IFAU
Institute for Evaluation of Labour
Market and Education Policy

Stereotypes of physical appearance and labor market chances

Mahmood Arai
Marie Gartell
Magnus Rödin
Gülay Özcan

WORKING PAPER 2016:20

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala. IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

More information about IFAU and the institute's publications can be found on the website www.ifau.se

ISSN 1651-1166

Stereotypes of physical appearance and labor market chances

Mahmood Arai*, Marie Gartell†, Magnus Rödin†and Gülay Özcan†‡

November 22, 2016

Abstract

Using an experimental setup involving 436 case workers at the Swedish Public Employment Service (SPES) as subjects and the profile photographs and recorded voices of 75 jobseekers as treatments, we report results indicating that male case workers tend to favor jobseekers perceived as having a stereotypical Swedish appearance when they select candidates to be recommended for labor market programs (LMP). This bias represents a roughly 50-percent greater chance of being selected if you compare the candidate with the highest score with regard to stereotypical Swedish looks (8/10) with the candidate with the lowest score (3/10) in our sample.

Keywords: Discrimination; Ethnic Discrimination; Racial Discrimination; Stereotypes; Noncognitive Attributes
JEL: J7; J15

*Stockholm University.

†Swedish Public Employment Services (SPES). The experiments were performed while Marie Gartell was at SPES. Gartell's current affiliation is Swedish National Audit Office.

‡Financial support by IFAU is gratefully acknowledged. We are grateful to seminar participants at at IFAU, Linnaeus university, and BFH Conference on Discrimination in the Labor Market for helpful comments. We are grateful to Helge Bennmarker, Magnus Carlsson and Olof Åslund for helpful suggestions.

1 Introduction

Do beliefs about stereotypical Swedish looks associated with a jobseeker's physical appearance affect his/her labor market chances? Although Sweden does not have census statistics related to self-reported race or ethnicity, as is the case in the US and the UK, it is not uncommon to observe references to expressions that link stereotypes of physical appearance to nationality. The notion of a "stereotypical Swedish look" or "a Swedish look" / "foreign look" are used in the Swedish context and can be seen as a racial narrative.[1]

The aim of this paper is to examine the effect of beliefs about a stereotypical Swedish look associated with the profile photographs of a jobseeker on the likelihood of being selected to attend labor market programs (LMP). Group be- longing signaled by a person's look, name or speech may trigger intergroup-stereotypes activation leading to discriminatory behavior.[2]

Previous studies have used names signaling group affiliation and associated beliefs about group characteristics.[3] We use photographs to focus on stereotypes of appearance related to Swedishness and associated attributes such as the degree of trustworthiness and agreeableness. We include these non-cognitive attributes to examine what subjects see in a stereotypical Swedish appearance.[4]

We ran a computerized experiment involving 436 caseworkers at the Swedish Public Employment Service (SPES) as subjects. We recruited 75 jobseekers at the employment offices, took their profile photographs and

[1]For Stereotypes of physical appearance related to Swedishness see, for example Jeff Werner & Tomas Björk (2014). In this paper, we are not dealing with historical roots and the dynamics of the construction of these narratives and details of their components but search to examine their role for the labor market chances in Sweden.

[2]There is great deal of literature in social psychology that deals with stereotyping, prejudice and discrimination. For early work on racial and ethnic stereotypes see Katz and Braly (1933). For an overview of this literature see Dovidio et al. (2010). In economic theory, differential treatment of individuals based on their perceived group belonging is explained in terms of employers' (or coworkers'/customers') taste (Becker 1957) or differences in group statistics (Arrow, 1973; Phelps, 1972).

[3]See Fershtman and Gneezy 2001; Riach & Rich (2002), Bertrand & Mullainathan (2003), Carlsson & Rooth (2007), Arai & Thoursie (2009), Ahmed (2010).

[4]For impact of non-cognitive attributes on labor market outcomes see Hamermesh & Biddle (1994), Jackson et al. (1995), Averett & Korenman (1996), Heckman (1999), Nyhus & Pons (2005), Borghans et al. (2006), Mobius & Rosenblat (2006), Rooth (2009), Mocan & Tekin (2010), Ruffle & Shtudiner (2010), Rooth (2010), Rödin & Özcan (2011), López Bóo et al. (2013) and Dechter (2015).

recorded their voices. We drew up jobseeker profiles with these 75 photographs and recorded voices in accordance with the layout template that is used for registered jobseekers at the SPES. We also included randomized information about the level of education, previous jobs and unemployment spells in these profiles.

The experiment consists of two separate parts; a *characterization part*, and a *program-assignment part*. In the characterization part involving 160 case workers, we measure scores associated with portraits/ photographs and recorded voices. For physical appearance, we focus on the stereotypical Swedish look and for recorded voices we focus on non-Swedish accents measured as a score associated with the probability of being perceived as having grown up abroad. In the program-assignment part, involving 276 other case workers, we measure scores associated with the jobseekers' ability to successfully complete an LMP, chances of success upon completion of a program and the candidates' chances of being recommended for an LMP when only one out of two candidates can be assigned to a program.

The probability of being assigned to a labor market program is greater for those who have higher scores associated with growing up abroad based on the recorded voice. Caseworkers are instructed to employ a positive bias in favor of those who are perceived as having reduced chances of getting a job when offering labor market programs. Our results indicate that both male and female caseworkers follow these instructions given that higher scores for growing up abroad is an indication of being associated with reduced chances of getting a job.

Our results indicate that men tend to favor jobseekers who more closely resemble the stereotypical Swedish look. Female caseworkers, on the other hand, do not seem to care about appearance when they choose between competing candidates for an LMP. These results are in line with previous studies indicating that discrimination is primarily practiced by men (see, for example, Fershtman and Gneezy, 2001). Notice that we found no significant correlation between the scores for the stereotypical Swedish look and the scores for growing up abroad based on the recorded voices implying that the effect of the stereotypical Swedish look and the effect of growing up abroad can be separated. Moreover, we find that male caseworkers see attributes like being trustworthy, masculine and agreeable in the stereotypical Swedish look when they assess male candidates but when they assess female candidates, they do not see much in the appearance of the female candidates, except the stereotypical Swedishness.

The remaining of the paper is organized as follows. Next Section describes the experimental setup. Section 3 discusses the determinants of the choice problem when assessing two candidates and you have to choose one candidate to recommend for a labor market program. The data generated in the experiment is described in Section 4, while Section 5 presents the results of estimating the impact of the stereotypical look and other attributes of candidates on the probability of being admitted to a labor market program. Finally, Section 6 summarizes the paper.

2 The experiment design

We limited our attention to two types of employment training; nursing and warehousing and logistics training including a forklift driver license course. In this way we studied a male-dominated and a female-dominated occupation.

Job seekers

We recruited 75 unemployed jobseekers (33 men and 42 women) at the SPES offices in the municipality of Stockholm. We went to the employment offices and asked jobseekers whether they wanted to participate in an experiment regarding labor market programs. The participation was voluntary and those who participated were rewarded with a giftcard worth 200 SEK. We explained that their participation was limited to letting us take a portrait/photograph of them to use in the study and to them being expected to read a few short messages.

