

Automatic Associations and Discrimination in Hiring - Real World Evidence^{*}

by

Dan-Olof Rooth[†]

Linnaeus University, CReAM and IZA

Abstract: This is the first study providing empirical support for automatically activated associations inducing discriminatory behavior among recruiters in a real-life hiring situation. Two different field experiments on ethnic discrimination in hiring are combined with a measure of employers' automatic attitudes and performance stereotypes toward Arab-Muslim men relative to Swedish men using the Implicit Association Test.

The results show that the probability to invite Arab-Muslim job applicants decreases by five percentage points when the recruiter has a one standard deviation stronger negative implicit association toward Arab-Muslim men. This suggests that automatic processes may exert a significant impact on employers' hiring decisions, offering new insights into labor market discrimination.

JEL classification: J64, J71

Key words: Implicit attitudes/stereotypes, discrimination, exit from unemployment

^{*} I thank Jens Agerström, Per Johansson, Jonas Lagerström, Lena Nekby, Inga Persson, Olof Åslund, participants at seminars at the Nordic Summer Institute in Labor Economics in Helsinki, at IFAU (Uppsala) and at IZA (Bonn) for valuable comments and helpful suggestions. Magnus Carlsson, Rickard Carlsson, Klara Johansson and Terese Johansson provided excellent research assistance. A research grant from the Swedish Council for Working Life and Social Research and another research grant from the Institute for Labor Market Policy Evaluation (IFAU) are gratefully acknowledged.

[†] Linnaeus University, e-mail: dan-olof.rooth@lnu.se.

1. Introduction

Increasing attention has been given in economics to factors previously expected to be unrelated to productivity, such as beauty, and to non-cognitive skills, such as self-discipline, and how employers value these attributes.¹ Following this empirical line of research the search for *theories* explaining why employers attach prices to such characteristics has increased. In particular, explanations taken from the field of social psychology have gained momentum in economics.

In this study yet another domain, i.e., the causes of ethnic discrimination, is explored from a social psychology perspective. Although much attention has been directed at the unequal treatment of ethnic minorities little is still known about what causes discriminatory behavior in the hiring situation (e.g., Bertrand and Mullainathan, 2004). Through the lens of “taste-based” and “statistical” discrimination theories economists, much like social psychologists, have assumed that this discrimination can be traced back to people’s consciously mediated perceptions of the discriminated target group (see Altonji and Blank, 1999; Devine, 1989). However, considerable evidence now supports the view that a person’s attitudes and stereotypes in fact often operate in an automatic, less conscious, mode (see Stanley, Phelps, and Banaji, 2008).² According to dual-process-theory³ people can process information using simple, low effort, readily available decision rules, referred to as *automatic, spontaneous, implicit or impulsive*

¹ See e.g., Hamermesh and Biddle (1994) and Heckman and Rubinstein (2001) for research on the importance of beauty and non-cognitive skills in the labor market, respectively.

² De Houwer, Thomas, and Baeyens (2001) show that people frequently form such associations without conscious intent. Hence, while explicitly stated prejudiced attitudes reflect racism, automatic ones do not.

³ See Chaiken and Trope (1999) for an overview of dual process models.

processes, as well as through the ordinary type of processing involving the active, effortful scrutiny of relevant information, referred to as *controlled, deliberate, explicit* and *reflective* processes, see Ranganath, Smith, and Nosek (2006).⁴

While people's explicit attitudes and stereotypes have for long been captured directly by self-reported reflections in response to attitude questions, the measurement of their automatic/implicit counterparts has a much shorter history. The latter are often achieved at by respondents taking the Implicit Association Test (henceforth IAT), a computerized response latency measure designed to tap individual differences in associations between concepts and attributes, see Greenwald, McGhee, and Schwartz (1998). The IAT thus indirectly measures an individual's automatic/implicit attitude or stereotype by assessing his/her relative differences in strengths of associations.⁵

Relative to explicit attitudes and stereotypes the IAT has been found to predict prejudiced behavior especially well in socially sensitive contexts (see Lane et al., 2007; Greenwald et al., in press).⁶ For example, McConnell and Leibold (2001) found that negative automatic associations

⁴ We interchangeably use the terms *automatic* and *implicit* throughout the article.

⁵ The design of the IAT is such that it can only assess relative difference in association. Thus, when we say that Arab-Muslims are associated with low productivity, this is not in an absolute sense. Rather, it means that Arab-Muslims are more easily associated with attributes reflecting low productivity *compared with* native Swedes. While the relative nature of the IAT can be a limitation in some cases, it is actually beneficial in the case of research on labor market discrimination, since it is the relative difference in how the groups are perceived and treated that is of interest.

