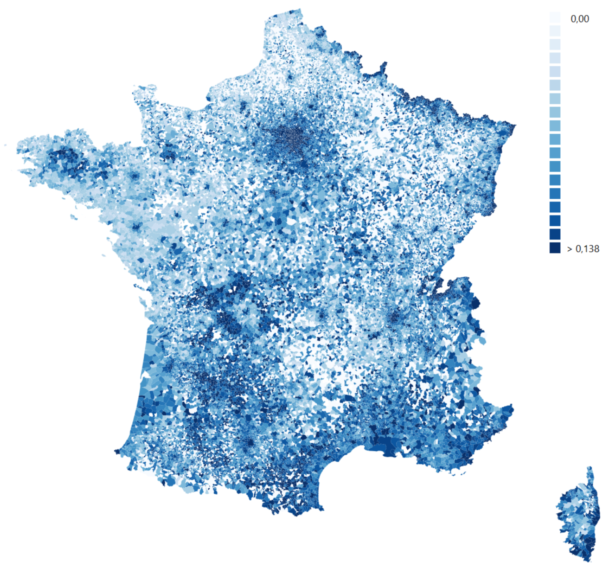
Using this information, the callback measure is a dummy variable equal to one if the raw data indicates the event (2), (4), (5), (6) and (7) and zero otherwise. There is also a more conservative measure that only regards receiving an invitation to an interview as a positive callback, expressed by a dummy variable equal to one only if the raw data indicates the event (2) and zero otherwise. I also define a dummy variable measuring "rejection" which is equal to one if the raw data indicates the event (3) and a dummy variable defining "no response" if the raw data indicates the event (1) or no outcome has been coded (yet)[22].

3.3.2 Spatial and within firm diversity measures

Three alternative approaches of measuring diversity in the municipality

To capture homophilious behaviour based on the composition of the neighbourhood, one would want the most precise measure of the composition of the population on the smallest possible local scale. It however turns out that researchers working on the concentration of ethnic/racial minorities in France face major restrictions in the form of *"France's universalistic, colorblind national model that does not recognize ethnic/racial groups"*, McAvay and Safi (2018). Therefore, a general draw-back is thus the fact, that data describing the situation of minorities in France based on ethnic origin is, by law, not allowed. This approach is the reason that generally the members of different generations of the immigrant population are statistically identified by their country of birth or the country of origin of their parents (Sabbagh and Peer (2008); Simon (2008)).

Figure 3.2: Geographical distribution of the share (in %) of immigrants (foreign borns) by municipality in France, 2017



Scrolling through the available data-bases on these measures, one generally faces a quality vs. quantity trade-off. The main limitations of the most rich databases in terms of content,

[22]Note that the experiment was still ongoing during my analysis of the data.

are their lack of statistical power. Qualitatively rich databases such as the Labour Force Survey or the TeO, which is *"one of the rare data sources, randomly pulled from the French census, in which immigrants and their offspring can be identified, and which contains precise information on national origin"* (McAvay and Safi (2018)), have by far not enough observations or lack the geographic precision to compute shares on the municipality level.

I rely on the quantitatively richest base, the French Census, following the empirical strategy of Edo et al. (2019). This measure reflects the level of inter-nationality as a whole in a municipality. As shown in Figure 3.2, this allows to capture the share of the foreign population on a very small local scale. Columns 3-6 in Table 3.3 display the grouping of my sample in tertiles, based on this measure. It can be seen that the municipalities in which the firms are located (Figure 3.1) indeed show a good variation in their population composition. The two main problems with this data source are however the lack of the information on the origin of the people counted as immigrants and that there is no information on descendants of immigrants.

I therefore use a second innovative index that has been recently developed by another Master student of PSE, Sirugue (2020). This index captures a fitting decomposition of the population, considering the experimental design at hand. Sirugue (2020) is aiming to overcome both the quantity vs. quality trade-off as well as the color blindness of French statistics by creating a measure that is based on the recognition of the origin of last names. To circumvent the above mentioned limitations he wrote an algorithm that attributes a probability of French origin and a probability of Arabic origin to last names. Individuals with a sufficiently high probability of Arabic origin and low probability of French origin for their last name are counted to be of Arabic origin. As a data source he is - so far - using the "last name file" (fichier patronymique, 1999) available for research purposes from the National Institute of Statistics and Economic Studies (INSEE). It gathers the last name and municipality of birth of every individual for whom INSEE collected a birth certificate, born between 1966 and 1990. Applying his algorithm to this database allows him to create a measure of the share of individuals with an Arabic last name per municipality of birth between 1966 and 1990, including municipalities that had at least 50 births during that period. Merging this information with my experimental data at hand leaves a sub-sample of 1560 (out of 2916)[23], for which this information is available.

Although the birth cohorts match exactly two of the applicants age groups in the experiment at hand, the measure only captures the location of where these people were born (and potentially grew up), which is not necessarily where they live today. I however believe it is a relevant and well suited proxy arguably overcoming the lack of inter-generational and ethnic origin identification problems in other studies.

My two main measures of diversity of the neighbourhood are summarized as: **1. Share of immigrants (foreign-borns) by municipality**, Data: French Census, Tableaux détaillés - Population par situation quant à l'immigration (Population by immigration status), 2017; and **2. Births with Arabic last name (1966 - 1990)**, Data: Sirugue (2020) and INSEE, Fichiers des noms patronymiques de 1891 à 1990 - Édition 1999.

[23]This reflects the number of applications to 224 municipalities out of 464 municipalities included in the experimental data.

Table 3.3: Group creation of neighbourhood diversity (measured in % by municipality)

Share of... by registered pop. in mun.				
...Foreign population (2017)	N	Min	Max	Measure
1	976	0	7.3	low
2	968	7.3	13.8	low
3	972	13.8	44.8	high
...Births with Arabic last name (1966-1990)	N	Min	Max	Measure
1	520	0	2.6	low
2	536	2.6	7.5	low
3	504	7.5	33.8	high
Share of... by no. of employees in mun.				
...French, born in France (2015)	N	Min	Max	Measure
1	988	31.0	85.2	low
2	960	85.3	89.8	low
3	968	89.9	1	high
...Non-European foreigners (2015)	Freq	Min	Max	Measure
1	1012	0	3.5	low
2	836	3.6	7.3	low
3	924	7.3	38.3	high
...European foreigners (2015)	N	Min	Max	Measure
1	956	0	1.2	low
2	896	1.3	2.8	low
3	920	2.8	18.9	high
...French, born in oversea depart. (2015)	N	Min	Max	Measure
1	972	0	0.6	low
2	972	0.6	1.3	low
3	972	1.3	5.7	high
...French, born abroad (2015)	N	Min	Max	Measure
1	928	0	4.4	low
2	1048	4.5	8.4	low
3	796	8.3	33.3	high

Data: French Census, *Tableaux détaillés - Population par situation quant à l'immigration, 2017* (Share of Foreign population, 2017); Sirugue (2020) and INSEE *Fichiers des noms patronymiques de 1891 à 1990 - Édition 1999* (Share of births with Arabic last name (1966-1990)); DADS, *Postes 2015* (Share of... by no. of employees in municipality); Cutoffs are defined splitting the sample in tertiles of which the highest is the "high share"

As a robustness test, I additionally created another proxy which does not capture the share of immigrants in the population but among all employees in a given municipality. To do so, I am using information from the DADS social security register (Déclaration annuelle de données sociales)[24]. The data consists of individual-level observations of all private sector workers, plus all hospital and local civil service workers employed in France. The aim of this exercise is to create a share of immigrants employed in the municipality where a firm is located. This data allows me to disentangle between European, Non-European, French (born in France), French (born in oversea departments) and French born abroad employees by municipality using combined information from the variables: (1) *"Position held by a foreign employee in year n (IND_ETRANGER)"* which indicates whether an employee was French; foreign, coming from the European Economic Area; or foreign, coming from outside that area. And (2) *"Employee's department of birth (DEP_NAISS)"*[25]. Using the municipality identifyer of the location on the company, I then define shares of employees with a given background by the total number of people employeed in that region. A reverse reading of the share of French employees, born in France is therefore the potentially closest to the share of immigrants living in a municipality (see Figure 1.1 in the introduction to see their correlation). A cross-correlation table of all measures used can be found in the Appendix, Table .4. I then merge this information with our experimental data and create subsets of the included firms located in "high" and "low" diversity share areas by splitting the sample in tertiles for each measure used, indicating the highest 30% of my sample as the "high" share municipalities. The categorization of all variables is displayed in Table 3.3. Note that all results regarding the measure of diversity by the share of "Europeans" and "French, born in oversea departments" have to be looked at cautiously, as there is not a lot of variation of these two measures in the sample.

Firm and job characteristics:

In order to investigate the extent of ethnic homophily on the firm, rather than on the spatial level, I create a proxy for the recruiter's identity. I do so using the name of the person that the application was addressed to[26]. I classify these names by origin (Arabic/African/Asian sounding names vs European sounding names) and gender.

Recruiter identity (gender and origin)

Table 3.4: Classification of recruiter identity by origin of name and gender

Origin recruiter	**Freq.**	**Percent**
European name	2,332	90.11
Non-European name (Arabic)	256 (192)	9.89 (7.42)
Total	2,588	100.00

Gender recruiter	**Freq.**	**Percent**
Male	1,088	41.21
Female	1,552	58.79
Total	2,640	100.00

Data: Names and salutation of people that the applications were addressed to (if provided in the vacancy description); origin was classified using the *Liste de Prénoms* (first names existing in France by language of use and gender of name (2014), https://www.data.gouv.fr/en/datasets/liste-de-prenoms/

The gender is deduced from first names and/or abbreviations before the last names

[24]Access to the DADS data was obtained through the CASD (Centre d'accès sécurisé aux données) dedicated to researchers authorized by the French Comité du secret statistique.

[25]Coded providing the INSEE (municipality) code, except for French overseas collectivities (98) and foreign countries (99).

[26]This corresponds to the recruiter's name that was mentioned on the job advertisement page.

like “M.” and “Mme”. To define the origin (and gender for those where no abbreviation was indicated), I matched the names with the "Liste de Prénoms" (see reference in the Table 3.4). This data-base provides the "language of use" for each first name, present in France as of 2014. Following their classification, an employer is considered to be of (i) non-European origin if his/her last name appears in Arabic/Asian/African language and of (ii) European origin otherwise. Roughly ten percent of recruiters have a non-European sounding name. Most of them (7.4%) have an arabic-originated name (see Table (3.4)).

Within-firm diversity measure

For a subset of 204 of the 729 firms I merged the experimental data with the firm-specific information on the composition of employees, using again the employee database DADS 2015.

Table 3.5: Group creation of firm diversity (measured in % of no. of total employees by firm and size of the firm)

Share of... by no. of employees in firm				
...French, born in France (2015)	N	Min	Max	Measure
1	272	0	87.8	low
2	276	88.1	96.9	low
3	268	97.1	1	high
...Non-European foreigners (2015)	Freq	Min	Max	Measure
1	428	0	0	low
2	116	0.3	2.2	low
3	272	2.3	1	high
...European foreigners (2015)	N	Min	Max	Measure
1	544	0	0.51	low
2	272	0.57	33	high
Size of the firm				
	N	Min	Max	Measure
1	284	1	11	low
2	264	12	38	low
3	252	40	1461	high

Data: DADS, *Postes 2015*

In order to be able to merge the experimental data with the DADS using a firm-identifier (instead of a municipality identifier as done in the previous section), I wrote and applied a program (API-finder)[27] to find out the official registration number of a firm ("siret"). This program allows an automated multi-criteria search of the data at hand using a new service of INSEE, called "API Sirene". This service provides access to information on companies and establishments registered in France, including their registration number. Using the API-finder, multiple search query results are requested from the "API-Sirene" and after batch processing them, all results are combined in a data-frame for further data analysis tasks.

[27] This program is still being improved as the "API-Sirene" is developing further, allowing for more variables and criteria to be tested.

The search was conducted in three steps: (1) Before applying the "API-finder", I used the most recent data-set on registered companies in France to attempt a direct merge based on the name of the firm indicated in the job offer and the name of the municipality. I did so, because the "API Sirene" does not yet include a function to search for the variable *"denominationUsuelleEtablissement"*, the name of a firm used in public, although that would be the preferred merging variable. (2) For the second step I created an API search function. It allows to apply a multi-criteria-search on the API with the variables "denominationUniteLegale" (official registered name of the legal entity) and the municipality code, by manually choosing a "fuzzyness" level of the match[28]. Running the experimental data through that function, searching for the indicated firm-names and their respective municipality, I get a final set of 301 "unique" matches[29]. I proceeded with my analysis using this subsample[30]. (3) Because the unique matches resulting from that search included some "fuzzy" ones, I did a last manual check, comparing the matches of the names indicated in our sample and those found on the "API-Siren". This manual check resulted in the identification and manual re-coding of a total number of 23 firms that were merged with a wrong company using the automated processes.

Table .1 in the Appendix displays summary statistics, comparing the sub-sample used in this analysis and the rest of the sample. One can see that the core difference between the municipalities found with the automated process and firms that I found no match or multiple matches for, is that the firms included in this sub-sample are located in much smaller municipalities. They are also significantly less likely to be located in the Paris area. This makes sense, as the likelihood of finding multiple-matches in larger cities and municipalities is potentially higher than in small municipalities. Names of firms might also be more distinct in smaller municipalities than in larger ones.

I then proceeded to merge this sub-sample with the DADS firm-level data using the identified firm registration number. This merge results in a total of 204 matches as there are 97 firms included in our data that were created after 2015. Unfortunately the DADS data already available for 2016 and 2017 does not include the necessary variables to identify the origin of employees, so the 2015 data is the most recent information I can use. I am then finally able to create a within-firm diversity measure following the same procedure as creating the variables by municipality, but now calculating the share of European, Non-European and French (born in France) employees within a firm. The classification in high and low categories also follows the same procedure as above and is displayed in Table 3.5. I also added an additional variable capturing the size of the firm, measured by the total number of employees. The grouping of this variable is displayed in the bottom three lines of Table 3.5. It shows that there is a large variation in the sample also given the size of the firms, ranging from one employee to 1461, with a third of the firms included in the sample employing between 12 and 38 people.

[28] If the search is done on several words, the Levenshtein distance is calculated with permutation of 2 words, otherwise with 2 letters. Note that allowing the Levenshtein distance to be larger, did not result in any more unique matches but more "multiple-matches".

[29] I find multiple matches for 280 firms and no match for 148 using this automated process.

[30] Given time-constraints related to the Covid19 pandemic I was not able to do a manual search for the right matches of the rest of the firms. As the new service of the "API-Sirene" is constantly improved, I believe one will be able to find more unique matches using additional criteria (such as the address of the establishment and the date of application) with some more time and an improvement of the function.

Table 3.6: Classification of occupations by "visibility", i.e. the proportion of contact with costumers

	French visibility measure (FQP, 2003) by Socioprofessional Categories (PCS, 2 digits)			French visibility measure (FQP, 2003) by Profession (PCS,2 digits)		US visibility measure (O*Net)	
			(1)		(2)		(3)
	Profession	Socio-profesisonal category	% Contact PCS, 2 digits	Profession	% Contact PCS, 3 digits	Profession	% Contact
1	Electrical and electronic equipment assembler	Skilled industrial workers	17.80	Electrical and electronic equipment assembler	12.41	Software and applications developer	23.5
2	Commercial sales representative			Administrative employee	39.77	Electrical and electronic equipment assembler	29
3	Industrial production managers	Engineers and company technical managers	45.60	Human resources specialist	48.35	Industrial production managers	34.8
4	Software and applications developer			Human resource managers	48.35	Human resource managers	40.3
5	General and operations managers			Commercial sales representative	50.00	Administrative employee	41
6	Restaurant managers			Software and applications developer	50.00	Controlling	41.5
7	Controlling	Administrative and commercial executives of companies	55.40	General and operations managers	66.02	Human resources specialist	44
8	Human resources specialist			Controlling	77.61	Commercial sales representative	47.2
9	Human resource managers			Restaurant managers	80.00	Stock clerk, sales floor	53.5
10	Cook	Skilled artisan type workers	57.90	Cook	80.67	Restaurant managers	55.7
11	Administrative employee	Administrative employees	62.90	Industrial production managers	93.02	Cook	57
12	Stock clerk, sales floor	Commercial employees	94.70	Stock clerk, sales floor	94.96	General and operations managers	71
	Classification = Low		Classification = High				

Classification of the 12 occupations included in the experiment by different measures of the visibility of the job. The proportions in **(1) and (2)** are based on the same source but aggregated on different occupation-digit-levels. *Data:* Laouénan (2013) and Combes et al. (2016) who use the FQP survey data, 2003 INSEE. *Index base:"Was your job in direct contact with the public?"*. **(3)** *Data:* Survey job task data from the Dictionary of Occupational Titles (DOT - US Department of Labor, Employment and Training Administration, 1977).*Index base: "Performing for people or dealing directly with the public. This includes serving customers in restaurants and stores, and receiving clients or guests."* (see also Laouénan (2017)).

