nature research
https://doi.org/10.1038/s41586-020-03136-0

Supplementary information

Monitoring hiring discrimination through online recruitment platforms

In the format provided by the authors and unedited

Supplementary Information:

Monitoring hiring discrimination through online recruitment platforms

Dominik Hangartner, [1,2,3,⋆] Daniel Kopp,[3,4,⋆] Michael Siegenthaler [3,4,⋆]

[1] *Public Policy Group, ETH Zurich, 8092 Switzerland*

[2] *Department of Government, London School of Economics and Political Science, London WC2A 2AE, UK*

[3] *Immigration Policy Lab, Stanford University, Stanford, CA 94305, and ETH Zurich, 8092 Zurich, Switzerland*

[4] *KOF Swiss Economic Institute, ETH Zurich, 8092 Switzerland*

⋆ *All authors contributed equally to this work*

Contents

Supplementary Methods

Background Information on Job-Room

Job-Room is the employment website of the Swiss State Secretariat for Economic Affairs. It provides employer with the profiles (online CVs) of a constantly changing stock of about 170,000 jobseekers registered with the Swiss Public Employment Service. In order to be able to register, jobseekers need a valid residence permit. The majority of jobseekers is registered because they claim unemployment benefits. Eligibility for benefits usually requires to have worked for at least 12 months within the last 24 months in Switzerland. For these reasons, the vast majority of jobseekers registered on Job-Room hold a residence permit (or Swiss citizenship) that does not create any additional costs for employers (see Extended Data Table 2). The characteristics of the jobseekers visible on the platform are entered by the responsible case worker of the public employment service, typically at the time when the jobseeker registers at the public employment service. Jobseekers can opt out from showing their CVs on the platform, but only approximately 15% choose to do so.

Recruiters can navigate through the standardised profiles of the jobseekers and contact jobseekers that they are interested in interviewing. On the website, recruiters can specify certain search criteria that jobseekers should meet. Typical entries are occupation and place of work. Extended Data Figure 1a shows a screenshot of this first step.

After specifying the search criteria, recruiters get a list with at most 100 candidates who match the criteria (see Extended Data Figure 1b for a screenshot of the result list). Only exact matches are shown. On average, 39 jobseeker profiles are returned in a search list. Each list entry contains a fairly limited amount of information about the respective jobseeker, namely the jobseeker's desired work volume (in full-time equivalents), gender, canton of residency, whether the jobseeker is "immediately available" to start work, and possible additional skills (see below). The information on Job-Room is presented in a structured (tabular) way. There is only one unstructured text field that contains information about additional skills of the candidate (e.g. "Experience in long-term care, experience in Alzheimer and dementia care", see Table 1 for additional examples). In order to include this information in a statistical model, we encoded it by a simple text-mining algorithm.[1] Only 61% of all candidates and their case workers make use of the possibility to enter additional skills. Robustness checks show that our results are robust to excluding jobseekers with additional skills entry (see Table 5).

Based on the information in the result list, recruiters can select the interesting candidates to see their full profiles (see Extended Data Figure 1c). On average, they select 42% of profiles for full view per search. The full profile comprises information about the occupations in which a candidate is searching for work, the level of work experience and highest educational attainment for each occupation, jobseeker's language skills, gender, the place of residence, whether he or she wants to work full- or part-time and the desired work region. Moreover, private recruitment agencies can register in order to get preferential access to a a limited set of protected information of jobseekers. In particular, registered users also see the nationality and—if the jobseeker consents when registering—the name of a jobseeker (84% of the profiles show show the name). Even though registered recruiters account only for 30% of all users, they are responsible for 73% of all searches on Job-Room.

[1] We assigned the statements in the text to different categories like additional experience, additional education, soft skills, IT skills, machine operating skills, leadership skills or additional language skills. If the jobseeker mentioned work experience we also extracted the respective duration.

The information about the jobseeker contains three markers of ethnicity: nationality, name (those two variables are only visible to registered users) and language (visible to all users). We classified the names of the jobseekers according to their ethnic origin with the name-classification algorithm Onolytics.[2]

Data

The data used for this study consists of click data tracking recruiters' behaviour on Job-Room matched with administrative data on the characteristics of the jobseekers registered with Job-Room from the unemployment register. The click data and the register data are collected and owned by the State Secretariat for Economic Affairs and were shared with us in anonymised form after signing a data sharing agreement. The study was conducted according to the guidelines of the ETH Zurich Ethics Committee, which exempted the study from a full review due to the secondary nature of the data used (approval RSETHZ 413).

The pre-processing of the raw (i.e. logged) data from the website—stored in a text file where each line corresponded to an event in the file format JSON—involved several steps. The most important of these steps was to link the click data logged from the website with jobseeker's characteristics from the unemployment register. This merge was done based on a unique person identifier but also based on date (since certain personal characteristics vary over time).

We imposed the following sample restrictions to our analysis sample:

- We discard a small number of bots (automated agents).[3]
- We exclude all searches where the user did not specify a search criteria (2.0% of all observations).
- We exclude all searches that take place within less than 10 seconds to the next search (3.4% of the remaining observations). In almost all of these cases, the recruiters do not select a candidate from the result list but go back to step 1 to change the search criteria.
- For the analyses that are conditional on profile view, we exclude all observations (profile views) where the recruiter spends less than two seconds on the profile (14.7% of all profile views). Most of these searchers are likely conducted by users with the goal to collect jobseekers'

[2] Onolytics (formerly Onomap) is a name-classification algorithm that draws upon a large database of the most common surnames and first names in the world, their frequencies by geographical area and the most likely ethnic origin. The names were compiled from a variety of publicly-available sources like census records, telephone directories, or name frequency statistics from civil registers. Onolytics has been validated [1] and is widely used to classify names into ethnic origins (see the list of publications on https://onolytics.com/publications). The methodology is outlined in detail in [2].

[3] Since the aim of automated agents is to gather data, such agents do not interact with the website in the same way as human users. In particular, they typically do not click on a search result to see the full candidate profile. Rather, they visit the URL of the profile directly. As a consequence, automated agents generate many search events but very few actual click events. Therefore, we classify agents as non-human if any of the following conditions is true: 1) less than 5% of the logged events are generated by interacting with the site in a human fashion (such as clicking on a profile in the result list); 2) more than 99% of the logged events of the user are searches; 3) the user conducted more than 10'000 searches during the observation period and less than 5% of these searches yielded no results. Condition 3 detects agents who search very regularly (e.g. every second) by going through the same combinations of search criteria known to generate results. Additionally, we deleted seven suspect agents that had huge spikes in search activity within a very short period of time (100 searches per minute). In total, we identified 65 bots (anonymised IP addresses), which we excluded from the analysis.

contact information. The results are similar if we include these observations, as shown in the robustness section below. We do not exclude these profiles in the gender regressions since gender can be inferred from the result list, i.e. *before* recruiters visit a profile.

Extended Data Table 3 reports the number of users (recruiters), the number of searches, the number of jobseekers in the results list, and the number of profile views with and without these sample restrictions.

The sample used for our main ethnicity regressions below is 5.4% smaller, the one for the main gender regressions 0.2%. The reason is missing values in specific covariates (particularly the origin of a candidate's name). Where feasible, however, we limited the loss in sample size due to missing values. We replaced the missing value with a fixed value not taken by any observation and estimated separate effects for these categories by adding indicator variables equal to one if the value on the original variable is missing.

Click data

The click data was collected by the State Secretariat for Economic Affairs and contains all typed entries and mouse clicks conducted by recruiters on Job-Room between March 6, 2017 and December 31, 2017. The dataset used in the empirical analyses contains the following variables:

Search parameters of the recruiters

- *Search ID:* Search identifier
- *Session ID:* Session identifier
- *Date:* Date at which the search was carried out
- *Timestamp:* Exact time (in milliseconds) at which the search was carried out
- *Recruiter ID:* Anonymised recruiter identifier
- *Registered:* Binary variable whether recruiter is registered or not
- *User language:* Language that the recruiter selected (German, French or Italian)
- *Occupation:* Occupation in which the recruiter searches (4-digit ISCO 08)
- *Firm location:* Canton in which the firm is located
- *Place of residence:* Desired canton of residence of the jobseeker
- *Work volume:* Desired work volume in full-time equivalents (10-100 percent)
- *Education:* Desired educational attainment (6 categories)
- *Origin of diploma:* Desired origin of diploma (4 categories: Swiss, Foreign but accepted in Switzerland, Foreign and not accepted in Switzerland, No information)
- *Work experience:* Desired years of work experience (4 categories: None, less than 1 year, 1–3 years, 4 years or more)

- *Driver license:* Desired category of driver license (15 categories)
- *Special working conditions:* Special working conditions: night work, shift work, work from home, work on Sundays and public holidays
- *Availability:* Binary variable whether candidate has to be immediately available or not
- *Additional skills:* 14 binary or categorical variables capturing information from the unstructured text field with additional skills
- *Language skills:* Desired language

Click events

- *Jobseeker ID:* Anonymous identifier for jobseeker in list with search results
- *Candidate clicked:* Recruiter clicks to see full profile of the jobseeker
- *Contact button clicked:* Recruiter clicks on the contact button to reveal contact information of a jobseeker
- *Click on email:* Recruiter clicks on the hyperlink of the e-mail address of the candidate hidden behind the contact button
- *Click on print button:* Recruiter clicks on the button to print out a jobseeker's profile information
- *Time spent on profile:* Time a recruiter has spent on the profile of a jobseeker until s/he either leaves the profile or clicks on the contact button

Data on jobseeker characteristics

All information on Job-Room about the characteristics of jobseekers is linked from the unemployment register. This information is entered by the case workers of the Swiss public employment service together with the jobseekers during their first meeting. We can expect case workers to truthfully report on jobseeker characteristics, including educational attainment and language skills, for several reasons. Case workers understand that the information they enter will be the basis for a range of decisions that directly impact the jobseeker. In addition to being fed into Job-Room and provided to recruiters, case workers also suggest candidates to employers with vacancies based on Job-Room profiles. In addition, this information is the basis for important decisions by the public employment service such as assigning jobseekers to different active labour market programs. Furthermore, case workers might be concerned that recruiters may stop using Job-Room if the information on jobseekers is inaccurate.