⁶ In fact, the IAT has been found to predict criterion measures ranging from interracial friendliness and impression formation to anxious and shy behaviors, consumer choices, and voting, see e.g., Uhlman and Cohen (2007), and Phelps (2006). However, the IAT does not stand without criticism. Especially important is the concern that the IAT

toward Blacks (as compared to Whites) successfully predicted more negative nonverbal behaviors, such as less speaking time and less smiling, during an interaction with a Black researcher relative to a White researcher. The explicit measure did not predict the behavior to the same extent.

The present study examines the extent to which recruiters possess more negative automatic associations in terms of individual productivity (*Experiment I*), and of attitudes (*Experiment II*), toward Arab-Muslim men relative to Swedish men and whether these associations predict discrimination in the context of deciding which job candidates to call for an interview.⁷ The extent to which implicit attitudes (or stereotypes) are in fact related to discriminatory behavior in the hiring situation *as such* has so far only been studied in laboratory settings. For instance, Ziegert and Hanges (2005) found that an implicit racist attitude, which interacted with a climate for racial bias, predicts discriminatory behavior, while an explicit counterpart does not. Thus, the present research is to our knowledge the first to study the predictive power of automatically

just measures knowledge of cultural attitudes/stereotypes rather than actual implicit bias. However, in a recent study Nosek and Hansen (2008), using 158 different samples, find a zero correlation between the IAT score and measures of cultural knowledge.

⁷ The attitude IAT is the “standard tool” in social psychology when studying implicit associations and was employed as a benchmark measure, while the performance stereotype IAT, which has much shorter history, was expected to better explain the results found in Carlsson and Rooth (2007) since ethnic discrimination varies across occupations. However, existing evidence indicates that the two measures are highly correlated. When we measure, say, the recruiters’ automatic stereotypes, it might also, to some extent, pick up their automatic attitudes, and the other way around.

activated associations in a real-life hiring situation.⁸ In doing so we combine two research projects, one in economics and the other in social psychology.

In the studies by Carlsson and Rooth (2007, 2008) a field experiment specifically designed to test for discrimination in recruitment was used. This so called correspondence testing methodology implies that the researcher sends two equal applications to advertised job openings with the only difference being the name of the applicant; in this case one with a Swedish native sounding male name and the other with an Arab-Muslim sounding male name.⁹ Ethnic discrimination is quantified by the difference in the probability of a callback for interview between the two groups. In Agerström, Carlsson, and Rooth (2007) two different IATs were developed that measure automatic attitudes and automatic work performance stereotypes toward Arab-Muslim men relative to Swedish native men. These were administered to the recruiters in the Carlsson and Rooth studies; one for each experiment. In the analyses the results from the IATs were used to predict the decisions made by the recruiters in the hiring process.

⁸ However, a recent study by Rooth (2008) also provides empirical support for automatically activated associations inducing unequal treatment against the obese among recruiters in a real-life hiring situation. In this study a field experiment on differential treatment against obese job applicants in hiring is combined with a measure of employers' automatic performance stereotype toward obese relative to normal weight using the implicit association test. Rooth finds a strong and statistically significant obesity difference in the correlation between the automatic stereotype of obese as being less productive and the callback rate for an interview.

⁹ The reason for using the combined category "Arab-Muslims" is because this group is found being discriminated the most in the Swedish labor market and that the names used belong to this group. Hence, we do not focus on prejudice towards Muslims, irrespective of origin, or towards all persons of an Arabic background, irrespective of religion.

To be more precise, Figure 1 shows that Experiment I and II were initiated over a year apart, first collecting the information on ethnic differences in callbacks for interview. At the second stage of the experiments recruiters were contacted in person and asked to participate in taking the IAT and answering a number of questions, in that order. All participants in Experiment I were given the performance stereotype IAT, while all participants in Experiment II were given the attitude IAT.

*** Figure 1 ***

The results of the first experiment focusing on automatic stereotypes show that recruiters find it significantly easier to associate Arab-Muslim males with incompetence, laziness, and inefficiency than with productivity, ambition, and efficiency. In fact, 77 percent have at least a slight, and 56 percent at least a moderate automatic performance stereotype. Also, 53 percent of the recruiters explicitly state that they would prefer hiring a native Swedish male over an Arab-Muslim male. Interestingly then, it is found that employers' job interview callback rates for applicants with Arab-Muslim sounding male names are negatively correlated in a statistically significant way with their automatic performance stereotype, but only weakly correlated with their explicit attitude and performance stereotype.