3.3.3 Visibility of the job - measure

As noted above, I use three different indexes to measure the "visibility", i.e. the importance of contact with customers. Two of them are based on a French classification and are retrieved from the papers Laouénan (2013) and Combes et al. (2016), while the third is using a classification based on US data.

For the first two, the papers at hand provide the measures by socio-professional category (Laouénan (2013)) and occupation (Combes et al. (2016)) in their appendices. In both classifications information from the FQP (Formation et Qualification Professionnelle) survey performed in 2003 by Insee is used to compute the fraction of contact jobs in each socio-professional category and by occupation. The FQP is an individual-level database derived from a survey over a representative sample (39,285 persons) of the French population. In face-to-face interviews, working individuals, or people who stopped working less than 5 years ago, give a yes or no answer to the question: *"Was your job in direct contact with the public?"*. Laouénan (2013) provides this measure, denoted as the percentage of contact jobs on the 2-digit occupation level, i.e. by socio-professional category (see columns three and four in Table 3.6).

As the 2-digit level might give a quite blurry measure of the actual importance of contact for each occupation (which is defined on the 4-digit PCS-level), I refer to the percentage of contact-jobs by occupation, classified on the 3-digit level in Table 3.6 (columns five and six) as my preferred one (Combes et al. (2016)).

In an attempt to have even more precision and as a robustness test, I use the "Code Rome", the job classification of the French job agency "Pôle Emploie", to define a third contact measure. The Code Rome defines distinct jobs that we apply to for each occupation. On the job portal, matching vacancies for the occupation "Controlling" (PCS 4-digits) are, for instance, selected for two types of jobs: (1) "Contrôle de gestion" (Code Rome: M1204), and (2) "Audit et contrôle comparable et financier" (Code Rome: M1202). Using official international conversion scales (ISCO), I get information on the corresponding "SOC" - the US classification of occupation measures (see Appendix, Table .2 in the Appendix for the correspondence PCS - Rome - ISCO - SOC). This allows me to use the measure of contact jobs from the Occupation Information Network (O*NET). The network provides more than 275 standardised descriptors of skills, knowledge, tasks, and requirements for 974 occupations. I use the index for how important *"Performing for people or dealing directly with the public"* is in a given occupation (see also Laouénan (2017)). As can be seen in Appendix, Table .2, there are many more sub-categories for each Code-Rome, when defining them by the US classification (SOC). The final index is therefore the mean of the contact measures (SOC) by job category (Code Rome) (displayed in Appendix, Table .3). Unfortunately, the experimental data base only provides me with the Code Rome for 513 of the 729 vacancies. I therefore impute the mean of the contact measures (SOC) by occupation category (PCS) (displayed in columns seven and eight in Table 3.6), whenever the Code Rome was unobserved.

Measure of "prejudiced" population

As a robustness check for detecting customer discrimination Combes et al. (2016) use the share of votes in favor of the far-right party (the Front National or FN) in the first-round of the 1995 French presidential election as a proxy for the share of "prejudiced" people in a locality. To adapt this method to my context, I use the share of votes in favor of the far-right party leader (Marine Le Pen, Front National) in the first-round of the 2017 French presidential election. The vote shares by municipality are taken as a proxy for the share of "prejudiced" people in a municipality. The higher the share, the greater might be the discrimination, as employers anticipate the discriminatory behavior of their community. The cut-off for high and low vote shares in my sample is found in the Table below:

Table 3.7: Group creation of votes for Marine Le Pen in first round of the France national elections in 2017

	N	Min	Max	Measure
1	972	3.61%	12.09%	low
2	972	12.17%	20.41%	low
3	972	20.44%	52.15%	high

4 Results

As I want to compare the variation in discrimination based on the share of foreigners residing or working in the municipality, I need to ensure that I am not capturing other differences which are, for instance, related to the population size of a municipality. As one can seen in Figure 3.2 the share of immigrants by municipality tends to be high around large urban areas and along land borders as well as at the Mediterranean sea border. Brutel (2016) highlights, that in 2012 eight out of ten immigrant inhabitants lived in major urban centres, while less than six out of ten non-immigrants did so. The report also establishes that immigrants reside less frequently in the outskirts of the major urban centres than non-immigrants (8.7% compared to 19.6%) and that a smaller proportion of immigrants live in medium or small areas (5.0% compared to 7.7% for non-immigrants).

Table 4.1 therefore provides the descriptive statistics on job- and firm-level characteristics as well as on municipality characteristics that will later be included as controls in all sub-sample analyses[31]. The table compares these summary statistics for the group of the sample in which the share of the foreign population in the location of the vacancy is defined as "low" with the group of the sample in which the share of the foreign population in the location of the vacancy is defined as "high", i.e. larger than 13.8% (as defined in Table 3.3)[32]. Column four shows the statistics for the full sample and column five the p-value of a mean-comparison test (ttest) for each respective variable. One can see that groups differ significantly in terms of both firm- and municipality level characteristics. While the share of recruiters with a non-European name is only 6.2% in the sample that I define as the "low diversity" sample, this share amounts to 17.2% for the "high diverse" sample.

[31] Note that this table does not include applicant's characteristics except for age, as they are randomly assigned and thus perfectly balanced across the applications, regardless the sub-sample. It also does not include the additional measures of the share of foreigners computed using the DADS, as that data cannot be downloaded for confidentiality reasons. For more information on those, please see the classification in low and high categories in Table 3.3 as well as in the cross-correlation table of the diversity proxies in the Appendix, Table .4.

[32] You can find the same descriptive table by the share of births with Arabic last names in the Appendix, Table .6.

Table 4.1: Descriptive statistics of the full sample and by share of foreigners living in the municipality of the firm's location

	Share = Low (N=1944)	Share = High (N=972)	Full sample Total (N=2916)	p value
Non-European recruiter (yes/no)				< 0.001
N-Miss	216	112	328	
Mean (SD)	0.062 (0.242)	0.172 (0.378)	0.099 (0.299)	
Range	0.000 - 1.000	0.000 - 1.000	0.000 - 1.000	
Occupations				< 0.001
Administrative employee	176 (9.1%)	64 (6.6%)	240 (8.2%)	
Commercial sales representative	124 (6.4%)	180 (18.5%)	304 (10.4%)	
Controlling	160 (8.2%)	76 (7.8%)	236 (8.1%)	
Cook	268 (13.8%)	72 (7.4%)	340 (11.7%)	
Electrical and electronic equipment assembler	152 (7.8%)	32 (3.3%)	184 (6.3%)	
General and operations managers	108 (5.6%)	64 (6.6%)	172 (5.9%)	
Human resource managers	32 (1.6%)	16 (1.6%)	48 (1.6%)	
Human resources specialist	168 (8.6%)	144 (14.8%)	312 (10.7%)	
Industrial production managers	200 (10.3%)	48 (4.9%)	248 (8.5%)	
Restaurant managers	124 (6.4%)	56 (5.8%)	180 (6.2%)	
Software and applications developer	128 (6.6%)	148 (15.2%)	276 (9.5%)	
Stock clerk, sales floor	304 (15.6%)	72 (7.4%)	376 (12.9%)	
Age group				0.078
Mean (SD)	1.636 (0.633)	1.679 (0.605)	1.650 (0.624)	
Range	1.000 - 3.000	1.000 - 3.000	1.000 - 3.000	
Municipality characteristics				
Annual Standard of Living[33] in Euro (2017)				< 0.001
N-Miss	0	4	4	
Mean (SD)	21674.753 (3026.899)	23044.132 (5883.006)	22129.959 (4246.017)	
Range	15550.000 - 38160.000	13590.000 - 44370.000	13590.000 - 44370.000	
Dep. is located at land border (yes/no)				< 0.001
N-Miss	8	0	8	
Mean (SD)	0.211 (0.408)	0.436 (0.496)	0.286 (0.452)	
Range	0.000 - 1.000	0.000 - 1.000	0.000 - 1.000	
Population (2017)				< 0.001
Mean (SD)	67310.677 (123246.585)	596500.362 (863639.041)	243707.239 (566419.575)	
Range	135.000 - 516092.000	115.000 - 2187526.000	115.000 - 2187526.000	
Unemployment rate (2017)				< 0.001
Mean (SD)	0.098 (0.029)	0.108 (0.027)	0.102 (0.029)	
Range	0.017 - 0.172	0.017 - 0.192	0.017 - 0.192	
Municipality in Paris (75056) (yes/no)				< 0.001
Mean (SD)	0.000 (0.000)	0.218 (0.413)	0.073 (0.260)	
Range	0.000 - 0.000	0.000 - 1.000	0.000 - 1.000	
Share of births with Arabic last name (1966-1990)[35]				< 0.001
N-Miss	992	364	1356	
Mean (SD)	0.037 (0.034)	0.114 (0.065)	0.067 (0.061)	
Range	0.000 - 0.262	0.000 - 0.338	0.000 - 0.338	
Share of votes for Marine Le Pen (1st tour)				< 0.001
Mean (SD)	18.933 (8.677)	13.346 (7.675)	17.071 (8.761)	
Range	5.560 - 52.150	3.610 - 34.270	3.610 - 52.150	
Share of votes for Marine Le Pen (2nd tour)				< 0.001
Mean (SD)	30.581 (12.355)	22.147 (11.011)	27.770 (12.568)	
Range	10.620 - 71.410	10.320 - 50.180	10.320 - 71.410	

The share of the sample of municipalities that is located in a department with a French land border is also significantly higher in the high diverse sample (43.6%) compared to the low diverse sample (21.1%). While both of these two trends are also true for the sample with a high share of births with an Arabic last name, the latter difference is lower, with 28.6% in the high share, compared to 14.1% in the low share municipalities (see Appendix, Table .6). Following the INSEE report from 2012 it is not surprising that the population size also varies substantially between the two samples[37]. Despite being significant, the unemployment rate (9.8% in low diverse, vs. 10.8% in high diverse sample)

[33] The standard of living is equal to the disposable income of the household divided by the number of consumer units. The consumer units are calculated using the modified OECD scale which allocates 1 consumer unit to the first adult in the household, 0.5 to the persons of 14 years or older and 0.3 to children under the age of 14 years.

[35] This measure captures the share of individuals with an Arabic last name per municipality of birth between 1966 and 1990 including only municipalities that had at least 50 births during that period (see Sirugue (2020)).

[37] Decreuse and Schmutz (2012) suggest that this is a result of the combination of immigrants having trouble entering the labor and housing market in small towns and their preferences for not moving their.

and the standard of living (higher in high share areas) do not vary much in their scope. Similar trends can be observed again for the descriptive statistics by the share of births with Arabic last-names.

While vacancies of firms located in the municipality of Paris make up for 7.3% in the full sample, they are all located in the high share sub-sample and make up 21% of that sample. I thereby include department dummies and a dummy capturing whether a firm is located within the municipality of Paris in all heterogeneous Probit regressions in order to avoid capturing a "Paris" effect rather than really the diversity proxied by the share of foreigners in a municipality[38].

It is further worth noting that the average share of births with an Arabic last name between 1966-1990 (which is my alternative measure of diversity) is substantially higher in the "high" (11.4%) compared to the "low" (3.7%) sample. Lastly, one can see that there was a tendency for higher support for the far-right party leader Marine Le Pen in the more homogenous municipalities (18.9%) compared to the heterogenous municipalities (13.3%) in the first tour of the 2017 French elections. This difference was even more distinct for the second tour results (30.6% vs. 22.1%). These differences are less pronounced when looking at the descriptive statistics by the share of births with Arabic last-names.

The p-values suggest significant differences between the groups regarding all character variables listed. This fact underlines the importance to control for these factors in the main regressions in order to avoid that results are driven by these differences rather than to capture the diversity effect.

I will now turn to the main outcome variable, the measure of discrimination. Table 4.2 displays the mean response rates for all potential outcomes defined in section 3.3.1., by origin of the applicant. We can see that applicants with French names elicit far more call-backs than those with Maghrebian names, with an overall rate of 37.4% against 25.4%. The overall discrimination ratio is equal to 1.5[39]. This ratio remains robust and increases slightly to 1.6 when looking at the more conservative definition of the invitation rate. Interestingly, Maghrebian applicants, despite receiving a positive response significantly less often, are not actively rejected more often than French applicants. They are however significantly more often left with no response at all (in 68.1% of cases) compared to French applicants (57.2%).

Table 4.2: Responses by origin

	French origin (N=1458)	Maghrebian origin (N=1458)	p value
Call-back			< 0.001
Mean (SD)	0.374 (0.484)	0.254 (0.436)	

[38]Note that I do not have the information on the share of births with an Arabic lastname for municipalities located in Paris (Appendix, Table .6).

[39]The positive callback ratio is calculated by dividing the total percentage of applications for which the candidate not invoking discrimination observed a positive callback by the total percentage of applications for which the candidate invoking discrimination received a positive callback. This measure can be interpreted as the ratio of applications a candidate invoking discrimination has to send to receive the same number of positive callbacks as a candidate not invoking discrimination. Note that those estimates are well in the range of 1.3 - 1.7, which Edo et al. (2019) refer to being the range of studies estimating ethnic discrimination around the world.

Table 4.2: Responses by origin

	French origin (N=1458)	Maghrebian origin (N=1458)	p value
Invitation			< 0.001
Mean (SD)	0.171 (0.377)	0.104 (0.306)	
Refusal			0.240
Mean (SD)	0.054 (0.226)	0.064 (0.246)	
No response			< 0.001
Mean (SD)	0.572 (0.495)	0.681 (0.466)	

Decomposition according to applicants' and occupation characteristics

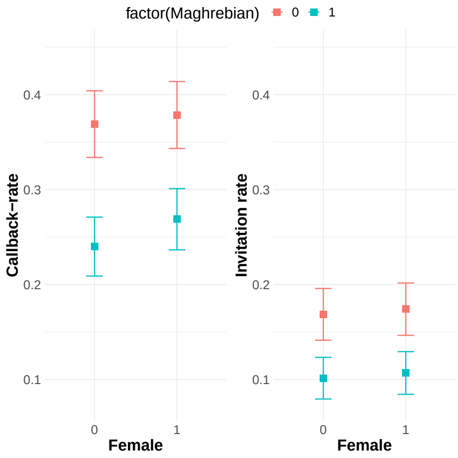
In Figure 4.1 positive response rates (call-back and invitation) are broken down by personal characteristics, which were randomly assigned to the applicants in each quartet of applications sent to one vacancy. Each of the graphs presents the average response rate observed among French and Maghrebian candidates according to the different signal, as well as the 95% confidence interval. Overall the observed difference in call- back rates related to the ethnic origin of the applicant seem very robust in indicating variations, while the observed differences in invitation rates are more volatile. Although the gap appears slightly smaller for women than for men, the gender does not change the gaps to be significantly different between French and Maghrebian candidates. They also do not suggest a significant difference between call-back or invitation rates of men compared to women, for neither French nor Maghrebian candidates[40]. While the differences sometimes become smaller when applications carry certain signals, these variations only lead to the difference becoming insignificant in some cases.

It does seem as if being inactive, while not changing the mean differences, leads to the effect becoming insignificant. The most striking effect is observed regarding the age dimension. Here the graphical display suggests that differences in call-back as well as in invitation rates are vanished for the oldest, and most experienced applicants. They are substantially lower in general compared to the younger categories.

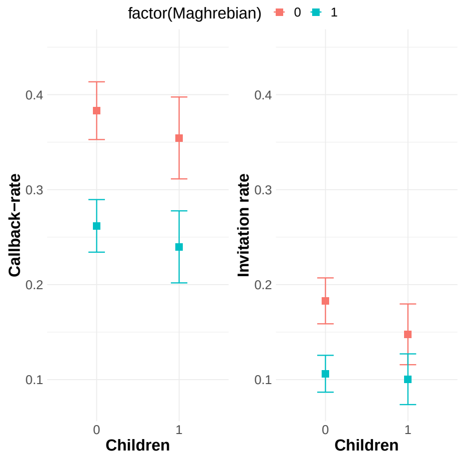
Figure 4.2 shows call-back rates for the different occupation classifications defined in Table 3.1. One can see that the visibility of the occupation, measured by the percentage of contact jobs, is both higher when defined on the socio-professional category and occupation level ((a) and (b)). High tension on the labour-market leads to generally higher call-back rates while lower tension leads to generally lower call-backs. Both however are indicating significant call-back differences between Maghrebian and French candidates.

[40]Note that I will refrain from a separate analysis by gender of the candidate in the realm of this thesis. It however remains an interesting path to explore especially when considering to add a gender-dimension to the diversity measure, captured on the municipality and within-firm level and its potential differential impact on discrimination.

