

# DO MINORITY CANDIDATES BENEFIT FROM ANONYMOUS JOB APPLICATIONS? EVIDENCE FROM A RANDOMIZED FIELD EXPERIMENT\*

LUC BEHAGHEL  
BRUNO CRÉPON  
THOMAS LE BARBANCHON<sup>†</sup>

August 19, 2013

This paper reports the results of a randomized experiment in which recruiters were assigned to receive either anonymous or standard resumes of real candidates. The experiment was designed to examine the effect of a law passed in France in 2006, which made anonymous resumes mandatory for all firms with more than 50 employees. The purpose of the law was to reduce discrimination against immigrants and people living in deprived neighborhoods, who usually have lower interview rates than majority candidates. The study's results draw on a sample of about 600 firms that participated in the experiment. The main finding is that anonymization widened both the interview and the hiring gaps between applicants from the majority and the minority. Anonymous resumes thus have the exact opposite effect to what policymakers hoped. Moreover the total number of interviews across both types and measures for recruitment success were unaffected. Using detailed information coded from the resumes, we find evidence that the impact of anonymous resumes is partly due to a change in the way recruiters evaluated the candidates' credentials in resumes. These changes in the way resumes are valued also imply strong redistribution in the population selected for interviews.

*Keywords:* Anonymous applications, discrimination, randomized experiment.

*JEL:* J71, J78.

*Word count:* 19,264

---

\*We thank participants at various conferences (SOLE 2012, ESPE 2012, ESEM 2012, EALE 2012, ASSA 2013, NORFACE 2013) and seminar participants at IZA, IAB, Crest, Centre d'Etudes de l'Emploi, Cergy University, Ined, and Paris School of Economics for their helpful comments. We are especially grateful to Kate Antonovics, David Blau, Laurent Davezies, Philip Oreopoulos, Roland Rathelot for comments on earlier drafts of this paper. We thank the French public employment service, Pôle Emploi, for its commitment in implementing this experiment. Pôle Emploi also provided access to data and financial support for this study. We thank Andrea Lepine, Julie Moschion and Pascal Achard for excellent research assistance, Abderrazak Chebira, Yann Algan, Corinne Prost and Hélène Garner for very helpful discussion on the measurement of applicants' origin.

<sup>†</sup>Address correspondence to thomas.le-barbanchon@ensae.fr, 15 Boulevard Gabriel Péri 92245 Malakoff Cedex France - Tel.: +33623786708 - Fax.:+33141176029

# I. INTRODUCTION

In correspondence testing experiments, researchers create fictitious job applications for real jobs in order to gauge the degree of discrimination among hiring firms. These studies have repeatedly provided compelling evidence that recruiters largely favor members of the racial or ethnic majority over minority candidates when choosing among otherwise-identical resumes (Riach and Rich, 2002; Bertrand and Mullainathan, 2004). This finding has often been interpreted as evidence of hiring discrimination against minority candidates.<sup>1</sup> This has led some policymakers in various countries (for instance Belgium, Germany, the Netherlands, Sweden and the United Kingdom) to advocate anonymous job applications as a way to reduce the hiring gap between minority and majority candidates. Anonymous resumes, however, have never been forcefully implemented on a large scale, and empirical evidence on their impact remains scarce.

In this paper, we analyze a large-scale experiment, in which about 600 recruiters were randomly assigned to receive either anonymous or standard resumes. The experiment was designed to examine the potential effects of a law passed in France in 2006 to make anonymous resumes mandatory for all firms with more than 50 employees, but yet never implemented. The law was passed in the wake of the 2005 riots in the greater Paris area in an attempt to attenuate discrimination against minority candidates, namely residents from deprived neighborhoods, immigrants and children of immigrants.

Although some studies already address the effect of anonymization on the recruitment process, they focus mostly on gender discrimination. Most existing results have been obtained using difference-in-difference strategies. Goldin and Rouse (2000), for example, study the introduction of screens concealing the identity of musicians in hiring auditions of American orchestras. Their difference-in-difference analysis shows that women benefit from blind auditions. Aslund and Skans (2012) analyze recruitment processes for 109 public jobs in Sweden that used anonymous applications. Their difference-in-difference analysis shows that women and foreigners benefit from anonymous applications.<sup>2</sup> The only existing randomized controlled trial was implemented at an economic

---

<sup>1</sup>The definition of “minority” candidates depends on the context (for instance, minority candidates can be characterized by an African-American sounding name in a particular correspondence study). As detailed below, minority candidates in our applications will be characterized by their foreign background and/or their residence in deprived neighborhoods.

<sup>2</sup> Other difference-in-difference results include Edin and Lagerström (2006). They take advantage of the possibility to apply anonymously on public employment service websites in Sweden. The callback rates for women increase with blind applications (in particular in male-dominated occupations) but there is no significant effect on applicants with foreign-sounding names. Lastly, Krause et al. (2012a) analyze anonymous applications in volunteering firms on 11

research institution (Krause et al., 2012b). They find that women are adversely affected by anonymous applications.

The public debate about anonymous resumes is quite contentious. Proponents of anonymous resumes put forward several arguments. Hiding whether the candidate belongs to the minority should have an immediate impact on who gets interviewed (since fewer resumes – or none – would be screened by origin) and possibly on who gets hired. In turn, even if minority candidates do not necessarily get hired, reaching the interview stage should boost their confidence and, in the long run, induce them to apply for more jobs (the so-called “calling effect”). Moreover, having recruiters meet minority candidates face-to-face might reduce their prejudice against them. On the other hand, opponents note that anonymous resumes only make one half of the hiring process “race blind.” In the remaining half, i.e. during the interview, the candidate’s minority status is revealed, so that one can hardly prevent recruiters from choosing according to their preferences. A related point is that recruiters may interview more candidates when all the resumes they receive are anonymous, in order to make sure to see at least one candidate of their preferred group. The resulting equilibrium would clearly be inefficient: an increase in hiring cost, but no reduction in the hiring gap. Another argument against anonymous resumes is the fact that “race blind” screening processes are not necessarily “race neutral.” If there is a correlation between minority status and other elements in the resume, it may not be difficult for recruiters to reconstruct missing information and continue to discriminate.<sup>3</sup>

Although the French anonymization law was passed in 2006, its implementation was long postponed. As the issue came up again in the public debate in 2009, the government asked the Public Employment Service (PES) to first provide evidence about what the effect of such a policy could be in terms of costs of hiring, productivity of the realized matches and the reduction in discrimination. As part of the evidence it intended to produce, the PES asked our team to design a randomized experiment, the results of which are reported in this paper.

The design was as follows: when a vacant job suitable for the experiment was identified and when the recruiter was ready to participate in the experiment, PES caseworkers first selected a suitable pool of candidates for the vacancy. The matched-up vacancy/candidate pool was then

---

hiring processes in Germany, and find heterogeneous impacts on minority candidates depending on the firm.

<sup>3</sup>See for instance Persico (2009). Similarly Autor and Scarborough (2008) study the conditions under which “race-blind” job testing are not necessarily “race-neutral.”

randomly assigned either to a group with a standard recruitment procedure or to a group where all candidates' resumes were made anonymous. We implemented two surveys designed to follow both applicants and firms. It is important to note that firms' participation in the experiment was not mandatory: we document the selection of firms into the experiment below, and discuss the consequences for external validity in our conclusion. While our experiment does not enable us to estimate the effect of anonymous resumes for any French firm, we argue that we still estimate a relevant treatment effect for public policy purposes, notably on those firms who would comply if the policy was encouraged but not strictly enforced (monitoring costs are likely to be important). It is also important to note that candidates and caseworkers did not know whether their resumes would be anonymized or not. This guarantees that the impact of anonymous resumes on firms' reactions is not confounded by changes in the pool of candidates, but it rules out some potential effects of the policy, such as a calling effect on minority candidates.

We first present the results of a model building on the statistical discrimination literature *à la* Aigner and Cain (1977) and Cornell and Welch (1996), in which we take into account the two stages of a typical hiring process (a tournament in which resume screening is followed by interviews to pick one candidate). The model builds upon the assumption that a given credential in the resume is not as informative to the recruiter when the candidate belongs to the minority as when he belongs to the majority. The model shows that the effect of anonymous resumes improves the access to interview for minority applicants, when the cost per interview is high, but that it can have the opposite effect, when the cost per interview is low. In the latter case, recruiters in the standard procedure interview a larger share of applicants from the minority due to the low information content of their resumes. The model also shows that using anonymous resumes should always improve the hiring rate of applicants from the minority, although the effect can be very small, especially when interview costs are small.

The experimental results draw on a sample of about 600 firms participating in the experiment. We compare the gap in interview and hiring rates between minority and majority candidates under anonymous and standard recruitment procedures. We find the surprising result that anonymization leads to a large and significant widening of the gap in interview rates: the interview rate of minority candidates decreases, while that of majority candidates increases. We also find that the hiring

rate gap worsens, but the effect is smaller than that on interview rates, and it is only significant at the 10% level. Anonymous resumes thus had the exact opposite effect to what policymakers hoped: minority candidates are adversely affected by anonymous resumes. It was thought that the effect could be small and insignificant; indeed, a large negative opposite effect was never expected. Interestingly, the total number of interviews and measures for recruitment success are unaffected.<sup>4</sup>

Interpreting these results is a challenge. To deepen our understanding of the treatment effect, we took all the resumes involved in the experiment, anonymized all of them and asked counselors from the PES to evaluate them in the way recruiters would do. They rated them and also coded some information from the resume: inactivity periods, consistency of applicant skills with the position, adequate experience, and so on. We used this coding to examine how resumes are valued in the standard procedure, paying attention to potential differences in the return to each of the information items between minority and majority candidates. Consistent with Aigner and Cain (1977) and our theoretical model, we indeed found that some items had a significantly different return. This is the case for inactivity periods, which are strongly negatively valued for majority candidates but not for minority candidates. We also found that recruiters examining anonymous resumes update the value they attach to a particular item by averaging across the minority and majority groups. As a consequence, inactivity periods, for example, are now negatively valued for minority candidates under anonymization, and penalize majority candidates less. We show that the change in the valuation of the resumes under treatment and control explains a large share of the change in the interview gap between majority and minority candidates. In addition, we show that it alters the relative positions of candidates within either of these two groups. However, a single theoretical model of statistical discrimination does not explain all the findings: confronted with anonymous resumes, recruiters do not always update their beliefs rationally, as they also start putting more weight on signals that they would not consider when reading name-bearing resumes of minority and majority candidates, even when those signals are not informative about the minority or majority status of the candidate.

The rest of the paper is structured as follows. Section 2 provides some information about the

---

<sup>4</sup>Importantly, we follow control firms after the experiment and document that their recruitment behavior is not different during the experiment. Thus, the results are not an artifact of the experiment due to a change in the behavior of the control group. Accordingly, had we conducted a correspondence study (in which firms do not know that they are observed) on the firms participating to the experiment, it is likely that we would have found that, in these firms, minority workers are more often interviewed than majority ones.

political context in which the experiment took place. Section 3 presents the main assumptions and predictions of the theoretical model. Section 4 presents the experimental design, data collection and description. Section 5 presents main results and the robustness check. Section 6 provides results about changes in the way resumes are valued. The last section concludes.

## II. INSTITUTIONAL BACKGROUND

The first significant mention of anonymous resumes in the French political discourse was due to an official report by Claude Bébéar (2004), the CEO of AXA, the main French insurance company, which had for several years been using this procedure for its own recruitment.<sup>5</sup> The recommendation to use anonymous resumes in this report came along with several other recommendations aimed at sensitizing recruiters and developing reporting tools on ethnic diversity within companies, despite the fact that collecting explicit data on race is forbidden by French law (even in surveys, individuals cannot be asked whether they are black or white, Caucasian or African, etc.). At that stage, the purpose of anonymous resumes is to prevent hiring discrimination against women and older workers as much as against ethnic minorities or residents of deprived neighborhoods. Following Bébéar’s report, Parliament debated and rejected a legislative amendment put forward by a conservative member that would have made anonymous resumes mandatory for firms with more than 250 workers. Bébéar himself did not support the amendment, arguing that “All the recommendations in [his] report are based on firms’ volunteering (...). A law might even be counterproductive.”

The debate on anonymous resumes was revived after the riots that took place in the greater Paris area in November 2005. A law passed in April 2006 (the Equality of Chances Act) made anonymous resumes mandatory for firms with more than 50 employees. The law, however, specified neither who was supposed to actually anonymize resumes and what it meant in practice, nor what sanctions would be levied against those who do not respect the law. The government turned to firms and worker unions to hash out the conditions of implementation and discuss more broadly ways to fight hiring discrimination. The measure faced strong resistance by firms, who stressed the costs of anonymization, and received limited support from worker unions. Anonymous resumes however remained on the political agenda, in particular due to pressure from associations fighting discrimination. A group of recruitment firms assembled in 2006 to promote the equal treatment

---

<sup>5</sup>In French: [http://www.larevueparlementaire.fr/pages/RP875/RP875\\_AN\\_enbref.htm](http://www.larevueparlementaire.fr/pages/RP875/RP875_AN_enbref.htm).

of all candidates produced a report in 2009 which sums up the main arguments, as of 2008-2009<sup>6</sup>. The report is based upon a survey of 40 recruitment firms, 50% of which are favorable to anonymous resumes and 50% against. The arguments in favor of the practice include that it would lead to equal treatment for all candidates, would relieve minority candidates from the need to self-censor, would increase diversity in firms, and protect them from discrimination lawsuits. As for the arguments against, anonymous resumes are criticized for treating the symptoms not the disease, depriving candidates from expressing part of their identity, being easily bypassed through other recruitment channels, being costly to implement (for recruiters or recruitment firms) and hard to enforce. Interestingly, the report also details the debate on how anonymization should be done. It advocates erasing the candidate's name, age, country of citizenship, gender, marital status, and native language<sup>7</sup>.

The experiment that is the basis of this research was officially launched in a speech by President Sarkozy in December 2008, with an interesting mix of voluntarism and caution: "I want anonymous resumes to become a reflex for employers. We have long known that recruitment for French people is not egalitarian. Everybody must have a chance to benefit from an interview, to be able to state their motivation, to put forward their skills. On this topic, we will not move forward by constraint, but by conviction and pragmatism. (...) In 2009, the government will invite 100 large firms to implement anonymous resumes. Results will tell us whether we should go further."

The French public employment agency, Pôle Emploi, was put in charge of the experiment "with 100 large firms" which, in the final design detailed below, ended up involving several hundred large and medium firms. Pôle Emploi is the public agency which, until 2005, had a legal monopoly on the placement of unemployed job seekers and, at least in theory, a key role in labor market intermediation (all vacancies were supposed to be posted by the public employment services until

---

<sup>6</sup>In French: <http://alaingavand.typepad.com/files/rapport-cv-anonyme-a-comp%C3%A9tence-egale-10-juin-2009.pdf>.

<sup>7</sup>The debate is more open on whether to remove information on past experience (the report suggests limiting the work history to the past 10 years). The report also discusses whether the candidate himself should send an anonymous resume, or fill out a form that would include only appropriate fields, or whether anonymization should be performed by labor market intermediaries (such as recruitment firms or the public employment agency), or even if it should take place within the hiring firm.

2005).