Since not all the information about the jobseekers is available on Job-Room, we can distinguish between jobseeker characteristics visible to recruiters and characteristics not visible to recruiters.

Jobseeker characteristics visible to recruiters

- *Jobseeker ID:* Identifier for jobseeker in list with search results
- *Nationality:* Nationality
- *Name origin:* Ethnic origin of name (9 categories, based on analysis with Onolytics)

- *Gender:* Binary indicator for female
- *Work volume:* Desired work volume in full-time equivalents (1-100 percent)
- *Education:* Highest level of educational attainment (16 categories)
- *Occupations:* Occupations in which a candidate is willing to work
- *Experience:* Level of work experience by occupation (4 categories: None, less than 1 year, 1-3 years, 4 years or more)
- *Origin of diploma:* Origin of diploma by occupation (4 categories: Swiss, Foreign in Switzerland accepted, Foreign in Switzerland not accepted, No information)
- *Last job in occupation:* Binary indicator: 1 if the jobseekers last job was in the same occupation in which a recruiter is searching
- *Work region:* Region (canton) in which a jobseeker is searching for a job
- *Unlimited contract:* Binary indicator: 1 if the jobseeker desires an unlimited contract
- *Availability:* Binary indicator: 1 if jobseeker is immediately available
- *Contact information:* Binary indicator: 1 if contact information (including name) are visible for registered recruiters
- *Special working conditions:* Jobseeker is willing to work under special working conditions (5 categories: none, night work, shift work, work from home, work on Sundays and public holidays)
- *Driver license:* Category of driver license (15 categories)
- *Canton of residence:* Canton of residence (26 cantons)
- *Language skills:* Language proficiency (4 proficiency levels, 40 different language dummies)
- *Additional skills:* 14 binary or categorical variables capturing information from the unstructured text field with additional skills proficiency (Experience mentioned, no. of years of work experience, (tertiary/vocational) education, training, soft-skills, general IT-skills, specialised IT-skills, machine-operating skills, leadership-skills, language skills (12 categories), no. of words in skill field, language of skill field). Table 1 provides examples of such entries.
- *Absolute rank:* Absolute rank of the jobseeker in a given search (100 dummies)
- *Decile of relative rank:* Decile of the relative rank (rank divided by number of search results) of the jobseeker in a given search
- *Profile seen before:* Number of times that the same recruiter has seen the profile in previous searches

Jobseeker characteristics not visible to recruiters

- *Age:* Continuous variable indicating the age of the jobseeker
- *Last insured wage:* Last insured wage (top coded at CHF 12,350 per month) prior to registering at the Swiss public employment service
- *Marital status:* Marital status of the jobseeker (4 categories: unmarried, married, divorced, widowed)
- *Residence permit:* Residence permit of the jobseeker (16 categories)
- *Native language:* Native language of the candidate (40 different language dummies)
- *Registration date:* Date at which the jobseeker registered at the Swiss public employment service
- *De-registration date:* Date at which the jobseeker is de-registered from the Swiss public employment service

Table 2 shows the coverage of our data across occupations, dividing occupations into five skill levels according to the ISCO skill classification. Column 1 shows the number of searches per occupational skill group. Column 2 relates the number of ISCO 4-digit occupations with a minimum of 25 profile views registered during the study period to the total number of ISCO 4-digit occupations (without armed forces). Column 3 reports the employment share of the ISCO 4-digit occupations covered by our data. The data shows that we register at least 25 profile views in 65% of all ISCO 4-digit occupations, thereby covering 94% of total employment in Switzerland.

Table 1.
Examples of additional skills provided in the unstructured text field.

German text	English translation
Baggerführer Maschinist mit über 10 Jahren Erfahrung. Vor allem gearbeitet mit Menzi Muck und Bagger Ãijber 20 Tonnen. Aushub, Umgebungen, Kanalarbeiten, Vermessung, Abbruch, Vorarbeiter. SUVA Staplerschein vorhanden.	Excavator operator with over 10 years of work experience. Mainly worked with Menzi Muck and excavators over 20 tons. Excavation, surroundings, canal work, surveying, demolition, foreman. SUVA forklift licence available.
Berufsabschluss als Coiffeuse. Erfahrung in Zupfen und Färben von Augenbrauen/Wimpern. Fachkenntnisse im Bereich Kundenberatung, Haarpflegeprodukte und Kosmetika.	Professional qualification as hairdresser. Experience in plucking and tinting eyebrows and eyelashes. Expertise in customer service, hair care products and cosmetics.
Ausbildung EFZ Radio- und Fernsehelektriker. Langjährige Berufserfahrung als Kundenberater, z. Zt auch Storemanager bzw. Stv. Abteilungsleiter. Hauptbereiche: Telekommunikation (Verkauf Handy- und Accessoires sowie Auftragswesen Reparaturservice) sowie Wohnen/Einrichtungen.	Swiss certificate of proficiency in radio and television electrician. Many years of professional experience as customer consultant, currently also store manager and deputy department manager. Main areas: Telecommunications (sale of mobile phones and accessories as well as order processing for repair services) as well as living/furnishing.
Kauffrau EFZ mit Weiterbildung zur Finanzplanerin mit FA. Hat auch Händlerdiplom ACI. Langjährige Berufserfahrung (30 Jahre) im Devisenhandel einer ausländischen Privatbank. Sucht neue Herausforderung in Vollzeit wiederum in diesem Berufs- und Tätigkeitsbereich. Ist auch offen fr̈ andere Tätigkeiten einer Bank (z.B. Privat-Banking, Execution im Bereich Fixed Income oder Immobilienabteilung, usw.). Muttersprachen: Deutsch & Türkisch, sehr gute Englischkenntnisse (w/s), Grundkenntnisse im Französisch (muss aufgefrischt werden).	Swiss certificate of proficiency in business administration with further training as financial planner (with advanced degree). Also has a ACI diploma. Many years of professional experience (30 years) in foreign exchange trading in a foreign private bank. Looking for a new full-time challenge in this profession and field of activity. Is also open for other activities of a bank (e.g. private banking, execution in the fixed income or real estate department, etc.). Native languages: German & Turkish, very good knowledge of English (w/s), basic knowledge of French (needs to be refreshed).
Ausgebildete Kauffrau mit Langjähriger Erfahrung als Sachbearbeiterin Buchhaltung. Verfügt über gute Kenntnisse in:SAP CRM, 'Wecare', Navision Financials/Windows NT, AS 400, Word und Excel mit Zertifikat, PowerPoint, Outlook, Europrint, Zeiterfassungssysteme Calitime.	Trained businesswoman with many years of experience as an accounting clerk. Has good knowledge in: SAP CRM, 'Wecare', Navision Financials/Windows NT, AS 400, Word and Excel with certificate, PowerPoint, Outlook, Europrint, time recording systems Calitime.

Notes: The table shows five comparatively long entries of jobseekers in the unstructured text field “additional skills”. The examples stem from individuals in our sample period (March–December 2017).

Table 2.
Occupational coverage of the Job-Room sample.

	(1) # searches	(2) # occupations	(3) Share in employment
Low skilled	18770	18/37	.93
Medium skilled	290936	145/216	.98
Technicians and assoc. prof.	42671	60/94	.93
Professionals	31227	82/110	.94
Managers	7362	18/38	.88
Total	390966	323/495	.94

The table shows the occupational coverage of the tracking data collected on the platform. To show the skill composition of the searches in the data, we divide occupations into five skill levels according to the ISCO skill classification. Column 1 reports the number of searches per skill level. Column 2 relates the number of ISCO 4-digit occupations with a minimum of 25 profile views registered during the sample period to the total number of ISCO 4-digit occupations (without armed forces). Column 3 reports the employment share of the ISCO 4-digit occupations covered by our data. Total employment by ISCO skill level in Switzerland is computed using the Structural Survey of the Federal Statistical Office.

<u>Statistical Model</u>

To estimate the effect of ethnicity on contact attempts, we use ordinary least squares regression and flexibly control for *all* other job-seeker characteristics visible to recruiters as well as search and rank fixed effects. Table 15 confirms that the main results are very similar when using binary logistic regression instead of OLS to model contact. We estimate the following model:

$$y_{i,s} = \sum_{k=1}^{K} \beta_k Ethnicity_{i,s}^k + \mathbf{x}_{i,s}^\top \boldsymbol{\gamma} + \mathbf{r}_{i,s}^\top \boldsymbol{\rho} + \delta_s + \epsilon_{i,s}, \quad i = 1, ..., N; \quad s = 1, ..., S. \tag{1}$$

In this linear probability model, the main outcome of interest, $y_{i,s}$, is a binary indicator that the recruiter in search s clicks on the contact button on the profile of candidate i in order to see the contact information. The probability that a recruiter clicks on the contact button conditional on the profile appearing in the search list is 10 percent. The probability that a recruiter clicks on the contact button conditional on viewing the profile is 42 percent. As a secondary outcome, we use the log time the recruiter spends on the profile of a jobseeker until s/he either leaves the profile or clicks on the contact button.