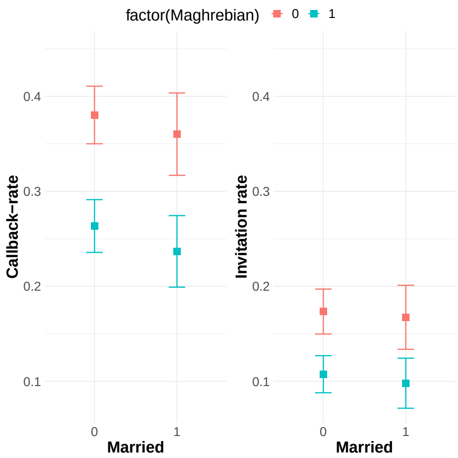
Figure 4.1: Positive response rates by origin and individual characteristics



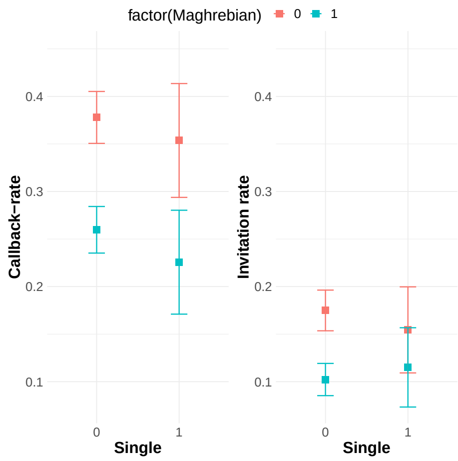
(a) 1 = being female



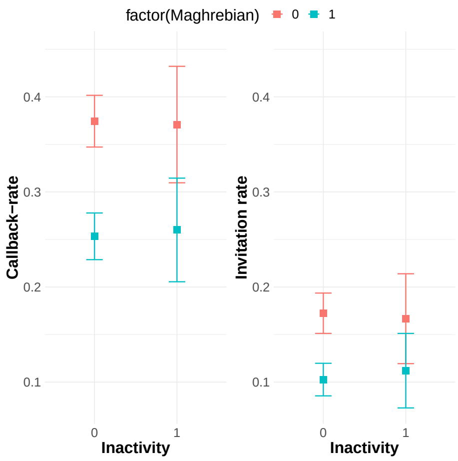
(b) 1 = having 2 children



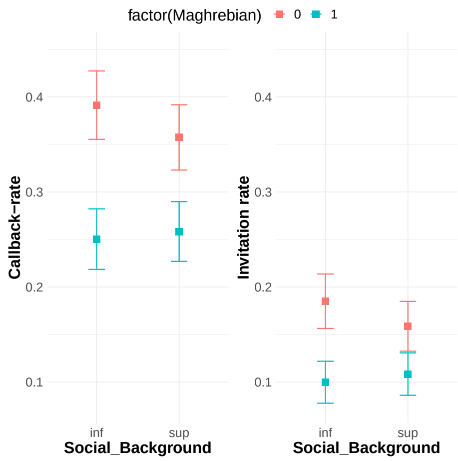
(c) 1 = being married



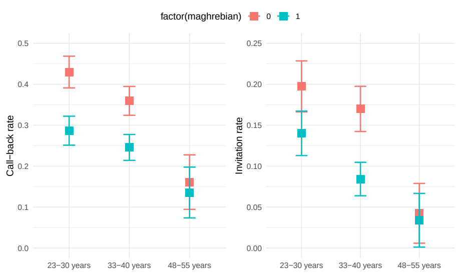
(d) 1 = being single



(e) 1 = a period of inactivity in CV



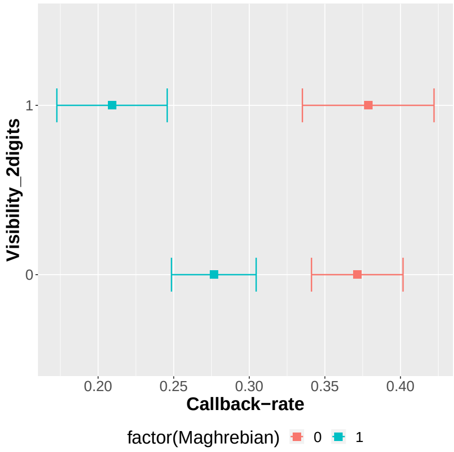
(f) 1 = disadvantaged social class



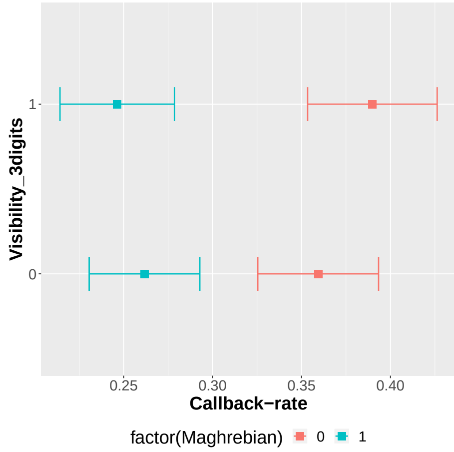
(g) CB and Invitation rates by age group

Notes: 23-30 years, (N = 1256); 33-40 years, (N = 1424); 48-55 years (N = 236)

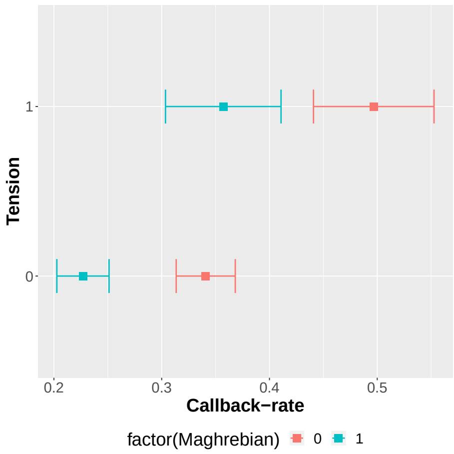
Figure 4.2: Call-back rates by origin and occupation characteristics



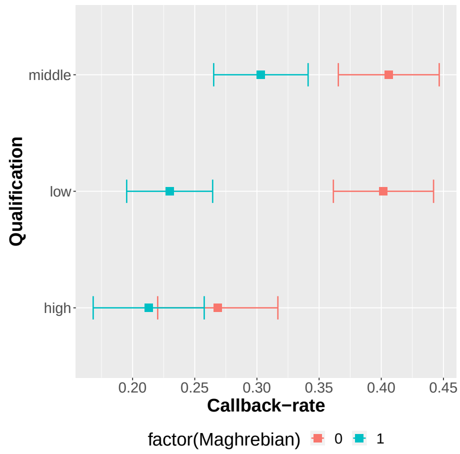
(a) 1 = high visibility,i.e. contact with costumers, of socio-professional category



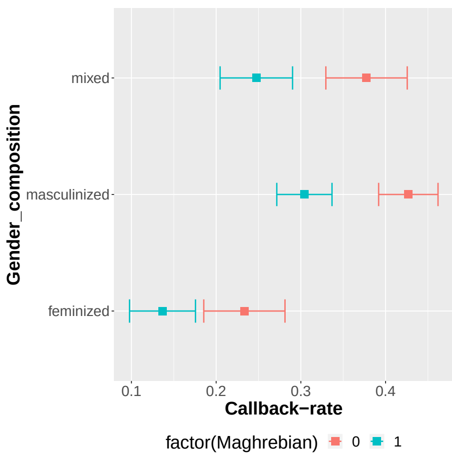
(b) 1 = high visibility,i.e. contact with costumers, of occupation



(c) 1 = high labour market tightness of occupation



(d) Rates by qualification level of occupation



(e) Rates by gender composition of occupation

Similarly to the age-group effect, the discrimination measure becomes insignificant for high skill occupations, which require a higher level of experience. This is in line with a stylized fact found in Aeberhardt et al. (2017), who report that highly educated workers experience lower employment gaps than less-educated ones in France[41]. The trend of the difference decreasing with qualification suggests that the longer the "track-record" and the higher the qualification level, the less likely one observes discrimination in hiring. One should note however that these occupations are also those with the lowest volume of vacancies as they are not as often posted on the web-page that is used so far[42]. Lastly discrimination appears to be lower for female-dominated occupations with generally lower call-back rates observed for those compared to mixed and male-dominated occupations.

Table 4.3: Number and outcome of responses by occupation

Origin	Occupation	CB(N)	Refusal(N)	No response(N)	CB-rate	N
French	Administrative employee	10	10	100	0.08	120
Maghrebian	Administrative employee	19	8	93	0.16	120
French	Electrical and electronic equipment assembler	31	2	59	0.34	92
Maghrebian	Electrical and electronic equipment assembler	48	1	43	0.52	92
French	Stock clerk, sales floor	40	5	143	0.21	188
Maghrebian	Stock clerk, sales floor	78	3	107	0.41	188
French	Cook	50	7	113	0.29	170
Maghrebian	Cook	84	5	81	0.49	170
French	Human resources specialist	31	11	114	0.2	156
Maghrebian	Human resources specialist	49	10	97	0.31	156
French	Commercial sales representative	46	16	90	0.3	152
Maghrebian	Commercial sales representative	65	9	78	0.43	152
French	Controlling	34	13	71	0.29	118
Maghrebian	Controlling	46	9	63	0.39	118
French	Software and applications developer	60	7	71	0.43	138
Maghrebian	Software and applications developer	69	9	60	0.5	138
French	Human resource managers	0	3	21	0	24
Maghrebian	Human resource managers	2	3	19	0.08	24
French	Industrial production managers	28	14	82	0.23	124
Maghrebian	Industrial production managers	39	16	69	0.31	124
French	General and operations managers	23	3	60	0.27	86
Maghrebian	General and operations managers	24	3	59	0.28	86
French	Restaurant managers	18	3	69	0.2	90
Maghrebian	Restaurant managers	22	3	65	0.24	90

One can further observe that both the overall call-back measures and the discrimination seems to vary not only by the classification but also the specific occupation tested (see 4.3 and 4.3). I will thus always include occupation dummies in the regressions in order to capture trends that are specific for certain occupations.

I will largely focus on comparisons between sub-samples when assessing the question on the extent to which homo- or heterogeneity of the population composition and discrimination are interconnected. As has been shown, these sub-samples differ on a range of other dimensions, so it is important to capture as much of the noise that may confound a precise distinction by diversity. Overall the data at hand has been shown to be a very rich source, that arguably is well representative in terms of the French society when it comes to the geographical spread of applications sent but also as it captures a large range of different socio-economic back-grounds of applicants. It covers twelve occupations that belong to six different socio-professional categories. It therefore allows to control for a lot of relevant factors when analysing call-back rates in hiring on sub-samples and therefore makes the inference of findings to the French labour market and society as a whole more valid. With this said, I will now turn to the presentation of the main results.

[41] Creating an "employability" index, the authors show that the higher the employability, the lower the unemployment gaps between North African and French men.

[42] As this is an ongoing experiment, the research group therefore decided to consult new sources in order to increase the number of applications for these categories and thus enable to more precisely distinct the effect by occupations

Figure 4.3: Call-back rates by occupation

4.1 Main regression results

In order to recover unbiased estimates of discrimination, I include variables which impact the probability of hiring, regardless of the ethnic origin. While the type of jobs do affect hiring their effects did vary with ethnicity, which makes them not a valid candidate to capture the variance. The individual signals such as the age of the applicant however turned out to affect hiring while not varying systematically with ethnicity, which makes them valid candidates to capture the variance in the unobservables. Table 4.4 presents the estimation results for different specifications of the homo- and heterogenous Probit model. I gradually add in individual and firm level characteristics (2 and 3) as well as recruiter identity (4 and 5); and finally municipality characteristics (6 and 7). In each model, the effect of ethnic background is captured by a dummy variable indicating Maghrebian origin. One can see that the sign and scope of the estimated effect is quite robust to the inclusion of different sets of control variables. It is especially noteworthy, that the marginal effects from the heteroscedastic specification remain very close to those from the Probit estimates, indicating that differences in the distribution of the unobservables do not seem to matter too much. The heteroscedastic estimates are always marginally higher, indicating that, if anything, a downward bias would be present. All models point to results very much in line with the call-back differences in means found above.

Table 4.4: Probit and Het. Probit, Full Sample

	(1) Probit	(2) Probit	(3) Het.Probit	(4) Probit	(5) Het. Probit	(6) Probit	(7) Het.Probit
maghre	-0.120***	-0.124***	-0.127***	-0.137***	-0.141***	-0.134***	-0.135***
	(-9.12)	(-9.18)	(-8.99)	(-7.80)	(-7.81)	(-9.15)	(-8.30)
$Log-lik.$	-1790.6	-1703.0	-1702.9	-1090.2	-1089.9	-1536.9	-1536.9
N	2916	2916	2916	1824	1824	2816	2816
Ind./Firm Char.	No	Yes	Yes	Yes+HR	Yes+HR	Yes	Yes
Mun. Char.	No	No	No	No	No	Yes	Yes

Marginal effects; t statistics in parentheses; (d) for discrete change of dummy variable from 0 to 1
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The table reports the marginal effects of having a Maghrebian vs. French origin on the prob. of eliciting a callback for the Homo- and Heteroscedastic Probit model. Standard errors are clustered on the vacancy level. Individual characteristics include, gender, the send order of the application, the age group, the signal block indicating marital status, having children and whether the respondent has been in inactivity. The identifying variables for the Het. Probit model are controls for the type of occupation. In models (4) and (5) additional controls on the gender and the origin of the recruiter are included. Municipality controls include the median household income, the unemployment rate, the population size, the mean net hourly wage by municipality, dummies for whether a municipality is located at a land border and department dummies.

4.2 Decomposition by diversity of the neighbourhood

A first look at the differences in mean call-back rates for the two sub-samples shows that the differences in call-back rates vary with the level of diversity both measured by the share of immigrants in the municipality of the firm location and by the share of births with an Arabic last name between 1966-1990. For both definitions the p-value of the comparison in means ttest in fact becomes insignificant for the high diversity samples.

Table 4.5: CB-rates by share of foreigners

	French origin	Maghrebian origin	p value
Share = low	(N=972)	(N=972)	
Mean (SD)	0.392 (0.488)	0.250 (0.433)	< 0.001
Share = high	(N=486)	(N=486)	
Mean (SD)	0.337 (0.473)	0.263 (0.441)	0.012

Table 4.6: CB-rates by share of Arabic last name

	French origin	Maghrebian origin	p value
Share = low	(N=528)	(N=528)	
Mean (SD)	0.424 (0.495)	0.277 (0.448)	< 0.001
Share = high	(N=252)	(N=252)	
Mean (SD)	0.321 (0.468)	0.254 (0.436)	0.095

One should further note that the difference in the scope of discrimination is not a result of Maghrebian candidates being called back less often in the samples with low population diversity, but French candidates being called back more often in the less diverse municipalities.

Running the above specified OLS model with interaction terms for the level of diversity and the dummy indicating a Maghrebian background, shows that the differences in the discrimination ratios are indeed significant. The difference in call-back rates of 14.2 percentage points appears to be reduced by around 6.8 percentage points when the candidates apply to a job in a municipality in which the share of the foreign population is above 13.8 percent. This result is significant at the five percent level and robust to the inclusion of municipality characteristics[43]. This is reassuring as it means that the differential treatment is not simply capturing any of the other established differences between the type of municipalities compared. The same holds when capturing diversity with the measure of the share of births with an Arabic last name. In the model without controls, the difference in call-back rates of 14.8 percentage points appears to be reduced by around 8.0 percentage points, when the firm is located in a more diverse municipality.

[43] Note that I also run the OLS with a continous measure of the share of the foreign population. Results are displayed in the Appendix, Table .7

This effect again is robust to the full set of control variables. Interestingly the interaction term of being female and of a Maghrebian origin also appears to have a significant positive effect in the subset of municipalities, for which the share of people with an Arabic last name is observable, although, on average, gender did not have a significant differential effect in the full sample. Let's recall that the municipalities not included are those with less than 50 births during that time-period and are also not including any vacancies in the municipality of Paris, so the sample is missing very small municipalities and those for whom the closest hospital to give birth in is presumably in a neighbouring municipality. It is also worth noting that the age group does not seem to significantly impact call-backs while the sending order, i.e. the rank of the application regarding it's timing (being the second, third or fourth application send) does have a significant negative effect of about 2 percentage points on the overall likelihood of being called-back for both French and Maghrebian candidates.