We took portrait/photographs of each jobseeker. These photographs were later harmonized regarding light, contrast etc. to avoid variation in picture quality. We also recorded their voices with the participants reading eight prewritten short personal messages saying that they wished to participate in an LMP. Male participants were assigned to reading messages related to the warehouse program while female participants were assigned to reading the messages for the nursing program.

These 75 participants were used as jobseekers in our experiment. We needed many participants to obtain variation in physical appearance as well as accents and to avoid the risk of relying on a few observations that might be specific in one way or another.

The characterization part

The objective of this part of the experiment was to measure stereotype beliefs about attributes that might be perceived as being embedded in these photographs and recorded voices. The attributes were: *Trustworthy, Orderly, Masculine / Feminine, Mature, Educated, Ambitious, Agreeable, Attractive, Stereotype Swedish Look* and *Age*, and the attributes for recorded voice were *Agreeable Voice, Extrovert, Motivated, Insecure, Tired, Sounds smart, Grown up in Stockholm* and *Grown up abroad*. In order to measure these stereotypes, we organized a characterization part where case workers guessed how employers would assign scores to photographs and recorded voices according to the dimensions listed above.

We recruited 160 caseworkers in Employment Service offices in the counties of Stockholm and Mälardalen excluding the offices where we had earlier recruited the jobseekers. We contacted the offices and ran our experiment during their ordinary morning meeting. In the introductory part of the experiment, caseworkers were informed that the aim of the study was to investigate caseworkers' recommendations for labor market programs. The caseworkers received a randomly assigned participation number and were informed that their responses would remain anonymous. Participation was voluntary and we told them that they could win lottery tickets.

To prepare for the characterization part, we asked two employers to assign values (1,2,...,10) associated with the various dimensions listed above for how they thought employers in general would perceive these photographs and recorded voices. This allowed us to have an incentive scheme and to reward caseworkers if their guesses matched the scores assigned by the employers. In this way, we were able to avoid answers based on socially desirable views and give incentives to subjects to put an effort into reporting what they believed were stereotypical views among employers. Our basic assumption is that these measures of stereotypical beliefs are unbiased representations of existing stereotypes in the labor market in general. Caseworkers were asked to guess how the employers would score the photographs and recorded voices of the jobseekers. A correct guess was rewarded with a lottery ticket. The caseworkers listened to messages that were randomly drawn from a pool of eight different recorded messages.

Photos were rated by 61 caseworkers and voices were rated by 99 caseworkers. Each caseworker rated photographs or recorded voices that were randomly drawn from the two clusters of 37/38 photos or clusters of 25/25/25

recorded voices. This construction of randomly drawing from randomly constructed clusters assured us that all photos/voices were rated the same number of times. The characterization part yielded on average around 30 observations on each of the 75 portraits and recorded voices for nine photo attributes and eight voice attributes.

Constructing jobseekers profiles

Photos, voice messages and the real names of the jobseekers were subsequently used to draw up profiles for all our 75 jobseekers. In this way a jobseeker was represented by his/her own look, voice and name. In constructing the profiles, we randomly assigned level of education (completed / not completed high school), date of registration at the SPES office indicating short or long current unemployment spells as well as dates for latest unemployment spell. Each profile was also assigned work experience in an occupation close to the LMP in question. To create variation in past experience, male candidates were randomly assigned experience in cleaning or dish washing. Female candidates had either experience in cleaning or childcare. A jobseeker was then represented with these randomly assigned characteristics together with the photograph, name, recorded voice message and a date of birth based on the average age associated with each photo from the characterization part.

The program-assignment part

For this part of the experiment, we recruited 276 caseworkers excluding the offices where we recruited the jobseekers. Our primary interest lay in measuring how caseworkers assigned scores to a profile in the following three dimensions: i) ability to successfully complete a program, ii) chances of success in the labor market upon completion of the program and iii) the need to participate in a program. The last measure was added because the caseworkers are supposed to prioritize jobseekers in most need of an LMP. We used our 75 profiles to run a program-assignment part in two steps.

In the first step, caseworkers were asked to guess how their colleagues had assigned scores to profiles in the three above-mentioned dimensions. We used the assigned scores from our pilot as correct answers to reward the

caseworkers. Correct guesses were rewarded with lottery tickets. Caseworkers considered eight profiles, one at a time, that were randomly drawn from our pool of 75 profiles. Note that no caseworkers saw the photograph of a jobseeker or listened to a voice message more than once.

Choice probabilities associated with job seeker profiles

Finally, in the second step of this part of the experiment, we were interested in knowing how caseworkers choose between two profiles if they are restricted to only recommending one jobseeker for an LMP. Due to the limited number of LMP slots, such situations are quite common in the everyday work of caseworkers in order to avoid choices systematically related to having a high school education or not, we let all the profiles in this step have the same level of education.[5] Results from this step gave us choice probabilities associated with each profile. The caseworkers were asked to guess who had the highest probability of being selected by their colleagues when comparing profiles in pairs of profiles that had been randomly drawn from the pool of our 75 candidates not drawn in the first step. This was repeated four times involving in total four pairs of profiles. As before correct guesses were rewarded with lottery tickets.

3 Probability of being selected

This setup allows us to use data from the characterization part of our experiment to estimate the probability of a profile being selected for an LMP in competition with other profiles as a function of the profile characteristics. We can estimate the effect of perceived stereotypical Swedishness, controlling for not having a foreign accent, and having a Swedish-sounding name holding the following variables constant: ability to successfully complete an LMP and chances of labor market success upon completion of the program associated with the profiles. Notice that work and unemployment experience were randomly assigned to the profiles. We also include these variables in our estimation to double check that the randomization has worked as intended.

[5]Due to a minor bug in our web application, in 8 of 1088 cases the competing candidates had different levels of education. Removing these observations or controlling for this difference in levels of education left our results unchanged.

Given that various attributes of a profile can be highly correlated, we focus on the perceived stereotypical Swedish look and then examine the effects of embedded attributes by replacing the variable for a stereotypical Swedish look with the variables measuring other attributes to assess what is seen in a look. The equation of interest is then as follows:

$$y_{(i|j)} = \alpha + \delta_1 \Delta_{ij} S_{LOOK} + \delta_2 \Delta_{ij} S_{SPEECH} + \delta_3 \Delta_{ij} S_{NAME} + \boldsymbol{\gamma} \Delta_{ij} \mathbf{X} + \varepsilon$$

Where y is a dummy variable measuring whether a profile i is selected when competing with another profile j to participate in a labor market program (LMP) or not. Data for y are from the second step in the program-assignment part of our experiment where case workers were instructed to choose one candidate when comparing pairs (i, j) of profiles. The pairs were drawn from the set of profiles that were not drawn and shown previously to the case worker. The variable $\Delta_{ij} S$ measures within pair difference in average scores of stereotypical Swedishness in the look, scores of not having a foreign accent generated in the characterization part of the experiment and whether the name is Swedish sounding or not.[6] The variable $\Delta_{ij} \mathbf{X}$ is a vector of between pair difference in average scores of *perceived ability of completing an LPM and the average score of chances of success after completing the LMP*. Data for $\mathbf{X}$ are generated in the first step of the program-assignment part. We do not have to include any other characteristics of a profile since they were all randomly assigned to the portraits and recorded voices. However, we include these characteristics to assure that the randomization has worked as expected.