### III. THEORETICAL BACKGROUND

To clarify the effect of anonymization, we consider a recruitment model in the spirit of Cornell and Welch (1996), involving two stages: resume screening and interviews.<sup>8</sup> Both resumes and interviews deliver independent binary signals,  $\eta$  and  $\theta$ . For instance, the signal can be thought as the work history displayed in the resume. When the candidate has some periods of inactivity, the signal could be interpreted as negative. Following models of screening discrimination, we consider two groups, a “majority” group ( $D = 0$ , with share  $\pi$  in the population), and a minority group ( $D = 1$ , with share  $1 - \pi$ ). The underlying productivity distribution  $y$  is the same in the two groups.<sup>9</sup> Individual ability to signal productivity differs from one group to another. Job seekers from the minority group are assumed to have less informative resumes than job seekers belonging to the majority. Formally we assume  $E[\eta = High|D = 0, y] = y$  in the majority group, but  $E[\eta = High|D = 1, y] = E[\eta = High] = 1/2$ . We consider however, that the quality of interview signals is the same in both groups. During interviews, recruiters observe the second signal ( $\theta$ ) which is equally informative in both groups:  $E[\theta = High|D = 0, y] = E[\theta = High|D = 1, y] = y$ .<sup>10</sup> Risk-neutral recruiters have rational expectations about candidates’ productivity and majority/minority distribution. In a first stage, recruiters screen resumes and decide whom to interview. In a second stage, which is costly, recruiters interview candidates, typically face-to-face, and decide to hire the best candidate.

Recruiters rank resumes according to candidates’ expected productivity. They decide to interview  $m$  candidates such that the marginal increase in expected productivity due to the last candidate called for an interview equals the cost per interview (constant and homogeneous across types). Intuitively, the number of interviews decreases with the cost per interview. It also depends on the composition of the initial pool of resumes. This composition affects the probability that the marginal candidate called for an interview actually wins the second-stage tournament during interviews and it also affects the corresponding increase in the productivity of the hired candi-

---

<sup>8</sup>The model is formally presented in the appendix.

<sup>9</sup>We implicitly assume that wages are constant and homogeneous across groups. Alternatively,  $y$  can be thought as the surplus, productivity net of wage.

<sup>10</sup>The discrimination behavior in this model can be labeled as statistical discrimination with heterogeneous signal quality.

date.<sup>11</sup> Recruiters may pay for interviews to compensate the low information content of minority resumes. In the standard procedure, recruiters observe the signals and minority/majority status of every candidate in the pool of resumes. They can thus interpret the signal depending on the type of candidate. They typically prefer majority candidates with good signals to minority candidates, and minority candidates to majority candidates with bad signals. When resumes are anonymous, recruiters only observe the signals of each candidate in the pool of resumes. As a consequence, majority and minority candidates with the same signals in their resumes have the same chance to be interviewed.

To gain insight on the impact of anonymization, we compute solutions of the recruiters' program for different values of the interview cost. We take a benchmark case inspired by our field experiment (see below). We fix the number of resumes to 8 and we solve the programs when there are 4 resumes with good signals and 4 with bad signals (the fraction of resumes with good signals is fixed). In such an environment there are 25 possible pools of resumes. The most frequent pool comprises 4 resumes from the majority (2 with good signals and 2 with bad) and 4 resumes from the minority (2 with good signals and 2 with bad). But the recruiter may also receive a pool of resumes exclusively from the majority. In the standard procedure, there are as many solutions as resume pools (25 in this case). We average outcomes over all possible pools. In the anonymous procedure, all resume pools are observationally equivalent and recruiters always have the same strategy.

Figure I shows the effect of anonymization on the number of interviews, the interview gap, the hiring gap and the productivity of the hired worker as function of cost per interview. The cost per interview is expressed as a percentage of maximal productivity  $y$ . In the reported simulations, it varies from almost 0 to 0.05. Annual labor cost at the minimum wage is 20,000 euros. If we consider this amount as a reference value for productivity, a 0.01 cost corresponds to 200 euros. When the cost is high, typically greater than half a percent of the maximal workers' productivity, anonymization increases the number of interviews. When the cost is lower, the number of interviews is hardly affected by anonymization (less than 25%). Anonymization may lead to a slight decrease in the number of interviews, when recruiters fear to interview some majority candidates with bad

---

<sup>11</sup>Let us compare two pools of candidates, the first composed of majority candidates with good signals, the second with only minority candidates. The number of interviews is lower in the first pool, because the chances that an extra candidate wins the tournament are lower (the chances that there is a top candidate among a few majority candidates with good signals are high). Besides, the increase in the productivity due to the hiring of a candidate with two good signals is lower in the first pool than among minority candidates whose resumes are uninformative.

signals.

INSERT FIGURE I HERE

Anonymization may increase or decrease the interview gap between minority and majority candidates; the effect depends on the cost per interview. The interview gap is mechanically equal to zero when resumes are anonymous. In the standard procedure, the interview gap is usually negative (minority candidates have a lower interview rate than majority candidates) and thus anonymization increases the gap, which is the case when the cost per interview is high. However, the interview gap in the standard procedure can also be positive when the cost is low. Then anonymization decreases the gap between minority and majority candidate. When the cost per interview is low, the only candidates who are not interviewed under the standard procedure are majority candidates with bad signals. This “reverse” discrimination phenomenon is a common feature of statistical discrimination models with heterogeneous signal quality when recruiters are not too picky (Aigner and Cain, 1977). When the interview cost is high, majority candidates with good signals are interviewed, while only a few minority candidates are interviewed. Then “usual” discrimination occurs (as found in typical correspondence studies).

In our simulations, anonymization always increases the hiring gap although how much depends on the cost per interview. For large values, the effect on the hiring gap is large: some minority candidates are interviewed because of anonymous applications; thus they may have a chance to be hired (this would surely be the case, for example, if the minority candidate ends up in interview pools with no other majority candidate). However, when the cost per interview decreases, two things occur: the number of applicants received in interviews increases in both the standard and anonymous procedures, and the effect of anonymization on the composition of the pool of applicants received in interview decreases. There are fewer applicants from the minority received in interview due to anonymization (and even more applicants from the majority received in interview when the cost of interviews is very low, i.e. when reverse discrimination would have taken place). The effect of anonymization is therefore weaker in this case. In the limit, when the cost is null, all the job seekers are received in interview under the two procedures and therefore the effect of anonymization on the hiring gap is zero.

Figure I also reports the effect on the productivity of the hired candidate. As can be seen,

there is no clear effect. It depends on whether recruiters adjust by paying for more interviews (to preserve the productivity of the hired candidate) or by hiring less productive candidates (to save on interview costs). Overall changes in hiring productivity are low in magnitude. We also computed the effect of anonymization on the firm’s profit, defined as the expectancy of  $y$  minus interview costs (not reported in the figure). As expected, the effect is negative. The loss of information in the first stage of the procedure is costly.

## IV. EXPERIMENT AND DATA COLLECTION

### IV.A. *Experimental design*

In this section, we present the experimental design used to measure the impact of anonymous resumes. The experiment was conducted over ten months in eight (out of 100) French *départements*, at branches of the public employment service (PES) located in urban areas. It proceeded as follows:

- a. *Firm entry in the experiment.* Firms posting vacancies at the PES have the option to ask for a PES agent to make a first screening of applicants based on their resume. In that case, the firm receives only selected resumes from the PES (from a couple to a dozen, in most cases), instead of having applicants contact them directly. This service is free. During the period of the experiment, all firms with more than 50 employees posting a job lasting at least three months and asking for this service are invited to enter the experiment. They are told that their vacant job will be randomly assigned to the anonymous or standard procedure, with probability 1/2. Firms are free to refuse; however, in order to induce positive responses, participation is presented as the default option. A given firm enters the experiment at most once: firms that have already entered the experiment are no longer asked to participate.<sup>12</sup>
- b. *Matching of resumes with vacant jobs.* The vacant job is posted by the PES on a variety of media, including a public website on which interested job seekers are asked to apply through the PES branch. On the webpage of the vacant job, one sentence informs job seekers that

---

<sup>12</sup>The main reason was the fear of the PES that repeated participation in the experiment and the corresponding surveys would have been too much trouble for firms. To maximize positive responses when inviting the firm to participate in the experiment, it was therefore clearly specified that the experiment would only concern the current job opening.

there is an experiment going on. The PES agent selects resumes from these applicants and from internal databases of job seekers. A first pool of resumes is thus matched with the vacant job.

- c. *Randomization and anonymization.* The resumes are sent to research assistants in charge of the randomization at the central PES offices. Job postings (and their first pool of resumes) are randomly assigned (using a random number generator) to treatment or to control group, with probability 1/2. If the offer is assigned to the treatment group, all the resumes are given a number and anonymized by the research assistant<sup>13</sup>; then they are sent back to the PES agent in charge of the vacant job follow-up.
- d. *Selection of resumes by the employer.* The employer selects the resumes of applicants she would like to interview. Control group employers contact the applicants directly; treatment group employers give the PES agent the resumes' numbers so that it is the PES agent who sets up the hiring interviews, in order to maintain applicant's anonymity.<sup>14</sup>

This experimental design calls for a few comments:

*Anonymization is limited.* Anonymization consisted in erasing the top part of the resume: name, address, gender, ID picture, age, marital status and number of children.<sup>15</sup> However, it did not imply any further standardization of the content of the resume. In particular, information on gender could be gathered from gender-specific terms used in the main part of the resume, neighborhood of residence could be partly inferred from information on where the applicant graduated from high school, and ethnicity could be spotted from foreign language skills. Going further would have implied much more complex logistics during the experiment, and it was felt that standardization would anyway not have been feasible if anonymous resumes had been made mandatory nationwide.

*Randomization occurs at the vacancy level.* For a given vacant job, all resumes transiting by the PES are treated identically (either anonymous, or standard). This level of randomization corresponds to the policy evaluated which would have all resumes anonymized, instead of some

---

<sup>13</sup>The degree of anonymization is described below.

<sup>14</sup>If the employer could not fill the position with the first pool of resumes, he/she requests additional pools. The PES sends a new pool of selected resumes with the same format as for the first pool. Our main analysis excludes the additional pools, as PES counselor knowing the format could have altered their selection.

<sup>15</sup>Job seekers could apply by sending a hard copy resume (pdf or doc) or fill in an online application form. As described below, only 10% of job seekers used the online solution.

anonymous resumes competing with standard resumes. However, the PES is not the only channel for recruitment: firms may also receive applicants from other sources, whose resumes are not anonymized. We measure below whether firms substitute these other channels for the PES in response to anonymization.

*Randomization occurs after matching resumes to vacant jobs.* Had the pre-screening of resumes by the PES occurred after randomization, the PES agent could have selected different applicants for vacant jobs with anonymous resumes (consciously or not). This would have affected the comparability of treatment and control applicants. To avoid this, a first pool of resumes was selected before randomization occurred. All analyses below are restricted to this first pool, as it contains resumes that are by construction statistically identical in the control and treatment group.

To summarize, the experimental design allows us to test one of the possible ways of implementing the law, in which anonymization of the resumes is performed by labor market intermediaries. Hence, the experiment misses a significant share of vacant jobs that the initial anonymization law aimed to include, as not all recruiters use the services of the public employment agency. Furthermore, the experiment was run in specific urban areas, and employers were allowed to opt out. Also, even for the vacant jobs entering the experiment, only applicants coming through the PES were concerned; the firms could keep using their other (non-anonymous) recruitment channels as well. As detailed below, the data collection strategy was adapted to assess the consequences of these features of the experiment.

During the 10 months of the experiment (November 2009 to September 2010), 1,005 vacant jobs entered the experiment out of a total of just over 6,000 eligible offers (each hiring firm could only participate once and counted for one offer). This low-entry rate into the experiment is due to losses at two steps. First, administrative data on all vacant jobs posted at the PES shows that only 25.5% of the eligible employers were invited to enter the experiment. It should be noted that the experiment took place at a time when inflows of job seekers were extremely large due to the recession, so that PES agents were very busy and some of them simply forgot or neglected to invite firms to participate. However, it is also likely that some PES agents preferred not to invite firms that they expected would refuse. Among firms invited to enter the experiment, the take-up rate amounted to 62.3%. Clearly, although only 37.7% of firms formally declined to participate, the

representativeness of the experimental sample is an issue, and it is analyzed in depth below. Among the 1,005 vacant jobs, some firms decided to stop their recruitment before the counselor had sent a first pool of resumes. In the end, and taking into account response rates, we can analyze around 600 vacant jobs split into a control and a treatment group.

#### *IV.B. Data collection*

We collect administrative and survey data. In addition to these two main sources, information available in the resumes was also coded. The administrative data covers all firms and all job seekers who used the public employment services in the experimental areas during (and after) the experiment. It includes basic information on the firm (size, sector), on the job position offered (occupation level, type of contract) and limited information on candidates (unless the candidate is filed as unemployed).<sup>16</sup> In what follows, the administrative data is mostly used to characterize the population of firms entering the experiment, compared with the broader population of firms interacting with the PES.

We conducted telephone interviews with all firms entering the experiment, as well as with a subsample of applicants to these firms. The data from these two surveys constitutes the core database used in the analysis. In addition, we interviewed a sample of firms that had refused to enter the experiment or had not been invited by PES agents despite the fact they were eligible for the experiment, and a subsample of applicants to these firms. Again, the goal is to check whether our core sample is representative of the target population of firms. Lastly, a subsample of applicants on vacant jobs from control group and non-participating firms *after the experiment* was also interviewed: as detailed below, the goal was to check whether control firms behaved differently because they knew they were in an experiment. The surveys used for applicants (respectively, for firms) were similar across subsamples. We will describe these two surveys briefly here, and present specific questions when they contribute to the subsequent analysis.<sup>17</sup>

The main goal of the survey of applicants is to provide a reliable measure of whether the applicant was interviewed for the job, and of all his/her characteristics that could lead to discrimination. We

---

<sup>16</sup>It also provides a follow-up of the recruitment process until the position is filled or the vacant job is withdrawn; however, the quality of that follow-up is weak, and some critical information is missing. (In particular, the data does not show whether the candidate was interviewed before the firm rejected his/her application).

<sup>17</sup>Survey tools are available upon request.

ask in particular for the country of birth and the citizenship at birth, both for the applicant and his/her parents.

The firm survey has two main functions. The first is to measure additional characteristics of the vacancy (that could be associated with discriminatory practices). Second, the survey includes detailed questions on the result of the recruitment, in particular when the recruitment was abandoned without filling the position, or when the hired candidate came from channels other than the PES (in which case he/she would not have been present in our survey of applicants). It measures the time to hiring and the match quality.

The initial population of applicants in the experiment (3,975 applicants) is partitioned in two ways: control vs. treatment; at risk of discrimination vs. other. At that stage, applicants at risk of discrimination are identified from the administrative information as those living in a deprived neighborhood or with an African or Muslim-sounding name. They are given higher sampling weights, in order to maximize statistical power. Table I details the sample of applicants. Overall, response rates are around 65-70%; even if they are lower in the control group, the difference is not statistically significant (the p-value is .27). The survey thus yields a total sample of 1,268 applicants, which constitutes our main sample of analysis. A subsample of candidates has resumes in pdf or doc formats that could be coded after the experiment.<sup>18</sup> The corresponding sample (1,140 candidates) is thoroughly analyzed in Section IV, below.