Table 4.7: OLS results, interaction of neighbourhood diversity (high/low) and origin of the applicant

	Share of foreigners		Share Arabic names	
	(1)	(2)	(3)	(4)
maghre	-0.142***	-0.152***	-0.148***	-0.180***
	(-8.44)	(-7.65)	(-6.91)	(-6.68)
Share of foreigners = high	-0.0545	0.0399		
	(-1.65)	(0.77)		
Share = high*maghre	0.0679**	0.0681**		
	(2.64)	(2.59)		
Female		0.00673		-0.0106
		(0.41)		(-0.47)
Female*maghre		0.0229		0.0676*
		(1.11)		(2.38)
Send order		-0.0157**		-0.0192*
		(-2.85)		(-2.45)
Age group		-0.0468		-0.0457
		(-1.75)		(-1.16)
Share of Arabic last names = high			-0.103*	-0.0946
			(-2.21)	(-1.35)
Share = high*maghre			0.0803*	0.0843*
			(2.13)	(2.15)
_cons	0.392***	0.577**	0.424***	0.653
	(19.46)	(2.77)	(15.43)	(1.84)
Ind./Firm Char.	No	Yes	No	Yes
Mun. Char.	No	Yes	No	Yes
N	2916	2904	1560	1556
adj. R^2	0.017	0.130	0.020	0.168

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

While these results are already revealing, my preferred model remains the non-linear heteroscedastic Probit. I therefore also run the heteroscedastic Probit both with and without municipality controls on the respective sub-samples. The results are shown in Table 4.8 below. Comparing the estimates for the heterogenous Probit models, one sees that a candidate with a Maghrebian name is indeed 18 percentage points less likely to be called back than a French-origin named candidate in an municipality with a low share of "births with an Arabic last name between 1966-1990". This is well above the discrimination estimated for the full sample. That effect becomes insignificant when the share is high (≥7.5%, which is the cutoff for the highest 30 percent in my sample). The finding for the diversity measure, capturing the share of all immigrants in a municipality remains around 5 percentage points below the discrimination estimate of the sample with a low share of immigrants. This difference is a bit smaller than the previously estimated differentials and one should note that the discrimination effect for the low-diversity sample now remains significant (compared to not being significant when just comparing the raw call-back means).

Table 4.8: Het. Probit by diversity of the neighbourhood

	(1) Low	(2) High	(3) Low	(4) High
measure = share of foreigners living in mun. (2017)				
maghre	-0.147***	-0.091***	-0.165***	-0.114***
	(-8.05)	(-4.20)	(-7.14)	(-4.04)
$Log-lik.$	-1133.3	-550.3	-971.5	-467.0
N	1944	956	1824	928
measure = share of births with Arabic lastname (1966-1990)				
maghre	-0.149***	-0.087*	-0.180***	-0.004
	(-6.30)	(-2.28)	(-5.75)	(-0.08)
$Log-lik.$	-616.7	-267.4	-484.2	-195.4
N	1056	496	968	476
Ind./Firm Char.	Yes	Yes	Yes	Yes
Mun. Char.	No	No	Yes	Yes

Marginal effects; t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Het. Probit regressions, with standard errors clustered on the vacancy level. Ind./Firm Char. include, gender, the send order of the application, the age group, marital status, having children and whether the respondent has been in inactivity as well as occupation dummies. Mun. Char. include the median household income, the unemployment rate, the population size, the mean net hourly wage by municipality, dummies for whether a municipality is located at a land border and department dummies.

Overall these results suggest a quite substantial relation between the level of diversity in a municipality and the discriminatory behaviour of firms. The next section assesses the robustness of these results to composition effects of the working population in municipalities.

4.2.1 Robustness Tests - A heterogeneity analysis by the composition of the work-force population in the neighbourhood

As defined in Table 3.3, the level of diversity (high/low) of the working population is broken down according to the origin of the employees in a municipality. Table 4.9 presents the results of heterogeneous Probit models like the one presented above, for each high and low sample respectively defined. The first information I use is the share of French employees that were born in France. This can be regarded as a more precise measure of the homogeneity of the (working) population compared to the share of non-immigrants (French), as those also include people that were foreign born. This split gives rise to a noticeable difference: While significant evidence of discrimination against applicants with

Maghrebian names is found in both samples, it is with 20.8 percentage points estimated to be substantially higher, when the share of French employees is high (above 89.9%), compared to 12.9 percentage points, when the share is low, i.e. when the employee composition in the municipality is more diverse.

Table 4.9: Het. Probit by diversity (low/high) of the work-force population in a given neighbourhood

	(1) Low	(2) High	(3) Low	(4) High
measure = share of French, born in France (2015)				
maghre	-0.114***	-0.157***	-0.129***	-0.208***
	(-6.95)	(-5.42)	(-6.91)	(-4.31)
$Log - lik.$	-1129.5	-552.3	-990.5	-433.3
N	1924	968	1864	840
measure = share of European foreigners (2015)				
maghre	-0.143***	-0.113***	-0.151***	-0.132***
	(-7.67)	(-4.80)	(-6.49)	(-4.59)
$Log - lik.$	-1079.8	-505.2	-935.7	-428.4
N	1852	900	1716	832
measure = share of Non-European foreigners (2015)				
maghre	-0.155***	-0.094***	-0.176***	-0.102**
	(-8.36)	(-3.60)	(-7.73)	(-3.28)
$Log - lik.$	-1041.4	-550.6	-881.1	-448.1
N	1848	912	1728	840
measure = share of French born in oversea depart. (2015)				
maghre	-0.128***	-0.130***	-0.130***	-0.139***
	(-7.21)	(-5.45)	(-6.02)	(-4.64)
$Log - lik.$	-1124.0	-559.2	-984.1	-481.0
N	1944	952	1820	900
measure = share of French, born abroad (2015)				
maghre	-0.133***	-0.114***	-0.146***	-0.152***
	(-7.51)	(-4.18)	(-6.72)	(-4.32)
$Log - lik.$	-1144.9	-446.9	-982.8	-389.8
N	1976	780	1876	760
Ind./Firm Char.	Yes	Yes	Yes	Yes
Mun. Char.	No	No	Yes	Yes

Marginal effects; t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Het. Probit regressions, with standard errors clustered on the vacancy level. Ind./Firm Char. include, gender, the send order of the application, the age group, marital status, having children and whether the respondent has been in inactivity as well as occupation dummies. Mun. Char. include the median household income, the unemployment rate, the population size, the mean net hourly wage by municipality, dummies for whether a municipality is located at a land border and department dummies.

A breaking down of the foreign employee population into non-European and European indicates that the differential treatment remains a robust finding for the former, but not necessarily for the share of European foreigners. The differential treatment by share of non-European foreigners is, in fact, the only measure of diversity for which I observe estimates that are of similar magnitude compared to the above findings. The split by people born in oversea departments or French who were born abroad does not give rise to any noticeable difference: significant evidence of discrimination against applicants with Maghrebian names is found in both areas with similar magnitude.

The difference between the discrimination measures between municipalities where the share of non-European employees is high vs low, is with 7.2 percentage points again similar to the difference observed when defining diversity by the share of births with an Arabic last-name. This gives support to the behaviour being driven by ethnic homophily, as both groups might be perceived as more different considering their ethnic origin, when they are less present in a given population.

As there appeared to be not much variation in the measure of European foreigners by municipality in my sample, this result has to be looked at with caution (high share $\geq$ 2.8% of European foreigners employed in municipality). Table .8 in the Appendix displays results for a more stringent cut-off on this dimension, i.e. the one for non-European foreigners (high share $\geq$ 7.3%). When the share of Europeans is high, following this definition, Maghrebian candidates are in fact 22.5 percentage points less likely to be called back compared to a 13.5 percentage point difference when the share of European foreigners is low in those municipalties. These results are thereby showing levels similar to those for when the share of French is high. This underlines the above stated hypothesis of homophilious behaviour driving the results, assuming that also other Europeans act in homophilious ways and seem to discriminate more against Maghrebian candidates when grouped together in high shares.

To take the analysis a step further I will now turn to an analysis on the firm level. If homophilious discrimination is indeed driving these results, it should also be observed when measured even more explicitly on the firm level. Homophilous discrimination should hence also occur when the diversity level within a firm is considered.

4.3 Decomposition by within-firm diversity

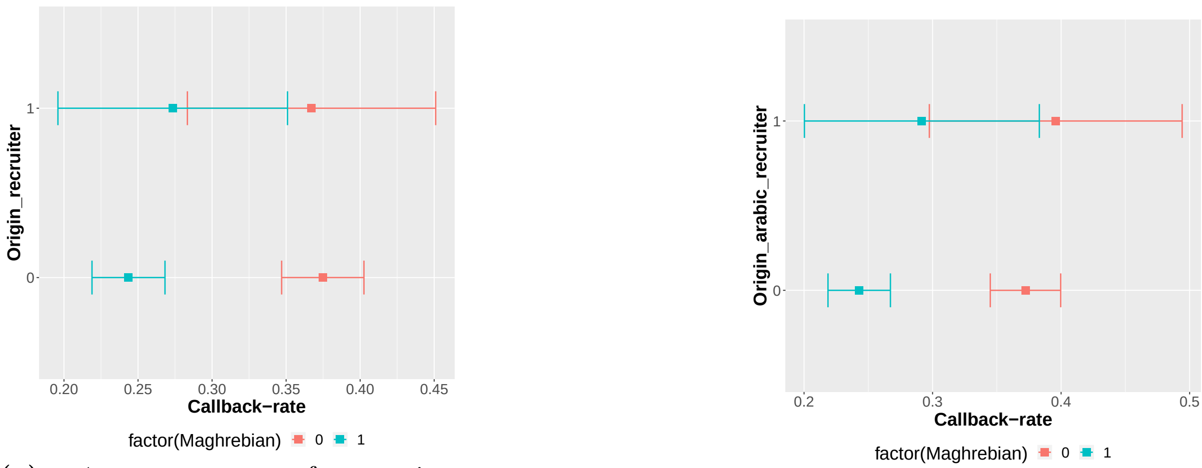
In a first attempt to capture the level of diversity within a firm and to directly observe homophilious behaviour, I *"rely on proxies of the employers' characteristics by using the observed identity of the person to whom we send the applications"*, following Edo et al. (2019). As noted by Edo et al. (2019) such covariates are *"by design, endogenous, since they result from hiring decisions by firms in the past. The aim of the decomposition is thus to substantiate the interpretation of the data based on correlations between recruiters' identity and applicants characteristics"*.

Table 4.10 and Figure 4.4 display differences in call-back means for firms where the recruiter had a European classified name compared to firms where the recruiter had a Non-European classified name. The results of discriminatory behaviour against Maghrebian candidates are found to be only significant for the set of recruiters with European names with a discrimination ratio of 1.5, similar to the one found for the full sample. Maghrebian candidates are also less likely to be called back, when the recruiter has a non-European name, the difference however is smaller than in the case of European recruiters (ratio around 1.3) and it is not significant. This is true for both, non-European names defined as names being used in Asian, African and Arabic language and when looking at the name being Arabic in particular.

Table 4.10: Call-back rates by the of the recruiter

	French origin	Maghrebian origin	p value
European name	(N=1166)	(N=1166)	
Mean (SD)	0.375 (0.484)	0.244 (0.429)	< 0.001
0.001			
Non-European name	(N=128)	(N=128)	
Mean (SD)	0.367 (0.484)	0.273 (0.447)	0.109

Figure 4.4: Call-back rates by origin of the recruiter



(a) 1 = name of recruiter Arabic/African/Asian

(b) 1 = name of recruiter Arabic

Table 4.11: Het. Probit by origin of the recruiter

	(1)	(2)	(3)	(4)
	Recruiter identity			
	European	Non-Eur.	European	Non-Eur.
maghre	-0.137***	-0.112	-0.148***	-0.112
	(-8.16)	(-1.46)	(-7.61)	(-0.47)
$Log - lik.$	-1335.2	-139.9	-1189.0	-75.5
N	2304	248	2216	180
Ind./Firm Char.	Yes	Yes	Yes	Yes
Mun. Char.	No	No	Yes	Yes

Marginal effects; t statistics in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

These results are robust to the inclusion of control variables when assessed in the heterogeneous Probit model. One should however note that the sub-sample for recruiters holding a non-European name is comparably small and therefore just like the one for European foreign employees not the most precise.

In the Appendix Table .10 and .1 I further include results that split the sample not by the origin but by the gender of the recruiter. In contrast to Edo et al. (2019), the call-back rate differences do not suggest that gender homophily is at play (no significant difference in call-backs for men vs women, given the recruiter's gender). Interestingly, the results suggest that discrimination against Maghrebian candidates is about 6 percentage points higher when the recruiter is a man (20.5) compared to when the recruiter is a women (14.5). These findings point to potential other paths to be explored when aiming to understand discriminatory behaviour and employment outcomes, which however go beyond the scope of this work.

The unique feature of merging the experimental data with firm-level data at least for a sub-sample of my observations, allows me to proxy the diversity level within a firm not only by the identity of the recruiter but also by observing the number of French vs foreign employees. The results on this dimension are presented in the next section.

4.3.1 A heterogeneity analysis by the composition of the workforce population within the firm

As explained in section 3.3.2, I merged the data with firm-level information for a sub-sample of 204 firms included in the experimental data. Results on this whole sample (displayed in Appendix, Table .11) suggest that overall, discrimination against Maghrebian candidates compared to French candidates is slightly higher (14.5) than in the full sample (13.5).

As defined in Table 3.5 the level (high/low) of diversity within the firm is broken down according to the origin of the employees in that firm. Table 4.12 below displays the results of heterogeneous Probit models for each of these measures. The results are very much in line with what has been found in the previous sections. Measuring the composition by the share of French employees that were born in France, results in the finding of substantial differences in the discrimination estimates between firms that employ more than 97.1% of this group compared to those employing a lower share. Significant evidence of discrimination against applicants with Maghrebian names is found in both samples. The estimates for the firms with the most homogeneous composition of employees is with 19.4 percentage points very close to the discrimination measured for firms who are located in municipalities with relatively high shares of such employees. ´While Maghrebian applicants are still 11.8 percentage points less likely to be called back when the share is low, i.e. when the employee composition in the firm is less homogeneous, the difference is with 8 percentage points the highest found among all specifications.

As for the diversity measure for the municipalities, the effect of the share of European employees also suffers from low variation in the within-firm measure (finding a cut-off was only possible by splitting the sample in two) and appears to be quite noisy. The results for the two groups do not display any noticeable difference and the evidence of discrimination against applicants with Maghrebian names is found in both samples with similar magnitude.

Discrimination becomes insignificant when looking only at the highest tertile of observations, measured by the share of non-European employees. This exercise further shows that the size of the firm seem to matter too. The estimate for discrimination becomes lower and insignificant when only looking at large companies (> 40 employees). This might have several explications. Larger firms are potentially more often found to be located in larger municipalities which in turn have a more diverse composition of the population and employees. It might also be that larger firms have more structured hiring processes that allow for less individual discriminatory behavior of recruiters as usually a whole human resources department would be responsible to coordinate the hiring process. Lastly this also fits the homophily argument if one assumes that in smaller firms group identity is higher than in larger firms.

Table 4.12: Het. Probit by diversity of the employees within the firm

	(1) Low	(2) High	(3) Low	(4) High
measure = share of French, born in France (2015)				
maghre	-0.132***	-0.211***	-0.118**	-0.194**
	(-3.67)	(-3.40)	(-3.26)	(-2.66)
$Log-lik.$	-282.4	-137.3	-270.1	-127.0
N	520	264	516	264
measure = share of European foreigners (2015)				
maghre	-0.160***	-0.146**	-0.171***	-0.162*
	(-4.97)	(-2.68)	(-5.31)	(-2.35)
$Log-lik.$	-284.7	-139.9	-268.7	-128.3
N	532	248	528	248
measure = share of Non-European foreigners (2015)				
maghre	-0.165***	-0.162*	-0.165***	-0.145
	(-5.12)	(-2.33)	(-4.93)	(-1.87)
$Log-lik.$	-300.8	-118.1	-291.4	-108.7
N	536	244	536	240
measure = size of the firm (2015)				
maghre	-0.142***	0.147	-0.142***	-0.089
	(-4.61)	(0.14)	(-4.47)	(-0.97)
$Log-lik.$	-294.9	-131.2	-281.6	-125.5
N	544	252	540	252
Ind./Firm Char.	Yes	Yes	Yes	Yes
Mun. Char.	No	No	Yes	Yes

Marginal effects; t statistics in parentheses

(d) for discrete change of dummy variable from 0 to 1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Het. Probit regressions, with standard errors clustered on the vacancy level. Ind./Firm Char. include, gender, the send order of the application, the age group, marital status, having children and whether the respondent has been in inactivity as well as occupation dummies and the size of the firm (for all but the last line of estimates). Mun. Char. include the median household income, the unemployment rate, the population size, the mean net hourly wage by municipality, dummies for whether a municipality is located at a land border and a dummy indicating whether the firm is located in the city of Paris.

Both the results for the municipality composition of firms, as well as for the more direct measure of homophily on the within-firm level indicate substantial differentials in discrimination depending on the composition of the locality and the within-firm environment. As hinted in the first two sections, the neighbourhood effect might not only be driven by homophily per-se but also by the fact that employers anticipate the homophilious and discriminatory behaviour of their customers. The next section is therefore devoted to analyse and disentangle such customer discrimination from the pure employer discrimination.