The results of the second experiment focusing on implicit attitudes show that recruiters find it significantly easier to associate Arab-Muslim males with unpleasant words than with pleasant words.¹⁰ In this experiment a much larger share, 93 percent, have at least a slight, and 82 percent at least a moderate implicit attitude. However, a smaller share, 33 percent, of the recruiters explicitly state that they would prefer hiring a native Swedish male over an Arab-Muslim male.

¹⁰ Examples of unpleasant/pleasant words used are misery, pain, happiness and wonderful.

Again, it is found that employers' job interview callback rates for applicants with Arab-Muslim sounding male names are negatively correlated in a statistically significant way with their implicit attitudes, but only weakly correlated with their explicit attitudes/performance stereotypes. Hence, these results show that an automatically activated association is a significant determinant in the hiring process, while explicitly stated attitudes and stereotypes seem to be controlled and not acted upon.

The remainder of this article is outlined as follows. Section 2 presents evidence of the explicit measure not being affected by measurement bias and that the IAT is not just a better measure of a person's attitudes and stereotypes compared to explicit measures, but rather measures related but different constructs. Section 3 presents the economic model of the probability of receiving a callback when recruiters have automatic/implicit and explicit attitudes and stereotypes. Section 4 presents the correspondence test and its results in more detail, while section 5 presents the implicit association test. Section 6 presents results where the main question concerns the extent to which the job interview callback rate can be predicted from the automatic associations (the IAT score), explicit attitudes and stereotypes, respectively, while section 7 concludes.

2. Implicit and explicit measures and processes – the same, related or distinct?

The IAT is of special value when measuring people's attitudes or stereotypes due to its high resistance to faking. In fact, participants typically cannot successfully fake their IAT score even if explicitly instructed to by researchers (see Kim, 2003; Steffens, 2004). However, this can not be expected to be the case for self-reported measures. In the context of prejudice people might be unwilling to report an evaluative response that comes to mind because (i) they do not want others

to know about it, or (ii) the feeling is unwanted in the sense that it is not endorsed or accepted as one's evaluation. However, while the former creates measurement error in a self-reported measure the latter does not, since it might be genuine in the sense that he/she believes the response is unfair and attempts to replace it with a more positive explicit response.¹¹ Interestingly, approximately half of the recruiters in this study explicitly report being prejudiced. This share is in fact much higher than found in surveys specifically designed to measure prejudice in Sweden and elsewhere.¹² Hence, there is less scope for expecting them not to have disclosed their true prejudiced attitudes.

Several meta-analyses have shown a weak correlation between IAT scores and explicit measures, see e.g., Greenwald et al., in press. In the present study, especially so for *Experiment I* but also to some extent for *Experiment II*, the share having at least moderate negative automatic

¹¹ This might be especially relevant for recruiters since ethnic discrimination is prohibited by law, and it might be a company strategy to have a diverse labor force, and the recruiter therefore consciously rethinks in a non-prejudiced direction.

¹² The level of explicit prejudice among the recruiters is far higher than what is found in other studies of explicit prejudice for Sweden. In the Eurobarometer (2006) 40 percent of the EU population state that immigrants contribute a lot to their country. Sweden had the highest agreement level of 79 percent, while countries like the UK, France and Germany had corresponding levels of 47, 44, and, 30 percent, respectively. In the 2003 European Social Survey a "Resistance to immigrants" index was calculated and again Sweden showed the lowest level of negative attitudes in that only 15 percent had such attitudes, while the corresponding numbers for the UK, and Germany was 51 and 38 percent, respectively. Finally, in the Pew Global Attitudes Project (2007) it was found that 57 percent in Sweden view immigration from the Middle East and North Africa as a good thing. Also, the share of the Swedish population having prejudiced attitudes found in Åslund and Rooth (2005) are even lower, varying between 20 to 30 percent. Hence, not in a single instance did the share with negative attitudes reach the level found among the recruiters in the present study.

associations of Arab-Muslims and the share that declare having explicit preferences for hiring a native Swedish male are similar in size and one might expect that they measure the same underlying construct. However, even though the correlation coefficient is significant in *Experiment I* its magnitude is only 0.23.¹³ In *Experiment II* the correlation is 0.06, and not statistically significant. Hence, either the implicit and explicit measures tap related but distinct constructs or there are measurement issues involved.