$Ethnicity_{i,s}^k$ is a binary indicator equal to one if jobseeker i in search s has ethnicity k, where $k \in 1, \ldots, K$. The coefficients β_k are the main coefficients of interest. They capture the effect of ethnicity k on the likelihood of contact. Recruiters that are not registered can infer ethnicity from jobseekers' language skills only. Recruiters that are registered, however, can infer ethnicity from language, nationality and name. If we estimate separate effects for each of these signals of ethnicity, we find that all three signals reduce contact attempts of registered users, suggesting discrimination along all three dimensions. Table 3 shows that the three indicators are highly correlated. For example, 85% of individuals that have a nationality from a Balkan country also speak a language and have a name that is typical for this region. These shares are between 65–94% for the other nationalities in our sample. To focus the analysis, we build a composite variable of ethnicity combining the three signals. A 'typical' member of an ethnic group unites all three signals of ethnicity, i.e. they are foreigners with a name that is not Swiss-sounding[4] and they speak a language from their origin region—often in addition to an official Swiss language. In the paper, we focus on the comparison of contact attempts for these 'typical' ethnic minorities to the reference group, which consists of jobseekers with Swiss citizenship, a Swiss-sounding name, and proficiency in at least one of the official Swiss languages (German, French, Italian or Romansh).[5] We keep 'atypical' members of ethnic groups in our estimation sample and control for their presence by estimating separate effects for these groups. Moreover, since knowledge of a language is not only a signal for ethnicity but also a productivity-relevant skill, we explicitly control for languages that can be considered as productivity-relevant in the Swiss context, namely German, Swiss-German, French, Italian and English. We do not control for other languages since in most jobs they are mainly a signal for ethnicity. This assumption is confirmed if we estimate separate effects for these languages: proficiency in those foreign languages reduces, rather than increases, contact attempts.

The vector of control variables $\mathbf{x}_{i,s}$ contains the characteristics of jobseeker i visible to the recruiter in search s. $\boldsymbol{\gamma}$ is the associated vector of coefficients. Our research design is based on the extent to which this vector controls for all information visible to the recruiter that correlates with both, the ethnicity indicators *and* recruiters' contact decisions. Important advantages of

[4] Jobseekers from Southern or Western Europe also include those with an Italian, German or French name.

[5] We define names as 'Swiss' if Onolytics indicates a German-, French-, Italian- or Romansh origin.

Job-Room in this regard is that candidate characteristics are presented in a structured (tabular) format, that it does not include a picture of the applicant, and only one unstructured text field. These features make it easier to statistically control for the relevant information using a series of mostly binary variables. Nevertheless, selecting the important set of covariates among a large set of potential covariates is a non-trivial task. We achieve this goal by controlling for all direct effects of the variables visible to recruiters, and by relying on supervised machine learning methods in order to select the relevant subset of first-order interactions among the high-dimensional set of all possible first-order interactions. In particular, we apply the 'post-double-selection' method, a recent approach for variable selection developed by [3, 4]. This method proceeds in three steps: First, we perform a Lasso regression of the outcome on all covariates and their first-order interactions, and retain those interactions that are selected by the Lasso estimator (using the default procedure of [4] to select the λ-penalty term). Second, we perform a series of Lasso regressions of each ethnicity indicator on all covariates and their first-order interactions, and again retain those interactions that are selected by the Lasso estimator. The third step is to estimate the ethnicity coefficients using the ordinary least-square regression model (equation 1) controlling for all baseline covariates and the union of the set of all interactions selected in the two selection steps.

Our regression models also control for search fixed effects δ_s—a separate regression intercept for each search. We thus account for all characteristics that are constant within a specific search request, such as the search criteria entered by the recruiter (e.g. the occupation) as well as observed and unobserved recruiter characteristics. This allows us to focus on comparisons of jobseekers within the same search that all fit the search criteria. Moreover, since the contact probability decreases with rank, we control for the absolute rank of the specific profile using a full set of rank fixed effects (a dummy for the rank of jobseeker i in search s). We also control for the decile of the relative rank of the profile in a given search, i.e. the rank relative to the total number of search results.[6] These rank controls are subsumed in the vector $\mathbf{r}_{i,s}$. The associated coefficient vector is $\boldsymbol{\rho}$. Lastly, we flexibly control for the number of times that the same recruiter has seen the profile of jobseeker i in previous searches.

Our approach to identify the extent of gender discrimination is isomorphic to our approach regarding ethnicity. The regression model is specified as follows:

$$y_{i,s} = \eta Female_{i,s} + \mathbf{x}_{i,s}^{\top}\boldsymbol{\gamma} + \mathbf{r}_{i,s}^{\top}\boldsymbol{\rho} + \delta_s + \epsilon_{i,s} \tag{2}$$

where $Female_{i,s}$ is a binary indicator whether jobseeker i in search s is female. η is the associated regression coefficient. Again, the control vector $\mathbf{x}_{i,s}$ contains the full set of baseline covariates visible to recruiters and selected first-order interactions. The only major difference in the empirical implementation of the gender regression is the sample restriction. Our ethnicity analyses are *conditional on profile view.* After all, the ethnicity indicators are only revealed once a recruiter sees the full profile. Ethnicity should not—and in fact does not—influence the decision of a recruiter whether or not to view the full profile. On the other hand, the gender of a candidate is revealed in the result list (step 2). The gender regressions thus encompass all individuals visible in the result list.

[6] Excluding these rank controls has no influence on the main results, as shown in Table 7.

Table 3.
Construction of ethnicity.

	(1) typical cat.	(2) non-typical cat.
Switzerland	.71	.29
Western & Northern Europe	.94	.064
Southern Europe	.91	.09
Central & Eastern Europe	.87	.13
Balkan	.85	.15
Middle East & North Africa	.87	.13
Asia	.85	.15
Sub-Saharan Africa	.65	.35
North & South America	.86	.14
Total	.79	.21
Observations	158319	41966

This table reports the share of jobseekers from a certain nationality-group that are classified as 'typical' (first column) and 'atypical' (second column) member of their ethnic group. A 'typical' member of an ethnic group unites all three signals of ethnicity (citizenship, name and language skills), i.e. they are foreigners (Swiss) with a name that is not Swiss-sounding (is Swiss-sounding) and they speak a language from their origin region (from Switzerland). For example, the table shows that 85% of individuals that have a nationality from a Balkan country also speak a language and have a name that is typical for this region.

Replication Materials

Researchers interested in replicating or extending our result can obtain the code to run all analyses form the publicly accessible Harvard Dataverse: https://doi.org/10.7910/DVN/GGENFB. See Code availability and Data availability statements in the main paper for further details on code and data access.

Placebo Tests

In this section, we present a series of placebo tests to probe the validity of our research design and the underlying identification assumption. The first, and most direct, of these placebo tests exploits that non-registered recruiters observe only language skills but neither nationality nor name. Hence, if our covariates sufficiently account for the relevant information visible to recruiters, nationality and name should have no impact on contact decisions of non-registered users once we flexibly control for jobseekers' language skills. Extended Data Figure 2a and Table 4 report the results of this placebo test. Column (1) of Table 4 shows the estimates from our baseline model (in percentage points). Column (2) reports the corresponding estimates when controlling for detailed language skills. Although the estimated ethnic penalties are somewhat smaller in size than in the baseline specification, they are still highly statistically significant and negative. Registered users thus discriminate against immigrant jobseekers based on name and nationality even if we account in detail for the ethnic information signalled by their language proficiency. In contrast, if we run the same regression using non-registered users (column 4 of Table 4 and Extended Data Figure 2a), all ethnicity coefficients are non-significant and close to zero. This placebo test provides direct evidence that the vector of covariates contains all correlates of nationality and name that affect contact rates.

The upper part of Table 4 explores the placebo "effects" of two other variables not visible to recruiters but visible to us researchers because we can access all data from the unemployment register: age and the wage insured by the unemployment insurance (which is equal to the wage earned at the previous employer, measured in CHF 1,000). Because they are correlated with productivity, we can assume that both variables would affect contact rates if recruiters could observe them. However, we find no significant effect of the insured wage on the contact rate conditional on the covariates in any of the regressions in Table 4, which supports the selection-on-observables-assumption underlying our research design. In contrast, the first two columns of Table 4 suggest that some recruiters are able to infer the age of the candidates. A plausible explanation is the unstructured text field containing information on additional skills of the candidates. This text sometimes contains indirect signals of jobseekers' age (e.g. statements such as "30 years of work experience"). Indeed, Column 3 shows that the age coefficients become smaller and insignificant once we restrict the sample to jobseekers who do not report additional skills and whose profile has been seen the first time by a recruiter (such that recruiters cannot infer the length of unemployment duration, which is correlated with age). Does the fact that recruiters can infer age (and possibly other information) from a candidate's additional skill entries question the validity our research design? In light of the evidence, we think the answer is no: our results are essentially unchanged if we focus solely on candidates without "additional skills" reported in the open text field, as we show in the robustness section below. The information from the open skill field is thus not systematically related to the estimated discrimination coefficients.[7]

In sum, the placebo tests presented in this section support the assumption that the Lasso-based double selection method allows us to control for all relevant jobseeker characteristics correlated with contact decisions and ethnicity and gender.

[7]This is also the case regarding age: the gender and ethnicity indicators are by and large uncorrelated with age (see Extended Data Table 2).

Table 4.

Placebo estimates (in percentage points) of the effects of ethnicity on the contact rate.