5 Mechanism analysis - Does the "visibility" of an occupation matter?

5.1 The theoretical implications of customer discrimination

As established above, the estimation strategy applied allows to measure the aggregate effect of statistical and taste-based discrimination of employers, denoted:

$$\underbrace{\mu}_{\text{aggregate discrimination}} = \underbrace{[E(Z|M=0) - E(Z|M=1)]}_{\text{statistical discrimination}} + \underbrace{\gamma}_{\text{taste-based discrimination}} . \tag{5.1}$$

This section is aiming at shedding some more light on the forces driving these two and more specifically the role of customer discrimination. For my theoretical predictions I follow the two-sector static matching model established in Combes et al. (2016) on which the following simplified and contextualized considerations are based. Note that the major difference is that I do not consider the preferences of the applicant pool, as applicants do not choose to send their CV's to a specific location in my set-up. The strength of my analysis is therefore that I truly capture the effect driven by the "demand" (employer discrimination and firm's interest) side. To do so, let's consider two groups of occupations: (1) with no (or little) consumer contact (e.g. Electrical and Electronic equipment assembler); and (2) with a high proportion of consumer contact (i.e. Stock clerk, sales floor). Job seekers are homogeneous except for their observable ethnic group and the total population is normalized to 1, with n being the share of French natives and $1-n$ the share of foreigners. Now, a_e is the proportion of available jobs whose corresponding employer has a taste for discrimination and discriminates as a result (we can think of it as the share explained by parameter γ for each municipality). Let me further introduce a_c, denoting the local share of consumers' racial prejudice. The predictions of the model then rely on two crucial assumptions: firstly, *"the population of French natives provides the pool of potential prejudiced consumers and employers"*[44]; and secondly *"that employer discrimination is not more prevalent in the contact job sector than in the rest of the economy"*. The authors crucially expect in their model specification that *a_e and a_c will depend on the proportion of French natives* in the location of the firm, so that for any n,

[44] Note that I, alike the Combes et al. (2016), *"rely on this assumption since they [I] cannot accurately measure the level of prejudice across local areas*, Laouénan (2017). A more ideal measure of prejudice would in fact be data on the share of people holding racist believes, similar to the one used in Laouénan (2017).

employer discrimination arises when $a_e(n) > 0$, and customer discrimination arises for $a_c(n) > 0$[45].

This can be mirrored to my context and I am thus able to test their main two predictions of the model in an alternative set-up.

Prediction 1: *If the ethnic differential unemployment probability is positively affected by the proportion of French-native residents (or negatively affected by the share of foreigners), then there is ethnic (either customer or employer) discrimination.*

In fact, this is what has been shown in the previous section, by establishing that discrimination is substantially higher in municipalities with a high proportion of French natives, compared to discrimination in municipalities with a low proportion of French natives.

Prediction 2: *There is customer discrimination if and only if there is ethnic discrimination and the ethnic differential probability of working in a contact job is negatively impacted by the proportion of French natives.*

In their context as well as in mine, *"the consideration of the unemployment rate (discrimination) differential, does not allow customer discrimination to be distinguished from employer discrimination"*. As they look at the labour market ex-post (i.e. at actual employment statistics in France), they rely on the condition that both minority and majority candidates are employed to disentangle customer discrimination. They thus look at the conditional probability of working in a contact job, q^A, as their main parameter of interest. Their model then predicts that *"employer discrimination is at work in both sectors and does not affect q^A, the conditional probability of working in a contact job for the minority group. Conversely, customer discrimination only occurs in such jobs and therefore affects q^A.[...] The proportion of French natives, therefore, negatively affects the differential conditional probability if and only if there is customer discrimination at the margin"*. The difference in my study is that I do not rely on the conditional of being employed to observe whether the composition of the population results in differentials in discrimination by the level of contact. With the data at hand, I can directly observe the difference in the likelihood of being employed in a high contact job vs a low contact job, by comparing the differences in hiring discrimination in high- and low-contact jobs. To do so I will split my dataset in four groups in a factorial way (1: Low diversity/Low contact, 2: Low diversity/High contact, 3: High diversity/Low contact, 4: High diversity/High contact). Following the above predictions I assume that discrimination should be found to be highest in the case of subgroup 2: Low diversity/High contact. The difference should be a result of the addition of a_c to the aggregate discrimination measure.

One should note that, again, the lines between taste-based and statistical discrimination remain blurry in this case. The authors even denote, that their basic model abstracts from statistical discrimination, *"as such discrimination is very likely and may vary across sectors"*, Combes et al. (2016). Edo et al. (2019) indeed show that explicitly mentioning language skills in the CV, reduces the degree of discrimination for foreign (including North-African) candidates. This shows that statistical discrimination, based on employers fearing a lack of communication skills, certainly comes into play. As those skills are more important in the high contact jobs, compared to the low contact jobs, I expect call-back

[45]This includes the assumption that, $a_e(0) = a_c(0) = 0$.

rates to also differ solely on this dimension. To test this, one would expect a difference in call-back rates already without considering the diversity level, by comparing discrimination for low and high contact jobs. Following the language skill argument, discrimination should be found to be higher in high-contact jobs, as this statistical discrimination factor weights more in such occupations than in the rest of the economy. In contrast to Combes et al. (2016), I would further also argue that statistical discrimination varies with n, since the higher the share of foreigners, the less likely it is that an employer is concerned with the lack of language abilities. This could, for instance, be due to the fact that the employer is more aware and experienced that language is actually not a concern or that language barriers are not so much a concern, when the customers themselves represent a linguistically diverse group.

Any observed differential in the likelihood of being employed in a contact job might thus be explained by either of the two discrimination patterns discussed in this section. In the following, I will first decompose discrimination according only to the visibility of the job (5.2) and then present the estimates based on the interaction of visibility and the level of diversity (5.3).

5.2 Decomposition according to visibility of the job

Table 5.1 shows the decomposition of the discrimination effect by the three different classifications of the contact jobs, defined in section 3.3.3. The results give very robust support for the above stated hypothesis. When the share of contact jobs by occupation (PCS, 3 digits) is classified as high, the likelihood of being called back is decreased by 17.5 percentage points for Maghrebian candidates compared to French candidates. This decrease is about 6.5 percentage points less, when the occupations have a low share of contact jobs.

This finding turns out to be very similar when I use the alternative O*Net definition of the importance of being in contact with customers, which allows to be measured on the Code rome level for each observation. The differential effect is even stronger if I use the measure of contact jobs on the socio-professional category level, which only defines low-skilled contact job measures as highly visible. This measure is therefore interesting but likely to capture both the proxy of the visibility but also the fact that the difference in call-back means was shown to be higher for the low qualification occupations (see Table 4.2). These findings support the hypothesis that statistical discrimination, is at play when one assumes that language skills are more relevant in highly visible occupations which require contact with costumers compared to this visibility being low. In a next step I will try to disentangle to what extend customer discrimination plays a role by interacting the level of visibility with different levels of my diversity measure.

5.3 Interaction of visibility and the diversity of the neighbourhood

In order to observe the difference in the likelihood of being employed in a high contact job compared to a low contact job, depending on the customers representing a potentially high concentration of the prejudiced group, I split the sample in four groups in a factorial way (1: Low diversity/Low contact, 2: Low diversity/High contact, 3: High diversity/Low contact, 4: High diversity/High contact). The results for the four coefficients are displayed in one figure for each different combination of the visibility and diversity measure used. I conduct this analysis using the four measures of the neighbourhood composition for which I found the most substantial differential results in discrimination. Figure 5.1 (a) presents the discrimination coefficients found when defining the customer population by their share of immigrants, while (b) takes my alternative definition of municipalities where the customer population is defined of municipalities with a low versus a high share of births with an Arabic last name.

Table 5.1: Het. Probit. by visibility (low/high) of the occupation

	(1) Low	(2) High	(3) Low	(4) High
Contact measure = PCS, 3 digits (Combes et al. (2016))				
maghre	-0.110***	-0.129**	-0.110***	-0.175***
	(-5.55)	(-2.65)	(-4.80)	(-6.06)
$Log-lik.$	-864.6	-829.6	-757.8	-672.1
N	1536	1380	1456	1276
Contact measure = PCS, 2 digits (Laouénan (2013))				
maghre	-0.098***	-0.172***	-0.108***	-0.244***
	(-5.96)	(-6.36)	(-5.57)	(-4.98)
$Log-lik.$	-1158.6	-527.0	-1033.2	-413.3
N	1960	956	1872	868
Contact measure = O*net (Laouénan (2017))				
maghre	-0.110***	-0.124*	-0.116***	-0.174***
	(-5.93)	(-2.29)	(-4.95)	(-5.72)
$Log-lik.$	-901.7	-791.0	-783.5	-652.1
N	1608	1308	1500	1220
Ind./Firm Char.	Yes	Yes	Yes	Yes
Mun. Char.	No	No	Yes	Yes

Marginal effects; t statistics in parentheses

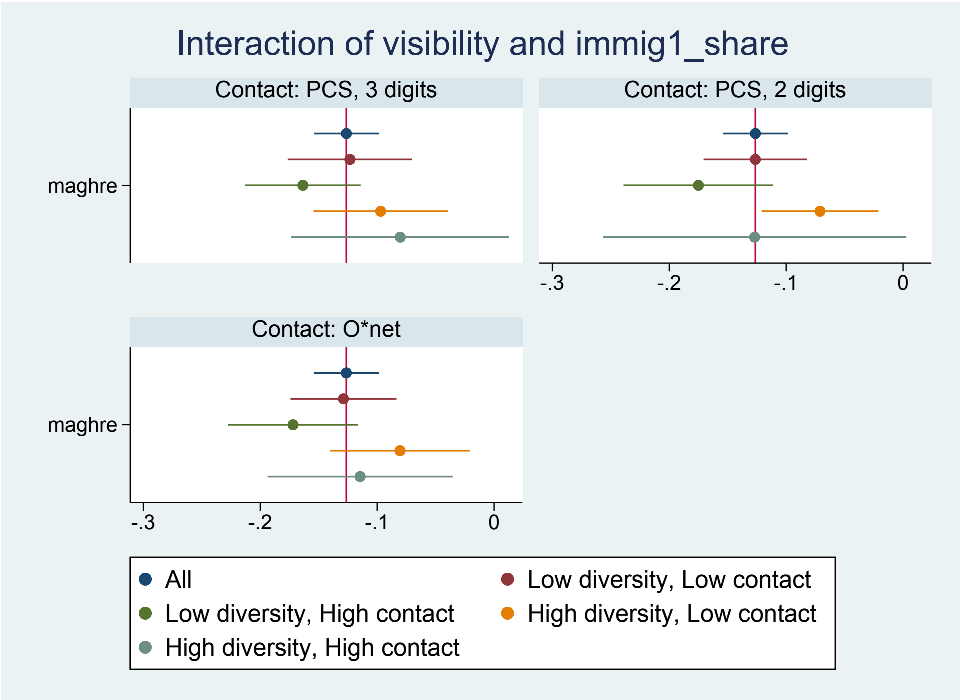
(d) for discrete change of dummy variable from 0 to 1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

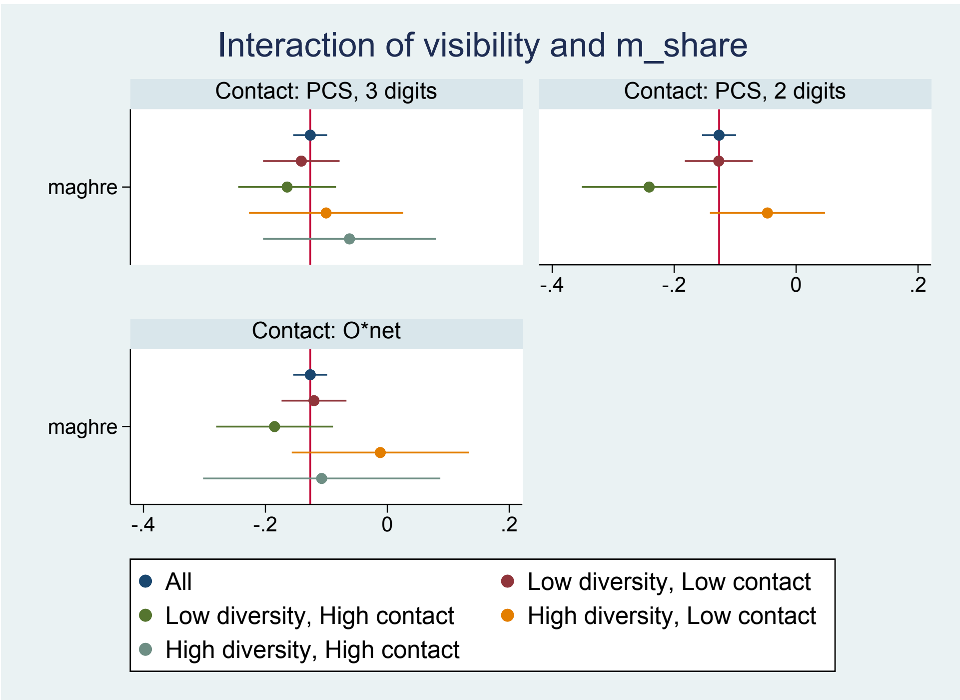
Het. Probit regressions, with standard errors clustered on the vacancy level. Ind./Firm Char. include, gender, the send order of the application, the age group, marital status, having children and whether the respondent has been in inactivity as well as occupation dummies and the size of the firm (for all but the last line of estimates). Mun. Char. include the median household income, the unemployment rate, the population size, the mean net hourly wage by municipality, dummies for whether a municipality is located at a land border and a dummy indicating whether the firm is located in the city of Paris.

Consistently for both diversity measures and in interaction with any of the three contact job measures, the marginal effect of discrimination is found to be the lowest, when the occupation has a high contact measure and the level of diversity in the neighbourhood is low (see Tables .12 ff. in the Appendix for all estimation results). The results thereby indicate that the chances of getting a job for an applicant holding a Maghrebian name compared to an applicant holding a French name are the lowest, when the job is requiring a high level of contact with costumers and the diversity in the neighbourhood is low. This differential should be devoted to the additionally anticipated customer discrimination, when the initial level of discrimination is, as stated above, assumed to be the same for both high and low visibility occupations.

Figure 5.1: Coef-plots of the interaction of visibility and the diversity of the neighbourhood for all three visibility measures and two main proxies of neighbourhood diversity

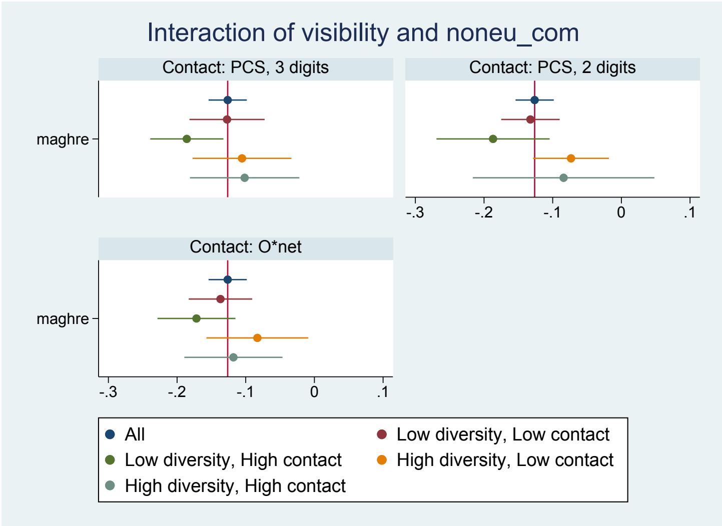


(a) immig1_share = Share of foreign population, Data: French Census 2017

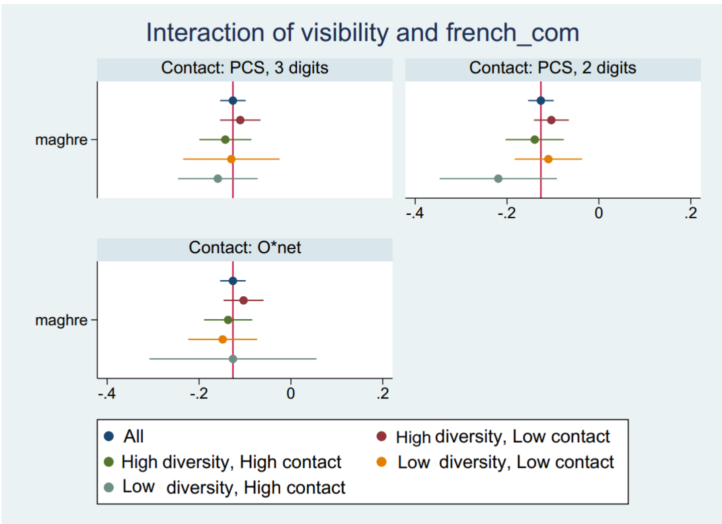


(b) m_share = Share of births with Arabic last name (1966-1990), Data:Sirugue (2020)

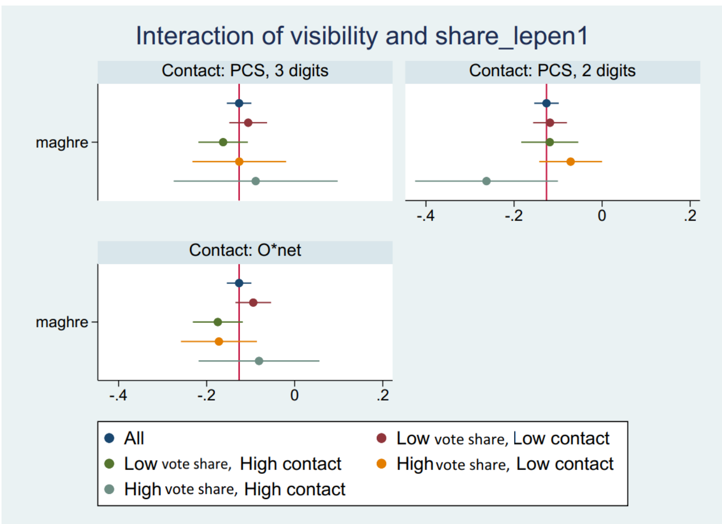
Figure 5.2: Coef-plots of the interaction of visibility and the diversity of the neighbourhood for all three visibility measures and other proxies of diversity (employee population) and prejudiced population (Vote share for far right party in 2017)



(a) noneu_com= Share of non-European employees, Data: DADS, 2015



(b) french_com = Share of French employees, born in France, Data: DADS, 2015



(c) share_lepen1 = Share of votes for Marine Le Pen, first round election 2017, Data: Voting registers

Figure 5.2 displays the results for alternative measures of the diversity levels of the customers. This trend is shown to be robust when capturing the composition of the pool of customers by the share of the non-European working-population of a given neighbourhood.