In such a setup our estimate of δ_1 will identify the effect of physical appearance on the probability of being selected. Then we replace this variable with each of non-cognitive scores associated with the profile portraits and recorded voices to extract information about what is seen in a look and how various non-cognitive attributes influence the choice probability of a profile through observing a photograph in a profile. Throughout the paper, we allow for the different behavior of male and female caseworkers. This is motivated

[6] Whether names would be considered as Swedish sounding or foreign sounding is measured subjectively and independently by all 4 of us. A Swedish sounding name score (0-4) is then generated by adding how many of us considered the name as Swedish sounding. In 16 percent of cases 2 out of 4 characterized the name as Swedish sounding and in 43 out of 75 cases at least two of us rated the name as Swedish sounding.

by the previous studies on the tendency of men to discriminate (see, e.g. Freshtman & Gneezy, 2001).

4 A description of the data

We had 436 case workers who participated in the two parts of our experiment. In the characterization part, 99 case workers characterized voices and 61 case workers characterized photographs.[7] In the program-assignment part of the experiment, they were 276 case workers.

In a few cases, some subjects chose not to assign a score and instead chose the response alternative: "I don't know". This generated a few missing values. These missing values are so few and were therefore disregarded when computing the average scores.[8]

Since we examine gender differences in behavior, we need to know whether a case worker was male or female. The information for this variable is extracted from the survey given to the case workers after the experiment. The response on gender is missing for one case worker. The observations for this case worker have been deleted which entails a loss of 4 out of 1104 observations in the program-assignment part. In the the estimations we have $1100(= 275 * 4)$ observations including 304 observations for 76 male case workers and 796 observations for 199 female case workers.

Figure 1 & Figure 2 give the distribution of scores for various attributes associated with profile photographs and recorded voices. Table 1. & Table 2. give the correlation for photo attributes and voice attributes. The correlations are listed under the diagonal and the $p - values$ for zero correlation are found above the diagonal.

The results indicate that the stereotypical Swedish look is strongly correlated with all attributes that are usually perceived as being positively correlated with productivity. A stereotypical Swedish look is correlated to our measured attributes: Orderly ($\rho = 0.66$), Mature ($\rho = 0.66$), Ambitious ($\rho = 64$), Agreeable ($\rho = 0.51$), Masculine/Feminine ($\rho = 0.41$) and At-

[7] In one case a case worker chose to stop participating in the characterization part the experiment. The reason she gave was that She "was not comfortable with it".

[8] Less than 1 percent of scores for portrait pictures and 1.7 percent for scores related to recorded voices were missing. We had a few missing values in scores related to ability to complete a LMP, we had 2 missing values and for chances of success in the labor market after completed program, we had 7 missing values.

tractive ($\rho = 0.14$). Stereotypical Swedish look is also correlated to being perceived as Educated ($\rho = 0.63$). This means that being perceived as having a stereotypical Swedish look is positively associated with many attributes that employers value on the labor market. Moreover, having no accent is correlated with a perceived greater ability to complete an LMP. The weakest correlations are found for Age and Attractiveness. Similar patterns, however, with much weaker correlations are observed for voice attributes.

Furthermore, we checked the correlation between scores for perceived as having grown up abroad based on the recorded voice and the scores for the stereotypical look. There is a negative correlation (-0.08) but the $p - value$ is as high as 0.48. This means that stereotypes of the Swedish look are not a proxy for being foreign born and does not capture things than can be heard in the recorded voice of a person.

Table 3 gives the main characteristics of the case workers in the characterization part and the program-assignment part related part. We see that the men in the program-assignment part are slightly older, have slightly greater seniority and have to a lesser degree Swedish as their mother tongue. The differences are not large but we will check whether these differences have any effects on our results. Our main assumption is that case workers in the program-assignment part have similar perceptions of how jobseekers are perceived by employers as those who characterized the photographs in the characterization part of the experiment. The characteristics of the case workers participating in the different parts of the experiment are similar. Moreover, we find that they have similar job assignments (not reported here).

Figure 3 depicts the distribution of the evaluation scores related to ability to successfully complete an LMP, the chances of success in finding a job upon completion of the program and the need for an LMP associated with a jobseekers' profile. This implies that there is a variation in how profiles are perceived. The question is what explains this variation. Notice that the photographs in the profiles are randomly matched with information about the duration of the last spell of unemployment, previous occupation and a high school degree. To describe what attributes are correlated to these perceived ability and success scores, we run regressions explaining each of these variables with photograph and voice attributes.

The results in Table 4 indicate that compared with male case workers, female case workers, are more generous in assigning scores to female jobseekers. Moreover, case workers consider the ability of those who have no accent to be higher than those who have foreign accents. Having a foreign accent

might signal a lower level of Swedish language proficiency but can also be associated with belonging to a non-Swedish ethnic group.

Regarding our main variable of interest, male case workers tend to assign a higher ability score to those with a stereotypical Swedish look when it comes to women applying for the nursing program. Female case workers do this for male jobseekers. There is a cross-gender positive bias in estimating the ability of individuals with a stereotypical Swedish look.

To better understand how scores for ability to complete LMPs are related to other profile attributes, we estimate separately for females and males, a model replacing the score for stereotypical look with other measured attributes, one at a time, controlling for accent and foreign sounding name. This approach is chosen due to very high correlations for various attributes implying that we will not be able to separate the effect of individual attributes in a meaningful way. However, we can measure an aggregate effect of physical appearance using an additive index of scores of all the measured attributes.

The results of these estimations for various attributes for male and female case workers are presented in Table 5 Let us first look at the results for the female case workers presented in panels *(a)* and *(b)*. These results indicate that the ability to complete the program for forklift drivers consisting only of men is estimated to be higher for those with a stereotypical Swedish look. The same is observed for Agreeableness and Trustworthiness and Ambition. For assistant nurses, a non-cognitive additive index seem to have a positive effect for the perceived ability of profiles that aim at the assistant nurse occupation. The same is true if the candidate is perceived to be attractive. Female case workers seem to assign lower ability score to men who are less stereotypically Swedish-looking but seem to put stress on other attributes when it comes to female jobseekers.

The results for male case workers are reported in panels *(c)* and *(d)* in Table 5. These results do not give much insight into how male case workers estimate the ability of male jobseekers. However, male case workers seem to be sensitive to almost all of the attributes of female jobseekers except attractiveness when estimating their ability to complete the assistant nurse program. The results for probability of success in the labor market presented in Table 6. have a similar message as in the case of estimated ability score. Using the need for an LMP (not shown here) gives very little insight into how the scores of need is related to profile attributes.