	(1)	(2)	(3)	(4)
Last insured wage (in 1'000)	0.0	0.0	0.1	0.0
	(0.03)	(0.03)	(0.06)	(0.05)
	[0.294]	[0.299]	[0.359]	[0.614]
Age				
$<$ 30 years	0.5	0.6	0.2	0.5
	(0.10)	(0.10)	(0.23)	(0.25)
	[0.000]	[0.000]	[0.290]	[0.065]
30-39 years	Ref-Cat.	Ref-Cat.	Ref-Cat.	Ref-Cat.
40-49 years	0.0	0.0	0.3	-0.1
	(0.09)	(0.09)	(0.25)	(0.23)
	[0.694]	[0.856]	[0.286]	[0.761]
50-60 years	-0.1	-0.3	0.1	-0.0
	(0.10)	(0.10)	(0.25)	(0.24)
	[0.162]	[0.012]	[0.811]	[0.921]
$>$ 60 years	0.4	0.2	0.5	0.1
	(0.17)	(0.17)	(0.38)	(0.35)
	[0.037]	[0.345]	[0.157]	[0.819]
Ethnicity				
Switzerland	Ref-Cat.	Ref-Cat.	Ref-Cat.	Ref-Cat.
Western & Northern Europe	-1.8	-1.6	-1.4	-0.4
	(0.18)	(0.18)	(0.41)	(0.36)
	[0.000]	[0.000]	[0.001]	[0.303]
Southern Europe	-0.2	-1.4	-1.0	0.1
	(0.24)	(0.21)	(0.48)	(0.41)
	[0.481]	[0.000]	[0.044]	[0.770]
Central & Eastern Europe	-2.6	-2.2	-2.2	0.2
	(0.30)	(0.35)	(0.80)	(0.79)
	[0.000]	[0.000]	[0.005]	[0.794]
Balkan	-5.3	-2.5	-3.0	-0.2
	(0.24)	(0.19)	(0.44)	(0.47)
	[0.000]	[0.000]	[0.000]	[0.625]
Middle East & North Africa	-5.7	-3.1	-2.7	-0.7
	(0.29)	(0.30)	(0.79)	(0.74)
	[0.000]	[0.000]	[0.001]	[0.360]
Asia	-7.8	-4.1	-4.3	-0.7
	(0.41)	(0.44)	(1.08)	(0.99)
	[0.000]	[0.000]	[0.000]	[0.478]
Sub-Saharan Africa	-7.2	-5.5	-6.3	0.2
	(0.39)	(0.40)	(1.06)	(1.02)
	[0.000]	[0.000]	[0.000]	[0.848]
North & South America	-2.7	-3.9	-1.5	-0.2
	(0.39)	(0.40)	(1.02)	(0.87)
	[0.000]	[0.000]	[0.149]	[0.836]
Mean dep.var.	0.42	0.42	0.44	0.18
P value F-stat ethnic = 0	0.00	0.00	0.00	0.93
Controls for all languages	No	Yes	Yes	Yes
Obs. with skills excluded	No	No	Yes	No
Only first profile visit	No	No	Yes	No
Sample	All	All	All	Not registered
No of recruiters	29834	29834	14941	18724
Observations	3,251,303	3,251,263	395,255	254,975

Notes: The table reports percentage point effects of ethnicity, age, previous wage and residence permit on the likelihood that the contact button has been clicked conditional on profile visit. Age, previous wage and residence permit are not visible on Job-Room. Controlling for detailed language skills, ethnicity is not visible for *non-registered* recruiters, the sample in column (4). Column (1) reports the estimates from our baseline model. Column (2) reports the estimates from the baseline model controlling for all language skills (not only the productivity relevant ones). Column (3) reports estimates from the placebo model when restricting the sample to jobseekers who do not report additional skills and whose profile has been seen the first time by a recruiter. Column (4) shows estimates from the placebo model when restricting the sample to non-registered recruiters. The ethnicity effects reported are the same as in Extended Data Figure 2a. Standard errors (shown in parentheses) are clustered at recruiter level. P-values of a two-sided t-test against the null hypothesis of no effect are shown in square brackets. No adjustments were made for multiple comparisons.

Validation of outcome variable

This section validates our main outcome variable: the click on the contact button on the profile page of jobseekers.

The first validation exploits that we can link jobseekers' to their entries in the Swiss unemployment register. We can thus estimate the causal effect of contact attempts on Job-Room on unemployment duration by comparing *observationally identical* jobseekers that appear in the same result list on Job-Room, but only a subset of them is contacted by the recruiter. Such cases exist mainly for two reasons. The first is pure luck—recruiters may contact only one of two identical jobseekers for idiosyncratic reasons. The second is that recruiters may simply not see some of the candidate profiles of the statistical twins. Indeed, this happens quite regularly because two jobseekers that fit equally may be listed on very different ranks in the result list, particularly if the list is long.

Extended Data Figure 7 shows the causal effect of a click on the contact button on Job-Room on the likelihood that a jobseeker leaves unemployment within 30, 60, 90, 120 and 150 days, respectively, after a search. The variable of interest is the contact attempt on Job-Room. The regressions control for search fixed effects and hence only compare candidates that appeared in the same search. We also control for all variables visible to recruiters on Job-Room, including the relevant first-order interactions. The regressions thus control for all factors that plausibly influence recruiters' contact attempts on the platform. This means that, conditional on covariates, a jobseeker with a contact attempt and a jobseeker without a contact attempt who appeared in the same search are identical except that one was contacted. Under selection on observables, the regression implies that each contact attempt increases the likelihood to leave unemployment within 90 days after the search by 2,1%. As we may expect given that recruitment processes usually take some time, the effect builds up in the first months after the contact attempt and stays constant thereafter.

The second validation exercise is presented in Extended Data Figure 2b. It shows that our results are similar if we use two additional outcome variables that are somewhat more concrete signals of contact attempts by recruiters: The blue estimates present our baseline outcome variable (a dummy equal to one if a recruiter clicked on the contact button). The yellow estimates are based on a binary indicator equal to one if a recruiter clicked on the contact button and subsequently stayed on the profile of the respective jobseeker for at least 60 seconds. This is a rather frequent event—it happens in 27.3% of all contact attempts and 11.5% of all profile visits—and is an indicator that the recruiter may have contacted the jobseeker immediately. The red estimates are based on a binary variable equal to one if a recruiter clicked on the hyperlink of the email address of the candidate or on the button "print candidate profile" on the profile page. These click events are rather rare and happen in only 3.3% and 4.8% of profile visits. Nevertheless, and despite the differences in the outcome measures, Extended Data Figure 2b shows that they all provide similar results. If anything, the ethnic penalties are slightly more pronounced for the contact click + 60 seconds and the click on the hyperlink and "print" button.

Robustness Tests

This section reports the results of several robustness tests. First, we test whether our results may be biased because we do not fully account for the information contained in the unstructured text field that contains the "additional skills" of jobseekers. Table 5 shows the ethnic penalties derived from our baseline regression model (column 1) and compares them to the penalties estimated using the sample of jobseekers who report *no* such skills (column 2). The coefficients of the two regressions are very similar, suggesting that the textual information in the skill field are uncorrelated to the estimated discrimination coefficients conditional on our control variables.

Second, we test whether our results may be biased because registered recruiters gather additional information about a jobseeker online, e.g. by conducting a parallel search on social media or Google using the jobseeker's name. We believe it is unlikely that this behaviour is widespread because recruiters spend on average less than ten seconds until taking a decision on the profile. Ten seconds is arguably not sufficient time to gather and process additional information on the candidate from other internet sources. Indeed, the placebo tests presented above would likely fail if parallel search on the internet were common. We nevertheless test for the possibility directly by restricting the sample to cases where recruiters spend less than 20 seconds on jobseekers' profiles until taking a decision. The comparison of columns 1 and 3 of Table 5 shows that the estimated ethnic penalties are very similar to our baseline results with this sample restriction, limiting remaining concerns about parallel search on the internet.

Third, Table 6 shows the effect of ethnic origin on the contact rate and time on profile if we include profiles visited for less than 2 seconds. We exclude these profiles from our main sample as these searches seem to stem from users (or automated bots) that solely collect jobseekers' contact information, rather than from recruiters that consciously process jobseekers' characteristics. Consistent with this view, including these visits lowers the estimated ethnic contact penalties slightly, but it also reduces the estimated effects of all other variables (such as the effect of experience). Importantly, including very short profile visits does not affect our findings regarding time on profile.

Fourth, we evaluate whether our results are influenced by the fact that candidates in the same applicant pool may be evaluated against each other. It is conceivable that jobseekers' opportunities are shaped by the presence of high-quality competitors.[8] To study this question, we calculate each jobseekers' overall "employability", i.e. the predicted probability that this jobseeker is contacted by employers. The index captures the characteristics of jobseekers, weighted by the occupation-specific effect of those characteristics on the probability to be contacted—a measure of recruiters' valuation of each signal. In this estimation, we incorporate all characteristics visible to recruiters except for the protected characteristics (i.e. gender and ethnicity) that are the main quantities of interest. Using this measure of jobseekers' employability, we can rank each jobseeker within the applicant pool and statistically control for his or her rank in a given search.

The index of jobseekers' employability is constructed as follows:

1. We split the sample randomly in a 50% training and a 50% test sample.

2. Using the training sample, we fit a model of job-seeker employability. For this, we regress a binary indicator whether a person was contacted by a recruiter, conditional on the recruiter seeing the profile, on the observed characteristics. To allow for heterogeneous effects of characteristics and skills by occupation, we interact each variable with one-digit occupation indicators (ISCO 1).

3. We use the coefficients of this regression to predict the occupation-specific contact probability of each jobseeker in the *test* sample. The predicted contact probability is our overall measure of jobseekers' employability for a given occupation.

[8]It is a distinct advantage of employment websites relative to extant methods to study hiring discrimination that they provide information on the entire pool of candidates available to recruiters.

Table 5.
Robustness test: Effect of ethnicity on contact rate excluding jobseekers with additional skills and with long profile views.

	(1)	(2)	(3)
Switzerland	Ref-Cat.	Ref-Cat.	Ref-Cat.
Western & Northern Europe	-4.2	-5.2	-4.2
	(0.42)	(0.73)	(0.43)
	[0.000]	[0.000]	[0.000]
Southern Europe	-0.4	0.3	-0.4
	(0.57)	(0.88)	(0.58)
	[0.481]	[0.713]	[0.538]
Central & Eastern Europe	-6.2	-5.4	-6.1
	(0.72)	(1.04)	(0.74)
	[0.000]	[0.000]	[0.000]
Balkan	-12.6	-11.8	-12.6
	(0.58)	(0.88)	(0.58)
	[0.000]	[0.000]	[0.000]
Middle East & North Africa	-13.5	-13.7	-13.3
	(0.69)	(1.22)	(0.71)
	[0.000]	[0.000]	[0.000]
Asia	-18.5	-19.6	-18.0
	(0.98)	(1.67)	(0.99)
	[0.000]	[0.000]	[0.000]
Sub-Saharan Africa	-17.1	-22.1	-17.2
	(0.92)	(1.80)	(0.93)
	[0.000]	[0.000]	[0.000]
North & South America	-6.4	-4.2	-6.1
	(0.93)	(1.98)	(0.94)
	[0.000]	[0.034]	[0.000]
Sample	Baseline	No skills	Only 2-20 Sec.
Mean dep. var.	0.42	0.43	0.44
No of recruiters	29834	17088	24192
Observations	3251303	890206	2886002

Notes: Column (1) of the table reports the baseline estimates of the effect of ethnicity on the contact rate from Figure 1, Panel A. Column (2) reports ethnicity estimates from the same model but restricts the sample to jobseekers that report no additional skills in the unstructured text field. Column (3) reports the ethnicity estimates when restricting the sample to profile views that lasted no longer than 20 seconds. Standard errors (shown in parentheses) are clustered at recruiter level. P-values of a two-sided t-test against the null hypothesis of no effect are shown in square brackets. No adjustments were made for multiple comparisons.