Intuitively one might expect automatic activations of general associations, as assessed by the IAT, and self-reported prejudice to be rooted in the same experiences and thus be directly related. Research in social psychology, and more recently in neuroscience, typically does not support this expectation. Instead, the evidence supports that automatic (as measured by the IAT) and explicitly measured attitudes are related, but distinct, constructs (for reviews see Fazio and Olson, 2003; Nosek and Smyth, 2007). First, the IAT score and corresponding explicit measures of attitudes show a low correlation even in domains where they are expected to be highly correlated, i.e., in a situation in which there is little reason to believe that participants would have been concerned with controlling their explicit attitudes, see Karpinski and Hilton (2001).¹⁴ Second, research in social psychology using subliminal priming methods have shown that automatically triggered attitudes can influence behavior without our awareness (Fazio, Dunton, and Williams, 1995; Greenwald and Banaji, 1995). Third, and perhaps most convincing, is the

¹³ If we combine all three explicit measures an overwhelming majority, 66 percent, express negative attitudes or stereotypes toward Arab-Muslim male applicants. Using this measure the correlation coefficient is no longer statistically significant and drops to 0.13.

¹⁴ In their experiment respondents' attitude towards candy bars and apples were measured both indirectly (by IAT) and directly. While both measures had similar means, they had a low correlation.

identification of distinct neural systems for the expression and regulation of automatic/implicit attitudes and the involvement of those systems in implicit and explicit processes (Stanley, Phelps, and Banaji, 2008). This neurological evidence suggests that implicit measures such as the IAT reflect more automatic than controlled reactions to social groups (Cunningham et al., 2004; Phelps and Banaji, 2005).

Could automatically activated associations then affect recruiters' willingness to hire minorities? Given that automatic associations are thought to be more influential in situations where there is a high degree of automaticity, and where deliberate and more time consuming cognitive processes are mitigated (Greenwald et al., in press), it may be that they play a more important role when, for example, recruiters have to do a quick initial screening among a number of different job applications, as well as when cognitive load, fatigue and "gut feelings" enter the equation. In (a small number of) interviews with recruiters we often came across the comment that they instantaneously get a "feeling" of the application.

Ideally, we would like to measure (i) under what conditions a negative automatic response is activated when reading the job application, and (ii) if and when more controlled processes override such initial automatic associations. However, in the context of a real hiring situation, being outside the black box of the hiring decision, we can only observe the reduced form of such a process, i.e., whether automatically activated associations correlate with discriminatory behavior. Hence, the ecological validity of the field experiment comes at the price of imperfect control over the hiring situation.

3. A model of explicit and automatic prejudice and discrimination in hiring

Discrimination in the hiring process could possibly be a result of both consciously controlled, as well as automatically activated negative associations of the minority applicant group. Reading the Arab-Muslim name on a job application might activate the recruiter's automatic generalized beliefs about that minority group. Given that these generalized beliefs toward the group in question (attitudes/stereotypes) are negative, the recruiter may attach a negative value automatically, e.g., a gut feeling, to this job application, which results in the Arab-Muslim applicant having a lower probability of being called for an interview compared to a native Swedish applicant. Bertrand, Chugh, and Mullainathan (2005) argue, based on results taken from social psychology, that automatically activated stereotypes/attitudes may be especially important determinants in the hiring situation when the employer is inattentive to the task, experiences time pressure and ambiguity, something that may characterize the hiring process in many instances. This study attempts to determine which form of prejudice, the consciously controlled or the automatically activated, that is the most important determinant in the hiring situation of ethnic minorities.

Assume that the probability of a recruiter to giving a callback for a job interview to an application with a certain ethnicity attached to it has the following relationship for recruiter i receiving an application from ethnic group j :

$$(1) \quad \Pr(\text{Callback} = 1)_{ij} = X\beta_j + \delta_j + \varepsilon_{ij}; j = \text{Arab-Muslim or native Swede.}$$

where X is a vector of characteristics in the job application, which by construction of the experiment is the same for both ethnic groups, β_j is the vector of prices that recruiters attach to

those characteristics for ethnic group j , while δ_j is an omnibus prejudice measure of all types of attitudes/stereotypes, explicit as well as implicit, towards group j . The aim of this study is to specifically test whether automatically activated forms of attitudes and stereotypes are important in the hiring situation. Hence, we can rewrite Equation 1 as:

$$(2) \quad \Pr(\text{Callback} = 1)_{ij} = X\beta_j + \delta_j^{EP} + \delta_j^{ES} + \delta_j^I + \varepsilon_{ij}$$

where the three *prejudice* terms express the recruiter's explicit preference, his/her explicit stereotype (only in *Experiment I*), and his/her implicit association towards group j , respectively.

4. Correspondence testing – ethnic discrimination in hiring

Carlsson and Rooth (2007, 2008) tested for ethnic discrimination in hiring using the correspondence testing method (see Riach and Rich, 2002). When “correspondence testing” in conjunction with ethnicity, the researcher typically sends two job applications containing applicants with identical skills to advertised job openings with the only difference being group membership, which is signaled by the name of the applicant – one randomly assigned a Swedish name and the other an Arab-Muslim name. The extent of ethnic discrimination is quantified by the relative callback rates for job interview between the two groups.