INSERT TABLE I HERE

There are five separate groups of sampled firms. 385 control and 366 treatment firms accepted the experiment and went through the randomization. 254 firms accepted the experiment but were not randomly assigned to treatment or control; they canceled or filled the job opening before a first pool of resumes was collected and randomization could take place. This underscores the fact that many firms actually fill their positions quickly without any help from the PES. 608 firms refused the experiment, and 4714 were not invited to participate. These last two groups of firms were sampled with lower sampling rates. Table I also presents the sample of firms. Their response rates are also somewhat lower, as could be expected. The response rate difference between control and treatment

---

<sup>18</sup>The other 10% filled an online application form that was not recorded during the experiment.

firms is not statistically significant.<sup>19</sup>

In addition to administrative and survey data, information available in the resumes was also coded. This is exactly the information set available to recruiters. All resumes (in pdf or doc formats) of the main sample of responding applicants were made anonymous after the experiment. This amounts to 1,140 resumes. We asked 16 PES counselors to act as recruiters and to code the resumes.<sup>20</sup> Anonymous resumes were randomly allocated to these PES recruiters, who then answered a dozen questions about the adequacy of the qualification, experience and education of the candidates. They were also asked whether the resume format was satisfactory (spelling mistakes, general appearance, etc.). Finally, they were asked to give an overall grade to the candidate and state whether they were certain or not about that grade (ambiguity).

#### *IV.C. Characteristics of firms and vacancies*

Table II presents descriptive statistics on firms and vacant jobs. Because of the experimental setting, firms in our sample recruit in the upper part of the labor market (statistics reported in column 7). Firms in the experiment are large: almost half have more than 200 employees. Most of them deliver services: 55% in the merchant sector and 24% in the non-merchant sector. Most firms in the experiment offer skilled positions with good contracts. Two thirds offer open-ended contracts, which is much higher than the average in employment listings in France (Cahuc and Postel-Vinay, 2002). Most jobs offered involve team work (81%) or frequent contact with customers (70%). Those characteristics are usually associated with taste-based discrimination. On average, recruiters receive around five resumes from the PES counselor.

INSERT TABLE II HERE

The first and second columns of Table II show that the balancing between control and treated firms is satisfactory. There are only two statistically significant differences (tests reported in the third column, labeled a-b): treated firms are less likely to post high-level jobs or jobs that involve teamwork. The fourth and fifth columns compare firms that counselors invited to participate and

---

<sup>19</sup>Sampling and response rates of candidates applying to firms that refused entry or were not invited are in Supplementary Table XIII. This table also displays sampling and response rates of candidates applying to vacant jobs posted after the experiment by the same participating or non-participating firms.

<sup>20</sup>The exact questions are available on request.

firms they did not (test of difference in the sixth column). There are some indications that firms invited to participate are less likely to belong to the non-merchant sector and that they post more skilled positions. There was some anecdotal evidence from the field that some counselors did not understand that non-merchant firms were also eligible (as long as the labor contract was standard). Columns 7 and 8 compare firms that accepted or refused to participate (among those that were asked to). Compared to accepting firms, refusing firms deliver more merchant services and post more jobs for workers with intermediate skills (skilled employees or skilled blue-collar workers).

In Table III, we compare interview rates of candidates applying to vacant jobs posted by firms with or without proposition to participate (panel A) and to vacant jobs posted by firms accepting or refusing the experiment (panel B).<sup>21</sup> To document selection, treated firms are excluded from both groups: firms with proposition and accepting firms. According to the first column, firms with and without any proposition to participate interview the same number of candidates (first and second line with difference in the third line). Neither are there any significant differences between the interview rates of minority (majority) candidates applying to firms with or without any proposition reported in the second column (third column). Recall that we define minority candidates as applicants living in deprived neighborhoods or immigrants or children of immigrants (see the next subsection for more details). The interview rate of minority candidates is 10.8% in firms with proposition, whereas that of majority candidates is 17.7%. The raw minority/majority gap is significant at the 10% level (fourth column). The difference in the minority/majority gap between firms with and without proposition, reported in the bottom right cell of the panel, is not statistically significant. In the lower panel, accepting and refusing firms interview the same number of candidates. However this hides considerable difference between the interview rates of minority and majority candidates. While the interview rates of minority and majority candidates are very similar in control firms accepting the experiment, the raw minority/majority gap is large in firms refusing the experiment (7.3% vs. 21%). The difference in raw minority/majority gap between refusing and accepting firms is thus large (14.3 points) and statistically significant at the 5% level.<sup>22</sup> To sum up, while the proposition margin (manipulated by the counselor) does not seem to select

---

<sup>21</sup>More precisely, we analyze vacant jobs during the experiment, but also vacant jobs after the experiment. Because the behavior of firms in those subgroups (control, refusing and not invited) does not evolve after the experiment (see below), it is acceptable to pool the sample over time.

<sup>22</sup>The difference persists when we control for firms' and candidates' characteristics. Results available upon request.

firms with particular discriminatory practices, the explicit refusal of some recruiters reveals potential discriminatory behaviors. This highlights the fact that we evaluate the impact of anonymous job applications on a specific population of firms, those complying with the policy. Recall that this may be the population of interest in the case that the program is extended, but not strictly enforced.

INSERT TABLE III HERE

#### *IV.D. Applicants' characteristics*

Table IV reports some descriptive statistics on candidates. Remember that initial candidate pools are exogenous, as randomization into treatment occurs after job counselors select candidates. As a consequence, we find no statistically significant difference between candidates applying to treated or control firms (first vs. second column), except in three dimensions (adequacy of skills and work experience and unusual foreign languages skills).<sup>23</sup> Half of the candidates are women, one third are younger than 26, and around 10% are over 50. Women and younger and older workers are disadvantaged groups on the labor market. In particular, there is some evidence of differential treatment of women in correspondence studies (Riach and Rich, 2002). However, women are not the main target group of the anonymization policy, and indications of the gender are likely to appear in the resume even when its upper part is erased. Age is also a characteristic that can be guessed – from the date of the last diploma obtained. As a consequence, in this paper we do not consider the effect of anonymization on the differential treatment based on gender or age, and instead focus on other discrimination dimensions. More precisely, we call “minority” individuals residing in deprived neighborhoods or having a foreign background. This is exactly the target group of the anonymization law that was implemented in response to the 2005 riots in Paris.<sup>24</sup>

INSERT TABLE IV HERE

---

<sup>23</sup>We also tested whether differential selection by the PES agent introduces systematic differences between applicants in the treatment and control groups for pools of resumes that were selected after randomization (as would be the case if the agent decided to over-select applicants at risk of discrimination for the anonymous procedure, for instance). There is, however, no evidence of this: control and treatment applicants remain comparable. More precisely, one does indeed note that resumes from the first pool differ from resumes of the subsequent pools, but the difference is the same for treatment and control vacant jobs. (Results omitted here.)

<sup>24</sup>We explored the impact of anonymous resumes on women as well and young or older workers. Introducing these dimensions interacted with the treatment in our estimations does not significantly alter our difference-in-difference results.

The first task at hand is how to characterize deprived neighborhoods of residence. In the US, Bertrand and Mullainathan (2004) use a variety of criteria based on the fraction of Whites, the fraction of college graduates, or the average per capita income. We use administrative classifications of neighborhoods defined to target subsidies or tax exemptions, since their boundaries closely match socioeconomic geographical disparities. Moreover, one of the alleged perverse effects of such administrative classifications is to create a stigma effect,<sup>25</sup> making them particularly relevant to assess the impact of anonymization.

The main issue is how to measure discrimination risk associated with foreign origin or ethnicity. French law forbids the use of ethnic categories that would label someone as White, Black, or African-French, for instance. Instead, we take a twofold approach. First, in the spirit of correspondence testing studies (Bertrand and Mullainathan, 2004), we code whether the applicants’ first names have a foreign-sounding origin. Following research by Felouzis (2003), we use the etymology of the applicant’s name. Muslim first names are identified from a database created by Chebira (2005). Second, we use the place of birth and the citizenship at birth. Immigrants are defined as those born abroad who did not have French citizenship at birth. Children of immigrants are those whose father was born abroad and did not have French citizenship at birth. Specific questions are used for the special case of individuals from former French colonies, who might declare they were French citizens at birth if they were born before independence; they are classified as foreigners if they took the citizenship of their new country at independence. The two approaches – based upon name or the migration status – are complementary. In some cases, a foreign-sounding name is the only signal that appears on the resume. But in other cases, immigrants may have a French-sounding name although their origin can be inferred from other signals on the resume (for instance, their last name or an ID picture).<sup>26</sup>

In Table IV, roughly one applicant in four lives in a deprived neighborhood; the same proportion has a Muslim or African-sounding name. One in five is an immigrant, and that proportion rises to two in five for immigrants or children of immigrants. The different measures of origin are

---

<sup>25</sup>They are known as “Zones urbaines sensibles” (ZUS) and “quartiers en contrat urbain de cohésion sociale (CUCS)”; these zoning schemes are comparable to “Enterprise zones” in the US.

<sup>26</sup>Alternative measures of origin include the applicant’s patronymic and his mother tongue. Measures using the applicant’s patronymic were hard to implement and did not seem, by cursory look at the resumes, to improve on the information yielded by the first name and the migration status. Moreover, in the French context, the mother tongue does not allow to capture immigrants well: according to Simon (1998), only 13% of second-generation Algerian youth declare their parents’ language as their mother tongue.

correlated. Of particular interest is the correlation between the name and the migration status: clearly, African or Muslim-sounding names correspond to applicants with a foreign origin (90%); however, a significant fraction of immigrants (including those from Africa) do not have African or Muslim-sounding names (45%). The variables based on immigration (as declared during the surveys) may better capture the risk of discrimination when that origin can be inferred from other signals in the resume. In the analysis, we compare the effects of using these alternative measures.

The lower panel of Table IV reports more detailed descriptive statistics on candidates. We analyze the subsample of candidates whose resumes were coded by job counselors.<sup>27</sup> As a result of randomization, there are very few statistically significant differences between the treatment group in the first column and the control group in the second. P-values are reported in the third. Skills of candidates in the treatment group are more in line with the job requirements. Their work experience is more often sufficient to take the vacant jobs. Candidates in the treatment group declare less often that they speak languages other than French, English, Spanish and German. Those languages are typical for majority candidates, as they are taught in French secondary schools.<sup>28</sup>

In Table IV, we also describe the resumes of the majority and minority groups (in the fifth and sixth columns, labeled c and d). Job counselors have rated resumes in their anonymous form. We thus interpret their ratings as a signal of candidates' productivity (independent from their minority or majority status). As the table shows, the overall picture is that the two populations are not so different. There are, of course, differences related to cultural background (location of education and work experience, languages). Also, the share of applicants holding a driving license is substantially lower in the minority group than in the majority, and a larger share of applicants from the minority are willing to work at the minimum wage. But the other characteristics are quite balanced across the two groups: frequency of inactivity periods, education and work experience relevant for the job, and overall rating given by the PES counselors. Last, the skills of minority candidates are only marginally more in line with the job's requirement. One should keep in mind that all this concerns a sample of candidates who have been pre-screened by PES agents. Differences between the minority and majority may be much larger in the whole population of candidates.

---

<sup>27</sup>In our main sample of 1,268 candidates, 128 candidates sent a resume generated by the PES website. Those standardized resumes have not been coded. Differences between the two subsamples are minor.

<sup>28</sup>We check below that the results are robust to controlling for these initial differences between treatment and control groups (table IX).

## V. MAIN RESULTS

To measure the impact of anonymous applications on interview and hiring rates of applicants from the minority, we consider a difference-in-difference analysis. We compute the change in the gap between interview (or hiring) rates of minority and majority applicants. If we consider a variable  $An \in \{0, 1\}$  corresponding to candidates applying to a vacant job with anonymous applications (treatment group) and a variable  $D \in \{0, 1\}$  corresponding to the candidate being from the minority group, the parameter of interest can be written as:

$$(1) \quad \delta = \left( \bar{Y}^{An=1, D=1} - \bar{Y}^{An=1, D=0} \right) - \left( \bar{Y}^{An=0, D=1} - \bar{Y}^{An=0, D=0} \right)$$

where  $Y$  is the variable measuring the fact that the applicant was received in interview, or was hired. It can also be measured using the usual difference regression equation:

$$(2) \quad Y_{ij} = \alpha_0 + \alpha_D D_i + \alpha_{An} An_j + \delta D_i \times An_j + \varepsilon_{ij}$$

where candidates are indexed by  $i$  and vacant jobs by  $j$ . Parameter  $\alpha_0$  measures the interview or hiring rate of job seekers from the majority group applying to a standard vacant job. Parameter  $\alpha_D$  is a measure of the gap in interview or hiring rates between job seekers from the minority and majority groups applying to standard vacancies. Parameter  $\alpha_{An}$  is a measure of the difference in interview or hiring rates of job seekers from the majority group when they apply to an anonymous vacant job compared with when they apply to a standard vacant job. This parameter is therefore a measure of the effect of anonymous applications on interview or hiring rates.

### *V.A. Interviews rates*

We first contrast the interview rate of candidates applying to control vacant jobs with that of candidates whose resumes were sent in anonymous form to recruiters. Due to the experimental design, this comparison can be interpreted as causal. Table V shows the interview rates for different samples (first line for the control group and second line for the treatment group). The average interview rate in the control group is around 10.5%. Recruiters are quite selective when they screen

resumes.<sup>29</sup> The average interview rate in the treatment group is 11.3%. The difference between the two interview rates is not statistically significant (standard errors are clustered at the vacant job level). This means that the average number of interviews per vacant job does not increase when resumes are anonymous.

INSERT TABLE V HERE

In column 2 of Table V, we report interview rates for the minority group. The interview rate drops from 9.3% in the control group to 4.7% in the treated group. The difference is statistically significant at the 5% level. Minority workers are actually harmed by the anonymization procedure. In column 3, the interview rate of majority workers increases from 11.6% to 17.7% due to anonymization. This increase is not statistically significant. Both effects imply that the gap in interview rates between minority and majority workers actually widens (column 4). The difference-in-difference estimate, reported in the third line of column 4, amounts to 10.7 percentage points and is statistically significant at the 5% level, implying that anonymous resumes led to an increase in the interview gap against minority candidates – the exact opposite impact to what policymakers aimed at.

Both results (on the average interview rates and on the minority majority gap) can be understood in the context of our theoretical model when the interview cost is low<sup>30</sup>. In the standard procedure, recruiters interview majority candidates with good signals and all minority candidates irrespective of their signals. Then anonymization prevents recruiters from distinguishing between minority and majority candidates among resumes with bad signals. The gap in interview rates between minority and majority candidates widens, while the number of interviews remains constant. According to this interpretation, there is reverse discrimination in the standard procedure.

Interestingly, the gap in interview rates in the standard procedure is small (−2.4 percentage points) and not statically significant. Of course, this cannot be interpreted as evidence that there is little discrimination (or of any type of differential treatment of candidates according to their type),

---

<sup>29</sup>Candidates from the PES represent a fraction of the whole pool of candidates, so that the average number of interviewed candidates is more than  $5 \times 10.5\%$ .

<sup>30</sup>The low interview rates that we observe (slightly around 10%) may seem to contradict the idea that interview costs are low: in the theoretical model's simulations, low interview costs lead to interview rates well above 50%. However, it should be noted that these high interview rates result mechanically from the simplifying assumption that resumes only contain a binary signal (see figure I where interview rates remain above 25% whatever the interview cost). Allowing for more heterogeneity across resumes could explain why recruiters screen a larger share of candidates based on their resumes, even when interview costs are low.

as the effect of the candidate’s type may be confounded by differences in the resumes’ credentials (observed by the recruiter and not by the econometrician). We explore this issue further in the next section, taking advantage of the fact that we have access to the resumes, which in principle allows us to code all the signals received by recruiters.