Table 6.
Effects of jobseekers' ethnicity, work experience and profile length on contact rate and time on profile including jobseekers that have been visited for less than 2 seconds.

	(1) Contact rate	(2) Contact rate	(3) Log time	(4) Log time
Switzerland	Ref-Cat.	Ref-Cat.	Ref-Cat.	Ref-Cat.
Western & Northern Europe	-3.4	-2.7	0.6	0.4
	(0.37)	(0.38)	(0.22)	(0.21)
	[0.000]	[0.000]	[0.003]	[0.047]
Southern Europe	-0.4	-0.2	3.3	3.3
	(0.53)	(0.54)	(0.26)	(0.27)
	[0.396]	[0.663]	[0.000]	[0.000]
Central & Eastern Europe	-5.4	-4.6	0.7	0.9
	(0.62)	(0.67)	(0.32)	(0.31)
	[0.000]	[0.000]	[0.020]	[0.003]
Balkan	-10.6	-9.9	-1.6	-1.3
	(0.50)	(0.51)	(0.24)	(0.24)
	[0.000]	[0.000]	[0.000]	[0.000]
Middle East & North Africa	-11.4	-10.8	-1.8	-1.6
	(0.60)	(0.62)	(0.35)	(0.34)
	[0.000]	[0.000]	[0.000]	[0.000]
Asia	-14.5	-13.7	-0.5	-0.5
	(1.06)	(1.15)	(0.60)	(0.56)
	[0.000]	[0.000]	[0.421]	[0.362]
Sub-Saharan Africa	-13.9	-13.2	-0.3	-0.1
	(0.78)	(0.77)	(0.47)	(0.46)
	[0.000]	[0.000]	[0.497]	[0.900]
North & South America	-5.1	-4.6	3.1	3.4
	(0.84)	(0.85)	(0.55)	(0.54)
	[0.000]	[0.000]	[0.000]	[0.000]
Experience				
None		Ref-Cat.		Ref-Cat.
< 1 year exp.		1.2		1.6
		(0.44)		(0.30)
		[0.007]		[0.000]
1-3 years exp.		3.7		1.1
		(0.42)		(0.27)
		[0.000]		[0.000]
At least 4 years exp.		9.2		0.9
		(0.49)		(0.31)
		[0.000]		[0.005]
Profile length				
11-16 rows		Ref-Cat.		Ref-Cat.
17-22 rows		-1.9		5.0
		(0.39)		(0.27)
		[0.000]		[0.000]
23-28 rows		-2.6		9.0
		(0.49)		(0.37)
		[0.000]		[0.000]
>28 rows		-4.2		12.8
		(0.52)		(0.41)
		[0.000]		[0.000]
Mean contact rate/time (in sec)	0.45	0.45	8.94	8.94
Model	Baseline	No interact.	Baseline	No interact.
Observations	3774476	3774476	3774476	3774476

Notes: The table reports the effects (in %) of jobseekers' ethnicity, work experience and the profile length on the contact rate and time spent on profile. The sample also comprises profiles of jobseekers that have been visited for less than 2 seconds. The standard errors (in parentheses) are clustered at the recruiter level. Columns (1) and (2) report the effects on contact rate, columns (3) and (4) on (log) time on profile. Columns (1) and (3) are estimated with our baseline model with all covariates and selected first order interactions. Columns (2) and (4) are estimated without first-order interactions in order to show the direct effects of some of the covariates. Standard errors (shown in parentheses) are clustered at recruiter level. P-values of a two-sided t-test against the null hypothesis of no effect are shown in square brackets. No adjustments were made for multiple comparisons.

4. Based on this measure, we rank all jobseekers within a given search according to their predicted employability. This yields the employability rank of each jobseeker within a given search. We then divide this rank by the number of candidates in the search (i.e. the number of search results) in order to get a measure of the relative position of a jobseeker within the employability distribution of the search. The motivation for this normalization is that a given absolute rank, say 10, has not the same effect if the candidate pool consists of 12 candidates rather than 100 candidates. By construction, the relative rank varies from 0 to 1.

We then enrich our baseline model with binary indicators for each decile of the relative employability rank of the candidate within the pool of candidates. As expected, the candidate's relative position within the pool of candidates has a large effect on recruiters' contact decisions *conditional* on observed skills: as shown in Table 7, the chances to be contacted are much lower for candidates that belong to the lower deciles of the relative employability index. Importantly, Table 7 shows that the ethnicity effects are virtually identical across the three specifications that include the rank controls from our baseline specification, no rank controls, and the relative employability rank.

Table 7.

Robustness of ethnicity results: Controlling for employability rank.

	(1)	(2)	(3)
Switzerland	Ref-Cat.	Ref-Cat.	Ref-Cat.
Western & Northern Europe	-4.1	-4.1	-4.1
	(0.52)	(0.52)	(0.52)
	[0.000]	[0.000]	[0.000]
Southern Europe	-0.5	-0.5	-0.5
	(0.63)	(0.63)	(0.63)
	[0.417]	[0.413]	[0.384]
Central & Eastern Europe	-6.4	-6.4	-6.2
	(0.85)	(0.85)	(0.85)
	[0.000]	[0.000]	[0.000]
Balkan	-12.8	-12.8	-12.8
	(0.62)	(0.62)	(0.62)
	[0.000]	[0.000]	[0.000]
Middle East & North Africa	-14.3	-14.3	-14.3
	(0.79)	(0.79)	(0.78)
	[0.000]	[0.000]	[0.000]
Asia	-20.3	-20.3	-20.2
	(1.20)	(1.20)	(1.20)
	[0.000]	[0.000]	[0.000]
Sub-Saharan Africa	-17.6	-17.6	-17.6
	(1.07)	(1.06)	(1.07)
	[0.000]	[0.000]	[0.000]
North & South America	-6.4	-6.3	-6.3
	(1.13)	(1.13)	(1.13)
	[0.000]	[0.000]	[0.000]
1st decile empl. index			Ref-Cat.
2nd decile empl. index			-2.1
			(0.45)
			[0.000]
3rd decile empl. index			-3.0
			(0.52)
			[0.000]
4th decile empl. index			-3.8
			(0.55)
			[0.000]
5th decile empl. index			-6.1
			(0.61)
			[0.000]
6th decile empl. index			-6.6
			(0.72)
			[0.000]
7th decile empl. index			-8.5
			(0.81)
			[0.000]
8th decile empl. index			-10.7
			(0.93)
			[0.000]
9th decile empl. index			-14.0
			(0.93)
			[0.000]
10th decile empl. index			-16.3
			(0.79)
			[0.000]
Mean dep. var.	0.42	0.42	0.42
Rank controls	Baseline	No rank	Empl. rank
Observations	1633833	1633833	1633833

Note: The table checks whether including a measure for the relative employability of a jobseeker compared to the other jobseekers in the search list changes our estimates. Column 1 reports our baseline estimates. They include detailed controls for the absolute rank of a candidate in the search list (up to 100 dummies) and controls for the relative rank, i.e. the ratio of the absolute rank and the number of results in the result list. Column (2) reports estimates without any controls for the rank of a candidate. Column (3) enriches the baseline model with binary indicators for each decile of the relative employability rank of the candidate within the pool of candidates (see text for details). For the sake of comparability, we use the *test* sample, used to construct jobseeker's employability rank, in all specifications. Standard errors (shown in parentheses) are clustered at recruiter level. P-values of a two-sided t-test against the null hypothesis of no effect are shown in square brackets. No adjustments were made for multiple comparisons.

Further Analyses and Extensions

In this section, we present several extensions to our main regression results to draw out further implications and shed light on potential mechanisms.

First, we assess whether recruiters become more or less discriminatory as the search drags on. To this end, we leverage variation in how long the same recruiter has been searching in the same session.[9] Since recruiters vary in their average session length, and these compositional differences can influence the results if, e.g., recruiters who in general search longer are less discriminatory, we focus our analysis on the first 30 minutes for sessions that last 30 minutes or longer. We measure the search duration, i.e. the time that a recruiter has been searching measured from the start of the session until the recruiter opens a jobseekers' profile view, and group this duration variable in four roughly equal-sized bins.

Extended Data Figure 4a shows the ethnic penalty for Europeans and non-Europeans immigrants relative to natives depending on search duration. While the differences in ethnic penalties across search duration are statistically significant for European immigrants, there are no significant differences for non-European immigrants. Overall, the results pattern suggests that, if anything, discrimination declines with longer sessions. One possible explanation is that recruiters (have to) become less picky, the longer it takes for them to fill a vacancy.[10]

Next, we investigate whether recruiters that spend more time screening the profiles of candidates exhibit more or less discrimination than recruiters that decide more quickly. To study this question, we first use a random subset of one third of searches from each recruiter to assign recruiters to one of four groups: fast deciders (lowest quartile), fairly fast deciders (second quartile), fairly slow deciders (third quartile) and slow deciders (highest quartile). To ensure that the training sample allows us to classify recruiters, we restrict this analysis to recruiters that conduct at least 6 searches and have at least 12 profile views over the study period. As expected, these four groups constructed from the 1/3 of searches in the training sample are highly predictive of recruiters' decision time in the 2/3 test sample: in the test sample, the average time that recruiters are looking at jobseekers' profiles are 5.7 seconds for fast deciders, 9.9 seconds for fairly fast deciders, 15.4 seconds for fairly slow deciders, and 31.7 seconds for slow deciders.