When considering the share of French employees, born in France as a measure for the composition of a municipality, one has to read the figures differently. Now, the light blue dot indicates the low diverse (i.e. high share of French, born in France) scenario paired with the visibility of the job being high. The same holds for the alternative measure of the vote share for the far right party leader Marine le Pen in the first tour of the French national elections of 2017. In these two cases, the results seem not as clear as in the above cases. In these two cases, only the visibility measure on the socio-professional category level is shown to be interacting in the predicted way with the measure of the composition of the customer population. One should recall that this measure however only represents occupations classified in the low qualification group. This effect is therefore likely to capture the higher discrimination in low-skill occupations paired with the high share of French and the high vote shares for the Front National. The unclear results for the composition measured by voting patterns is again in line with findings from Combes et al. (2016) who also do not find customer discrimination to occur when the share of votes for the same party in the first round of the 1995 elections was considered. They argue, that *one possible explanation is that the FN local vote share is not a relevant measure of prejudice. [...] if FN politicians are clearly prejudiced, it does not mean that all their voters are. For instance, during the 2002 French presidential election, the far-right leader Jean Marie Le Pen arrived second behind conservative candidate Jacques Chirac but ahead of Lionel Jospin from the Socialist Party who almost everyone assumed to be Chirac's strongest challenger. This surprising result was explained by Perrineau (2003), and many political analysts as a 'protest vote' and not a 'racist vote'*, Combes et al. (2016).

The overall

6 Discussion and Conclusions

Over all my findings are very much in line with the predictions stated under section 2, following the homophily theory. The results can be summarized in three main findings:

First, the results provide evidence for a substantial differential in discrimination that lays between 5 and 7 percentage points for high compared to low diverse municipalities in France. These findings are especially relevant in the light of their influence in shaping segregation on the municipality level. As stated above this differential can be considered as the part of segregation that is shaped by the demand side, i.e. by hiring discrimination. The existence of such a differential therefore indicates that, assuming the preferences and qualifications of candidates for working in a certain location being equal, their chances of getting a job are lower when this location is considered to have a low diversity level. This not only confirms that it is a potential driver of persistent segregation by McAvay and Safi (2018) but points to discrimination and segregation systematically reinforcing each other. While this is the case for measuring diversity with the share of immigrants this effect is even stronger, when capturing the ethnic composition of the neighbourhood in a more precise manner, i.e. by the share of births with an Arabic last name. Measuring the ethnic composition more precisely, I can show that there is no significant discrimination in the municipalities that are considered to be highly diverse.

These findings therefore very much show that such homophilious driven discrimination reinforces segregation. As both factors are found to be interdependent of each other one could rephrase this as them creating a "vicious circle of segregation and discrimination".

Second, I can further show that this is also true for the within-firm level. The results even mount to a differential of 8 percentage points when the share of French employees in a firm is above 97.1%. This indicates that ethnic homophilious behaviour leads to a reinforcement of segregation not only on the municipality but also on the firm level. My data further confirms the evidence shown by Edo et al. (2019) for the Paris region, for the whole of France. The most direct measure of homophilious behaviour indicates that recruiters with European names are more likely to call back French named applicants while there is no significant effect on discrimination when the recruiter has a non-European name.

Third, I provide first evidence of direct customer discrimination by showing that employers systematically discriminate the most, when the occupation is classified as one with a high amount of customer contact and the diversity of the neighbourhood is low.

Data limitations and potential extensions

While I do include a lot of different measures I have to emphasize yet again that one of the biggest caveats is the lack of data measuring the ethnic composition of the population. While the measure provided by Sirugue (2020) content wise displays exactly what one would want to observe, there are some caveats that still come with it. Although the birth cohort matches the applicants ages, the measure only captures diversity in the locality of where these people were born in, which is not necessarily where they grew up in or where they live today. The ideal data-set would capture the occurrence of Arabic names in the current composition of the society. An updated version of the measure of Arabic names by municipality will be provided by Sirugue (2020) using centralized electoral register (with last name and address of every French voter). Despite such a measure only being availabe by 2021 or 2022, defining the ethnic composition with such an updated measure would certainly improve the study. This leaves room for an updated analysis using an even more promising and relevant measure in the future. Alternatively one would want more precise variables describing the share of specific national origins in neighborhoods in order to at least capture the share of foreigners with more distinction.

While the selection of occupations included in the study turned out to be quite fitting to my analysis, it would be interesting to conduct a study, where the selection of occupations would have been selected on the basis of the contact job measure. One might further want a more updated data source on such a visibility measure, as job tasks and occupations are known to be transforming and changing fast with increasing trends of automation and digitalisation of jobs (Berger and Frey (2016)).

Another caveat is the sometimes quite small sample size especially of the sub-sample analyses in section 5. An obvious extension to improve the precision and power of my results would therefore be an analysis with the final experimental data set. As the data at hand only provides a third of the data that will ultimately be collected, using the full data source would certainly improve the analysis on that front. The same holds of course for the firm level data. While a first step would be to complete the merge of the whole experimental data with the firm level data, this analysis could not only improve in quantity but also in the quality of what is measured. As the DADS data is quite detailed one might be able to distinct the people in charge of recruitment not only by the name on the application webpage but by identifying them and their nationality in the firm-level data. In general, including more firm-level data would certainly allow for a more thorough analysis and to include more relevant control variables such as the economic and financial situation of firms. As for example the firm size has been shown to matter, one might even go a step further and test whether homophilious behaviour is larger in, for example, small or large groups of employees by interacting the size of the firm and the share of foreign employees[46].

Policy perspectives

The finding of this study, that I find discrimination and segregation to reinforce each other, has policy implications. If one believes that either segregation of ethnic groups or the discrimination based on the suggestive ethnic origin is normatively bad, it is valuable information that by breaking this vicious circle on any of the above shown levels, one might decrease the magnitude of one or the other. As the detected type of discrimination is prohibited by law in France, the findings further give valuable information on where the French government should target policies when aiming to fight ethnic discrimination. Ironically, *"French policies promoting immigrant assimilation have [..] been traditionally conducted by urban authorities mainly through colorblind actions in favor of employment and education in "priority neighborhoods" defined by socioeconomic characteristics"*, McAvay and Safi (2018). The urban authorities might therefore reconsider the targeting of neighbourhoods and put a focus on employment programs promoting diversity in the work-force or diversity trainings for managers and human resource departments in municipalities and firms that have an initially low share of diversity. While fighting discrimination is hard, policies targeted at increasing the diversity of a neighbourhood through, for instance, targeted housing programs in municipalities with a low share of immigrants might be one promising avenue[47]. A policy maker or firm manager wanting to improve the level of diversity within a firm might further want to consider what Stoll et al. (2004) proposed for the U.S: The authors suggest that moving more blacks (and/or in my case people with a Maghrebian background) into positions with hiring authority within firms might help to alleviate the persistent unemployment difficulties of African Americans. Such a policy approach would however demand from the government to acknowledge that France should stop it's "color-blind" policy approach in order to rightly target policies such as the promotion of ethnically diverse labour environments. A last valuable insight is the fact that I can show that discrimination is especially a concern in highly visible occupations. This could serve as a guide when prioritizing areas to tackle the fight in discrimination by sector or occupation.

[46]Another interesting sub-sample would be to check discrimination by date of creation and age of recruiters which was unfortunately beyond the scope of this thesis as time working on DADS was too constrained by the Covid-19 measures.

[47]Some might argue that such policies might cause a so-called "white (native) flight" effect, which implies a sudden or gradual large-scale migration of white people from areas becoming more racially or ethno-culturally. However this effect was found to be historically low in France and findings even tend to discredit the hypothesis of a "white flight" pattern in residential mobility dynamics in France (Rathelot and Safi (2014)).

References

Adida, C. L., Laitin, D. D., and Valfort, M.-A. (2014). Muslims in France: identifying a discriminatory equilibrium. *Journal of Population Economics*, 27(4):1039–1086.

Aeberhardt, R., Coudin, E., and Rathelot, R. (2017). The heterogeneity of ethnic employment gaps. *Journal of Population Economics*, 30(1):307–337.

Ahmed, A. M. and Hammarstedt, M. (2008). Discrimination in the rental housing market: A field experiment on the internet. *Journal of Urban Economics*, 64(2):362–372.

Ai, C. and Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics letters*, 80(1):123–129.

Alba, R. D. (2009). *Remaking the American mainstream: Assimilation and contemporary immigration*. Harvard University Press.

Baldini, M. and Federici, M. (2011). Ethnic discrimination in the Italian rental housing market. *Journal of Housing Economics*, 20(1):1–14.

Becker, G. S. (2010). *The economics of discrimination*. University of Chicago press.

Berger, T. and Frey, C. B. (2016). Structural transformation in the OECD: Digitalisation, deindustrialisation and the future of work.

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

Bertrand, M. and Mullainathan, S. (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.

Bonnet, F., Lalé, E., Safi, M., and Wasmer, E. (2016). Better residential than ethnic discrimination! Reconciling audit and interview findings in the Parisian housing market. *Urban Studies*, 53(13):2815–2833.

Bosch, M., Carnero, M. A., and Farre, L. (2010). Information and discrimination in the rental housing market: Evidence from a field experiment. *Regional science and urban Economics*, 40(1):11–19.

Brutel, C. (2016). La localisation géographique des immigrés: Une forte concentration dans l'aire urbaine de Paris. *Insee Premiére*, (No. 1591).

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

Currarini, S., Jackson, M. O., and Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045.

Currarini, S., Jackson, M. O., and Pin, P. (2010). Identifying the roles of race-based choice and chance in high school friendship network formation. *Proceedings of the National Academy of Sciences*, 107(11):4857–4861.

Decreuse, B. and Schmutz, B. (2012). Residential mobility and unemployment of African immigrants in France: A calibration approach. *Annals of Economics and Statistics/Annales d'Éeconomie et de Statistique*, pages 51–91.

Edo, A., Jacquemet, N., and Yannelis, C. (2019). Language skills and homophilous hiring discrimination: Evidence from gender and racially differentiated applications. *Review of Economics of the Household*, 17(1):349–376.

Ewens, M., Tomlin, B., and Wang, L. C. (2014). Statistical discrimination or prejudice? A large sample field experiment. *Review of Economics and Statistics*, 96(1):119–134.

Giuliano, L., Levine, D. I., and Leonard, J. (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.

Gobillon, L., Rupert, P., and Wasmer, E. (2014). Ethnic unemployment rates and frictional markets. *Journal of Urban Economics*, 79:108–120.

Goldin, C. and Rouse, C. (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American economic review*, 90(4):715–741.

Heckman, J. (1998). Detecting discrimination. *Journal of economic perspectives*, 12(2):101–116.

Heckman, J. and Siegelman, P. (1993). The Urban Institute audit studies: Their methods and findings. *Clear and Convincing Evidence: Measurement of Discrimination in America, Washington*, 187:258.

Holzer, H. J. (1996). *What employers want: Job prospects for less-educated workers*. Russell Sage Foundation.

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

Jacquemet, N. and Yannelis, C. (2012). Indiscriminate discrimination: A correspondence test for ethnic homophily in the Chicago labor market. *Labour Economics*, 19(6):824–832.

Laouénan, M. (2013). Un cas particulier de discrimination sur le marché du travail: l'accès aux emplois en contact avec le public. *Economie et statistique*, 464(1):35–58.

Laouénan, M. (2017). 'Hate at first sight': Evidence of consumer discrimination against African-Americans in the US. *Labour Economics*, 46:94–109.

Laouénan, M. and Rathelot, R. (2020). Can information reduce ethnic discrimination? evidence from airbnb. *Revise and Resubmit, American Economic Journal, Applied Economics.*

McAvay, H. and Safi, M. (2018). Is there really such thing as immigrant spatial assimilation in france? Desegregation trends and inequality along ethnoracial lines. *Social science research*, 73:45–62.

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

Mouw, T. (2002). Are black workers missing the connection? The effect of spatial distance and employee referrals on interfirm racial segregation. *Demography*, 39(3):507–528.

Neumark, D. (2012). Detecting discrimination in audit and correspondence studies. *Journal of Human Resources*, 47(4):1128–1157.

Pager, D. and Shepherd, H. (2008). The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annu. Rev. Sociol*, 34:181–209.

Petersen, T., Saporta, I., and Seidel, M.-D. L. (2000). Offering a job: Meritocracy and social networks. *American Journal of Sociology*, 106(3):763–816.

Quillian, L., Heath, A., Pager, D., Midtbøen, A. H., Fleischmann, F., and Hexel, O. (2019). Do some countries discriminate more than others? Evidence from 97 field experiments of racial discrimination in hiring. *Sociological Science*, 6:467–496.

Raphael, S., Stoll, M. A., and Holzer, H. J. (2000). Are suburban firms more likely to discriminate against African-Americans? *Journal of Urban Economics*, 48(3):485–508.

Rathelot, R. and Safi, M. (2014). Local ethnic composition and natives' and immigrants' geographic mobility in France, 1982–1999. *American Sociological Review*, 79(1):43–64.

Royster, D. A. (2003). *Race and the invisible hand: How white networks exclude black men from blue-collar jobs.* Univ of California Press.

Sabbagh, D. and Peer, S. (2008). French color blindness in perspective: The controversy over "Statistiques Ethniques". *French Politics, Culture & Society*, 26(1):1–6.

Simon, P. (2008). The choice of ignorance: The debate on ethnic and racial statistics in France. *French Politics, Culture & Society*, 26(1):7–31.

Simon, P. and Safi, M. (2013). Les discriminations ethniques et raciales dans l'enquête Trajectoires et Origines: représentations, expériences subjectives et situations vécues.

Sirugue, L. (2020). Inequality of opportunity in France: The role of spatial segregation in the ethnic gap. *Master Thesis*, Paris School of Economics.

Stoll, M. A., Raphael, S., and Holzer, H. J. (2004). Black job applicants and the hiring officer's race. *ILR Review*, 57(2):267–287.

Toma, S. (2016). The role of migrant networks in the labour market outcomes of Senegalese men: How destination contexts matter. *Ethnic and Racial Studies*, 39(4):593–613.

Waldinger, R., Lichter, M. I., et al. (2003). *How the other half works: Immigration and the social organization of labor.* Univ of California Press.

Williams, R. (2009). Using heterogeneous choice models to compare logit and probit coefficients across groups. *Sociological Methods & Research*, 37(4):531–559.