### *V.B. Hiring rates*

Table V also reports hiring rates for minority and majority candidates. The difference between the first and second line can be interpreted as causal, because these are unconditional hiring rates.<sup>31</sup> The hiring rate of minority workers is 2.3% in the standard procedure. It decreases to 1.7% in the anonymous procedure. The difference is not statistically significant. The hiring rate of majority workers increases from 2.1% in the control group to 5.2% in the treatment group. Again the difference is not significant. However, the gap in hiring rates between minority and majority candidates, which is not statistically different from 0 in the standard procedure, is statistically significant in the anonymous procedure. The difference-in-difference estimate,  $-3.7$  percentage points, is statistically significant at the 10% level. This suggests that the negative relative impact of anonymous resumes on minority candidates persists until the recruitment stage, even though it is less precisely estimated. This finding goes against the argument of opponents to anonymous applications who argue that anonymous resumes will only postpone discrimination behavior to later stages of the recruitment, when the applicant’s type is revealed. It confirms the non-experimental evidence found by Goldin and Rouse (2000) and Aslund and Skans (2012).

Our evidence on interview rates is consistent with our theoretical model when the interview cost is low. So our theoretical model predicts that there should be hardly any effect on hiring rates. When the interview cost is low, many majority candidates are interviewed, and it is very likely that one of them at least will display a good signal during the interview and be preferred over candidates who are interviewed because of anonymization. Our empirical results on hiring rates show that effects persist to the hiring stage. This contradicts our theoretical model when interview cost is low.

---

<sup>31</sup>Difference between hiring rates computed on the population actually interviewed can be biased as anonymization changes the composition of the pool of interviews.

### *V.C. Recruitment success*

Thanks to the recruiters' survey, we have proxy measures for the quality of the candidate eventually hired ( $y$  in the theoretical model). Each new contract starts with a trial period (for up to three months). We asked recruiters whether the recruitment was confirmed at the end of the trial period and whether they are satisfied with the performance of the hired candidate over that period. In the bottom panel of Table VI, we report the impact of anonymous resumes on the corresponding variables. Around 80% of recruiters hire the candidate past their trial period and recruiters rate candidates on average 7 (on a 1-10 scale). The anonymous procedure does not significantly affect indicators for the quality of the match.

INSERT TABLE VI HERE

The wage paid to the hired candidate is another potential measure of  $y$ , insofar as it is positively related to the surplus generated by the job. Wages paid to new hires do not seem to be affected by anonymous resumes; the same fraction of hired candidates are paid at the minimum wage in the treatment and control groups. For those paid above minimum wage, the median wage is around 1,700 euros in both groups.

Lastly, the first two panels of Table VI show the time needed to fill the vacancy, and the share of vacancies that remain unfilled. This may provide indirect evidence on whether anonymous resumes affected the firms' profits. We do not find any effect, however. Four recruitments out of five are successful in both treatment and control groups, and half of successful recruitments lasted less than 48 days, whatever the procedure.

The absence of match quality effects is consistent with our theoretical model when the interview cost is low. As explained above, anonymization of resumes does not affect hiring when the interview cost is low, as differential treatment is only postponed to later stages.

### *V.D. Robustness checks*

We perform a variety of tests to check the robustness of our key results to potential threats to the experimental design, as well as to alternative measurement and model specifications.

*A John Henry effect?*

As noted below, a possible issue with the experimental design is the fact that participating firms knew that they were part of an experiment. This in itself could affect their behavior. The risk is particularly acute for control firms: they know they were observed; they also know the identity of the applicants. They may therefore artificially select more minority applicants in order to signal to the PES that they do not discriminate. This type of effect is known as a “John Henry” effect, by which the control group makes extra efforts to perform well. Such an effect could explain why treated firms appear less favorable to minority applicants: the negative difference would not be due to a negative impact of anonymous resumes on treatment firms, but to the positive impact of monitoring firms in the control group.

Our strategy to test for the presence of such an effect is to look at control firm hiring behavior *after* the experiment. The idea is the following: randomization ensures that control and treatment firms are comparable, but *during the experiment*, control firms may change their behavior, to the point that they are not a valid counterfactual. However, unless having once been part of an experiment (without being treated) has surprisingly lasting effects, control firm behavior should not be distorted after the experiment. We therefore ran the survey of applicants on a subsample of applicants to 148 vacant jobs that were posted by control firms after the experiment.<sup>32</sup> The test for the presence of a John Henry effect is very simple: we ask whether the interview and hiring gaps between applicants of different groups were different before and after the experiment.

INSERT TABLE VII HERE

Table VII shows no evidence of a John Henry effect. If anything, control firms were more, rather than less, favorable to minority applicants when the experiment stopped. Of course, a John Henry effect may persist over time (once firms know they have been observed, they are durably more cautious); but one would expect the effect to subside. This is not the case.<sup>33</sup>

*Imperfect anonymization?*

Another issue with the experiment is imperfect anonymization. The treatment consists of re-

---

<sup>32</sup>Note that relying on applicants for information on interview and hiring decisions made by the firm removes any concern that firms become aware that we continue observing them.

<sup>33</sup>Instead of focusing on vacant jobs posted by control firms after the experiment, one could have looked at vacant jobs posted before the experiment that would be fully exempt from any John Henry effect. However, this turns out not to be feasible. Indeed, administrative information being insufficient, we would need to run a survey of applicants on these past vacant jobs. These surveys would often occur with a significant delay – to identify control firms, one needs to wait for them to enter into the experiment! – which would create memory bias among respondents.

moving the upper part of the resume, leaving its main part unaltered. As a consequence, one may spot minority candidates from three characteristics on the resumes: languages skills in Arabic, foreign education, or work experience abroad. One may think of such imperfect anonymization as a case where the treatment is only implemented on a subset of resumes, meaning that the effect of the full treatment (perfect anonymization) is *underestimated*. In that case, the effect we observe is only due to candidates with perfectly anonymized resumes (the treatment makes no difference for resumes that are imperfectly anonymized). However, the response of recruiters to imperfectly anonymized resumes may be more subtle: they could be suspicious about candidates whom they perceive as trying to hide their ethnicity, and avoid interviewing them. They may even feel *more* free to discriminate them, as they could claim that they did not know the identity of the candidate. In that case, the adverse impact of anonymization could be artificially large on imperfectly anonymized resumes, so that the effect of full anonymization would be *overestimated*.

To shed light on this issue, we test in Table VIII whether the impact of anonymous resumes on the interview rates of minority candidates changes with the presence of information that allows the reader to infer that the candidate belongs to the minority, even when the resume is anonymous. We do not find any differential effect. A caveat is that interaction effects are imprecisely estimated. The point estimates, however, remain very small compared to the direct effect of anonymous applications. This is consistent with the view that recruiters do not fully use the information that is available to them to infer the candidates' type/productivity, a finding that is reminiscent of Oreopoulos (2011).

INSERT TABLE VIII HERE

#### *Alternative measurement of minority status*

In our main analysis, the applicant's background enters in a quite specific way, exerting the same effect for being of foreign background (immigrant or child of immigrant), residing in deprived neighborhoods, and cumulating the two characteristics. We decomposed our minority population into three sub-populations: being of foreign background and residing in a deprived neighborhood, being of foreign background and not residing in a deprived neighborhood, and residing in a deprived neighborhood but not being of foreign background. The estimated impacts on both interview and hiring gaps for these three populations were found homogeneous.<sup>34</sup>

---

<sup>34</sup>Results are shown in the Supplementary Table XIV of the online appendix.

It is also not obvious how foreign background should be measured. Names are removed from anonymous resumes, so that this may be the relevant measures. However, coding whether family names denote a foreign background is not immediate. And even if first names and surnames do not denote a foreign background, a picture ID might. In that case, measuring foreign background with the applicant’s migration status may be more relevant. We tested three possible alternative measures of foreign background – being an immigrant, being the child of an immigrant, or having a Muslim or African-sounding first name. They yield similar results to our preferred measure (which groups immigrants and children of immigrants). Point estimates, however, tend to be lower, and the effect is no longer statistically significant when considering only Muslim or African-sounding name.<sup>35</sup>

*Other specification issues*

Table IX displays additional robustness checks on interview rates and hiring rates. In column 1, we report our main results (estimation of equation 2). In column 2 we add as covariates the whole set of variables listed in Tables II and IV. The difference-in-difference parameter on interview rates is marginally lower; it is still significant though at the 10% level. In column 3, we check whether removing sampling weights makes a difference. The coefficient on  $T \times D$  becomes smaller and marginally significant only. This may be due the fact that, among applicants from a foreign background, applicants with a Muslim or African-sounding name have been oversampled (this was the only information on foreign background available at the time of sampling). The lower point estimate suggests that the negative effect of anonymous resumes could be smaller on that group. The difference, however, is far from significant. In column 4, we introduce vacant job fixed effects. This restricts the estimation of the  $T \times D$  coefficient to vacant jobs where minority and majority candidate actually compete for the same job. The point estimate is very close to our baseline difference-in-difference estimate. This means that ethnic segmentation is not such an issue in our experiment. Actually 49.5% of candidates apply to mixed vacant jobs with both minority and majority candidates. In column 5, we check that using a logit model rather than a linear probability specification does not affect the results. Last, we try to document the influence of selection in the experiment on the treatment effect. We estimate a propensity score based on the covariates in Table 2 and weight the observations with the inverse of the acceptance probability in the last column.

---

<sup>35</sup>Results are shown in the Supplementary Table XV of online appendix.

Treatment effects are very close to the baseline estimation. This is no surprise, as there are not much difference in observable characteristics of firms accepting and firms refusing to participate in the experiment. Indeed selection seems to occur on unobservables: there are much more differences in interview rates gap between accepting and refusing firms.

INSERT TABLE IX HERE

Overall, the experiment yields four main findings on the impact of anonymous resumes (compared to the standard procedure):

- a. Recruiters interview the same number of candidates.
- b. Recruiters call relatively more majority candidates to interviews.
- c. Recruiters hire relatively more majority candidates.<sup>36</sup>
- d. The productivity of the hired candidates and other measures of the recruitment success are not impacted.

In what follows, we focus on the interpretation of the first two facts, for two reasons. First, these are the results for which our analysis has stronger statistical power. Second, we can hope to use the same information set as the recruiters when it comes to modeling the decision to interview candidates, as the only information that recruiters have on candidates comes through resumes that we coded thoroughly.

## VI. HOW ARE RESUMES VALUED UNDER STANDARD AND ANONYMOUS PROCEDURES?

To further interpret the effects of anonymization on interview rates, we use the detailed information coded from the resumes (PES counselors coded all the resumes in an anonymous format). We examine how signals in the resumes affect the interview decision. We document that some signals, such as interruptions in the work history, are not valued equally in minority and majority groups' resumes. Recruiters appear to update rationally the value they attribute to each of these

---

<sup>36</sup>The difference is significant at the 10% level only.

signals when resumes are anonymous. Uniformization of value attached to each item between the two groups induces large changes in the way resumes are evaluated globally. We show that this accounts for a large share of the change in the interview gap between the two groups.

The additional evidence provided in this section is more tentative than the key results above: it is not directly based on the experimental design, so that there may be unobserved confounding factors driving estimates that we tend to interpret causally. The fact that we have access to the same source of information as recruiters (namely, the resumes) mitigates this concern, but there may remain coding and specification issues.

### VI.A. *Econometric model*

We denote by  $X$  the information set available to the recruiter in the resume, besides minority status ( $D$ ). The recruiter screens resumes based on  $X$  when the resume has been treated anonymously, and based on  $(X, D)$  in the standard procedure. We can then define potential valuations of resumes associated to the standard and anonymous procedure when individual  $i$  applies to vacant job  $j$  as follows:

$$(3) \quad V(St)_{ij} = (1 - D_i) \times X_i \beta^0(St) + D_i \times X_i \beta^1(St) + D_i \alpha(St) + c(St)_j + \nu_i$$

$$(4) \quad V(An)_{ij} = X_i \beta(An) + c(An)_j + \nu_i$$

where  $\beta(An)$  is the vector of values of characteristics  $X$  in the anonymous procedure and  $\beta^0(St)$  and  $\beta^1(St)$  are the two vectors of values in the standard track for majority and minority candidates. The returns  $\beta(An)$ ,  $\beta^D(St)$  are assumed homogenous across recruiters. However, we allow for vacancy fixed effects  $c(St)_j$  and  $c(An)_j$  whose distribution may be affected by the treatment.<sup>37</sup> Introducing fixed effects is possible given the fact that there are several resumes for each job offer. It is important to introduce such fixed effects as some job seekers could be assigned to job offers because of a specific knowledge the caseworker could have about recruiters' preferences. One important assumption for the consistency of estimates is that the  $X$  characteristics we consider involve all the information available to the recruiters. The  $X$  characteristics we consider include almost all "intermediate" items introduced in the grid used to code resumes (adequate skills, adequate work experience,

---

<sup>37</sup>When  $X_i$  is unidimensional and binary, the above valuation equation can represent the theoretical model introduced above (then  $V(\cdot)$  is the expected productivity of the worker conditional on the binary signal).

inactivity period, information content of the resume).<sup>38</sup> We also include the overall rating made by the caseworkers evaluating the resumes. Doing so allows to capture all relevant information missing in our grid.<sup>39</sup>

If characteristics are rationally valued by recruiters then we expect the parameter in the anonymous equation to be a weighted mean of the two parameters in the standard procedure:  $\beta(An) = \beta^0(St)P(D = 0|X) + \beta^1(St)P(D = 1|X)$ . However, as highlighted in Bertrand et al. (2005), the way characteristics are valued can be a complicated process, especially when it regards discrimination issues. It may be context-dependent and it is a deep context change that occurs when switching from the standard to anonymous procedure. Some characteristics can see their valuation deeply changed in one context compared to another.

We consider that the interview decisions  $I(\cdot)_{i,j} = 1$  are given by a simple decision rule  $I(\cdot)_{i,j} = \mathbb{1}(V(\cdot)_{ij} > s(\cdot)_j)$  where  $s(\cdot)_j$  are vacant job-specific thresholds.<sup>40</sup> To keep interpretation simple and estimation transparent and easily linked to results in the previous section, we consider linear probability models. We therefore estimate simple models of the form:

$$(5) \quad E(I(St)_{i,j}|X_i, D_i, j) = (1 - D_i) \times X_i\beta^0(St) + D_i \times X_i\beta^1(St) + D_i\alpha(St) + c(St)_j$$

$$(6) \quad E(I(An)_{i,j}|X_i, D_i, j) = X_i\beta(An) + c(An)_j$$

## VI.B. Results

Table X presents the estimation results of models 5 and 6. Columns 1 and 2 present the estimation result for the valuation of resumes in the standard recruitment procedure, respectively for majority and minority candidates. The third column presents the p-value for the test that the coefficients of columns 1 and 2 are equal. The fourth reports the estimation for resumes in the anonymous procedure (pooling minority and majority candidates). The last column reports the results of the test of equality of the three previous estimated coefficients.