To sum up, our results indicate that there is a cross-gender pattern of

favoring the ability to complete a program and probability of labor market success upon completion of a program for those who have a stereotypical Swedish look. It should be noted that while the probability of success in the labor market might partly reflect the anticipated discrimination of the caseworkers based on stereotypical Swedish looks, the perceived ability of successfully completing an LMP should not involve expectations about labor market discrimination.

After this description of the data and the correlation patterns in attributes and estimated ability and success measures, we examine the effect of profile photo attributes and voice attributes on the probability of being recommended for an LMP in the following Section.

5 Profile picture attributes and chances of being selected

The results reported in Tables 7 and 8 indicate that those who have foreign accent are associated with greater chances of being recommended for an LMP.[9] Regarding physical appearance, we find that female caseworkers are not sensitive to other attributes given the ability to complete a program and the probability of labor market success. But, male caseworkers are sensitive to the physical appearance of candidates.

The results clearly indicate that even though there is a cross-gender favoring of jobseekers with stereotypical Swedish looks in the evaluation of ability and chances of success, it is only men that prefer those with stereotypical Swedish looks. This is the case both when they choose between two male candidates or between two female candidates. The estimated effect represents a roughly 50-percent greater chance of being selected for the most stereotypical Swedish-looking (8/10) candidate as compared to the candidate who has the lowest score for stereotypical Swedish look (3/10) in our sample.

Since various picture attributes are highly correlated, it is not meaningful to include these variables in the same regression. This is basically related to the fact that it is not possible to separate the stereotypical Swedish look from

[9]Those with a foreign accent might be given higher priority due to observed higher risks of unemployment for the foreign born. This is consistent with the general policy of the SPES aiming at giving priority to jobseekers with the least chances of finding employment. See Regeringens proposition för 2012, 2011/12:1, utgiftsområde 14, pp 34 (Swedish Budget Bill: Budget Statement, 2012, Ministry of Finance

the components that are in the perception of that look. We instead try to see what the caseworkers see in the stereotypical Swedish look. To examine this question, we estimated the probability of being selected to participate in an LMP as a function of differences in scores of various profile photo characteristics controlling for ability and accent but replacing the variable for Stereotypical Swedish Look with each picture attribute at the time.

Inspecting the results in Table 9 we find that the positive appreciation of the stereotypical Swedish look in female candidates of male caseworkers is not related to any other profile photo attribute. For male candidates, however, many other positive profile photo attributes exhibit a similar effect on the choice probability as the stereotypical Swedish look. Men see attributes like trustworthiness, masculinity and agreeableness in stereotypical Swedish looks when they assess male candidates but when they assess female candidates, they do not see much in the appearance of the female candidates apart from a stereotypical Swedishness.

We run similar estimations as in Table 9 focusing on voice attributes instead of look. Results are reported in Table 10. The probability of being assigned to a labor market program is lower for those who do not have a foreign accent for both male and female caseworkers. Our results indicate that female caseworkers are sensitive to the voice attributes of male candidates in a compensating manner. Male caseworkers do not seem to react to any other voice attributes. Caseworkers are instructed to employ a positive bias in favor of those who are perceived as having reduced job chances when offering labor market programs. Our results indicate that both men and women succeed in following these instructions using foreign accent in speech as an indication of being associated with lower job chances.

6 Summary

We find that male caseworkers are positively sensitive to stereotypical Swedish looks when choosing candidates to recommend for labor market programs. This bias represents a roughly 50-percent greater chance of being selected for the candidate with the highest score for stereotypical Swedish looks (8/10) as compared to the candidate who has the lowest score for stereotypical Swedish looks (3/10) in our sample. Such a difference is similar to call-back differences in corresponding tests using Swedish-sounding and Arabic-sounding names on resumés.

What explains this bias? Any explanation that would suggest that the bias we observe is related to characteristics that are correlated to stereotypical Swedish looks but not due to physical appearance as such has to be accompanied with a description of why such a bias is observed among male decision-makers but not among female decision-makers. What do men see that women do not? We cannot find any information-based description consistent with the different behavior of male and female caseworkers reported above.

Our results are in line with previous studies indicating that discrimination is primarily practiced by men (see, for example, Fershtman and Gneezy, 2001). Therefore, our interpretation is that this bias must be due to discriminatory behavior based on stereotypes of appearance. The SPES has no policy on the risk of bias associated with stereotypes concerning physical appearance. In the absence of such a policy/awareness, there is room for discrimination in favoring jobseekers who more closely resemble the stereotypes of a Swedish look leaving room for bias based on national/racial stereotypes.

References

Ahmed, Ali, 2009, "What is in a surname ? The role of ethnicity in economic decision making ", Applied Economics , 42 (1): 1- 9.

Arai, Mahmood, Bursell, Moa and Lena Nekby, 2015, "The Reverse Gender Gap in Ethnic Discrimination: Employer Stereotypes of Men and Women with Arabic Names", International Migration Review, DOI: 10.1111/imre.12170.

Arai, Mahmood, and Peter Thoursie, 2009. "Renouncing Personal Names: An Empirical Examination of Surname Change and Earnings," Journal of Labor Economics, University of Chicago Press, vol. 27(1): 127-147.

Arrow, Kenneth, 1973, "The theory of discrimination" , Princeton University Press, 30(2): 3-33.

Averett S. and Korenman S., 1996, "The economic reality of beauty myth", Journal of Human Resources 31: 304-330.

Aigner, Dennis. J. and Glen G. Cain, 1977, "Statistical theories of discrimination in labor markets", Industrial and Labor Relations Review, 175-187.

Becker, Gary S.,1971, "The economics of discrimination" , University of Chicago Press.

Bertrand Marianne and Sendhil Mullainathan, 2004, "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," American Economic Review, 94: 991-1013.

Borghans, Lex, Duckworth, Angela L., Heckman, James. J. and ter Weel, Bas, 2006, The Economics and Psychology of Personality Traits, "Journal of Human Resources",43:972-1059.

Regeringens propositionen för 2012, Prop. 2011/12:1, Finansdepartemnetet, Stockholm, 2011.

Bursell, Moa, 2007, "What's in a name? A field experiment test for the existence of ethnic discrimination in the hiring process", SULCIS Working

Paper 2007:7.

Carlsson, Magnus and Dan-Olof Rooth, 2007. "Evidence of ethnic discrimination in the Swedish labor market using experimental data," Labour Economics, Elsevier, 14: 716-729.

Dechter, Evgenia K., 2015, "Physical appearance and earnings, hair color matters", Labour Economics, 32: 15-26.

Fershtman , Chaim and Uri Gneezy, 2001, " Discrimination in a segmented society: An experimental approach", Quarterly Journal of Economics, 116 (1) :351 - 377.

Hamermesh, Daniel S. and Biddle, Jeff E., 1994. "Beauty and the Labor Market.", American Economic Review, 84: 1174–94.