Extended Data Figure 4b shows the estimated ethnic penalties for the fast- and slow-decision groups based on the test sample. Interestingly, the figure provides little evidence that ethnic penalties vary depending on the time that a recruiter usually spends on a profile before taking a decision. When comparing different recruiters, fast deciders are thus not necessarily more discriminatory than slow deciders. One possible explanation for this pattern is that time to decision is a function of recruiters' experience with the platform: the more experienced you are, the easier you find the relevant information on the platform. Similarly, fast deciders likely search in occupations with simpler (shorter) CVs, and these occupations might exhibit generally more/less discrimination. These considerations suggest that this cross-sectional comparison between fast and slow deciders is not necessarily informative regarding the effect of decision time on discriminatory behavior.[11]

[9] A session is defined as the time from when a recruiter logs on to Job-Room until she or he logs off or closes the window.

[10] Note that these within-search findings do not contradict the results that discrimination increases before noon and towards the end of the workday, when recruiters are arguably more exhausted from conducting multiple search sessions for different vacancies.

[11] Note that the absence of a correlation between ethnic penalties and time to decision between recruiters does not contradict our result that discrimination varies depending on the time of the day. The latter results are based on within-recruiter variation, where we find that the same recruiter discriminates more during those hours when s/he is

Next, we leverage the information provided in the unstructured text field "additional skills" to gauge the role of statistical discrimination. In many cases, this field contains additional information on the skills and competencies of jobseekers, for instance regarding responsibilities and roles in prior jobs (e.g. leadership roles), on non-cognitive skills and competencies (e.g. leadership qualities and team-work abilities), and more details on work experience and educational credentials (see Table 1). The extent to which this text field is filled out varies between jobseekers. These differences are not only a function of the skills and experience of the jobseekers, but, importantly, also of varying efforts of case workers to elicit this information from jobseekers (and enter it into the database) during the enrolment interview. As a consequence of the latter, there are differences between similar jobseekers in terms of the amount of information on merit-based characteristics provided to recruiters.

Extended Data Figure 6b shows the effects of ethnicity on the probability of being contacted, where we interact the ethnic penalties with the length of the unstructured text field of each jobseeker. Following [5], researchers have exploited such variation to study the relevance of statistical discrimination in accounting for discrimination in correspondence studies. Perhaps surprisingly, the figure does not provide strong evidence that a longer unstructured text reduces discrimination for most groups. The exception are the two groups that are most discriminated against: jobseekers from Asia and Sub-Saharan Africa.

Extending these analyses on statistical discrimination, we can exploit the link to jobseekers' entries in the Swiss unemployment register to explore whether unobserved ability can explain the pattern of discrimination that we uncover on the platform. In particular, we observe jobseekers' age, marital status, and wage insured at the public employment service—a good proxy of the jobseeker' monthly earnings prior to unemployment. These characteristics are unobserved by recruiters on Job-Room but, particularly in the case of age and prior wage, very likely to influence workers' attractiveness to employers. Using these unobserved productivity measures, we construct an index of "unobserved employability" and estimate separate ethnicity penalties for individuals with low, medium and high unobserved employability.

The construction and analysis of the index involves several steps.

1. First, we split the sample of jobseekers randomly in a 50% training and 50% test sample to separate index construction from index analysis.

2. Second, we regress a dummy variable that the jobseeker leaves unemployment within the first four months after registering at the Swiss public employment service on the unobserved characteristics. We interact the unobserved variables with 9 occupation dummies because the effects of the unobserved characteristics on the employability of a candidate likely differ by occupation.[12] For instance, we thus allow that workers' age has an occupation-specific effect on employability. The regression controls for occupation-period and canton fixed effects.

3. We use the occupation-specific regression coefficients for the "unobserved employability" index to predict the likelihood that jobseekers in the test sample leave unemployment within four months after registration. Extended Data Table 2 shows the average unobserved employability across ethnic groups. Importantly, we find that ethnic minorities have, if anything, *higher* unobserved employability than Swiss jobseekers.

deciding faster.

[12] We use the first-level of the Swiss occupational classification scheme, which closely mirrors the occupational categories used on the platform.

4. Based on this prediction, we assign jobseekers in the test sample to one of three equally-sized groups (splitting the test sample at the 33rd and 66th percentile): low, medium, and high unobserved employability. Validating our measure, we find that the unobserved employability index explains a substantial part of the variation in the unemployment duration of jobseekers in the test sample. On average, a jobseeker with high unobserved employability faces a probability of 37% to leave unemployment within 4 months. The corresponding probability decreases to 20% for jobseekers in the middle group and to 10% for jobseekers with low unobserved employability.

5. We reestimate our baseline model using the test sample and estimate the ethnicity penalty separately for the three unobserved employability groups by subsetting the sample.

Table 8 shows the effect of the unobserved employability index on unemployment duration for all origins, Swiss, Europeans, and non-Europeans. To calculate these estimates, we split the full sample in training and test subsamples. We calculate the unobserved employability index on the training sample. Using the test sample, we regress a binary indicator for whether a person leaves unemployment within 4 months after registration on a binary indicator equal to one if a jobseeker belongs to a specific tercile of the unobserved employability index. The first column of Table 8 shows the effect of medium and high unobserved employability relative to low unobserved employability for the test sample covering all nationalities. We see that jobseekers with high unobserved employability have a 17 percentage points higher probability to leave unemployment within 4 months (controlling for all jobseeker characteristics visible to recruiters on Job-Room) relative to the low employability group. Columns 2-4 show that this effect is fairly similar for Swiss, European, and Non-European jobseekers.[13] Overall, these results indicate that the characteristics subsumed in the index are positively related to workers' chances to find a job, and that the direction and strength of this relationship is similar across all origin groups.

Extended Data Figure 6a shows ethnic penalties for individuals with low, medium or high unobserved employability. We find that the differences in ethnic penalties between the three unobserved employability groups are substantively very small and despite the large sample size for most origin regions not statistically significant (the p-values from F-tests that the ethnic penalties are the same across groups are: p=0.07 for Western and Northern Europe; p=0.04 for Southern Europe; p=0.33 for Central and Eastern Europe; p=0.02 for the Balkan countries; p=0.60 for MENA countries; p=0.80 for Asia; p=0.97 for Sub-Saharan Africa, and p=0.23 for North- and South America). Together, this suggests that ethnic minority jobseekers with favorable unobserved characteristics face a similar disadvantage as those with medium and low employability. Furthermore, the finding that, on average, all ethnic minority groups possess the same or higher unobserved employability than Swiss natives, also suggests that our results are unlikely to be explained by statistical discrimination alone.

Next, we provide additional heterogeneity analyses by grouping different types of recruiters according to patterns in their search and click behavior. We distinguish between "power users" who use the platform frequently and "light" users that only conduct a few searches over the study period. Extended Data Figure 3a shows the effects of ethnicity on the contact rate for recruiters with one search, 1-10 searches, 11-50 searches and more than 50 searches during the nine-month

[13] Note, that we assigned job seekers to the different ethnic groups along the lines outlined in the paper. I.e. they have to unite all three signals of ethnicity on the platform: nationality, name and language. This is the reason why the number of observations from columns 2-4 do not sum up to the number of observations in column 1.

Table 8.
Effect of unobserved employability on probability of leaving unemployment within 120 days.

	(1) All	(2) Swiss	(3) Europe	(4) Non-Europe
Low unobserved employability	Ref-Cat.	Ref-Cat.	Ref-Cat.	Ref-Cat.
Medium unobserved employability	0.0464	0.0547	0.0531	0.0264
	(0.0021)	(0.0036)	(0.0044)	(0.0055)
	[0.000]	[0.000]	[0.000]	[0.000]
High unobserved employability	0.1704	0.1911	0.1606	0.1549
	(0.0029)	(0.0049)	(0.0056)	(0.0068)
	[0.000]	[0.000]	[0.000]	[0.000]
Canton and occup times month FE	Yes	Yes	Yes	Yes
Characteristics visible on Job-Room	Yes	Yes	Yes	Yes
Observations	156828	63879	43962	22167

Notes: The table shows the percentage points effect of unobserved employability on the likelihood of leaving unemployment for different ethnicities using the test sample. The dependent variable is a dummy equal to one whether a person leaves unemployment within 4 months after registration. The explanatory variables are dummy variables equal to one if a jobseeker belongs to a specific tercile of the index of unobserved employability. Column (1) shows the effects for all jobseekers in the test sample while columns (2)-(4) restrict the sample to Swiss (2), European (3) and Non-European (4) jobseekers. We control in all regressions for canton fixed effects (FE) and occupation times month fixed effects as well as for all jobseeker characteristics visible to recruiters on Job-Room. Standard errors are shown in parentheses. P-values of a two-sided t-test against the null hypothesis of no effect are shown in square brackets. No adjustments were made for multiple comparisons.

study period. To facilitate comparisons, the coefficients are normalized with the mean contact rate of the respective category. The figure provides little evidence that discriminatory preferences differ between "power users" and "light" users of the platform. Instead, the results suggest that discrimination is as likely among more experienced and/or professional recruiters as it is among those that conduct only two handful of searches.

Another way to look at heterogeneity between different types of recruiters is to compare them by registration status. As discussed above, registered users are typically private recruitment agencies who can register in order to get preferential access on Job-Room, which includes a limited set of additional, protected information about jobseekers. In particular, registered recruiters also see the nationality and—if the jobseeker consents when registering—their names. Extended Data Figure 3b shows that registered users discriminate at least as much as non-registered users. For the interpretation of these results, it is important to keep in mind that non-registered recruiters can only infer ethnicity from jobseekers' language skills. They can thus only identify the ethnic origin of jobseekers who registered languages that are typical for a certain region. This will make it hard to differentiate between jobseekers from certain countries, for example in North America and Sub-Saharan Africa, where the most common language is English or South America, where the most common language is Spanish. It is therefore not surprising that non-registered users are not (able to) significantly discriminate against jobseekers from these regions.