For the present study, the field experimental data from Carlsson and Rooth (2007, 2008) served as the behavioral criterion variable on which the employers' IAT scores were regressed. The behavioral data were collected between May 2005 and February 2006 and between August 2006 and April 2007 for *Experiment I* and *II*, respectively, by sending applications to job openings pertaining to twelve different occupations in the Stockholm and Gothenburg labor

market areas. The selected occupations were both skilled and semi/unskilled and included a relatively high as well as a low ratio of immigrants. The selected occupations were computer professionals, business sales assistants, four categories of teachers (preschool, math/science and language¹⁵ in upper level compulsory school, and upper secondary school), accountants, nurses, construction workers, restaurant workers, shop sales assistants and motor-vehicle drivers.

As stated in the introduction, the reason for focusing on the Arab-Muslim ethnic minority is that Swedish studies indicate that discrimination is worst against individuals with such an ethnic background (see Lange, 2000, and Rooth and Ekberg, 2003). Since ethnicity is only signaled by means of changing the name on the application, the choice of names used in the experiment is crucial. Fortunately, there is a clear distinction between typical Swedish male names and Arab-Muslim male names. The most frequent Swedish and Arab-Muslim male names occurring in Sweden were selected from Statistics Sweden's name register and randomly assigned to résumés.

The applications used in the experiment had to be realistic and yet not refer to any real persons. Thus, applicants had identical human capital within occupations and were of the same age (varied between 25-30 years old across occupations), had the same amount work experience (varied between two to four years across occupations), and had obtained their education in the same type of school, but at different Swedish locations. Further, the application consisted of a

¹⁵ Not included in the correspondence test of Experiment II. Hence, only eleven occupations were included in this experiment.

quite general biography on the first page and a detailed curriculum vitae, including education and work experience, on the second page.¹⁶

In total Carlsson and Rooth (2007, 2008) replied to 1,552 and 1,314 job ads, respectively, posted at the home-page of the Swedish Employment Agency, see Table A1a and A1b in appendix. These tables show that the callback rate of applications with a Swedish-sounding male name was nine and seventeen percentage points higher than for applicants with an Arab-Muslim sounding name in the correspondence test used in *Experiment I* and *II*, respectively.

4.1 Participants in the IAT experiment

The IAT data were collected among the recruiters also participating in the correspondence test experiments discussed in the previous section between August 2006 and January 2007 for *Experiment I* and between March and June 2007 for *Experiment II*. However, Lane et al. (2007) show that an individual's IAT scores are strongly correlated over time ($r > 0.5$).

The first step was to locate the firm's recruiter and, more exactly, the person who was responsible for selecting candidates to invite for an interview for exactly the job we applied for in the field experiment. This was achieved by calling the firm at the number included in the job ad. This was an extremely laborious task, reaching approximately only four recruiters a day. When reached they were informed that we were interested in this specific job vacancy and that we had followed its progress. However, we did not reveal that we had sent fictitious applications. We then informed them that our project intended to study the recruitment process in general and

¹⁶ It should be pointed out that written job applications being sent by e-mail are very common in Sweden and we have used job application exemplars available at the web page of the Swedish Employment Agency as templates, see the appendix for an example of a job application in the computer specialist occupation.

that their participation included first taking a “sorting test” on the computer and then answering a short, also computer-based, questionnaire, without revealing that we intended to measure their attitudes towards/stereotypes of Arab-Muslim minority males.

In an attempt to increase participation and to ensure their focus on the task employers/recruiters were offered a participation reward of 300 SEK (approximately 30 Euros or 39 dollars). They were then told that the reward for participation implies a total devotion to our study for approximately a total of ten minutes during which they were not allowed to be disturbed. They were allowed to choose whether to participate on a computer at work or at home. Still, recruiters were found to be a very busy group and were hard to convince to participate in the study.

4.2 Selective participation

A total of 729 and 811 employers/recruiters were invited to participate in *Experiment I* and *II*, respectively, see Table 1.¹⁷ Approximately fifty percent of the recruiters were not located, stated that the recruiting decision was not his/hers or not his/hers exclusively or were simply not interested in participating.¹⁸ In the first experiment 337 employers agreed to participate. However, in the end only 193 employers, or twenty-six percent, finally completed the IAT and the questionnaire. If focusing on each group/category (rows 1 through 4) separately we find that it is a similar share, around 30 percent, of the firms that have participated in taking the IAT,

¹⁷ Attempts were made to contact all employers in the first three categories (*only Swedish applicant, only Arab-Muslim applicant* and *both*), while, due to cost restrictions, only a random sample of employers were selected for contacting in the last group (*neither invited*).