Jackson Linda A., John E. Hunter and Carole N. Hodge, 1995, "Physical Attractiveness and Intellectual Competence: A Meta-Analytic Review", Social Psychology Quarterly, 58: 108-122.

Heckman, James 1999, Policies to forster Human Capital NBER Working paper 7288.

Katz, D.; Braly, K, 1933, "Racial stereotypes of one hundred college students", The Journal of Abnormal and Social Psychology, Vol 28(3): 280-290.

López Bóo Florencia , Rossi Martín A., Urzúa Sergio S., 2013, "The labor market return to an attractive face: Evidence from a field experiment," Economics Letters, Elsevier, 118: 170-172.

Mobius, M. M. and Rosenblat, T. S., 2006, "Why beauty matters", American Economic Review, 96: 222–35

Mocan, N. and Tekin, E., 2010, "Ugly criminals", The Review of Economics and Statistics, 92: 15–30.

Nyhus, Ellen K. & Pons, Empar, 2005, "The effects of personality on earnings," Journal of Economic Psychology, 26(3): 363-384.

Phelps, Edmund. S. ,1972, "The statistical theory of racism and sexism" , The American Economic Review, 62: 659-661.

Riach, Peter A. and Rich Judith, 2002, "Field experiments of discrimination in the market place", The Economic Journal 112: 480-518.

Rooth , D., 2009, "Obesity, attractiveness and differential treatment in hiring - a field experiment" , Journal of Human Resources, 44: 710-735

Rooth , Dan-Olof, 2010, "Automatic associations and discrimination in hiring : Real world evidence" , Labour Economics , 17: 523-534.

Rödin, M. and Özcan, G., 2011, "It How You Look or Speak That Matters? - An Experimental Study Exploring the Mechanisms of Ethnic Discrimination", Research Papers in Economics, Department of Economics, Stockholm University, 2011(12).

Ruffle, B. and Z. Shtudiner, 2014, "Are Good- Looking People More Employable ? " Management Science, 61: 1760 - 1776.

Tajfel, Henri and Turner, John C., 1979, "An Integrative Theory of Intergroup Conflict". In W. G. Austin & S. Worchel (Eds.), The Social Psychology of Intergroup Relations. Monterey, CA: Brooks-Cole.

Jeff Werner and Tomas Björk, 2014, Blond och blåögd : vithet, svenskhet och visuell kultur (Blond and blue-eyed : whiteness, Swedishness, and visual culture), Göteborgs konstmuseum, Göteborg.

Appendix: Figures and tables

Figure 1: Average scores of 75 profile picture attributes.

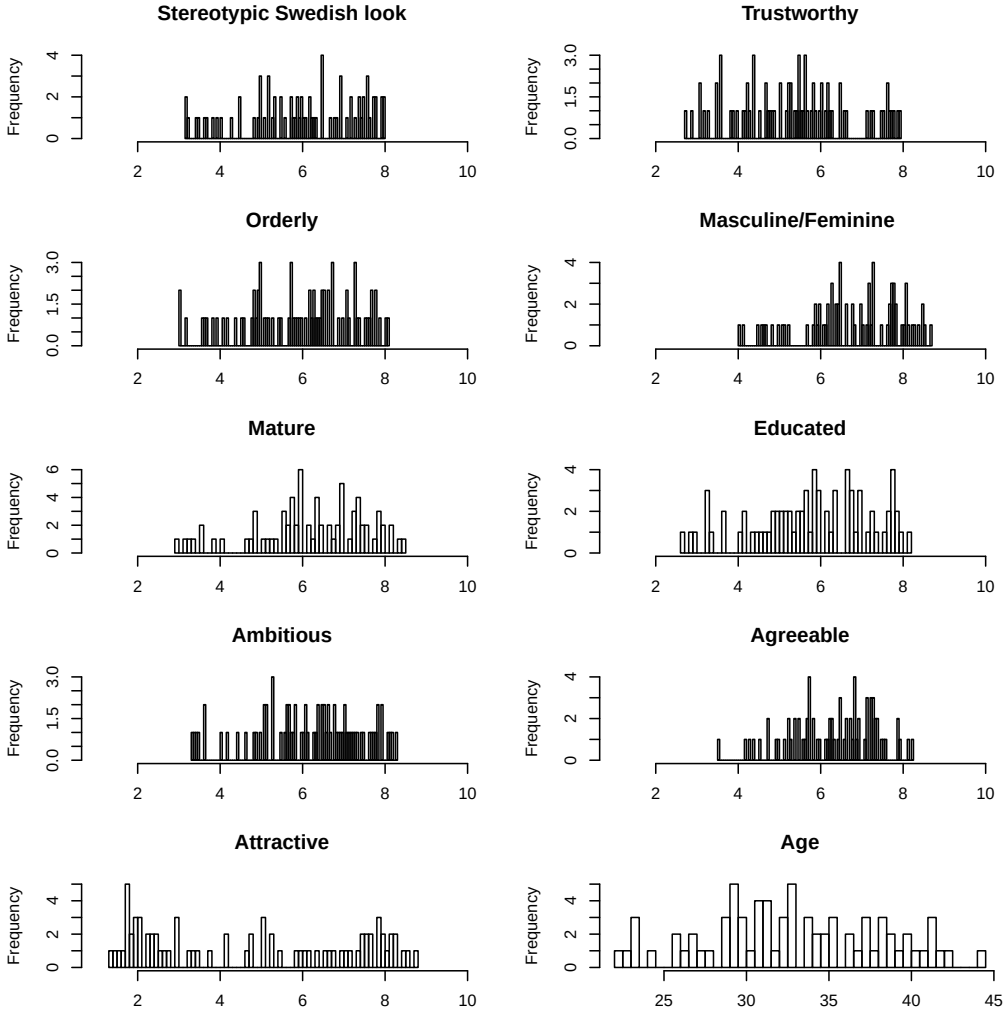


Figure 2: Average scores of 75 recorded voice attributes.