In addition, we provide further benchmarks to assess the substantive size and implications of the estimated ethnic penalties. We can benchmark the estimated ethnic penalties against the effect of possessing "very good" German language skills (in comparison to not having German language skills). Table 10 reveals that ethnic penalties that, e.g., candidates from the Balkans face relative to natives, are more than twice as large (in absolute terms) as the benefits from acquiring "very good" (compared to no) German language skills.

Lastly, we can use our ethnic penalty estimates for a back-of-the-envelope calculation of the direct costs of discrimination on Job-Room on search duration and earnings. By necessity, such an analysis has to rely on a range of additional assumptions. Our estimates imply a reduction in contact attempts for non-European candidates relative to natives by -12.9% on average. Each contact attempt, in turn, increases the likelihood to leave unemployment within 90 days after the search by about 2.1% (c.f. Extended Data Figure 7). On average, a candidate receives 1.4 contact attempts per month while registered on Job-Room and the average registration duration is about 10 months. Under the assumptions of a constant job arrival rate and that discrimination on Job-Room is the same as on other search channels, this implies that the search duration for Non-European jobseekers is 39 days longer because of discrimination. To estimate the effect on earnings, we assume that the new job would have paid 94.6% of the salary of the last job before the unemployment spell.[14] Since the average, unemployed Non-European immigrant registered on the platform earns CHF 2,643 in the last month prior to unemployment, the discrimination-induced longer search duration suggests a reduction in earnings at the order of CHF 3,197 (= 2643 x 0.946 x 39/30.5) per jobseeker and unemployment spell (the CHF/USD exchange rate is about 1).

[14] This wage reduction is based on the average unemployment duration of unemployed jobseekers in our study and the pre-/post unemployment wage differential estimated from a similar sample[6].

Supplementary Tables

Table 9.
Examples of public and private employment websites.

Platform	Country	Platform owner
Privately-owned		
Monster.com	International	Monster Worldwide, Inc.
CareerBuilder.com	International	Apollo Global Management
cv-library.co.uk	United Kingdom	Lee Biggins
jobstreet.com	Southeast Asia	JobStreet Corporation Berhad
Government-affiliated		
USAJobs.gov	USA	United States Office of Personnel Management
ec.europa.eu/eures	Europe	European Employment Services
jobbank.gc.ca	Canada	Employment and Social Development Canada
workindenmark.dk/Employer	Denmark	Danish agency for labour market and recruitment
entreprise.pole-employ.fr	France	French public employment service
jobboerse.arbeitsagentur.de	Germany	German public employment service
empleate.gob.es	Spain	Spanish ministry of labour, migrations and social security
arbetsformedlingen.se	Sweden	Swedish public employment service

Notes: The table shows examples of large privately-owned and government-affiliated employment websites all over the world. On these websites, recruiters can search for candidates for their open vacancies.

Table 10.
Effects of jobseekers' ethnicity, work experience, German skills and profile length on contact rate and time on profile.

	(1) Contact rate	(2) Contact rate	(3) Log time	(4) Log time
Switzerland	Ref-Cat.	Ref-Cat.	Ref-Cat.	Ref-Cat.
Western & Northern Europe	-4.2	-3.3	0.5	0.2
	(0.42)	(0.42)	(0.23)	(0.22)
	[0.000]	[0.000]	[0.047]	[0.298]
Southern Europe	-0.4	-0.1	2.7	2.7
	(0.57)	(0.56)	(0.26)	(0.26)
	[0.481]	[0.877]	[0.000]	[0.000]
Central & Eastern Europe	-6.2	-5.5	0.2	0.3
	(0.72)	(0.77)	(0.34)	(0.33)
	[0.000]	[0.000]	[0.526]	[0.311]
Balkan	-12.6	-11.7	-1.7	-1.6
	(0.58)	(0.60)	(0.25)	(0.25)
	[0.000]	[0.000]	[0.000]	[0.000]
Middle East & North Africa	-13.5	-12.8	-2.1	-2.0
	(0.69)	(0.71)	(0.36)	(0.35)
	[0.000]	[0.000]	[0.000]	[0.000]
Asia	-18.5	-17.6	-0.4	-0.6
	(0.98)	(0.99)	(0.63)	(0.60)
	[0.000]	[0.000]	[0.560]	[0.338]
Sub-Saharan Africa	-17.1	-16.1	-0.4	-0.0
	(0.92)	(0.90)	(0.52)	(0.50)
	[0.000]	[0.000]	[0.472]	[0.973]
North & South America	-6.4	-5.5	2.6	2.7
	(0.93)	(0.94)	(0.56)	(0.56)
	[0.000]	[0.000]	[0.000]	[0.000]
Experience				
None		Ref-Cat.		Ref-Cat.
< 1 year exp.		1.5		1.2
		(0.49)		(0.28)
		[0.003]		[0.000]
1-3 years exp.		4.8		0.8
		(0.42)		(0.26)
		[0.000]		[0.003]
At least 4 years exp.		11.5		0.7
		(0.50)		(0.28)
		[0.000]		[0.009]
German skills				
No German skills		Ref-Cat.		Ref-Cat.
Very good German skills		5.5		0.8
		(0.35)		(0.19)
		[0.000]		[0.000]
Profile length				
11-16 rows		Ref-Cat.		Ref-Cat.
17-22 rows		-2.1		5.0
		(0.39)		(0.30)
		[0.000]		[0.000]
23-28 rows		-2.8		8.6
		(0.55)		(0.40)
		[0.000]		[0.000]
>28 rows		-4.4		11.9
		(0.59)		(0.46)
		[0.000]		[0.000]
Mean contact rate/time (in sec)	0.42	0.42	10.26	10.26
Model	Baseline	No interact.	Baseline	No interact.
Observations	3251303	3251303	3191818	3191818

Notes: The table reports the estimates corresponding to Figure 1 in the main text and contains the effects (in %) of jobseekers' ethnicity, work experience, German skills, and profile length. Columns (1) and (2) report the effects on contact rate, columns (3) and (4) on (log) time on profile. Columns (1) and (3) are estimated with our baseline model with all covariates and selected first-order interactions. Columns (2) and (4) are estimated without first-order interactions in order to show the main effects of the covariates. Standard errors (shown in parentheses) are clustered at recruiter level. P-values of a two-sided t-test against the null hypothesis of no effect are shown in square brackets. No adjustments were made for multiple comparisons.

Table 11.
Variation in ethnic penalties over the work day.

	(1) Europe	(2) Non-Europe	(3) Time on profile
8-9	-1.7	-11.6	10.5
	(0.89)	(1.04)	(0.19)
	[0.049]	[0.000]	[0.000]
9-10	-2.0	-11.5	10.5
	(0.68)	(0.88)	(0.15)
	[0.004]	[0.000]	[0.000]
10-11	-0.9	-12.1	10.2
	(0.65)	(0.85)	(0.12)
	[0.171]	[0.000]	[0.000]
11-12	-3.1	-13.4	9.7
	(0.72)	(0.94)	(0.11)
	[0.000]	[0.000]	[0.000]
12-13	-0.5	-15.8	11.9
	(2.06)	(2.97)	(0.27)
	[0.805]	[0.000]	[0.000]
13-14	-2.0	-11.5	10.8
	(0.74)	(0.93)	(0.18)
	[0.008]	[0.000]	[0.000]
14-15	-1.7	-12.1	10.5
	(0.65)	(0.86)	(0.14)
	[0.008]	[0.000]	[0.000]
15-16	-2.4	-12.6	10.4
	(0.71)	(0.93)	(0.13)
	[0.001]	[0.000]	[0.000]
16-17	-4.1	-13.6	10.1
	(0.83)	(0.99)	(0.12)
	[0.000]	[0.000]	[0.000]
17-18	-3.0	-14.7	9.5
	(1.09)	(1.23)	(0.14)
	[0.006]	[0.000]	[0.000]
Observations	3341209	3341209	3281297

Notes: The table reports the estimates corresponding to Figure 2 in the main text. It shows the ethnic penalties (in %) for jobseekers from Europe (Column (1)) and outside of Europe (Column (2)) compared to natives over the work day. Column (3) shows the estimates for the average time a recruiter spends on the profile of a jobseeker. Standard errors (shown in parentheses) are clustered at recruiter level. P-values of a two-sided t-test against the null hypothesis of no effect are shown in square brackets. No adjustments were made for multiple comparisons.

Table 12.
Heterogeneity of ethnic discrimination by length of skill field.

	(1) No skills	(2) 1-10 words	(3) >10 words
Switzerland	Ref-Cat.	Ref-Cat.	Ref-Cat.
Western & Northern Europe	-4.8	-3.7	-3.3
	(0.52)	(0.69)	(0.65)
	[0.000]	[0.000]	[0.000]
Southern Europe	-0.7	-0.5	0.3
	(0.66)	(0.77)	(0.79)
	[0.306]	[0.545]	[0.750]
Central & Eastern Europe	-5.6	-5.5	-8.7
	(0.81)	(1.04)	(1.07)
	[0.000]	[0.000]	[0.000]
Balkan	-12.5	-12.6	-12.9
	(0.71)	(0.71)	(0.75)
	[0.000]	[0.000]	[0.000]
Middle East & North Africa	-13.1	-14.1	-13.8
	(0.89)	(1.03)	(1.05)
	[0.000]	[0.000]	[0.000]
Asia	-18.9	-20.7	-15.3
	(1.26)	(1.75)	(1.74)
	[0.000]	[0.000]	[0.000]
Sub-Saharan Africa	-17.6	-19.3	-13.9
	(1.25)	(1.46)	(1.45)
	[0.000]	[0.000]	[0.000]
North & South America	-6.6	-8.3	-3.9
	(1.28)	(1.42)	(1.77)
	[0.000]	[0.000]	[0.028]
Mean contact rate	0.42	0.42	0.42
Observations	3251263	3251263	3251263

Notes: The table reports the estimates corresponding to Figure 6b in the extended data section. It shows estimates for the ethnic penalties by length of the unstructured text field containing additional skills of jobseekers. Standard errors (shown in parentheses) are clustered at recruiter level. P-values of a two-sided t-test against the null hypothesis of no effect are shown in square brackets. No adjustments were made for multiple comparisons.