¹⁸ Unfortunately, since the information was not collected we can not distinguish these cases in the data.

except for the category *Neither invited* for which only 19 percent participated.¹⁹ In the second experiment 389 employers agreed to participate, but in the end only 158 employers, or nineteen percent, finally completed the IAT and the questionnaire. If again focusing on each group separately we find that approximately twenty percent of the firms have participated in taking the IAT for the categories *Only Swedish applicant* and *Both*. Again, the lowest participation rate, fourteen percent, is found for the category *Neither invited*, while for the category *only Arab-Muslim applicant invited* as much as fifty percent participated.

Since a much smaller fraction of the group *neither invited* were invited to participate in the studies we need to use weights in the analysis of the data. These weights are calculated according to the within group occupational distribution of the field experiments where we also correct for participation differences across groups and occupations. In the appendix we also present a sensitivity analysis using weights based only on the sampling design and when not using weights.

*** Table 1 ***

Since a large fraction of the sampled recruiters never participated in taking the IAT it is possible that participation is selective, which could bias the estimated correlation between the implicit/explicit measures and the callback rate.²⁰ To what extent participation is selective is difficult to address empirically. However, non-participation seems to be unrelated to the employer/recruiter knowing what we intended to study. First, we failed to locate or reach approximately twenty-five percent of the recruiters (we stopped trying to contact them if not

¹⁹ This could be expected since this group also was found to be especially difficult to get in contact with.

²⁰ Although, given the results it is difficult to see what type of selectivity that would generate the data.

reached within one month's time) and another twenty-five percent directly stated, without us having informed them about the project, that they did not have time to participate. Hence, these groups were probably unaware of our purpose, but could of course still be selective.

Second, what about those who agreed to participate but never did? It is clear that no one started taking the IAT (which is needed in order to realize what the test is about) and then decided to withdraw, which is observed by viewing the output files. Instead, almost half of the recruiters not participating stated that they experienced problems with starting the program for firm Internet security reasons.²¹ Hence, since they did not start taking the IAT, they were also unaware of our intentions, but could again of course still be selective.²²

4.3 Explicit attitudes and stereotypes

The explicit measures used in the experiment were developed based on the measures used in Nosek, Greenwald, and Banaji (2005), and included the *feeling thermometer*, the *hiring preferences* task, and the *performance stereotype rating* task. The two former are expected to capture recruiters' explicit attitudes, while the latter their explicit performance stereotype, of

²¹ Many companies have Internet firewalls that do not allow for plug-ins as required by the web-based IAT.

²² Another strategy would be to compare the distribution of observable characteristics of participants and non-participants. However, since the information we have is very scarce, such a strategy is not very convincing. Even so, the ethnic difference in callbacks for interview is of about the same magnitude in the participants and non-participants sample and so is the occupational distribution. These results are available upon request.

Arab-Muslim men compared to Swedish men.²³ These questions appeared on the screen after the IAT had been completed.

The *feeling thermometer* asked the participants to rate their positive or negative feelings on a ten-point scale (1 = very negative feelings, 10 = very positive feelings) toward Arab-Muslim men and Swedish men. A difference between the two scales was calculated and used as the measure.

The *Hiring preference rating* asked the participants to choose which group, native Swedes versus Arab-Muslim, they prefer when hiring people. Rather than being directly related to the IAT, hiring preferences were supposed to measure actual explicit preferences in hiring, which could be related to discrimination. The employers had to choose one of five alternatives (e.g. “When hiring staff I strongly prefer Swedish men to Arab-Muslim men”) ranging from strong preference for Arab-Muslim relative to Swedish men (coded as -2), to a strong preference for Swedish relative to Arab-Muslim men (coded as +2), with a neutral intermediate alternative (coded as 0).

The *performance stereotype rating* task was constructed in the same manner as the *hiring preference* task, thus consisting of five alternatives (e.g. “Swedish men perform much better at work than Arab-Muslim men”) with a neutral intermediate, but this time the participants were instructed to indicate how well Arab-Muslim and Swedish men perform at work compared with each other.

²³ The choice of explicit measures is a difficult one. We decided to use these questions since they are very direct and could not be misunderstood. If using more subtle questions, as for example situation based ones, then we would move toward questions being somewhat implicit in character.