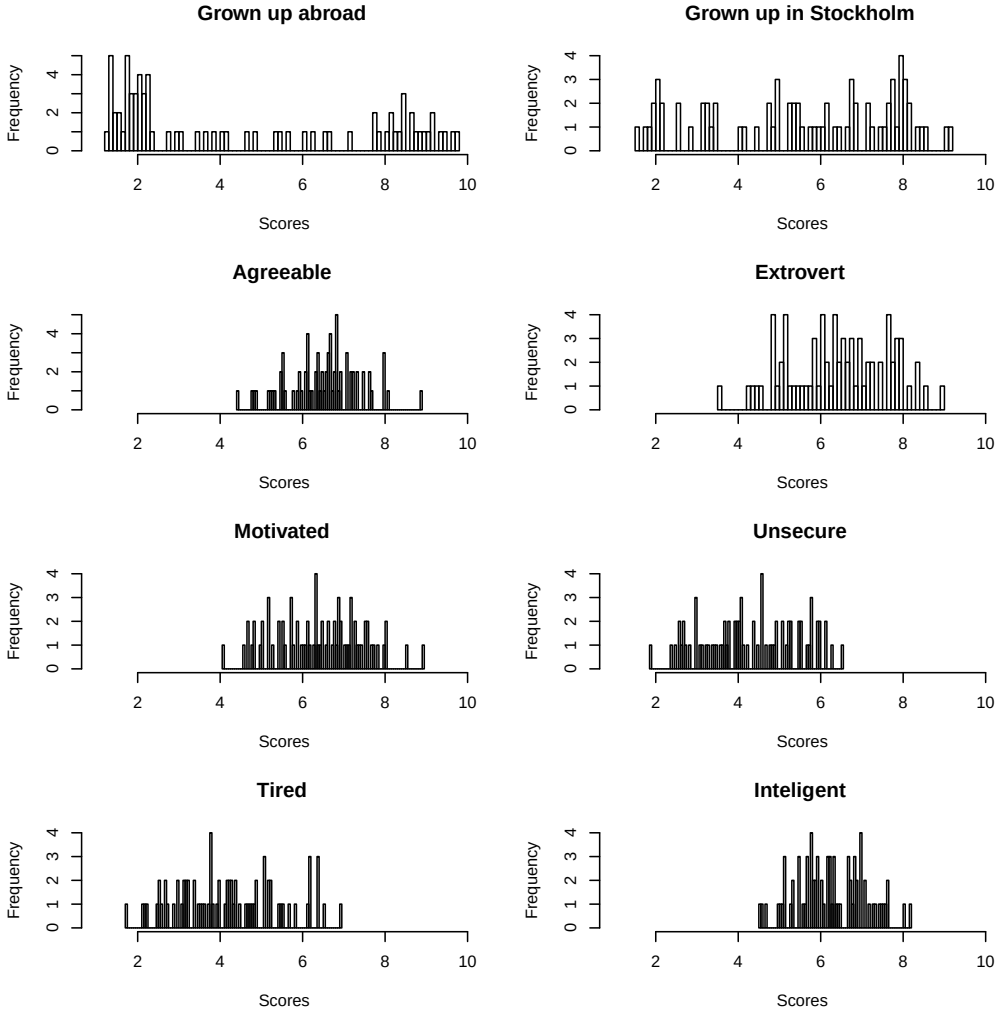


Table 1: Correlations of the picture attributes. Above-main-diagonal cells represent $p - values$ for $H0 : r = 0$ of corresponding below-main-diagonal cells.

	Trustworthy	Orderly	MasculFem	Mature	Educated	Ambitious	Agreeable	Attractive	StypeSwed	Age
Trustworthy	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04
Orderly	0.57	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Masculine/Feminine	0.64	0.49	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04
Mature	0.52	0.81	0.45	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Educated	0.49	0.78	0.40	0.76	1.00	0.00	0.00	0.00	0.00	0.00
Ambitious	0.55	0.82	0.46	0.82	0.80	1.00	0.00	0.00	0.00	0.00
Agreeable	0.62	0.73	0.54	0.70	0.63	0.70	1.00	0.00	0.00	0.02
Attractive	0.12	0.26	0.11	0.22	0.25	0.23	0.24	1.00	0.00	0.13
Stereotypical Swedish look	0.39	0.66	0.41	0.66	0.63	0.64	0.51	0.14	1.00	0.00
Age	-0.04	0.18	0.04	0.15	0.16	0.15	0.05	0.03	0.42	1.00

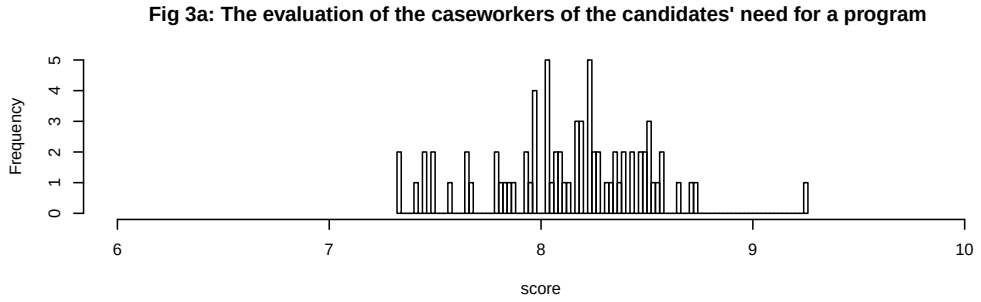
Table 2: Correlations of the voice attributes. Above-main-diagonal cells represent $p - values$ for $H0 : r = 0$ of corresponding below-main-diagonal cells.

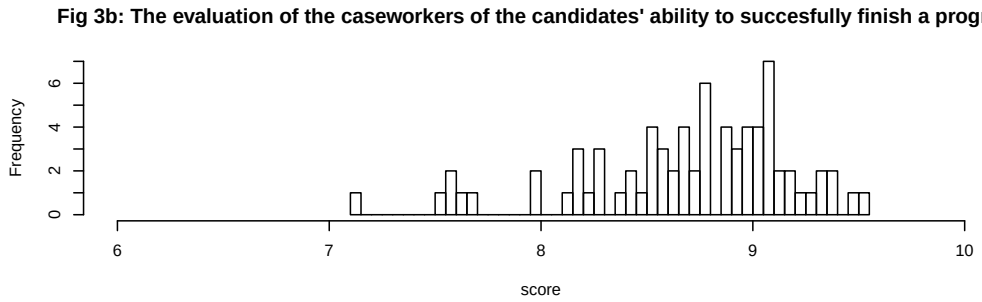
	AgreeableVoice	Extrovert	Motivated	Unsecure	Tired	Intelligent	GrownSt	GrownAbroad
Agreeable Voice	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Extrovert	0.59	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Motivated	0.64	0.70	1.00	0.00	0.00	0.00	0.00	0.00
Unsecure	-0.39	-0.54	-0.49	1.00	0.00	0.00	0.00	0.00
Tired	-0.41	-0.52	-0.56	0.55	1.00	0.00	0.00	0.00
Sounds smart	0.65	0.64	0.70	-0.45	-0.46	1.00	0.00	0.00
Grown upp in Stockholm	0.11	0.20	0.10	-0.20	-0.10	0.15	1.00	0.00
Grown up abroad	-0.08	-0.16	-0.09	0.21	0.12	-0.08	-0.55	1.00

Table 3: Characteristics of case workers participating in the program-assignmnet part and participating in the characterization part.

	Characterization		Program-Assignment	
	Men	Wom	Men	Wom
Age	43.48	45.34	48.55	43.03
Seniority	2.21	2.45	2.91	2.37
Mother Tongue Swedish	0.81	0.80	0.67	0.75
Number of Case Workers	48.00	116.00	75.00	197.00

Figure 3: Average evaluation scores of entire profile of candidates.