INSERT TABLE X HERE

---

<sup>38</sup>Adequate education, driving license and attractiveness of the resume were excluded because of potential multi collinearity.

<sup>39</sup>As a robustness test, we verify in Supplementary Table XVI that excluding this summary rating does not affect our estimation result.

<sup>40</sup>Note that this rule is simpler than that of the model developed in the theoretical section.

The different variables in the information set  $X$  are derived from the coding of resumes by PES counselors (see section *IV.D.*). The overall rating summarizes the relevance of the applications (see descriptive statistics in Table IV). We construct a dummy for best applicants, rated A and B (vs. C to E). The return for the minority group is significant (at the 10% level), its magnitude is quite large: having a good grade increases the chance of majority candidates to be interviewed by 7.4 percentage points. Even though the return is not significant for the majority group, its magnitude is similar (7.8 percentage points).

Job counselors were asked whether they were certain/confident in their rating of each candidate. Their answer is a proxy for the level of information conveyed by the resumes. They were told to declare their rating “uncertain” when they found they had insufficient information to unambiguously rate the candidate. We find no statistically significant effect of the “uncertainty” variable in the minority and majority groups. The estimates are not significantly different.

We summarize labor market history by a dummy indicating whether the candidate has some periods of inactivity in his resume.<sup>41</sup> The effect of such unexplained inactivity is statistically insignificant in the majority and the minority groups. However the parameter estimate is negative for the majority group and positive for the minority group, so that their difference is statistically significant (at the 10% level). This is suggestive evidence in favor of the statistical discrimination model with heterogeneous signals. It also corroborates the findings by Kroft et al. (2012) that recruiters attach less importance to unemployment spells when they can be explained by adverse labor market conditions (in our case, high unemployment rates in deprived neighborhoods).

Job counselors were asked whether the resumes showed evidence that the candidate had adequate skills to perform tasks advertised in the vacant job and whether the candidate had already performed such tasks in previous work experience. The return to adequacy in skills is large in magnitude and significant at the 5% level for majority candidates: having adequate skills increases the interview rate by 12% percentage points (this is the largest point estimate in the model). The return is smaller and not statistically significant for minority candidates. As a consequence, the difference in returns is significant at the 10% level. Lastly the returns to adequate work experience are low and not statistically significant.

---

<sup>41</sup>Excluding periods of inactivity that are explained by education or maternity leave.

Conditioning on the resumes' signal and on vacant jobs, the minority effect ( $\alpha$ ) amounts to  $-0.014$  points. It is not significant and similar to the unconditional minority/majority gap estimated in our main analysis (see above).

In standard screening discrimination model, the returns to signals when resumes are anonymous should be a weighted average of the returns in the minority and majority groups when the procedure is standard. We find some evidence in line with this prediction when we consider the return to unexplained inactivity. The average of the returns between minority and majority candidates is around  $-0.025$  ( $= -0.09 \times 0.5 + 0.04 \times 0.5$ , where  $0.5$  is the actual fraction of minority and majority candidates), which is close to the actual return estimated in the anonymous sample ( $-0.04$  in the column 4). Some minority candidates are actually harmed by anonymization, because they have an interrupted labor market history. However, differences between returns in the anonymous sample and the nominative (meaning, name-bearing) sample are not statistically significant when we consider unexplained inactivity or other signals, with the important exception of "uncertainty." The return to uncertainty is significantly negative in the treated sample (anonymous procedures), while coefficients were positive in the control group (standard procedure, both for minority and majority candidates). This goes against the screening discrimination model. In that dimension, anonymous resumes do not seem to be read as a weighted average of nominative resumes.

### *VI.C. Decomposing the overall effect on interview rates*

We have shown that minority candidates are adversely affected by anonymization because some negative signals in their resumes, such as inactivity, are not attenuated as they are in the standard procedure. Minority candidates can also be adversely affected because recruiters reject candidates with low-information resumes when resumes are anonymous. If the resumes of minority candidates are less informative than those of the majority candidates, the former would be more adversely affected by anonymization than the latter. To quantify the extent of the effects going through signal extraction, we perform a decomposition exercise. We use estimates reported in the previous table to decompose the change in the interview gap. Let us consider first anonymous resumes. The

gap in interview rates can be expressed:

$$\begin{aligned}
Gap(An) &= E(I_{ij}|An = 1, D = 1) - E(I_{ij}|An = 1, D = 0) \\
&= (E(X_i|An = 1, D = 1) - E(X_i|An = 1, D = 0)) \beta(An) + \\
&E(c_j|An = 1, D = 1) - E(c_j|An = 1, D = 0) + \\
&E(\nu_{ij}|An = 1, D = 1) - E(\nu_{ij}|An = 1, D = 0)
\end{aligned}$$

There are therefore three components. The first is the effect of differences in endowments of characteristics  $x$ , valued using the estimated  $\beta$ 's. The second is the vacancy effect: job seekers from the minority may apply to vacant jobs that in average receive more (or less) candidates than vacant jobs to which candidates from the majority are applying. The last component is the contribution of the residuals: there may be characteristics missing that are correlated with minority.

A similar decomposition can be produced for the standard procedure. The difference is, however, that the valuation of the various items entering the information set of the recruiter can be made group-specific.

$$\begin{aligned}
Gap(St) &= E(I_{ij}|An = 0, D = 1) - E(I_{ij}|An = 0, D = 0) \\
&= \alpha + (E(x_i|An = 0, D = 1)\beta^1(St) - E(x_i|An = 0, D = 0)\beta^0(St)) + \\
&E(c_j|An = 1, D = 1) - E(c_j|An = 1, D = 0)
\end{aligned}$$

There are several differences in this decomposition compared to the previous one. First, the effect of belonging to the minority ( $\alpha$ ) directly enters the decomposition; second, the returns to each item are allowed to differ between minority and majority candidates; and last, there is no residual effect. All these changes reflect the fact that the recruiter knows the minority/majority type of the candidate.

Results are reported in Table XI. The first column reports elements of the decomposition of the interview gap under anonymous procedure, the second presents the decomposition of the interview gap under the standard procedure, and the third gives the decomposition of the difference-in-difference. For each variable, as well as for the vacant job fixed effect, we compute a net effect by subtracting column 2 from column 1.

INSERT TABLE XI HERE

The first line provides the two interview gaps as well as the difference between them. Closely related to the previous results presented in Table V, the difference-in-difference to explain is a widening of 7 points.<sup>42</sup> The contributions of the different signals to the minority/majority gap in the anonymous sample are low (column 1). They are all negative and below 0.5 percentage points, so that their sum amounts to  $-0.7$  percentage points. This is not surprising, as there are few differences in the signals distribution between minority and majority candidates (see Section IV.D descriptive statistics on candidates). The contribution of vacant job fixed effects is larger, almost 7 percentage points. This may reflect the self-selection of minority candidates (who tend to apply to jobs offers where interview rates are low) or selection by the PES counselors in charge of the initial matching of vacant jobs with candidates. Lastly, the contribution of residuals is lower, but still important:  $-2.4$  percentage points. Even if PES counselors coded the resumes that were sent to recruiters, some signals correlated to the minority status are still missing from our analysis.

In contrast to the decomposition in the anonymous sample, the contributions of the different signals in the nominative sample (column 2) are high: the total amounts to 4.7 percentage points (to be compared to an average interview rate of about 10%). This mainly reflects the fact that returns to signals differ between the majority and minority candidates. For example, the attenuation of the negative signal conveyed by periods of inactivity leads to an increase by 3.9 points in the interview rates of minority candidates. Vacancy fixed effects have the same contribution in the nominative sample than in the anonymous sample, and the contribution of residuals is by construction equal to 0 in the nominative sample. The contribution of the constant  $\alpha$  (the minority effect in the table) is low:  $-1.4$  percentage points.

As a consequence, it appears in column 3 that the main component of our difference-in-difference estimate is due to signal extraction, more precisely to the uniformization of returns between majority and minority groups following anonymization. Out of the  $-7$  points of our difference-in-difference estimate,  $-5.3$  can be explained by difference in signal valuation. The contribution of vacant job fixed effects vanishes as it is the same in both nominative and anonymous samples. Finally, a relatively

---

<sup>42</sup>The difference between our estimate in the previous section and this estimate is due to the restriction of the sample from 1268 to 1140 individuals. We exclude candidates who did not send their resumes as PDFs or docs, but went through an online application form on the website.

small fraction of our difference-in-difference estimate is left unexplained. The contribution of the residual (-2.4) is low and may reflect unobserved heterogeneity. The contribution of the constant 1.4 may also reflect signals unobserved to the econometrician, or it might capture homogeneous reverse discrimination (arising in standard taste-based discrimination models).

#### VI.D. *Reshuffling the cards: redistributive effects*

We are interested in the changes in the potential value of each individual’s resume under the anonymous and standard procedures. This includes both the value parameters associated to each characteristic  $X$  and the firm-specific effect. One problem in computing the expected value of each of these potential values is the firm-specific effects which are estimated under either the anonymous or the standard procedure. We are, however, mainly interested in the value of resumes relative to other candidates. We can therefore circumvent the previous problem by considering within-firm values. We define the predicted potential values of resumes of individual  $i$  applying to vacant job  $j$

$$(7) \quad \widehat{V}(An, X_i, D_i, j)_{i,j} = W_j(X_i) \widehat{\beta}(An)$$

where  $W(X_i)$  is the within component of characteristic  $X_i$  relative to vacant job  $j$  and  $\widehat{\beta}(An)$  is the estimated value of coefficient  $\beta(\cdot)$  in equation 6. Similarly, we define for potential value under the standard procedure:

$$(8) \quad \widehat{V}(St, X_i, D_i, j)_{i,j} = W_j((1 - D_i) \times X_i) \widehat{\beta}^0(St) + W_j(D_i \times X_i) \widehat{\beta}^1(St)$$

One way to look at the redistributive effect of the anonymous procedure is to look at the correlation between the two predicted potential values. Table XII presents these correlations – in the whole sample but also in the minority and majority population. As the table shows, the correlation is rather weak. In the whole sample it is 0.24 but when considering the two specific populations we see that it is only 0.38 even among the majority and even weaker 0.12 for the minority population. The table also presents p-values for the test of perfect correlation which are performed bootstrapping the whole estimation procedure. The assumption is clearly rejected in the whole sample but also in

both populations. The change implied by the anonymous procedure on the values and consequently ranking of resumes has therefore deep redistributive implications both between and within the two populations. To illustrate the magnitude of the effects, we compute the fraction of the candidates who are ranked first in one procedure but not in the other. This fraction is substantial given the fact that we have information to rank on average about two candidates per vacant job (around 23% with standard errors around 2 points)<sup>43</sup>.

INSERT TABLE XII HERE

To sum up, the anonymous procedure does not only widen the gap between majority and minority candidates; it also substantially alters the relative positions of candidates within either of these two groups.

## VII. CONCLUSION

We have analyzed the first large-scale randomized experiment on the effects of anonymous applications. Among 600 firms advertising their vacant jobs through the French public employment agency, 50% received resumes where the upper block was suppressed. In anonymous resumes, recruiters could not read names, addresses, nationality, or other identifying items. Policymakers thought this manipulation would reduce hiring discrimination against minority candidates (residing in deprived neighborhood or immigrants or child of immigrants). The estimated effect is exactly the opposite. Results show that minority candidates are actually harmed by anonymous applications. The gap in interview rate between minority and majority candidates widens by around 10 points when resumes are made anonymous. The effect seems to persist beyond resumes' screening, as the hiring gap also widens by 4 points (only significant at the 10% level). Interestingly, anonymous resumes do not make recruiters interview more candidates, and the overall quality of the job match does not seem to be affected.

An important qualification of these results is that they pertain to a selected sample of firms that were using the public employment services and did not refuse to take part to the experiment. Our surveys allow us to document the behavior of these firms, which seem more favorable to minority

---

<sup>43</sup>See Supplementary Table XVII in the online appendix.

candidates than the average firm. Hence, our results do not allow us to predict what would be the average impact of making anonymous resumes mandatory for all hirings. This may be less of a problem with the evaluation than it appears, for two reasons. First, if the law had actually been enforced, it remains unclear which firms would have used anonymous referral procedures, as it would have depended upon the monitoring capacities of the administration or the deterring power of judges. In a scenario with stringent enforcement, all firms would comply; but with less stringent implementation, it may well be the case that the pool of complying firms would not differ much from the pool of firms that entered the experiment. Second, even if one should not extrapolate them to the whole population of firms, the unexpected results we find are a clear warning against making anonymous resumes universally mandatory. Locally, at least, this would be counterproductive.

This finding was disturbing enough for policymakers that the 2006 law was never enforced, and was finally abolished in 2010 following the public release of our evaluation report. Our take on this is that, even though we do not want to claim that the law would have had a negative average impact, the results strongly suggest that policymakers would have needed a much clearer enforcement strategy in order to reach positive effects.

While anonymous resumes are certainly no magic bullet for policymakers, the experiment turns out to be a particularly useful setting to illuminate how recruiters extract information throughout the hiring process. Analyzing the actual signals conveyed by resumes, we find suggestive evidence that recruiters value signals in minority and majority resumes differently. For example, periods of inactivity are interpreted as negative signals when the candidate is from the majority, whereas they have no impact for minority candidates. This is consistent with standard models of screening discrimination. Overall, differences in signal valuation benefit minority candidates when resumes show names, and these differences are quantitatively important, as they increase the interview rate of minority applicants by almost 5 points. But we also find suggestive evidence that recruiters do not necessarily update their beliefs about candidates' quality in a rational way. When receiving anonymous resumes, they put more weight on signals that they would not consider when reading name-bearing resumes of minority and majority candidates. Even though this second mechanism is quantitatively less important to understand the impact of anonymous resumes, it is in line with several papers in the correspondence testing literature showing that standard models of discrimi-

nation and signal extraction cannot fully explain recruiters' behavior (Bertrand and Mullainathan, 2004; Oreopoulos, 2011; Jacquemet and Yannelis, 2012). Understanding this mix of rational and non-standard decision making is a promising direction for future research.

PARIS SCHOOL OF ECONOMICS - INRA

CREST

CREST

## References

- Aigner, Dennis J. and Glen G. Cain, “Statistical theories of discrimination in labor markets,” *Industrial and Labor Relations Review*, 30 (1977), 175–187.
- Aslund, Olof and Oskar Nordstrom Skans, “Do anonymous job application procedures level the playing field?” *Industrial and Labor Relations Review*, 65 (2012), 82–107.
- Autor, David H. and David Scarborough, “Does Job Testing Harm Minority Workers? Evidence from Retail Establishments,” *The Quarterly Journal of Economics*, 123 (2008), 219–277.
- Bertrand, Marianne, Dolly Chugh, and Sendhil Mullainathan, “Implicit Discrimination,” *American Economic Review*, 95 (2005), 94–98.
- Bertrand, Marianne and Sendhil Mullainathan, “Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination,” *American Economic Review*, 94 (2004), 991–1013.
- Cahuc, Pierre and Fabien Postel-Vinay, “Temporary jobs, employment protection and labor market performance,” *Labour Economics*, 9 (2002), 63–91.
- Chebira, Abderrazak, “L’indispensable manuel orthographique des prénoms français d’origine arabe et musulmane,” Technical report, Ed APIC (2005).
- Cornell, Bradford and Ivo Welch, “Culture, Information, and Screening Discrimination,” *Journal of Political Economy*, 104 (1996), 542–71.
- Edin, Per-Anders and Jonas Lagerström, “Blind dates: quasi-experimental evidence on discrimination,” Working paper series, IFAU - Institute for Evaluation of Labour Market and Education Policy (2006).
- Felouzis, Georges, “La ségrégation ethnique au collège et ses conséquences,” *Revue Française de Sociologie*, 44 (2003), 413–447.
- Goldin, Claudia and Cecilia Rouse, “Orchestrating impartiality: the impact of Blind auditions on female musicians,” *American Economic Review*, 90 (2000), 715–741.