In *Experiment I* half (54%) of the employers explicitly state that they prefer hiring, or have more positive feelings towards (45%), a majority Swedish male over an Arab-Muslim minority male, while a clear majority (77 percent) state that there are no productivity differences between the two, see Table A2. In *Experiment II* a third of the employers explicitly state that they prefer hiring (33%), while half (52%) of them have more positive feelings towards a majority Swedish male over an Arab-Muslim minority male. Hence, these answers indicate the existence of quite strong explicit negative attitudes toward the Arabic minority, but less of a negative explicit productivity stereotype toward the same group. As argued in section 2, since the share reporting negative attitudes toward the Arab-Muslim minority is quite high we expect measurement error in this variable to be less of an issue when interpreting the empirical results in section 6.

5. The IAT – measuring implicit attitudes and stereotypes

The Implicit Association Test (IAT) has been developed within social psychology to measure automatic/implicit attitudes and stereotypes towards various groups. It is a computer based test designed to specifically measure individual differences in relative associations between two concepts. It was first introduced by Greenwald, McGhee, and Schwartz (1998) and has since become a widely used measure in order to capture individual's implicit attitudes toward and stereotypes of social groups.²⁴

The specific IATs used in the present research were the Arab-Muslim performance stereotype IAT and the Arab-Muslim attitude IAT found in Agerström, Carlsson and Rooth (2007).²⁵ These tests are identical except for the attribute categories (see below), i.e., the use of

²⁴ See <http://implicit.harvard.edu> for a test version of the IAT.

²⁵ A condensed version is found in Agerström and Rooth (in press).

unpleasant/pleasant words or words being associated with high/low work productivity. The following is a step by step description of the Arab-Muslim performance stereotype IAT.

In this computer based version participants first classify, as fast as possible, Swedish and Arab-Muslim sounding names appearing in the middle of the screen according to the (target) category to which they belong, “Arab-Muslim men” or “Swedish men”. In the next part of the IAT the participants now instead classify words that are found being associated with high and low work productivity. These include words such as “lazy”, “slow”, “efficient” and “hard-working”. In the third step names and high/low work productivity words appear interchangeably. The intuitive idea is that it will be easier, and hence, go faster, to classify names and words that are compatible than those that are incompatible. The IAT measures every latency in response to the presented stimulus. For example, when two categories are “easily” associated in terms of their nominal features (Arab-Muslim names + low work productivity and Swedish names + high work productivity) the participant classifies the stimuli much faster and with fewer errors than when they are not associated. A total of sixty stimuli are presented for the compatible and incompatible part, respectively. The difference in response latencies, or rather a recalculation of this difference gives the association strength, e.g. the IAT effect/score, between the compatible and incompatible parts.

Greenwald, Nosek, and Banaji (2003) have given bounds for the power of the IAT score, where below 0.15 is non-existent, 0.15-0.35 slight, 0.35-0.65 moderate, and an IAT-score above 0.65 is a strong effect.²⁶ The IAT scores of the 193 and 158 recruiters participating in *Experiment I* and *II*, respectively, show that a very clear majority associate words signaling negative productivity or unpleasant words with belonging to the Arab-Muslim minority. In fact, as many

²⁶ These bounds closely correspond to a conservative version of the bounds of the more familiar Cohen's d.

as 77 and 93 percent have an IAT score above 0.15 in the two experiments, respectively (see Figure 2a and 2b).

6. Results

The empirical analysis starts by investigating the correlation between the implicit association measures, i.e., the standardized IAT score, and the explicit attitude and stereotype measures, measuring relative preferences for and expected work productivity of majority Swedes and the Arab-Muslim minority, see Table 2. We find a slight positive correlation, ranging between 0.05 and 0.23, between the IAT score and the explicit statements. The correlations among the explicit measures are somewhat stronger and vary between 0.21 and 0.48. Hence, these correlations indicate that the Arab-Muslim performance stereotype IAT and the Arab-Muslim attitude IAT measures something that is related to, but far from the same as, the explicit measures.²⁷

*** Table 2 about here***

The next step is to analyze to what extent the implicit and explicit measures correlate with behavior, that is, the employers' decisions regarding whom to call for an interview, according to the model in Section 3. In doing so we analyze ethnic differences in the probability of a callback for interview using probit regressions (reporting marginal effects) and how this difference varies when introducing the implicit and/or explicit measure(s) into the regression. The data include

²⁷ It should be mentioned that the correlation between the implicit stereotype score and the explicit measures also is low ($r < 0.2$) in a student sample, a group that has less incentives to misreport their explicit statements, see Agerström, Carlsson, and Rooth (2007).

193/158 observations for Arab-Muslim and Swedish sounding applications, respectively, which are stacked together into the 386/316 applications being analyzed in *Experiment I* and *II*. All estimations are clustered on the level of the firm and weighted according to the strategy in section 4.1.²⁸ The implicit measure, as well as the explicit measures, were standardized before included in the analysis. However, we have also elaborated with using binary versions of the explicit measures, see Table A2 in the appendix for their construction.