Table 13.
Gender-based hiring disparities across occupations.

	Penalty	SE	P-value	Fem. share results	Fem. share occup.
Skilled forestry, fishery and hunting workers	-50.49	21.36	0.02	2.47	1.61
Assemblers	-27.53	4.06	0.00	4.43	19.43
Skilled agricultural workers	-23.34	2.48	0.00	15.51	14.54
Agricultural, forestry and fishery labourers	-22.29	4.45	0.00	12.08	14.14
Food process., garment, other craft and related trades workers	-14.61	2.03	0.00	13.49	30.89
Stationary plant and machine operators	-12.41	3.31	0.00	7.69	27.02
Cleaners and helpers	-11.97	6.26	0.06	68.70	87.41
Production and specialised services managers	-11.48	9.01	0.20	26.03	31.67
Drivers and mobile plant operators	-11.14	2.12	0.00	1.93	6.19
Metal, machinery and related trades workers	-8.52	1.76	0.00	1.97	7.76
Building and related trades workers, excl. electricians	-6.76	1.50	0.00	2.16	6.43
Science and engineering associate prof.	-5.96	1.61	0.00	11.16	15.67
Hospitality, retail and other services managers	-4.11	11.75	0.73	44.20	58.03
Electrical and electronic trades workers	-2.26	2.57	0.38	1.14	2.12
Numerical and material recording clerks	-1.27	4.95	0.80	26.62	27.94
Information and communications technology prof.	0.76	2.51	0.76	10.24	12.03
Information and communications technicians	1.51	6.83	0.83	11.05	35.10
Business and administration professionals	2.06	2.84	0.47	27.12	41.74
Sales workers	6.03	2.72	0.03	52.97	71.54
Science and engineering professionals	6.12	2.26	0.01	26.82	22.13
Legal, social, cultural and related associate prof.	8.22	6.76	0.22	40.08	57.78
Health professionals	9.17	4.35	0.04	74.36	67.54
Personal care workers	9.65	5.53	0.08	88.98	88.70
Health associate professionals	10.15	3.16	0.00	78.12	87.22
Administrative and commercial managers	10.51	6.18	0.09	30.42	36.98
Handicraft and printing workers	10.59	2.93	0.00	29.68	41.08
Personal service workers	11.38	3.33	0.00	48.01	61.88
Chief executives, senior officials and legislators	12.04	13.50	0.37	20.87	23.31
Refuse workers and other elementary workers	12.22	5.94	0.04	6.05	32.02
Business and administration associate prof.	12.34	2.22	0.00	51.13	62.24
Teaching professionals	13.48	9.50	0.16	66.37	65.86
Protective services workers	14.99	6.85	0.03	10.68	20.35
Customer services clerks	15.63	2.66	0.00	62.28	73.62
Other clerical support workers	15.75	9.24	0.09	82.93	57.42
Mining, construction, manufacturing and transport	19.33	5.01	0.00	9.12	25.82
Street and related sales and service workers	19.62	25.47	0.44	40.20	100.00
Legal, social and cultural professionals	29.75	17.00	0.08	53.91	56.97
General and keyboard clerks	33.26	6.63	0.00	76.72	84.65

Notes: The table reports the estimates corresponding to Figure 3 in the main text. It shows the gender penalties (first column) by (two-digit) ISCO-occupations (n=17,369,372), their respective standard errors (second column), the share of women in the search list (third column) and the share of women in the occupation (fourth column), calculated using the Swiss earnings structure survey.

Table 14.
Language skills by ethnicity.

	Northwest Europe	Southern Europe	Eastern Europe	Balkan	MENA	Asia	Africa	America
German	0.72	0.16	0.20	0.34	0.31	0.15	0.09	0.16
French	0.28	0.21	0.05	0.06	0.19	0.04	0.29	0.20
Italian	0.02	0.43	0.03	0.02	0.02	0	0.02	0.10
Rumansh	0	0	0	0	0	0	0	0
English	0.37	0.09	0.22	0.04	0.08	0.22	0.30	0.25
Spanish	0.03	0.20	0.02	0	0.01	0	0.01	0.62
Portuguese	0.01	0.49	0	0	0	0	0.06	0.33
Turkish	0.02	0	0	0.01	0.62	0.01	0	0
Greek	0	0	0	0	0	0	0	0
Hungarian	0	0	0.21	0	0	0	0	0
Polish	0.01	0	0.29	0	0	0	0	0
Czech	0	0	0.11	0	0	0	0	0
Serbo-Croatian	0.01	0	0.02	0.11	0	0	0	0
Dutch	0.03	0	0	0	0	0	0	0
Flemish	0	0	0	0	0	0	0	0
Arabic	0.02	0	0	0	0.33	0	0.06	0
Hebrew	0	0	0	0	0	0	0	0
Russian	0.02	0	0.14	0	0	0.01	0	0
Swedish	0.01	0	0	0	0	0	0	0
Japanese	0	0	0	0	0	0.02	0	0
Chinese	0	0	0	0	0	0.08	0	0
Slovenian	0	0	0.07	0	0	0	0	0
Croatian	0	0	0.01	0.07	0	0	0	0
Danish	0.01	0	0	0	0	0	0	0
Tamil	0	0	0	0	0	0.40	0	0
Albanian	0.02	0.02	0.01	0.65	0	0	0	0
Kurdish	0	0	0	0	0.21	0	0	0
Swiss German	0.07	0.07	0.02	0.14	0.14	0.05	0.03	0.05
Serbian	0.01	0	0.02	0.16	0	0	0	0
Macedonian	0	0	0	0.08	0	0	0	0
Bosnian	0	0	0.01	0.06	0	0	0	0
Bulgarian	0	0	0	0.01	0	0	0	0
Norwegian	0	0	0	0	0	0	0	0
Slovakian	0	0	0.20	0	0	0	0	0
Lithuanian	0	0	0.01	0	0	0	0	0
Thai	0	0	0	0	0	0.05	0	0
Finnish	0	0	0	0	0	0	0	0
Khmer	0	0	0	0	0	0	0	0
Vietnamese	0	0	0	0	0	0.04	0	0
Romanian	0.01	0	0.09	0.01	0	0	0	0

Notes: The table displays the percentage of jobseekers from a given origin region (columns) who reported very good knowledge of a given language (rows). Note that "very good knowledge" of a language is the highest of four levels that can be entered on Job-Room. The other levels are: "No knowledge", "basic knowledge", and "good knowledge". The table contains *all* languages shown to recruiters on Job-Room. The percentages are rounded to two decimal places. The table also illustrates why non-registered users, who do not see the jobseeker's name or nationality and can only infer ethnicity based on language skills, will find it hard to identify individuals from Western and Northern Europe, Sub-Saharan Africa, and North and South America. Jobseekers from these regions often speak languages that are not specific to these regions.

Table 15.
The effect of ethnicity on the contact rate (in %) estimated with OLS and logistic regression.

	(1) OLS	(2) OLS	(3) Logit
Switzerland	Ref-Cat.	Ref-Cat.	Ref-Cat.
Western & Northern Europe	-4.2	-3.3	-3.3
	(0.42)	(0.42)	(0.24)
	[0.000]	[0.000]	[0.000]
Southern Europe	-0.4	-0.1	-0.1
	(0.57)	(0.56)	(0.23)
	[0.481]	[0.877]	[0.537]
Central & Eastern Europe	-6.2	-5.5	-5.3
	(0.72)	(0.77)	(0.34)
	[0.000]	[0.000]	[0.000]
Balkan	-12.6	-11.7	-11.3
	(0.58)	(0.60)	(0.30)
	[0.000]	[0.000]	[0.000]
Middle East & North Africa	-13.5	-12.8	-12.7
	(0.69)	(0.71)	(0.43)
	[0.000]	[0.000]	[0.000]
Asia	-18.5	-17.6	-17.1
	(0.98)	(0.99)	(0.72)
	[0.000]	[0.000]	[0.000]
Sub-Saharan Africa	-17.1	-16.1	-15.8
	(0.92)	(0.90)	(0.62)
	[0.000]	[0.000]	[0.000]
North & South America	-6.4	-5.5	-5.2
	(0.93)	(0.94)	(0.59)
	[0.000]	[0.000]	[0.000]
Model	Baseline	No interact	No interact
Observations	3251303	3251303	2797417

Notes: The table compares estimates of the ethnicity effect based on OLS with estimates based on logistic regression. Column (1) reports the baseline estimates of the ethnicity effect on the contact rate from Figure 1, Panel A, estimated with OLS. Column (2) reports the OLS estimates based on a model without interaction terms as reported in Table 10. Column (3) reports the average marginal effects of ethnicity estimated with logistic regression. The model contains the same set of covariates as the model in column (2). Standard errors (shown in parentheses) are clustered at recruiter level for the OLS estimates. P-values of a two-sided t-test against the null hypothesis of no effect are shown in square brackets. No adjustments were made for multiple comparisons.

References

[1] Lakha, F., Gorman, D. & Mateos, P. Name analysis to classify populations by ethnicity in public health: validation of Onomap in Scotland. *Public Health* **125**, 688–696 (2011).

[2] Mateos, P. Names, ethnicity and populations. *Advances in Spatial Science* (2014).

[3] Belloni, A., Chen, D., Chernozhukov, V. & Hansen, C. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* **80**, 2369–2429. (2012).

[4] Belloni, A., Chernozhukov, V. & Hansen, C. Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies* **81**, 608–650 (2014).

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

[6] Ecoplan. Auswirkungen der Arbeitslosigkeit auf Einkommen und Erwerbsbiografien. Tech. Rep., SECO (2013).