Appendix

Table .1: Descriptive statistics of the full sample and sub-sample used for the within-firm diversity analysis

	0 (N=1712)	1 (N=1204)	Total (N=2916)	p value
diversity				0.832
Mean (SD)	0.332 (0.471)	0.336 (0.472)	0.333 (0.471)	
Range	0.000 - 1.000	0.000 - 1.000	0.000 - 1.000	
age_group				0.003
Mean (SD)	1.621 (0.628)	1.691 (0.616)	1.650 (0.624)	
Range	1.000 - 3.000	1.000 - 3.000	1.000 - 3.000	
metier_noms				< 0.001
Administrative employee	132 (7.7%)	108 (9.0%)	240 (8.2%)	
Commercial sales representative	196 (11.4%)	108 (9.0%)	304 (10.4%)	
Controlling	152 (8.9%)	84 (7.0%)	236 (8.1%)	
Cook	136 (7.9%)	204 (16.9%)	340 (11.7%)	
Electrical and electronic equipment assembler	120 (7.0%)	64 (5.3%)	184 (6.3%)	
General and operations managers	104 (6.1%)	68 (5.6%)	172 (5.9%)	
Human resource managers	36 (2.1%)	12 (1.0%)	48 (1.6%)	
Human resources specialist	196 (11.4%)	116 (9.6%)	312 (10.7%)	
Industrial production managers	132 (7.7%)	116 (9.6%)	248 (8.5%)	
Restaurant managers	104 (6.1%)	76 (6.3%)	180 (6.2%)	
Software and applications developer	184 (10.7%)	92 (7.6%)	276 (9.5%)	
Stock clerk, sales floor	220 (12.9%)	156 (13.0%)	376 (12.9%)	
MED17				0.902
N-Miss	4	0	4	
Mean (SD)	22121.827 (4023.626)	22141.495 (4544.557)	22129.959 (4246.017)	
Range	15350.000 - 44370.000	13590.000 - 44370.000	13590.000 - 44370.000	
border				0.023
N-Miss	4	4	8	
Mean (SD)	0.302 (0.459)	0.263 (0.441)	0.286 (0.452)	
Range	0.000 - 1.000	0.000 - 1.000	0.000 - 1.000	
P17_POP				< 0.001
Mean (SD)	304753.724 (641755.326)	156903.598 (422846.634)	243707.239 (566419.575)	
Range	115.000 - 2187526.000	135.000 - 2187526.000	115.000 - 2187526.000	
taux_chomage				0.901
Mean (SD)	0.102 (0.029)	0.102 (0.030)	0.102 (0.029)	
Range	0.017 - 0.185	0.029 - 0.192	0.017 - 0.192	
paris				< 0.001
Mean (SD)	0.098 (0.298)	0.037 (0.188)	0.073 (0.260)	
Range	0.000 - 1.000	0.000 - 1.000	0.000 - 1.000	
m_share				< 0.001
N-Miss	800	556	1356	
Mean (SD)	0.060 (0.055)	0.076 (0.068)	0.067 (0.061)	
Range	0.000 - 0.338	0.000 - 0.338	0.000 - 0.338	
share_lepen1				0.001
Mean (SD)	16.628 (9.025)	17.700 (8.334)	17.071 (8.761)	
Range	3.610 - 52.150	3.610 - 47.480	3.610 - 52.150	
share_lepen				< 0.001
Mean (SD)	27.092 (12.929)	28.734 (11.975)	27.770 (12.568)	
Range	10.320 - 69.590	10.320 - 71.410	10.320 - 71.410	
vis				0.040
Mean (SD)	56.125 (18.901)	57.555 (18.007)	56.715 (18.547)	
Range	17.800 - 94.700	17.800 - 94.700	17.800 - 94.700	
mean_contact_soc				< 0.001
Mean (SD)	45.726 (14.329)	47.840 (14.157)	46.599 (14.294)	
Range	23.500 - 79.000	23.500 - 79.000	23.500 - 79.000	
originRH				0.791
N-Miss	236	92	328	
Mean (SD)	0.100 (0.300)	0.097 (0.296)	0.099 (0.299)	
Range	0.000 - 1.000	0.000 - 1.000	0.000 - 1.000	

Table .2: Conversion table PCS - Rome - ISCO - SOC

	Occupation - PCS	Rome	Rome descr.	ISCO	SOC	%	SOC detail	SOC descr. detail
1	Administrative employee	D1401	Assistanat commercial	3322	41-1012	47	41-1012.00	First-Line Supervisors of Non-Retail Sales Workers
2	Administrative employee	D1401	Assistanat commercial	3322	41-3099	20	41-3099.01	Energy Brokers
3	Administrative employee	D1401	Assistanat commercial	3322	41-4011	50	41-4011.07	Solar Sales Representatives and Assessors
4	Administrative employee	D1401	Assistanat commercial	3322	41-4011	28	41-4011.00	Sales Representatives, Wholesale and Manufacturing, Technical and Scientific Products
5	Administrative employee	D1401	Assistanat commercial	3322	41-4012	54	41-4012.00	Sales Representatives, Wholesale and Manufacturing, Except Technical and Scientific Products
6	Administrative employee	M1607	Secrétariat	4120	43-6014	47	43-6014.00	Secretaries and Administrative Assistants, Except Legal, Medical, and Executive
7	Electrical and electronic equipment assembler	H2605	Montage et câblage électronique	8212	51-2021	26	51-2021.00	Coil Winders, Tapers, and Finishers
8	Electrical and electronic equipment assembler	H2605	Montage et câblage électronique	8212	51-2022	32	51-2022.00	Electrical and Electronic Equipment Assemblers
9	Electrical and electronic equipment assembler	H2605	Montage et câblage électronique	8212	51-2023	12	51-2023.00	Electromechanical Equipment Assemblers
10	Electrical and electronic equipment assembler	H2605	Montage et câblage électronique	8212	51-2093	42	51-2093.00	Timing Device Assemblers and Adjusters
11	Electrical and electronic equipment assembler	H2605	Montage et câblage électronique	8212	51-9194	33	51-9194.00	Etchers and Engravers
12	Electrical and electronic equipment assembler	H2602	Câblage électrique et électromécanique	8212	51-2021	26	51-2021.00	Coil Winders, Tapers, and Finishers
13	Electrical and electronic equipment assembler	H2602	Câblage électrique et électromécanique	8212	51-2022	32	51-2022.00	Electrical and Electronic Equipment Assemblers
14	Electrical and electronic equipment assembler	H2602	Câblage électrique et électromécanique	8212	51-2023	12	51-2023.00	Electromechanical Equipment Assemblers
15	Electrical and electronic equipment assembler	H2602	Câblage électrique et électromécanique	8212	51-2093	42	51-2093.00	Timing Device Assemblers and Adjusters
16	Electrical and electronic equipment assembler	H2602	Câblage électrique et électromécanique	8212	51-9194	33	51-9194.00	Etchers and Engravers
17	Stock clerk, sales floor	D1507	Mise en rayon libre-service	9334	43-5081	86	43-5081.01	Stock Clerks, Sales Floor
18	Stock clerk, sales floor	D1507	Mise en rayon libre-service	9334	43-5081	54	43-5081.02	Marking Clerks
19	Stock clerk, sales floor	D1507	Mise en rayon libre-service	9334	43-5081	52	43-5081.04	Order Fillers, Wholesale and Retail Sales
20	Stock clerk, sales floor	D1507	Mise en rayon libre-service	9334	43-5081	22	43-5081.03	Stock Clerks- Stockroom, Warehouse, or Storage Yard
21	Cook	G1602	Personnel de cuisine	5120	35-1012	67	35-1012.00	First-Line Supervisors of Food Preparation and Serving Workers
22	Cook	G1602	Personnel de cuisine	5120	35-2012	43	35-2012.00	Cooks, Institution and Cafeteria
23	Cook	G1602	Personnel de cuisine	5120	35-2013	60	35-2013.00	Cooks, Private Household
24	Cook	G1602	Personnel de cuisine	5120	35-2014	56	35-2014.00	Cooks, Restaurant
25	Cook	G1602	Personnel de cuisine	5120	35-2015	59	35-2015.00	Cooks, Short Order
26	charge_recrut	M1502	Développement des ressources humaines	2423	13-1071	51	13-1071.00	Human Resources Specialists
27	charge_recrut	M1502	Développement des ressources humaines	2423	13-1075	36	13-1075.00	Labor Relations Specialists
28	charge_recrut	M1502	Développement des ressources humaines	2423	13-1141	17	13-1141.00	Compensation, Benefits, and Job Analysis Specialists
29	charge_recrut	M1502	Développement des ressources humaines	2423	21-1012	72	21-1012.00	Educational, Guidance, School, and Vocational Counselors
30	Commercial sales representative	H1102	Management et ingénierie d'affaires	1221	11-2021	45	11-2021.00	Marketing Managers
31	Commercial sales representative	H1102	Management et ingénierie d'affaires	1221	11-2022	68	11-2022.00	Sales Managers
32	Commercial sales representative	D1407	Relation technico-commerciale	3322	41-1012	47	41-1012.00	First-Line Supervisors of Non-Retail Sales Workers
33	Commercial sales representative	D1407	Relation technico-commerciale	3322	41-3099	20	41-3099.01	Energy Brokers
34	Commercial sales representative	D1407	Relation technico-commerciale	3322	41-4011	50	41-4011.07	Solar Sales Representatives and Assessors
35	Commercial sales representative	D1407	Relation technico-commerciale	3322	41-4011	28	41-4011.00	Sales Representatives, Wholesale and Manufacturing, Technical and Scientific Products
36	Commercial sales representative	D1407	Relation technico-commerciale	3322	41-4012	54	41-4012.00	Sales Representatives, Wholesale and Manufacturing, Except Technical and Scientific Products
37	Commercial sales representative	M1706	Promotion des ventes	1221	11-2021	45	11-2021.00	Marketing Managers
38	Commercial sales representative	M1706	Promotion des ventes	1221	11-2022	68	11-2022.00	Sales Managers
39	Controlling	M1204	Contrôle de gestion	1211	11-3031	58	11-3031.02	Financial Managers, Branch or Department
40	Controlling	M1204	Contrôle de gestion	1211	11-3031	29	11-3031.01	Treasurers and Controllers
41	Controlling	M1202	Audit et contrôle comptables et financiers	2411	13-2011	51	13-2011.02	Auditors
42	Controlling	M1202	Audit et contrôle comptables et financiers	2411	13-2011	20	13-2011.01	Accountants
43	Controlling	M1202	Audit et contrôle comptables et financiers	2411	13-2031	13	13-2031.00	Budget Analysts
44	Controlling	M1202	Audit et contrôle comptables et financiers	2411	13-2082	78	13-2082.00	Tax Preparers
45	Software and applications developer	M1805	Études et développement informatique	2519	15-1199	41	15-1199.04	Geospatial Information Scientists and Technologists
46	Software and applications developer	M1805	Études et développement informatique	2519	15-1199	41	15-1199.05	Geographic Information Systems Technicians
47	Software and applications developer	M1805	Études et développement informatique	2519	15-1199	39	15-1199.09	Information Technology Project Managers
48	Software and applications developer	M1805	Études et développement informatique	2519	15-1199	30	15-1199.03	Web Administrators
49	Software and applications developer	M1805	Études et développement informatique	2519	15-1199	24	15-1199.10	Search Marketing Strategists
50	Software and applications developer	M1805	Études et développement informatique	2519	15-1199	23	15-1199.02	Computer Systems Engineers/Architects
51	Software and applications developer	M1805	Études et développement informatique	2519	15-1199	21	15-1199.12	Document Management Specialists
52	Software and applications developer	M1805	Études et développement informatique	2519	15-1199	18	15-1199.11	Video Game Designers

53	Software and applications developer	M1805	Études et développement informatique	2519	15-1199	15	15-1199.08	Business Intelligence Analysts
54	Software and applications developer	M1805	Études et développement informatique	2519	15-1199	15	15-1199.06	Database Architects
55	Software and applications developer	M1805	Études et développement informatique	2519	15-1199	9	15-1199.01	Software Quality Assurance Engineers and Testers
56	Software and applications developer	M1805	Études et développement informatique	2519	15-1199	6	15-1199.07	Data Warehousing Specialists
57	Human resource managers	M1502	Développement des ressources humaines	2423	13-1071	51	13-1071.00	Human Resources Specialists
58	Human resource managers	M1502	Développement des ressources humaines	2423	13-1075	36	13-1075.00	Labor Relations Specialists
59	Human resource managers	M1502	Développement des ressources humaines	2423	13-1141	17	13-1141.00	Compensation, Benefits, and Job Analysis Specialists
60	Human resource managers	M1502	Développement des ressources humaines	2423	21-1012	72	21-1012.00	Educational, Guidance, School, and Vocational Counselors
61	Human resource managers	M1503	Management des ressources humaines	1212	11-3111	19	11-3111.00	Compensation and Benefits Managers
62	Human resource managers	M1503	Management des ressources humaines	1212	11-3121	45	11-3121.00	Human Resources Managers
63	Human resource managers	M1503	Management des ressources humaines	1212	11-3131	42	11-3131.00	Training and Development Managers
64	Industrial production managers	H2502	Management et ingénierie de production	1321	11-3051	43	11-3051.06	Hydroelectric Production Managers
65	Industrial production managers	H2502	Management et ingénierie de production	1321	11-3051	25	11-3051.04	Biomass Power Plant Managers
66	Industrial production managers	H2502	Management et ingénierie de production	1321	11-3051	22	11-3051.01	Quality Control Systems Managers
67	Industrial production managers	H2502	Management et ingénierie de production	1321	11-3051	22	11-3051.02	Geothermal Production Managers
68	Industrial production managers	H2502	Management et ingénierie de production	1321	11-3051	21	11-3051.03	Biofuels Production Managers
69	Industrial production managers	H2502	Management et ingénierie de production	1321	11-3051	20	11-3051.00	Industrial Production Managers
70	Industrial production managers	H1301	Inspection de conformité	3257	29-9012	39	29-9012.00	Occupational Health and Safety Technicians
71	Industrial production managers	H1301	Inspection de conformité	3257	45-2011	57	45-2011.00	Agricultural Inspectors
72	Industrial production managers	H1301	Inspection de conformité	3257	53-1031	58	53-1031.00	First-Line Supervisors of Transportation and Material-Moving Machine and Vehicle Operators
73	Industrial production managers	H1301	Inspection de conformité	3257	53-6051	63	53-6051.01	Aviation Inspectors
74	Industrial production managers	H1301	Inspection de conformité	3257	53-6051	35	53-6051.08	Freight and Cargo Inspectors
75	Industrial production managers	H1301	Inspection de conformité	3257	53-6051	13	53-6051.07	Transportation Vehicle, Equipment and Systems Inspectors, Except Aviation
76	General and operations managers	D1509	Management de département en grande distribution	1420	11-1021	71	11-1021.00	General and Operations Managers
77	General and operations managers	D1504	Direction de magasin de grande distribution	1420	11-1021	71	11-1021.00	General and Operations Managers
78	General and operations managers	D1301	Management de magasin de détail	1420	11-1021	71	11-1021.00	General and Operations Managers
79	Restaurant managers	G1401	Assistance de direction d'hôtel-restaurant	1411	11-9081	79	11-9081.00	Lodging Managers
80	Restaurant managers	G1402	Management d'hôtel-restaurant	1411	11-9081	79	11-9081.00	Lodging Managers
81	Restaurant managers	G1403	Gestion de structure de loisirs ou d'hébergement touristique	1431	11-9071	79	11-9071.00	Gaming Managers
82	Restaurant managers	G1403	Gestion de structure de loisirs ou d'hébergement touristique	1431	11-9199	70	11-9199.11	Brownfield Redevelopment Specialists and Site Managers
83	Restaurant managers	G1403	Gestion de structure de loisirs ou d'hébergement touristique	1431	11-9199	70	11-9199.10	Wind Energy Project Managers
84	Restaurant managers	G1403	Gestion de structure de loisirs ou d'hébergement touristique	1431	11-9199	61	11-9199.07	Security Managers
85	Restaurant managers	G1403	Gestion de structure de loisirs ou d'hébergement touristique	1431	11-9199	57	11-9199.02	Compliance Managers
86	Restaurant managers	G1403	Gestion de structure de loisirs ou d'hébergement touristique	1431	11-9199	44	11-9199.08	Loss Prevention Managers
87	Restaurant managers	G1403	Gestion de structure de loisirs ou d'hébergement touristique	1431	11-9199	38	11-9199.09	Wind Energy Operations Managers
88	Restaurant managers	G1403	Gestion de structure de loisirs ou d'hébergement touristique	1431	11-9199	32	11-9199.04	Supply Chain Managers
89	Restaurant managers	G1403	Gestion de structure de loisirs ou d'hébergement touristique	1431	11-9199	26	11-9199.03	Investment Fund Managers
90	Restaurant managers	G1403	Gestion de structure de loisirs ou d'hébergement touristique	1431	11-9199	14	11-9199.01	Regulatory Affairs Managers
91	Restaurant managers	G1404	Management d'établissement de restauration collective	1412	11-9051	75	11-9051.00	Food Service Managers

Table .3: O*Net contact measure by Code Rome classification of jobs

	Occupations	Rome	PCS	% Contact O*Net
1	Software and applications developer	M1805	388a	23.5
2	Industrial production managers	H2502	385b	25.5
3	Electrical and electronic equipment assembler	H1504	622c	29
4	Electrical and electronic equipment assembler	H2602	622c	29
5	Electrical and electronic equipment assembler	H2605	622c	29
6	Administrative employee	D1401	542b	39.8
7	Commercial sales representative	D1407	388d	39.8
8	Controlling	M1202	373a	40.5
9	Administrative employee	M1608	542b	41
10	Controlling	M1204	373a	43.5
11	Human resources specialist	M1502	372c	44
12	Human resource managers	M1502	372c	44
13	Administrative employee	M1607	542b	47
14	Stock clerk, sales floor	D1507	551a	53.5
15	Commercial sales representative	H1102	388d	56.5
16	Cook	G1602	636d	57
17	General and operations managers	D1301	374a	71
18	General and operations managers	D1504	374a	71
19	General and operations managers	D1509	374a	71
20	Restaurant managers	G1404	377a	75
21	Restaurant managers	G1401	377a	79
22	Restaurant managers	G1402	377a	79

The highlighted occupations (14-22) are classified as "high" contact jobs, while (1-13) are classified as low. Note that the mean of several Code-Rome classifications can be the same (which, for instance, is the case for all Code-Rome jobs in the occupation "Electrical and electronic equipment assembler". This occurs, when the SOC classifications happen to be the same for each Code Rome, see .2 for the detailed conversions).