- Jacquemet, Nicolas and Constantine Yannelis, “Indiscriminate discrimination: A correspondence test for ethnic homophily in the Chicago labor market,” *Labour Economics*, 19 (2012), 824–832.
- Krause, Annabelle, Ulf Rinne, and Klaus Zimmermann, “Anonymous job applications in Europe,” *IZA Journal of European Labor Studies*, 1 (2012a), 1–20.
- Krause, Annabelle, Ulf Rinne, and Klaus F. Zimmermann, “Anonymous Job Applications of Fresh Ph.D. Economists,” *Economics Letter*, 117 (2012b), 441–444.
- Kroft, Kory, Fabian Lange, and Matthew J. Notowidigdo, “Duration Dependence and Labor Market Conditions: Theory and Evidence from a Field Experiment,” Working Paper 18387, National Bureau of Economic Research (2012).
- Oreopoulos, Philip, “Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes,” *American Economic Journal: Economic Policy*, 3 (2011), 148–71.
- Persico, Nicola, “Racial Profiling? Detecting Bias Using Statistical Evidence,” *Annual Review of Economics*, 1 (2009), 229–54.
- Riach, P. A. and J. Rich, “Field Experiments of Discrimination in the Market Place,” *Economic Journal*, 112 (2002), 480–518.

Table I: Sample size and response rates in the applicant and firm surveys

	Population size (a)	# sampled for survey (b)	# of respondents (c)	# of resumes coded (d)	Sampling rate (b/a)	Response rate (c/b)	Coding rate (d/c)
	A. Applicant survey						
Control	2035	1003	660	586	0.49	0.66	0.89
Treatment	1930	954	608	554	0.49	0.64	0.91
Total	3965	1957	1268	1140	0.49	0.65	0.90
	B. Firm survey						
Control	385	385	229		1.00	0.60	
Treatment	366	366	212		1.00	0.58	
Refused the experiment	608	335	146		0.55	0.44	
Not invited	4714	542	281		0.12	0.52	

**Notes:** The upper panel displays population and sample size among applicants to vacant jobs entering the experiment. The first two lines distinguish applicants according to whether resumes were anonymous (treatment) or not (control); the last line presents the total size. The lower panel displays population and sample size among vacant jobs eligible for the experiment (one vacant job per plant). The first two lines display vacant jobs handled with standard resumes (control vacant jobs) or anonymous resumes (treatment vacant jobs). The last two lines correspond to plants that refused the experiments or that were not invited to participate, despite the fact they were eligible. Supplementary Table XIII displays samples of candidates applying to vacancies outside the experiment (refusing or not invited firms or vacancies posted after the experiment).

Table II: Firms' and vacant jobs' characteristics

	Treatment		Diff-test		Invited		Not invited		Diff-test		Accepted		Refused		Diff-test	
	a	b	p-value	a-b	c	d	p-value	c-d	e	f	p-value	e-f				
Administrative data																
Firm with less than 100 employees	30.9	29.0	0.56	30.7	28.5	0.13	30.0	31.6	0.52							
Firm with 100 to 200 employees	15.1	15.6	0.85	15.9	16.7	0.47	15.3	16.6	0.52							
Firm with more than 200 employees	54.0	55.5	0.69	53.4	54.8	0.39	54.7	51.8	0.28							
Non-merchant services	24.7	23.2	0.64	22.9	30.2	0.00	24.0	21.5	0.29							
Merchant services	55.6	54.4	0.74	57.5	54.9	0.08	55.0	60.7	0.03							
Manufacturing	13.8	16.9	0.23	13.6	8.5	0.00	15.3	11.5	0.04							
Construction	3.4	3.6	0.90	3.2	3.4	0.73	3.5	2.8	0.48							
Upper occupations	9.9	6.3	0.07	6.7	5.4	0.08	8.1	4.9	0.02							
Intermediary occupations	24.4	26.0	0.63	23.3	21.1	0.09	25.2	20.9	0.06							
Skilled white or blue collar	55.3	58.7	0.34	60.0	58.8	0.46	57.0	63.7	0.01							
Unskilled white or blue collar	10.4	9.0	0.53	10.1	14.7	0.00	9.7	10.5	0.63							
Indefinite duration contract	66.5	62.6	0.26	63.7	59.5	0.00	64.6	62.7	0.47							
Contract for more than 6 months	86.0	82.2	0.16	83.6	83.3	0.79	84.2	82.9	0.53							
Nb of resumes sent by the PES	5.5	5.5	0.90													
Nb of obs.	385	366		1 359	4,714		751	608								
Firm survey																
Involves teamwork	85.2	75.4	0.01	81.1	84.2	0.26	80.4	83.0	0.50							
Frequent customer contact	71.9	67.9	0.41	70.2	75.8	0.09	70.1	70.4	0.95							
Nb of obs.	229	212		587	281		441	146								

**Source:** PES administrative file (upper panel) and firm survey (lower panel). All eligible vacant jobs.

**Notes:** First three columns present mean values of each variable on the population of firms assigned to treatment and control as well as the p-value of the test of identity. Next three columns compare the population of firms invited to participate in the experiment or not. The last three columns compare firms accepting to participate or not (among invited firms). There is no comparable data for the number of resumes sent by the PES to firms outside the experiment (outside the experiment there is no distinction between the first and following pools of resumes).

Table III: Interview rates in firms inside and outside the experiment

	All	Minority (D)	Majority (ND)	Gap (D-ND)
A. Firms with vs. without proposition to participate				
Proposition (p)	0.145	0.108	0.177	-0.069*
(control and refusing firms)	(0.021)	(0.019)	(0.031)	(0.037)
No proposition (n)	0.119	0.103	0.132	-0.029
	(0.028)	(0.031)	(0.037)	(0.041)
Difference (p-n)	0.026	0.005	0.045	-0.040
	(0.035)	(0.036)	(0.050)	(0.055)
Nb of candidates	1684	932	752	1684
Nb of vacant jobs	629	462	455	629
B. Accepting vs refusing firms				
Accepting (control firms) (c)	0.143	0.146	0.140	0.006
	(0.015)	(0.021)	(0.031)	(0.031)
Refused to participate (r)	0.146	0.073	0.210	-0.137**
	(0.038)	(0.028)	(0.059)	(0.063)
Difference (r-c)	-0.004	0.073**	-0.070	0.143**
	(0.041)	(0.035)	(0.063)	(0.070)
Nb of candidates	1378	759	619	1378
Nb of vacant jobs	507	374	376	507

**Source** : Candidates' survey. **Note** : This table displays interview rates of candidates applying to firms inside and outside the experiment. Panel A analyzes the effect of the counselors' decision to propose the experiment: firms with proposition are in the first line and firms without in the second. The difference between the mean interview rates of firms with proposition and without proposition is in the third line. Firms with proposition include firms that accept the experiment and are drawn in the control group and firms that refuse the experiment. Panel B analyzes the recruiters' decision to accept to participate: control firms which represent accepting firms (first line of panel B) and refusing firms (second line). Again, the difference in interview rates between the two types of firms is in the third line. In the first column minority and majority candidates are pooled together, while the second column restricts to minority candidates (D) and the third to majority candidates (ND). In the last column, we compute the difference between minority and majority candidate. Thus, the difference in the minority/majority gap between firms with and without proposition can be found in the third line, last column. Survey sampling weights are used. Standard errors are clustered at the vacant job level. We pool candidates applying during and after the experiment to different vacant jobs posted by eligible firms. For control firms, we pool candidates applying before or after randomization on the same vacant job. For those two reasons, the pool of candidates analyzed on control firms is different from that of our main sample of analysis. However, the minority/majority gap in interview rates is the same as in our main sample of analysis. More details on the sample characteristics can be found in Supplementary Table XIII.

Table IV: Candidates' characteristics

	Treatment	Control	Diff-test p-value	Majority	Minority	Diff-test p-value
	a	b	a-b	c	d	c-d
Candidates' Survey						
Women	0.488	0.520	0.531	0.568	0.435	0.001
Less than 26 years old	0.299	0.294	0.885	0.276	0.318	0.281
More than 50 years old	0.093	0.134	0.156	0.102	0.123	0.451
Deprived neighborhood (1)	0.241	0.224	0.557		0.484	
Immigrant (2)	0.218	0.215	0.926		0.449	
Child of immigrant (3)	0.157	0.169	0.661		0.337	
(1), (2) or (3)	0.474	0.492	0.649			
African or Muslim-sounding name (4)	0.225	0.228	0.922	0.017	0.451	0.000
At least upper secondary education	0.747	0.718	0.498	0.753	0.713	0.256
Long-term unemployed	0.281	0.340	0.122	0.306	0.311	0.890
Reservation wage is min wage	0.569	0.606	0.404	0.506	0.669	0.000
Resumes coding						
Very good candidate (A)	0.152	0.178	0.401	0.183	0.145	0.180
Good candidate (B)	0.291	0.336	0.247	0.330	0.294	0.333
Fair candidate (C)	0.187	0.211	0.515	0.160	0.237	0.035
Bad candidate (D)	0.258	0.198	0.101	0.237	0.222	0.695
Very bad candidate (E)	0.112	0.077	0.130	0.090	0.101	0.619
Uncertainty in rating	0.530	0.542	0.790	0.512	0.560	0.261
Adequate skills	0.480	0.619	0.000	0.532	0.559	0.512
Adequate work experience	0.538	0.634	0.018	0.605	0.560	0.272
Adequate education	0.637	0.647	0.803	0.638	0.646	0.844
Driving license	0.588	0.554	0.416	0.635	0.507	0.004
Inactivity periods	0.324	0.336	0.751	0.337	0.322	0.702
Attractive resume	0.432	0.448	0.692	0.438	0.441	0.940
Foreign education	0.146	0.138	0.789	0.081	0.205	0.000
Foreign work experience	0.194	0.188	0.868	0.146	0.238	0.007
English speaking	0.564	0.566	0.955	0.527	0.604	0.065
Unusual foreign languages	0.179	0.113	0.020	0.050	0.249	0.000
Arabic speaking	0.072	0.063	0.617	0.001	0.137	0.000
Nb of observations	586	554		636	504	

**Source:** Candidates' survey and resumes coding. **Notes:** First three columns present mean values of each variable on the population of candidates applying to vacant jobs assigned to treatment and control as well as the p-value of the test of identity. The three next columns compare majority and minority candidates in the experiment.

Table V: Impact of anonymous applications on interview and hiring rates gaps between minority and majority applicants

	All	Minority (D)	Majority (ND)	gap (D-ND)
Interview rates				
Standard (c)	0.105 (0.016)	0.093 (0.017)	0.116 (0.026)	-0.024 (0.031)
Anonymous (t)	0.113 (0.017)	0.047 (0.011)	0.177 (0.030)	-0.130*** (0.032)
Effect (t-c)	0.008 (0.020)	-0.046** (0.020)	0.061 (0.040)	-0.107** (0.045)
Nb of candidates	1268	696	572	1268
Nb of vacant jobs	598	418	385	598
Hiring rates				
Standard (c)	0.022 (0.006)	0.023 (0.008)	0.021 (0.009)	0.002 (0.012)
Anonymous (t)	0.035 (0.010)	0.017 (0.007)	0.052 (0.017)	-0.035** (0.018)
Effect (t-c)	0.013 (0.011)	-0.006 (0.010)	0.031 (0.019)	-0.037* (0.021)
Nb of candidates	1268	696	572	1268
Nb of vacant jobs	598	418	385	598

**Source:** Candidates' survey. **Notes:** The upper panel considers interview rates, the lower panel considers hiring rates. For the upper panel, for example, the first column gives the interview rate in standard vacant jobs (first line), in vacant jobs with anonymous applications (second line) and the difference between the two (third line). The second column (D) does the same computations but for applicants from the minority while the third column (ND) does it for applicants from the majority. The last column considers the difference between column (D) and column (ND). In this table, the coefficient  $\alpha_0$  of equation (2) can be found on the first line, column (ND); the coefficient  $\alpha_D$  can be found on first line, last column; the coefficient  $\alpha_{An}$  can be found on third line, column (ND); and the coefficient  $\delta$  can be found on third line, last column. Survey sampling weights are used. Standard errors are clustered at the vacant job level.

Table VI: Impact on the recruitment process and match quality (from the recruiters' point of view)

	Standard application	Treatment effect	Nb of vacant jobs
Recruitment process status			
Recruitment canceled	0.14*** (0.02)	-0.02 (0.03)	441
Recruitment in progress	0.08*** (0.02)	0.03 (0.03)	441
Recruitment completed	0.78*** (0.03)	-0.01 (0.04)	441
Time to hiring			
Mean in days	48.1*** (4.4)	-1.7 (5.6)	307
First quartile (in days)	19.0*** (2.0)	1.0 (2.9)	307
Third quartile (in days)	76.0*** (6.6)	-13.0 (8.2)	307
Match quality			
Successful trial period	0.82*** (0.03)	-0.02 (0.05)	240
Recruiter's satisfaction (1-10 scale) about early tasks	7.32*** (0.16)	0.06 (0.22)	220
Hired candidate paid the minimum wage	0.22*** (0.04)	0.01 (0.05)	268
Median wage (except min wage earners)	1715*** (57)	-15 (78)	177

**Source:** Firm's survey. **Notes:** The upper panels of the table provide information about the recruitment status and its duration. The lower panel provides information about the characteristic of the match on completed recruitment. In each row, we regress one dependent variable in the recruiter survey on the treatment dummy. Thus the first column gives the mean of vacant jobs with standard applications while the second column gives the difference between vacant jobs with anonymous applications compared to vacant jobs with standard applications. The number of observations in the lower panel can be lower than the recruitment success rate time the number of total vacant jobs, because of partial non-response.

Table VII: Interview rates in control firms during and after the experiment

	Minority (D)	Majority (ND)	Gap (D-ND)
During the experiment (c)	0.123 (0.028)	0.137 (0.029)	-0.014 (0.039)
After the experiment (a)	0.166 (0.033)	0.145 (0.036)	0.021 (0.051)
Difference (a-c)	-0.043 (0.044)	-0.008 (0.048)	-0.035 (0.065)
Nb of candidates	442	365	807
Nb of vacant jobs	220	218	296
Nb of firms	134	132	146

**Source:** Candidates' survey. **Notes :** The table as the same difference-in-difference structure as Table V. It only involves control firms which received standard resumes during the experiment. It compares interview rates for vacant jobs during the experiment, first line (i.e. the experimental vacant jobs analyzed in Table V) and vacant jobs posted later (second line). Only control firms that resubmitted a vacant job to the public employment services in the months following the experiment are included (explaining the minor difference with Table V). Survey sampling weights are used. Standard errors are clustered at the vacant job level. More details on the sample characteristics can be found in Supplementary Table XIII.