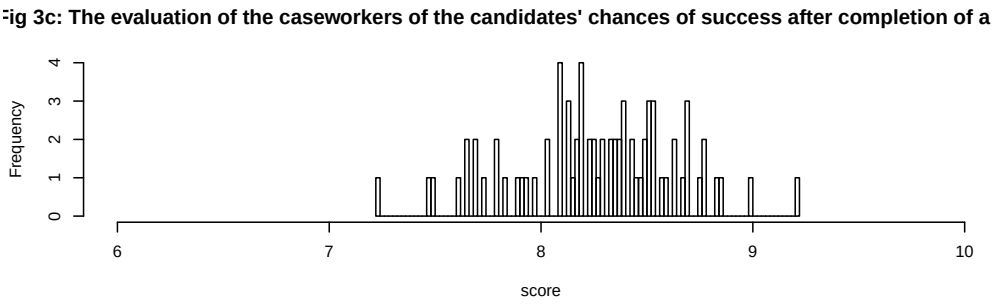


Table 4: Ability, no accent and look for Forklift Driver Program ($N = 1096$) and Assistant Nurse Program ($N = 1097$). Dependent variable is ability score average. Standard errors in parenthesis.

	Forklift Driver			Assistant Nurse		
	(1)	(2)	(3)	(1)	(2)	(3)
Stereotypical Swedish Look	-0.058	-0.064	-0.076	0.346*	0.253*	0.246*
	(0.080)	(0.076)	(0.077)	(0.079)	(0.079)	(0.079)
Look X Female Case Worker	0.195*	0.196*	0.191*	-0.227*	-0.187*	-0.185*
	(0.095)	(0.091)	(0.092)	(0.096)	(0.094)	(0.094)
Female Case Worker	-0.565	-0.547	-0.527	1.301*	1.117*	1.113*
	(0.405)	(0.384)	(0.386)	(0.489)	(0.480)	(0.479)
No accent		0.302*	0.395*		0.370*	0.404*
		(0.048)	(0.113)		(0.051)	(0.078)
Foreign-Sounding Name			0.105			0.048
			(0.113)			(0.072)
Adj. R-squared	0.018	0.059	0.059	0.035	0.096	0.095

NOTE: Swedish look and no accent are measured as mean score and are standardized. Other controls are education, unemployment and past occupation. Case worker cluster-robust standard errors. $p < 0.05$ indicated with an *.

Table 5: Ability score evaluated by case workers and standardized profile photo attributes. Dependent variable is ability score average. Each column represents the estimate of the impact of each attribute at the time.

Stereotype Swedish look	Agreeable	Trustworthy	Educated	Orderly	Attractive	Mature	Ambitious	Non-cognitive Look Index
Panel *(a)* Female Case Workers Evaluations of Forklift Driver candidates, $N = 792$								
0.119*	0.107*	0.176*	0.098	0.101	0.057	0.093	0.115	0.145*
(0.060)	(0.051)	(0.057)	(0.060)	(0.058)	(0.064)	(0.058)	(0.059)	(0.053)
Panel *(b)* Female Case Workers Evaluations of Assistant Nurse candidates, $N = 793$								
0.056	0.147*	0.106*	0.110*	0.126*	0.204*	0.143*	0.133*	0.123*
(0.052)	(0.054)	(0.045)	(0.054)	(0.054)	(0.075)	(0.058)	(0.054)	(0.062)
Panel *(c)* Male Case Workers Evaluations of Forklift Driver candidates, $N = 304$								
-0.078	-0.092	0.069	-0.104	-0.123	-0.171	-0.129	-0.107	-0.035
(0.090)	(0.085)	(0.081)	(0.094)	(0.085)	(0.118)	(0.088)	(0.090)	(0.085)
Panel *(d)* Male Case Workers Evaluations of Assistant Nurse candidates, $N = 304$								
0.241*	0.398*	0.293*	0.277*	0.341*	0.215	0.307*	0.341*	0.450*
(0.085)	(0.110)	(0.092)	(0.097)	(0.096)	(0.112)	(0.099)	(0.098)	(0.096)

NOTE: These estimations also include controls for accent, foreign sounding name, previous occupation, unemployment history, level of education and voice messages. Case worker cluster-robust standard errors. $p < 0.05$ indicated with an *.

Table 6: Chances of success in the labor market, no accent and look for Forklift Driver program ($N = 1096$) and Assistant Nurse Program ($N = 1097$). Dependent variable is average score for chances of success after completing a LMP. Standard errors in parantheses.

	Forklift Driver			Assistant Nurse		
	(1)	(2)	(3)	(1)	(2)	(3)
Stereotype Swedish look	-0.101	-0.104	-0.114	0.246*	0.196*	0.200*
	(0.084)	(0.082)	(0.082)	(0.068)	(0.070)	(0.070)
Look X Female Case Worker	0.162	0.162	0.158	-0.231*	-0.209*	-0.210*
	(0.098)	(0.095)	(0.095)	(0.083)	(0.083)	(0.083)
Female Case Worker	-0.073	-0.062	-0.045	1.312*	1.212*	1.215*
	(0.405)	(0.397)	(0.398)	(0.408)	(0.406)	(0.406)
No accent		0.171*	0.253*		0.200*	0.182*
		(0.043)	(0.095)		(0.050)	(0.071)
Foreign-Sounding Name			0.092			-0.026
			(0.098)			(0.074)
Adj. R-squared	0.052	0.064	0.064	0.014	0.032	0.031

NOTE: Swedish look and no accent are measured as mean score and are standardized. Other controls are education, unemployment and past occupation. Case worker cluster-robust standard errors. $p < 0.05$ indicated with an *.

Table 7: Probability of being selected to participate in a LMP and differences in looks including case worker characteristics for Forklift Driver program ($N = 544$). Dependent variable measures selected/not selected when compared with another candidate. Standard errors in parantheses.

	(1)	(2)	(3)	(4)
Stereotype Swedish look	0.089*	0.085*	0.076*	0.081*
	(0.019)	(0.019)	(0.019)	(0.026)
Look X Female Case Worker	-0.068*	-0.069*	-0.076*	-0.078*
	(0.022)	(0.022)	(0.023)	(0.023)
Female Case Worker	0.055	0.062	0.053	0.063
	(0.046)	(0.046)	(0.049)	(0.050)
Ability	-0.225*	-0.058	-0.024	-0.019
	(0.050)	(0.064)	(0.070)	(0.071)
Chances of Success	0.115	0.086	0.030	0.025
	(0.060)	(0.061)	(0.066)	(0.068)
No accent		-0.030*	-0.013	-0.014
		(0.007)	(0.015)	(0.016)
Swedish-Sounding Name			0.059	0.055
			(0.044)	(0.044)
Shorter recent unemployment			-0.085	-0.084
			(0.066)	(0.067)
Shorter history of uenmployment			-0.138*	-0.140*
			(0.062)	(0.063)
Shorter recent & history of unempl.			0.109	0.112
			(0.093)	(0.094)
Profile 1 has Occupation 2			-0.085	-0.079
			(0.062)	(0.064)
Profile 2 has Occupation 2			-0.084	-0.083
			(0.062)	(0.062)
Both have Occupation 2			0.091	0.084
			(0.089)	(0.090)
Look X CW Mother Tongue Swedish				-0.004
				(0.026)
CW Mother Tongue Swedish				-0.032
				(0.051)
Seniority of Case Worker				0.013
				(0.020)
Age of Case Worker				0.000
				(0.003)

NOTE: Variables are measured as the difference in the two candidates' mean score. Columns 3 and 4 include controls for 8 times 8 combinations of various voice messages read by the two competing candidates as well as the order shown in the first, second, ... fourth pair. Case worker cluster-robust standard errors. $p < 0.05$ indicated with an *.