Table .4: Cross-correlation table of municipality-level diversity proxies

Variables	immig1_share	eu_com	noneu_com	dom_tom_com	abroad_french_com	french_com	taux_chomage	m_share	share_lepen1	share_lepen
immig1_share	1.00									
Nb. Obs.										
eu_com	0.49	1.00								
	(0.00)									
Nb. Obs.	2772									
noneu_com	0.46	0.35	1.00							
	(0.00)	(0.00)								
Nb. Obs.	2772	2772								
dom_tom_com	0.57	0.32	0.24	1.00						
	(0.00)	(0.00)	(0.00)							
Nb. Obs.	2916	2772	2772							
abroad_french_com	0.73	0.63	0.55	0.54	1.00					
	(0.00)	(0.00)	(0.00)	(0.00)						
Nb. Obs.	2772	2772	2772	2772						
french_com	-0.64	-0.70	-0.85	-0.33	-0.86	1.00				
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)					
Nb. Obs.	2916	2772	2772	2916	2772					
taux_chomage	0.31	-0.17	0.16	-0.05	0.02	-0.04	1.00			
	(0.00)	(0.00)	(0.00)	(0.01)	(0.23)	(0.02)				
Nb. Obs.	2916	2772	2772	2916	2772	2916				
m_share	0.68	0.50	0.43	0.37	0.72	-0.55	0.10	1.00		
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)			
Nb. Obs.	1560	1508	1508	1560	1508	1560	1560			
share_lepen1	-0.40	-0.16	-0.19	-0.53	-0.24	0.22	0.15	0.03	1.00	
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.18)		
Nb. Obs.	2916	2772	2772	2916	2772	2916	2916	1560		
share_lepen	-0.42	-0.11	-0.19	-0.54	-0.24	0.20	0.11	0.01	0.98	1.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.66)	(0.00)	
Nb. Obs.	2916	2772	2772	2916	2772	2916	2916	1560	2916	

Table .5: Cross-correlation table of firm-level diversity proxies and share of foreigners in municipality

Variables	immig1_share	eu_siret	noneu_siret	abroad_french_siret	french_siret
immig1_share	1.00				
eu_siret	0.06	1.00			
	(0.09)				
noneu_siret	0.35	0.07	1.00		
	(0.00)	(0.04)			
abroad_french_siret	0.29	0.01	0.28	1.00	
	(0.00)	(0.88)	(0.00)		
french_siret	-0.40	-0.37	-0.72	-0.78	1.00
	(0.00)	(0.00)	(0.00)	(0.00)	
Nb. obs. : 812					

Table .6: Descriptive statistics of the sample for which share of births with an Arabic last name is observed and by low/high sub-samples

	0 (N=1056)	1 (N=504)	Total (N=1560)	p value
originRH				< 0.001
N-Miss	88	36	124	
Mean (SD)	0.074 (0.263)	0.179 (0.384)	0.109 (0.311)	
Range	0.000 - 1.000	0.000 - 1.000	0.000 - 1.000	
metier_noms				< 0.001
Administrative employee	80 (7.6%)	32 (6.3%)	112 (7.2%)	
Commercial sales representative	88 (8.3%)	64 (12.7%)	152 (9.7%)	
Controlling	76 (7.2%)	44 (8.7%)	120 (7.7%)	
Cook	116 (11.0%)	48 (9.5%)	164 (10.5%)	
Electrical and electronic equipment assembler	48 (4.5%)	24 (4.8%)	72 (4.6%)	
General and operations managers	60 (5.7%)	52 (10.3%)	112 (7.2%)	
Human resource managers	12 (1.1%)	8 (1.6%)	20 (1.3%)	
Human resources specialist	136 (12.9%)	72 (14.3%)	208 (13.3%)	
Industrial production managers	132 (12.5%)	16 (3.2%)	148 (9.5%)	
Restaurant managers	72 (6.8%)	40 (7.9%)	112 (7.2%)	
Software and applications developer	100 (9.5%)	64 (12.7%)	164 (10.5%)	
Stock clerk, sales floor	136 (12.9%)	40 (7.9%)	176 (11.3%)	
age_group				0.647
Mean (SD)	1.667 (0.648)	1.651 (0.622)	1.662 (0.639)	
Range	1.000 - 3.000	1.000 - 3.000	1.000 - 3.000	
MED17				0.004
Mean (SD)	21039.848 (4195.230)	20386.111 (4277.761)	20828.641 (4231.762)	
Range	15550.000 - 44370.000	13590.000 - 37240.000	13590.000 - 44370.000	
border				< 0.001
N-Miss	4	0	4	
Mean (SD)	0.141 (0.348)	0.286 (0.452)	0.188 (0.391)	
Range	0.000 - 1.000	0.000 - 1.000	0.000 - 1.000	
P17_POP				< 0.001
Mean (SD)	93648.170 (108233.575)	183463.111 (247050.767)	122665.305 (171420.256)	
Range	1242.000 - 479553.000	3735.000 - 863310.000	1242.000 - 863310.000	
taux_chomage				< 0.001
Mean (SD)	0.114 (0.024)	0.120 (0.028)	0.116 (0.026)	
Range	0.031 - 0.172	0.052 - 0.192	0.031 - 0.192	
paris				NaN
Mean (SD)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	
Range	0.000 - 0.000	0.000 - 0.000	0.000 - 0.000	
immig1_share				< 0.001
Mean (SD)	0.101 (0.048)	0.187 (0.072)	0.128 (0.070)	
Range	0.009 - 0.302	0.028 - 0.448	0.009 - 0.448	
share_lepen1				< 0.001
Mean (SD)	15.542 (7.543)	17.300 (7.886)	16.110 (7.697)	
Range	3.610 - 52.150	5.840 - 44.560	3.610 - 52.150	
share_lepen				< 0.001
Mean (SD)	25.517 (10.659)	27.523 (11.581)	26.165 (11.002)	
Range	11.220 - 71.410	11.110 - 64.710	11.110 - 71.410	

Table .7: OLS results, interaction of neighbourhood diversity (continuous variable) and origin of the applicant

	Share of foreigners		Share Arabic last name	
	(1)	(2)	(3)	(4)
maghre	-0.153***	-0.163***	-0.163***	-0.194***
	(-6.36)	(-6.11)	(-6.40)	(-6.40)
Share of foreigners	-0.117	0.558		
	(-0.54)	(1.43)		
Share of foreigners*maghre	0.295	0.294		
	(1.84)	(1.77)		
Female		0.00674		-0.0104
		(0.41)		(-0.46)
Female*maghre		0.0229		0.0672*
		(1.11)		(2.36)
Send order		-0.0156**		-0.0187*
		(-2.84)		(-2.40)
Age group		-0.0449		-0.0444
		(-1.67)		(-1.13)
Share Arabic last names			-0.912**	-0.604
			(-2.90)	(-1.05)
Share Arabic last names*maghre			0.616*	0.629*
			(2.46)	(2.40)
_cons	0.387***	0.551**	0.452***	0.653
	(12.92)	(2.64)	(14.09)	(1.70)
Ind./Firm Char.	No	Yes	No	Yes
Mun. Char.	No	Yes	No	Yes
N	2916	2904	1560	1556
adj. R^2	0.016	0.130	0.023	0.168

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table .8: Het. Probit by diversity (low/high) of the work-force population in a given neighbourhood with higher cut-off for share of European foreigners

	(1) Low	(2) High	(3) Low	(4) High
measure = share of French, born in France (2015)				
maghre	-0.114***	-0.157***	-0.129***	-0.208***
	(-6.95)	(-5.42)	(-6.91)	(-4.31)
$Log-lik.$	-1129.5	-552.3	-990.5	-433.3
N	1924	968	1864	840
measure = share of European foreigners (2015) cutoff = 7.3% (Non-EU)				
maghre	-0.131***	-0.106	-0.135***	-0.225*
	(-7.67)	(-1.86)	(-7.66)	(-2.23)
$Log-lik.$	-1554.5	-112.9	-1387.5	-85.7
N	2688	204	2592	184
measure = share of Non-European foreigners (2015)				
maghre	-0.155***	-0.094***	-0.176***	-0.102**
	(-8.36)	(-3.60)	(-7.73)	(-3.28)
$Log-lik.$	-1041.4	-550.6	-881.1	-448.1
N	1848	912	1728	840

Marginal effects; t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Het. Probit regressions, with standard errors clustered on the vacancy level. Ind./Firm Char. include, gender, the send order of the application, the age group, marital status, having children and whether the respondent has been in inactivity as well as occupation dummies. Mun. Char. include the median household income, the unemployment rate, the population size, the mean net hourly wage by municipality, dummies for whether a municipality is located at a land border and department dummies.

Table .9: Probit and Het. Probit full sample, outcome = invitation rate

	(1) Probit	(2) Probit	(3) Het. Probit	(4) Probit	(5) Het.Probit
maghre	-0.067***	-0.065***	-0.061***	-0.069***	-0.061***
	(-7.10)	(-7.03)	(-5.91)	(-7.11)	(-4.37)
$Log-lik.$	-1155.5	-1091.5	-1091.2	-954.5	-954.2
N	2916	2916	2916	2692	2692
Ind./Firm Char.	No	Yes	Yes	Yes	Yes
Mun. Char.	No	No	No	Yes	Yes

Marginal effects; t statistics in parentheses; (d) for discrete change of dummy variable from 0 to 1

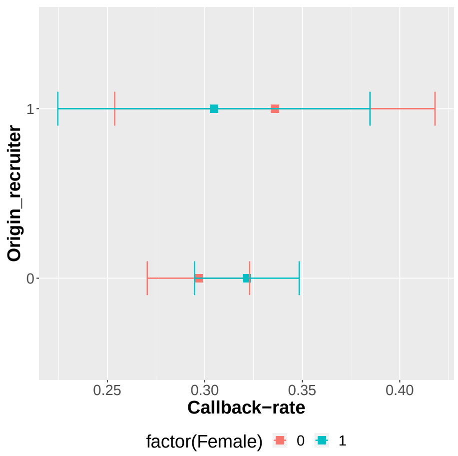
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table .10: Het. Probit by origin and gender of the recruiter

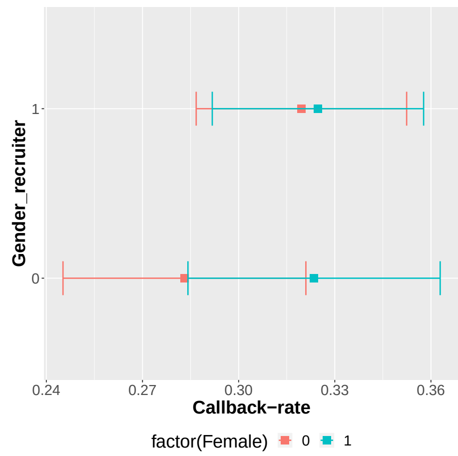
	(1)	(2)	(3)	(4)
	0	1	0	1
maghre	-0.172***	-0.115***	-0.205***	-0.145***
	(-6.61)	(-5.73)	(-5.19)	(-5.97)
N	1080	1532	964	1444
Ind./Firm Char.	Yes	Yes	Yes	Yes
Mun. Char.	No	No	Yes	Yes

Marginal effects; t statistics in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure .1: CB-rates by gender of applicant and identity of recruiter



(a) 1 = Non-European recruiter



(b) 1 = Female recruiter

Table .11: Probit and Het. Probit by origin, full sample for which within-firm diversity measure is observed

	(1)	(2)	(3)	(4)	(5)
	Probit	Probit	Het. Probit	Probit	Het.Probit
maghre	-0.143***	-0.157***	-0.143***	-0.158***	-0.145***
	(-6.12)	(-6.40)	(-5.66)	(-6.28)	(-5.61)
$Log-lik.$	-502.7	-446.8	-446.	-434.4	-433.8
N	816	804	804	800	800
Ind./Firm Char.	No	Yes	Yes	Yes	Yes
Mun. Char.	No	No	No	Yes	Yes

Marginal effects; t statistics in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table .12: Interaction diversity (low/high div.) and visibility (low/high vis.), div. measured by share of immigrants (immig_1_share in Figure 5.1)

	(low div.,low vis.)	(low div.,high vis.)	(high div.,low vis.)	(high div.,high vis.)
		contact = PCS, 3 digits		
maghre	-0.123***	-0.163***	-0.097***	-0.080
	(-4.53)	(-6.49)	(-3.31)	(-1.69)
N	888	1048	628	324
		contact = PCS, 2 digits		
maghre	-0.126***	-0.175***	-0.071**	-0.127
	(-5.59)	(-5.36)	(-2.79)	(-1.92)
N	1196	740	748	204
		contact = O*Net		
maghre	-0.129***	-0.172***	-0.080**	-0.115**
	(-5.57)	(-6.05)	(-2.64)	(-2.84)
N	1048	888	540	412

Marginal effects; t statistics in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table .13: Interaction diversity (low/high div.) and visibility (low/high vis.), div. measured by share of births with Arabic last name (m_share in Figure 5.1)

	(low div.,low vis.)	(low div.,high vis.)	(high div.,low vis.)	(high div.,high vis.)
		contact = PCS, 3 digits		
maghre	-0.141***	-0.164***	-0.100	-0.062
	(-4.40)	(-4.02)	(-1.55)	(-0.86)
N	524	528	308	188
		contact = PCS, 2 digits		
maghre	-0.127***	-0.241***	-0.047	0.000
	(-4.46)	(-4.27)	(-0.97)	(.)
N	724	328	376	104
		contact = O*Net		
maghre	-0.120***	-0.185***	-0.011	-0.108
	(-4.43)	(-3.78)	(-0.15)	(-1.08)
N	604	448	264	232

Marginal effects; t statistics in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table .14: Interaction diversity (low/high div.) and visibility (low/high vis.), div. measured by share of non-European employees (noneu_com in Figure 5.2)

	(low div.,low vis.)	(low div.,high vis.)	(high div.,low vis.)	(high div.,high vis.)
		contact = PCS, 3 digits		
maghre	-0.127***	-0.186***	-0.105**	-0.102*
	(-4.56)	(-6.84)	(-2.87)	(-2.50)
N	880	956	524	388
		contact = PCS, 2 digits		
maghre	-0.132***	-0.187***	-0.073**	-0.084
	(-6.10)	(-4.45)	(-2.60)	(-1.25)
N	1196	640	640	272
		contact = O*Net		
maghre	-0.137***	-0.172***	-0.083*	-0.118**
	(-5.80)	(-5.93)	(-2.19)	(-3.23)
N	976	860	516	396

Marginal effects; t statistics in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table .15: Interaction diversity (low/high div.) and visibility (low/high vis.), div. measured by share of French, born in France employees (french_com in Figure 5.2)

	(high div.,low vis.)	(high div.,high vis.)	(low div.,low vis.)	(low div.,high vis.)
		contact = PCS, 3 digits		
maghre	-0.110***	-0.143***	-0.130*	-0.159***
	(-4.92)	(-4.94)	(-2.42)	(-3.60)
N	1108	816	400	556
		contact = PCS, 2 digits		
maghre	-0.104***	-0.140***	-0.110**	-0.219***
	(-5.34)	(-4.33)	(-2.94)	(-3.36)
N	1348	576	588	368
		contact = O*Net		
maghre	-0.103***	-0.137***	-0.149***	-0.126
	(-4.64)	(-5.10)	(-3.89)	(-1.36)
N	1048	876	532	424

Marginal effects; t statistics in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table .16: Interaction vote shares for far right party leader (low/high vote) and visibility (low/high vis.), vote shares for far right party leader in first round elections, 2017 (share_lepen1 in Figure 5.2)

	(low vote,low vis.)	(low vote,high vis.)	(high vote,low vis.)	(high vote,high vis.)
	contact = PCS, 3 digits			
maghre	-0.106***	-0.163***	-0.126*	-0.089
	(-4.82)	(-5.67)	(-2.32)	(-0.93)
N	1128	804	400	568
	contact = PCS, 2 digits			
maghre	-0.118***	-0.119***	-0.071*	-0.263**
	(-6.04)	(-3.60)	(-1.96)	(-3.17)
N	1376	556	580	388
	contact = O*Net			
maghre	-0.094***	-0.175***	-0.172***	-0.081
	(-4.56)	(-6.03)	(-3.91)	(-1.16)
N	1104	828	492	472

Marginal effects; t statistics in parentheses

(d) for discrete change of dummy variable from 0 to 1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$