Table VIII: Impact of anonymous applications on minority applicants, by quality of anonymization

	Interview rate				
Intercept	0.096*** (0.017)	0.112*** (0.023)	0.107*** (0.020)	0.108*** (0.020)	0.099*** (0.019)
Anonymous (T)	-0.046** (0.020)	-0.051* (0.028)	-0.051** (0.024)	-0.048* (0.025)	-0.045* (0.023)
Resume imperfectly anonymized (C)		-0.044 (0.035)			
T x C		0.008 (0.041)			
Has studied abroad (C)			-0.053 (0.039)		
T x C			0.022 (0.045)		
Has worked abroad (C)				-0.049 (0.039)	
T x C				0.011 (0.045)	
Speak Arabic (C)					-0.033 (0.044)
T x C					0.001 (0.047)
vacant job controls	Yes	Yes	Yes	Yes	Yes
Observations	696	597	617	636	649

**Source:** Candidates' survey. **Notes :** This table restricts the sample to minority candidates. In the first column, we estimate the average treatment effect controlling for vacant job characteristics (note that there are almost no difference with estimates in table V, where we do not control for any covariates). In the following column, we introduce one-by-one some characteristics of the resumes to explore treatment effect heterogeneity. In column 3 to 5, the characteristics analyzed are easy to read from the resume. In the second column, we compute a score of minority status based on all signals available in the resume. When the score is over a certain threshold, the resume is said imperfectly anonymized. Survey sampling weights are used. Samples are smaller in columns 2 to 5, because not all resumes were available as hard copy and, as already discussed, coding is thus partial. Standard errors are clustered at the vacant job level.

Table IX: Robustness analysis: different specifications

	Baseline	With controls	Without sampling weight	vacant jobs fixed effect	Logit	With participation weights
Interview rates						
Intercept	0.116*** (0.026)	0.114*** (0.027)	0.135*** (0.020)		-0.357*** (0.087)	0.114*** (0.025)
Minority	-0.024 (0.031)	-0.010 (0.034)	-0.028 (0.026)	0.027 (0.036)	-0.023 (0.030)	-0.021 (0.031)
Anonymous (T)	0.061 (0.040)	0.056 (0.042)	0.058* (0.033)		0.045 (0.030)	0.060 (0.039)
T × minority	-0.107** (0.045)	-0.089* (0.049)	-0.073* (0.040)	-0.105** (0.053)	-0.086*** (0.025)	-0.106** (0.044)
Nb. of obs.	1268	1140	1268	1268	1268	1268
Hiring rates						
Intercept	0.021** (0.009)	0.013 (0.009)	0.035*** (0.087)		-0.523*** (0.431)	0.023** (0.010)
Minority	0.002 (0.012)	0.011 (0.013)	-0.002 (0.014)	0.020 (0.021)	0.002 (0.014)	.0005 (0.013)
Anonymous (T)	0.031 (0.019)	0.038* (0.020)	0.022 (0.017)		0.025 (0.016)	0.030 (0.020)
T × minority	-0.037* (0.021)	-0.039* (0.022)	-0.026 (0.021)	-0.057* (0.031)	-0.024** (0.011)	-0.037* (.022)
Nb. of obs.	1268	1140	1268	1268	1268	1268

**Source:** Candidates' survey. **Notes:** The table provides different estimations of equation 2. The "Baseline" column considers estimation of the equation using sampling weights. It provides the reference results also reported in the difference-in-difference Table V. The second column adds the whole set of variables listed in Tables II and IV as control variables; it thus restricts the sample to candidates whose resumes were coded. The third column estimates equation 2 without sampling weights. The fourth column adds vacant job fixed effects. The fifth column reports marginal effects of a logit estimation of equation 2. The last column weights observations by the inverse of the probability to accept the experiment. Survey sampling weights are used. Standard errors are clustered at the vacant job level.

Table X: Effects of different elements of the resume on the interview decision

	Standard application		Test	Anonymous	Test
	Majority	Minority	p-value	All candidates	p-value
	$\beta^0(St)$	$\beta^1(St)$	$\beta^0(St) = \beta^1(St)$	$\beta(An)$	$\beta^0(St) = \beta^1(St) = \beta(An)$
	(1)	(2)	(3)	(4)	(5)
Minority effect ( $\alpha$ )		-0.014 (0.090)			
High overall rating	0.078 (0.066)	0.074* (0.038)	0.961	0.044 (0.044)	0.853
High uncertainty	0.0636 (0.055)	0.076 (0.047)	0.862	-0.098** (0.050)	0.024**
Has inactivity periods	-0.090 (0.060)	0.038 (0.046)	0.070*	-0.044 (0.040)	0.156
Adequate skills	0.120** (0.056)	0.044 (0.045)	0.304	0.032 (0.040)	0.428
Adequate work experience	-0.056 (0.059)	0.007 (0.049)	0.416	0.031 (0.035)	0.444
Nb of candidates	252	334		554	
Nb of vacant jobs		283		270	

**Source:** Candidates' survey and resumes' coding. **Notes :** We estimate the effects of resumes' signals on the interview rate in a model with vacant job fixed effect. Standard errors are clustered at the vacant job level. Columns 1 and 2 show returns to signals  $x$  when resumes bear names (estimation of  $I_{ij} = (1 - D_i) \times X_i \beta^0(St) + \alpha D_i + D_i \times X_{ij} \beta^1(St) + c_j + \nu_{ij}$ , where  $i$  indexes candidates,  $j$  indexes vacant jobs and  $D$  indicates minority status), and column 4 when resumes are anonymous (estimation of  $I_{ij} = X_i \beta(An) + c_j + \nu_{ij}$ ). Column 1 concerns majority candidates (results for  $\beta^0(St)$ ), column 2 minority candidates (results for  $\alpha$  and  $\beta^1(St)$ ). In column 3, we report the p-value of the test of equality in returns between columns 1 and 2. In column 5, we report the p-value of the test of equality between columns 1, 2 and 4. For example, when nominative resumes show periods of inactivity, the interview rate of majority candidates decreases by 9 points and that of minority candidate increases by 3.8 points.

Table XI: Decomposition of interview gaps between minority and majority candidates

	Anonymous (a)	Standard (n)	Difference-in-difference (a)-(n)
Raw interview gap	-0.098*** (0.033)	-0.028 (0.027)	-0.070* (0.042)
Minority effect (constant)	0.000 (0.000)	-0.014 (0.090)	0.014 (0.090)
Signals/Characteristics ( $X$ ):			
High overall rating	-0.005 (0.005)	-0.003 (0.037)	-0.002 (0.037)
High uncertainty	-0.001 (0.004)	0.010 (0.036)	-0.011 (0.036)
Has inactivity periods	-0.000 (0.002)	0.039* (0.022)	-0.039* (0.022)
Adequate skills	-0.001 (0.002)	-0.037 (0.037)	0.036 (0.037)
Adequate work experience	-0.000 (0.002)	0.037 (0.045)	-0.038 (0.045)
<i>Subtotal over X</i>	<i>-0.007</i> <i>(0.008)</i>	<i>0.047</i> <i>(0.08)</i>	<i>-0.053</i> <i>(0.08)</i>
Vacancy fixed effects	-0.067** (0.026)	-0.061** (0.031)	-0.006 (0.039)
Residuals	-0.024 (0.017)	-0.000 (0.000)	-0.024 (0.017)

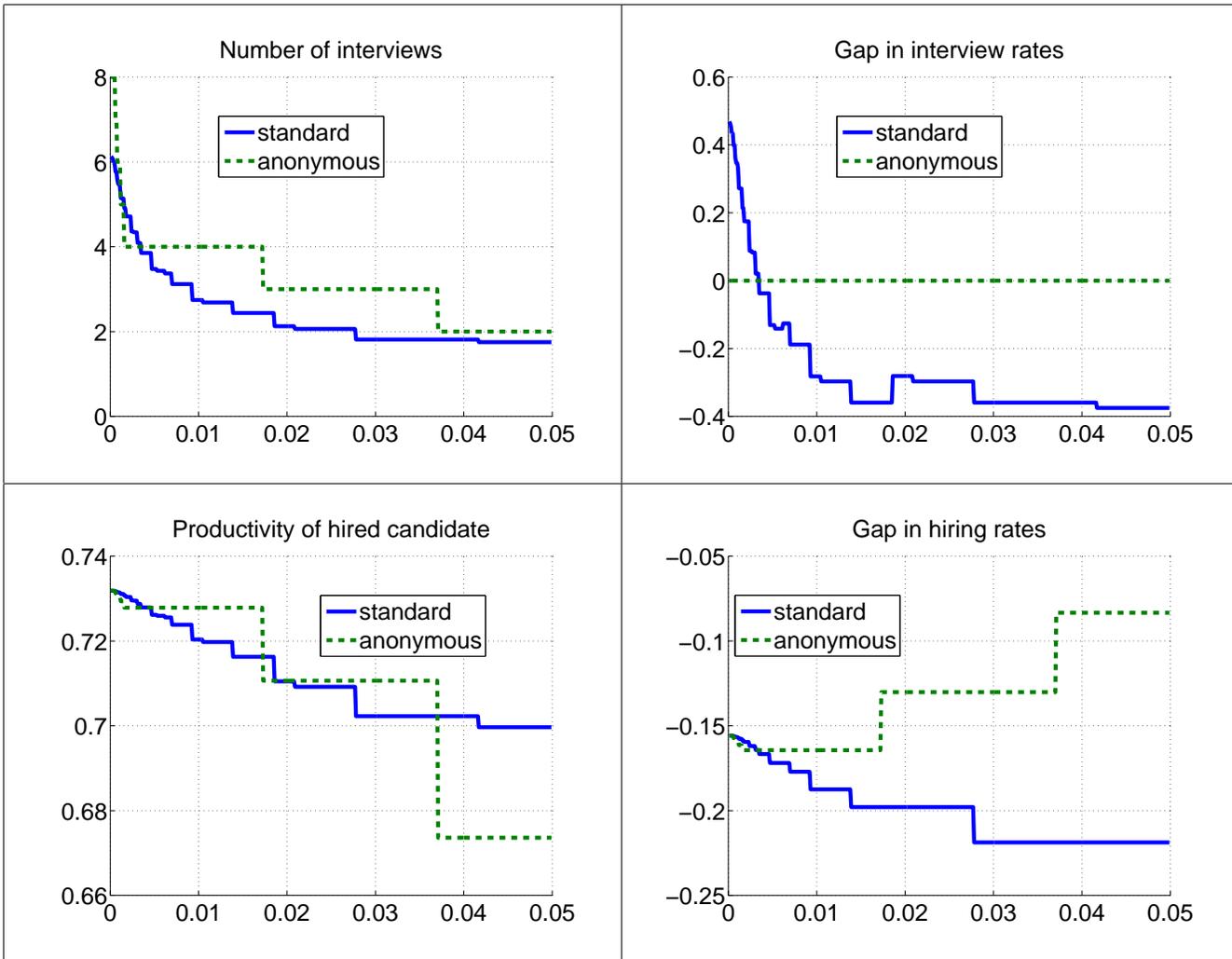
**Source:** Candidates' survey and resumes' coding. **Notes:** We report the minority/majority gap in interview rates when resumes are anonymous in column 1 and when resumes are nominative in column 2. In column 3, we compute the change in minority/majority gap due to anonymization. The raw gap in interview rates (line 1) is decomposed between contributions due to difference in signal valuations (from line 3 to 7), difference in type of vacant jobs and unexplained difference (line 2 and last line). Contributions are obtained from estimates in Table X. To obtain standard errors, we bootstrap the decomposition 500 times clustering at the vacant job level.

Table XII: Correlation of resume valuations under anonymous and standard procedure

Sample	Correlation	Standard error	P-value
	$\rho$		$\rho = 1$
Whole Sample	0.24	(0.22)	0.001
Majority	0.38	(0.28)	0.013
Minority	0.12	(0.25)	0.002

**Source:** Candidates' survey and resumes' coding. **Notes:** We compute the correlation of resume valuations predicted with the anonymous model and predicted with the standard model. Models used to predict valuations are those estimated in Table X. To obtain standard errors, we bootstrap 500 times the whole procedure (estimation and predictions of resume valuations) clustering at the vacancy level. P-values are for the test of perfect correlation.

Figure I: Model simulation



**Notes:** Each figure provides the result of model simulation. Upper left panel gives the number of interview, upper right panel gives the interview rate gap between minority and majority type workers, the lower panel left gives the quality of the hire and the lower panel right the hiring rate gap between minority and majority type workers. The horizontal axis is the cost of interview as a maximal productivity. By comparison with annual cost at the minimum wage level, 1% would represents roughly 180 euros.

## A ANALYTICAL APPENDIX

This appendix solves the formal model developed in section 2. Recall that this is a screening/statistical discrimination model with two stages: resume screening and interviews. The productivity of the candidate ( $y$  homogenous across minority and majority groups) is not perfectly observed. The recruiter receives two binary signals, one at each stage  $\eta$  and  $\theta$ . The resumes' signal is not informative when the candidate is from the minority ( $D = 1$ ). Formally recall that:

$$E[\eta = High | D = 0, y] = y$$

$$E[\eta = High | D = 1, y] = 1/2$$

$$E[\theta = High | D = 0, y] = E[\theta = Low | D = 1, y] = y$$

In the standard procedure recruiters observe resumes' types and signals. Their information set can be summarized as  $(\bar{n}_0, \bar{n}_1, \underline{n}_1, \underline{n}_0)$  where  $\bar{n}_0$  ( $\bar{n}_1$ ) is the number of resumes of majority (minority) candidates with a good signal and  $\underline{n}_0$  ( $\underline{n}_1$ ) is the number of resumes of majority (minority) candidates with a bad signal. In the anonymous procedure, recruiters only observe signals. Their information set is  $(\bar{n}, \underline{n})$  where  $\bar{n} = \bar{n}_0 + \bar{n}_1$  and  $\underline{n} = \underline{n}_0 + \underline{n}_1$ . We perform our analysis in a restricted environment where  $(\bar{n}, \underline{n})$  is fixed across resumes' pool. The recruiter is certain to receive a fixed number of good and bad resumes. However the types of the good resumes and bad resumes may vary from one pool to the other. This restriction simplifies our computation and mimics what happens in our empirical experiment. Based on the information about the resumes, recruiters choose to interview a pool of candidates. The cost per interview,  $c$ , is independent from the candidate's type. The objective of the recruiter of course is to hire the best candidate internalizing the interviews costs.

We solve the model backward:

- a. what is the optimal hiring given a set of interviewed candidates?
- b. what is then the optimal set of candidates to call for an interview ?