Table 8: Probability of being selected to participate in a LMP and differences in look including case worker characteristics for Assistant Nurse program ($N = 544$). Dependent variable measures selected/not selected when compared with another candidate. Standard errors in parantheses.

	(1)	(2)	(3)	(4)
Stereotype Swedish look	0.063*	0.057*	0.064*	0.092*
	(0.022)	(0.022)	(0.025)	(0.032)
Look X Female Case Worker	-0.054*	-0.053*	-0.050	-0.044
	(0.026)	(0.025)	(0.027)	(0.028)
Female Case Worker	-0.037	-0.053	-0.030	-0.027
	(0.047)	(0.046)	(0.049)	(0.050)
Ability	-0.189*	-0.063	-0.080	-0.087
	(0.047)	(0.062)	(0.067)	(0.068)
Chances of Success	0.091	0.058	0.063	0.069
	(0.063)	(0.064)	(0.068)	(0.068)
No accent		-0.025*	-0.028*	-0.027*
		(0.007)	(0.010)	(0.010)
Swedish-Sounding Name			-0.040	-0.039
			(0.029)	(0.028)
Shorter recent unemployment			-0.055	-0.053
			(0.067)	(0.067)
Shorter history of uenmployment			0.006	0.015
			(0.065)	(0.065)
Shorter recent & history of unempl.			-0.131	-0.138
			(0.094)	(0.093)
Profile 1 has Occupation 2			-0.023	-0.018
			(0.067)	(0.067)
Profile 2 has Occupation 2			0.040	0.043
			(0.065)	(0.066)
Both have Occupation 2			0.038	0.030
			(0.085)	(0.086)
Look X CW Mother Tongue Swedish				-0.044
				(0.028)
CW Mother Tongue Swedish				-0.027
				(0.050)
Seniority of Case Worker				-0.020
				(0.019)
Age of Case Worker				0.002
				(0.003)

NOTE: Variables are measured as the difference in the two candidates' mean score. Columns 3 and 4 include controls for 8 times 8 combinations of various voice messages read by the two competing candidates as well as the order shown in the first, second, ... fourth pair. Case worker cluster-robust standard errors. $p < 0.05$ indicated with an *.

Table 9: Look: Probability of being selected to participate in an LMP and the differences in attributes associated with the photograph of the candidates. Dependent variable measures selected/not selected when compared with another candidate. Each row represents the estimates for an attribute at a time. Standard errors in parentheses.

Stereotypical Swedish look	Trustworthy	Orderly	Educated	Ambitious	Attractive	Agreeable	Mature	Masculine/Feminine	Non-Cognitive Look Index
Panel *(a)* Male case workers choosing Forklift Drivers, ($N = 152$)									
0.062*	0.111*	0.056	0.034	0.054	-0.003	0.090*	0.052	0.107*	0.061
(0.026)	(0.042)	(0.029)	(0.028)	(0.031)	(0.02)	(0.037)	(0.028)	(0.040)	(0.035)
Panel *(b)* Male case workers choosing Assistant Nurses, ($N = 152$)									
0.063*	0.009	0.038	0.019	0.036	-0.016	0.005	0.061	0.030	0.021
(0.030)	(0.031)	(0.039)	(0.034)	(0.039)	(0.019)	(0.050)	(0.037)	(0.037)	(0.051)
Panel *(c)* Female case workers choosing Forklift Drivers, ($N = 398$)									
0.000	-0.018	-0.009	-0.013	-0.011	-0.011	-0.016	-0.006	0.005	-0.015
(0.014)	(0.019)	(0.014)	(0.013)	(0.014)	(0.011)	(0.018)	(0.014)	(0.022)	(0.017)
Panel *(d)* Female case workers choosing Assistant Nurses, ($N = 398$)									
0.004	-0.004	-0.008	-0.009	-0.009	-0.007	-0.008	-0.002	-0.005	-0.017
(0.016)	(0.013)	(0.019)	(0.017)	(0.019)	(0.010)	(0.025)	(0.018)	(0.015)	(0.027)

NOTE: Variables are measured as between-candidate differences in the mean score. Estimations include controls for accent, ability to complete an LMP, previous unemployment, occupation, 8 times 8 combinations of various voice messages read by the two competing candidates as well as the order shown in the first, second, ... fourth pair. Case worker cluster-robust standard errors. $p < 0.05$ indicated with an *.

Table 10: Voice: Probability of being selected to participate in a LMP and differences in attributes associated with the recorded voice of the candidates. Dependent variable measures selected/not selected when compared with another candidate. Each row represents the estimates for an attribute at a time. Standard errors in parantheses.

No accent	Motivated	Extrovert	Agreeable Voice	Unsecure Voice	Tired Voice	Sounds Smart	Non-Cognitive Voice Index
Panel *(a)* Male caseworkers choosing Forklift Drivers ($N = 152$)							
-0.027	0.023	0.014	0.005	-0.025	-0.035	-0.002	0.026
(0.019)	(0.045)	(0.039)	(0.049)	(0.043)	(0.037)	(0.048)	(0.046)
Panel *(b)* Male caseworkers choosing Assistant Nurses ($N = 152$)							
-0.038	0.000	-0.013	0.031	0.016	0.046	0.052	-0.018
(0.021)	(0.046)	(0.048)	(0.055)	(0.046)	(0.042)	(0.068)	(0.052)
Panel *(c)* Female caseworkers choosing Forklift Drivers ($N = 398$)							
-0.036*	-0.057*	-0.049*	-0.001	0.063*	0.05*	-0.038	-0.055*
(0.008)	(0.022)	(0.019)	(0.023)	(0.020)	(0.016)	(0.026)	(0.022)
Panel *(d)* Female caseworkers choosing Assistant Nurses ($N = 398$)							
-0.022*	-0.001	-0.012	0.006	0.012	0.020	-0.016	-0.012
(0.009)	(0.021)	(0.018)	(0.025)	(0.017)	(0.018)	(0.029)	(0.021)

NOTE: Variables are measured as the difference in the two candidates' mean score. Controls are ability to complete a LMP, previous unmeployment, occupation, 8 times 8 combinations of various voice messages read by the two competing candidates as well as the order shown in the first, second, ... fourth pair. The voice index adds voice scores including Tired and Unsecure with a negative sign. Case worker cluster-robust standard errors. $p < 0.05$ indicated with an *.