We compute the expected productivity of each candidate during the interview (when all information is known). Those expectations are the same in both procedures.

$$E[y|D = 0, \eta = H, \theta = H] = 3/4$$

$$E[y|D = 1, \theta = H] = 2/3$$

$$E[y|D = 0, \eta = H, \theta = L] = E[y|D = 0, \eta = L, \theta = H] = 1/2$$

$$E[y|D = 1, \theta = L] = 1/3$$

$$E[y|D = 0, \eta = L, \theta = L] = 1/4$$

Note that obviously  $E[y|D = 0, \eta = H, \theta = H] > E[y|D = 1, \theta = H] > E[y|D = 0, \eta = H, \theta = L] > E[y|D = 1, \theta = L] > E[y|D = 0, \eta = L, \theta = L]$

After interviews, recruiters rank the candidates in the following preferred order: (i) majority candidates with two high signals, (ii) minority candidates with high interview signal, (iii) majority candidates with mixed signal, (iv) minority candidates with low interview signal, (v) majority candidates with two bad signals. Denote  $\bar{m}_0$  ( $\bar{m}_1$ ) the number of interviews of majority (minority) candidates with a good resume signal and  $\underline{m}_0$  ( $\underline{m}_1$ ) that of majority (minority) candidates with a bad resume signal. We can compute  $S(\bar{m}_0, \bar{m}_1, \underline{m}_1, \underline{m}_0)$  the expected surplus of the recruitment, i.e. the expected productivity of the hired candidate among the set of interviewed candidates. This depends on the prediction that the recruiter can make about the value of the second signal given the information about the first signal. Let us denote  $P_0 = P(\theta = L|D = 0, \eta = H)$ ,  $P_{0,low} = P(\theta = L|D = 0, \eta = L)$  and  $P_1 = P(\theta = L|D = 1)$ . We can show that:

$$P_0 = 1/3$$

$$P_{0,low} = 2/3$$

$$P_1 = 1/2$$

For a given number of interviews  $m$ , it is always a better strategy to interview a majority candidate with a good signal instead of a minority candidate and to interview a minority candidate

in place of a majority candidate with a bad signal. Recruiters are indifferent between minority candidates with a good or bad resumes' signal. In a context where there is at least one candidate with a good resume signal, it is then sufficient to consider three quadruplets of interviews. Their expected values are the following:

$$\begin{aligned} S(\bar{m}_0, 0, 0, 0) &= \left(1 - P_0^{\bar{m}_0}\right) E[y|D = 0, \eta = H, \theta = H] \\ &+ P_0^{\bar{m}_0} E[y|D = 0, \eta = H, \theta = L] \end{aligned}$$

$$\begin{aligned} S(\bar{m}_0, \bar{m}_1, \underline{m}_1, 0) &= \left(1 - P_0^{\bar{m}_0}\right) E[y|D = 0, \eta = H, \theta = H] \\ &+ P_0^{\bar{m}_0} \left(1 - P_1^{\bar{m}_1 + \underline{m}_1}\right) E[y|D = 1, \theta = H] \\ &+ P_0^{\bar{m}_0} P_1^{\bar{m}_1 + \underline{m}_1} E[y|D = 0, \eta = H, \theta = L] \end{aligned}$$

$$\begin{aligned} S(0, \bar{m}_1, \underline{m}_1, \underline{m}_0) &= \left(1 - P_1^{\bar{m}_1 + \underline{m}_1}\right) E[y|D = 1, \theta = H] \\ &+ P_1^{\bar{m}_1 + \underline{m}_1} \left(1 - P_{0,low}^{\underline{m}_0}\right) E[y|D = 0, \eta = L, \theta = H] \\ &+ P_1^{\bar{m}_1 + \underline{m}_1} P_{0,low}^{\underline{m}_0} E[y|D = 1, \theta = L] \end{aligned}$$

#### *I.A. Standard procedure*

We note  $S(\bar{n}_0, \bar{n}_1, \underline{n}_1, \underline{n}_0, m)$  the expected surplus when  $m$  candidates are optimally chosen among the resume pool in the standard case. This can be expressed as a function of the above surplus. In the standard case, the recruiter can maximize for every quadruplet of resumes the expected productivity net of interview costs.

$$\max_m S(\bar{n}_0, \bar{n}_1, \underline{n}_1, \underline{n}_0, m) - c \times m$$

We denote  $m^*(\bar{n}_0, \bar{n}_1, \underline{n}_1, \underline{n}_0)$  the optimal number of interviews. Some exact formulas follow.

When  $m^*$  is such that  $m^* < \bar{n}_0$  (this happens when  $c > S(\bar{n}_0, 1, 0, 0) - S(\bar{n}_0, 0, 0, 0)$ )

$$\begin{aligned} m^* &= \lfloor 1/\log(1/P_0)(-\log(c) + \log(\log(1/P_0))) \\ &+ 1/\log(1/P_0) \log(E[y|D=0, \eta=H, \theta=H] - E[y|D=0, \eta=H, \theta=L]) \rfloor \end{aligned}$$

When  $m^*$  is such that  $\bar{n}_0 < m^* < \bar{n}_1 + \underline{n}_1$  ( $c < S(\bar{n}_0, 1, 0, 0) - S(\bar{n}_0, 0, 0, 0)$ )

$$\begin{aligned} m^* &= \lfloor 1/\log(1/P_1)(-\log(c) + \log(\log(1/P_1)) + \bar{n}_0 \log(P_0/P_1)) \\ &+ 1/\log(1/P_1) \log(E[y|D=1, \theta=H] - E[y|D=0, \eta=H, \theta=L]) \rfloor \end{aligned}$$

When  $m^*$  is such that  $m^* > \bar{n}_1 + \underline{n}_1$ , (this may happen when  $\bar{n}_0 = 0$ )

$$\begin{aligned} m^* &= \lfloor 1/\log(1/P_{0,low})(-\log(c) + \log(\log(1/P_{0,low})) + (\bar{n}_1 + \underline{n}_1) \log(P_1/P_{0,low})) \\ &+ 1/\log(1/P_{0,low}) \log(E[y|D=0, \eta=L, \theta=H] - E[y|D=1, \theta=H]) \rfloor \end{aligned}$$

Then the mean number of interviews when there are  $\bar{n}$  candidates with high resumes' signal writes:

$$E[m(\bar{n}_0, \bar{n}_1, \underline{n}_1, \underline{n}_0) | \bar{n}, \underline{n}] = \sum_{j=0 \dots \bar{m}} \sum_{k=0 \dots \underline{m}} \pi(j, \bar{n}) \pi(k, \underline{n}) m(\bar{n}_0, \bar{n}_1, \underline{n}_1, \underline{n}_0)$$

where  $\pi(j, \bar{n})$  is the probability that, among  $\bar{n}$  candidates,  $j$  are from the majority group (recall that  $\pi$  is the proportion of majority candidate in the economy, we have  $\pi(j, \bar{n}) = \binom{\bar{n}}{j} \pi^j (1 - \pi)^{\bar{n}-j}$ ). Note that, in the above sum, the dominant term corresponds to the balanced applicant pool (when there are as many minority and majority candidates,  $\pi(j, \bar{n})$  is maximal).

### *I.B. Anonymous applications*

In the anonymous case, the expected surplus associated with interviewing  $(\bar{m}, \underline{m})$  candidates among  $(\bar{n}, \underline{n})$  resumes is:

$$S^a(\bar{m}, \underline{m}) = \sum_{j=0 \dots \bar{m}} \sum_{k=0 \dots \underline{m}} \pi(j, \bar{m}) \pi(k, \underline{m}) S(j, \bar{m} - j, \underline{m} - k, k)$$

Given the optimal strategy of recruiters to first interview good candidates, we can define unambiguously  $S^a(\bar{n}, \underline{n}, m)$  the surplus associated with interviewing  $m$  candidates. Then the recruiters' program writes:

$$\max_m S^a(\bar{n}, \underline{n}, m) - cm$$

The analytical expression of the solution is complicated, we prefer to simulate solutions.

### *I.C. Simulations*

We solve both programs numerically. We choose an environment representative of our experiment: there are eight candidates from the resume pool (four with good signals and four with bad). The minority and majority groups are equally sized ( $\pi = 0.5$ ). We solve the recruiters' programs for different cost values ranging from 0.001 to 0.10. Very low values of the cost are reasonable as the mean productivity of the hired candidate is below unity.

## **B** SUPPLEMENTARY TABLES

Table XIII: Sample size and response rates in the applicants' survey for the analysis of external validity and John Henry effect

	Population size (a)	# sampled for survey (b)	# of respondents (c)	Sampling rate (b/a)	Response rate (c/b)
A. Sample for Table III (applicants in firms inside and outside the experiment)					
Participating (control firms)	4451	1651	981	0.37	0.59
Refused to participate	2389	717	397	0.30	0.55
Not invited to participate	2115	654	306	0.31	0.47
Total	8955	3022	1684	0.34	0.56
B. Sample for Table VII (applicants in control firms with job postings after the experiment)					
During the experiment	1903	815	420	0.43	0.52
After the experiment	2548	836	387	0.33	0.46
Total	4451	1651	807	0.37	0.49

**Notes :** The sample for Table VII only uses responding candidates in firms that have responding candidates during and after the experiment, so as to have a balanced panel of firms. This explains why only 807 candidates are considered as respondents (instead of 981). However, the results of Table VII are qualitatively unchanged if one uses 981 responding candidates and an unbalanced panel of firms.

Table XIV: Robustness analysis: interaction between foreign background and residence status

	Interview rates		Hiring rates	
	Baseline	Interaction	Baseline	Interaction
	(1)	(2)	(3)	(4)
Intercept	.116*** (.026)	.116*** (.026)	.021** (.009)	.021** (.009)
Minority (I or D)	-.024 (.031)		.002 (.012)	
Immigrant (or child of) (I)		-.034 (.035)		.006 (.015)
Deprived neighborhood (D)		.017 (.050)		.002 (.023)
I $\times$ D		-.019 (.062)		-.013 (.028)
Anonymous (T)	.061 (.040)	.061 (.040)	.031 (.019)	.031 (.019)
T $\times$ (I or D)	-.107** (.045)		-.037* (.021)	
T $\times$ I		-.111** (.048)		-.053** (.023)
T $\times$ D		-.148** (.061)		-.038 (.029)
T $\times$ I $\times$ D		.197** (.079)		.088** (.042)
Nb. of obs.	1268	1268	1268	1268

**Source:** Candidates' survey. **Notes :** In this table, we decompose our minority indicator in its two main components: foreign background (immigrant or child of immigrant) and residence status. In column 1 (3), we recall our baseline difference-in-difference estimation for the interview (hiring) rates. In column 2 and 4, we distinguish candidates with foreign background (line 3), those residing in deprived neighborhood (line 4) and those who cumulate both (line 5). For example, in the control group, candidates with foreign background but not residing in deprived neighborhood have an interview rate that is -3.4 points lower than majority candidates. The effects of anonymization on the interview (hiring) rates of candidates with only one "discriminatory" characteristics are not statistically different whether the candidate has a foreign or lives in a deprived neighborhood: -11.1 vs. -14.8 points (-5.3 vs. -3.8 points). OLS estimation without any covariates. Survey sampling weights are used. Standard errors are clustered at the vacant job level.

Table XV: Robustness analysis: different measures of foreign status

	Immigrant or child of immigrant	Immigrant	Child of immigrant	Muslim or African sounding names
Interview rates				
Intercept	0.116*** (0.026)	0.111*** (0.022)	0.109*** (0.022)	0.116*** (0.023)
Minority (D)	-0.024 (0.031)	-0.016 (0.030)	-0.013 (0.030)	-0.029 (0.030)
Anonymous (T)	0.061 (0.040)	0.037 (0.034)	0.037 (0.032)	0.031 (0.034)
T × D	-0.107** (0.045)	-0.075* (0.041)	-0.087** (0.040)	-0.063 (0.041)
Observations	1,268	1,268	1,268	1,268
Hiring rates				
Intercept	0.021** (0.009)	0.025*** (0.008)	0.019*** (0.007)	0.022*** (0.008)
Minority (D)	0.002 (0.012)	-0.008 (0.011)	0.007 (0.013)	-0.000 (0.012)
Anonymous (T)	0.031 (0.019)	0.018 (0.016)	0.022 (0.015)	0.022 (0.016)
T × D	-0.037* (0.021)	-0.015 (0.019)	-0.027 (0.020)	-0.024 (0.020)
Observations	1,268	1,268	1,268	1,268

**Source:** Candidates' survey. **Notes :** Survey sampling weights are used. Standard errors are clustered at the vacant job level. We verify the robustness of our analysis when we change how the foreign status is measured. In the first column, we recall our baseline estimate where an individual is considered foreign when he is immigrant or child of immigrant. In column 2 (3), foreign status is restricted to immigrant (to child of immigrant). In column 4, we consider as foreigners candidates with Muslim or African sounding first names. In each column, a candidate belongs to the minority if he has a foreign status or if he lives in a deprived neighborhood.

Table XVI: Robustness: Effects of different elements of the resume on the interview decision

	Standard application		Test	Anonymous	Test
	Majority	Minority	p-value	All candidates	p-value
	$\beta^0(St)$	$\beta^1(St)$	$\beta^0(St) = \beta^1(St)$	$\beta(An)$	$\beta^0(St) = \beta^1(St) = \beta(An)$
	(1)	(2)	(3)	(4)	(5)
Minority effect ( $\alpha$ )		-0.019 (0.087)			
High uncertainty	0.040 (0.052)	0.069 (0.048)	0.678	-0.106** (0.050)	0.030**
Has inactivity periods	-0.094 (0.061)	0.028 (0.046)	0.088*	-0.045 (0.040)	0.198
Adequate skills	0.144** (0.058)	0.073 (0.048)	0.360	0.054 (0.037)	0.419
Adequate work experience	-0.034 (0.051)	0.028 (0.046)	0.367	0.034 (0.035)	0.524
Nb of candidates	252	334		554	
Nb of vacant jobs		283		270	

**Source:** Candidates' survey and resumes' coding. **Notes :** As in table X, we estimate the effects of resumes' signals on the interview rate in a model with vacant job fixed effect. As a robustness check, we exclude the overall rating. Standard errors are clustered at the vacant job level. Columns 1 and 2 show returns to signals  $x$  when resumes bear names (estimation of  $I_{ij} = (1 - D_i) \times X_i \beta^0(St) + \alpha D_i + D_i \times X_{ij} \beta^1(St) + c_j + \nu_{ij}$ , where  $i$  indexes candidates,  $j$  indexes vacant jobs and  $D$  indicates minority status), and column 4 when resumes are anonymous (estimation of  $I_{ij} = X_i \beta(An) + c_j + \nu_{ij}$ ). Column 1 concerns majority candidates (results for  $\beta^0(St)$ ), column 2 minority candidates (results for  $\alpha$  and  $\beta^1(St)$ ). In column 3, we report the p-value of the test of equality in returns between columns 1 and 2. In column 5, we report the p-value of the test of equality between columns 1, 2 and 4. For example, when nominative resumes show periods of inactivity, the interview rate of majority candidates decreases by 9.4 points and that of minority candidate increases by 2.8 points.

Table XVII: Correlation of ranking (first candidate) under anonymous and standard procedure

Ranked as first		In the anonymous procedure	
		No	Yes
Whole sample			
In the standard procedure	No	0.430 (0.016)	0.110 (0.009)
	Yes	0.126 (0.010)	0.334 (0.012)
Majority candidates			
In the standard procedure	No	0.427 (0.023)	0.117 (0.014)
	Yes	0.105 (0.015)	0.351 (0.019)
Minority candidates			
In the standard procedure	No	0.432 (0.023)	0.104 (0.012)
	Yes	0.126 (0.016)	0.321 (0.015)

**Source:** Candidates' survey and resumes' coding. **Notes:** We use models estimated in Table X to predict ranks within vacancy under the two procedures. Correlations are computed for the whole sample in the upper panel, for the majority candidates in the intermediate panel and for the minority in the lower panel. For example, 33.4% of candidates are ranked first using both valuation models, in the standard and anonymous procedure. To obtain standard errors, we bootstrap 500 times the whole procedure (estimation and predictions of resume valuations).