When we regress the callback dummy on only the ethnic indicator variable we find an almost identical result as in Carlsson and Rooth (2007, 2008) using the full sample of observations, indicating that the weighting scheme is correct. Applications with an Arab-Muslim name attached to them have ten and seventeen percentage points lower probability of being called for interview compared to applications with a Swedish name, respectively.

To investigate which, if any, of the implicit and explicit attitude and stereotype variables are correlated with the ethnic difference in callbacks, we regress the callback dummy on the explicit and implicit measures and the interaction of those measures with the ethnic dummy, see Model A through F in Table 3a and 3b. The extent to which the included measures are associated with ethnic differences in callbacks is captured by the interaction effect, reported as a marginal effect. Model D through F control for occupation fixed effects and the interaction of these with the ethnic dummy. One could expect that the inclusion of occupation fixed effects would weaken the association strength of the implicit and explicit stereotype measures and the callback rate since stereotypes might be more important in certain occupations.

²⁸ An analysis using unweighted data and data weighted only according to the sampling design is found in Table A3 in appendix. Basically the magnitude of the estimates is the same, but becomes somewhat less precise.

*** Table 3a-b about here***

Experiment I

The estimate of the implicit stereotype measure in Model A implies that recruiters with more negative stereotypes of Arab-Muslims appear to treat applications with an Arab-Muslim name more negatively when hiring. The probability of a callback for interview is close to zero, and statistically insignificant, for Swedish named applications, but five percentage points lower (-0.051) and statistically significant for Arab-Muslim-named applications, when comparing recruiters with a one standard deviation difference in the implicit stereotype score.²⁹ The economic interpretation of this estimate is that the probability to invite job applicants with Arab-Muslim male names is, on average, five percentage points lower for recruiters that have at least a moderate implicit low-performance stereotype toward Arab-Muslim men in Sweden as compared to recruiters who have no such stereotypes.

This result is very robust, and is basically unchanged (the coefficient being reduced to -0.044), when the explicit attitude and stereotype measures are included into the model, see Model C, which is the specification that corresponds to Equation 2 in section 3. The estimates are also unaffected when occupation fixed effects are included, see Models D and F with a coefficient of -0.049 and -0.044, respectively. This implies that the association between the implicit stereotype measure and the ethnic difference in callbacks for interview is the same within as between occupations. However, in Model F the estimate of -0.044 is no longer statistically significant.

²⁹ How the means of the implicit and explicit measures vary between the four different callback categories are found in Table A4.

In the second column, Model B, only the explicit attitude and stereotype measures are included. While the point estimates for some of the interacted explicit measures are large, indicating that these variables might be economically important, they are estimated with a low precision. In fact, none of the interacted explicit measures are statistically significant in Table 3a.³⁰ These results for the explicit measures are unchanged when we include the implicit stereotype measure (Model C) and include occupation fixed effects (Model E and F).

Experiment II

The estimate of the implicit attitude measure in Model A is of the same size as in the previous experiment and implies that recruiters with more negative implicit attitudes appear to treat applications with an Arab-Muslim name more negatively. The probability of a callback for interview is close to zero, and statistically insignificant, for Swedish named applications, but five percentage points lower (-0.052) and statistically significant for Arab-Muslim-named applications, when comparing recruiters with a one standard deviation difference in the implicit attitude score.

Again this result is very robust, with the coefficient being increased only marginally to -0.058 when the explicit attitude and stereotype measures are included, see Model C. This is also the case when occupation fixed effects are included, see Models D and F with a coefficient of -0.055 and -0.066, respectively. However, in Model D the estimate is not statistically significant.

In the second column, Model B, only the explicit measures are included. While the point estimates for the interacted *hiring preference* measure is large, indicating that this variable might be economically important, it is estimated with a low precision. Using the binary version of this

³⁰ This does not change if we include the explicit variables one by one.

variable it becomes statistically significant (p-value 0.037), however, without affecting the interacted IAT variable.³¹

Finally, a question arises to why the estimate of the implicit measures, measuring automatic attitudes or stereotypes, are so similar in the two experiments. One possible explanation is that they are difficult to disentangle empirically since they are constructed in a similar way. But it could also be that they measure related but different constructs having a similar impact on the hiring decision.

7. Conclusion

This study is the first, to the best of our knowledge, to examine the relation among automatically activated associations, explicit attitudes and stereotypes, and discriminatory behavior in a real hiring situation. We find strong and consistent negative correlations between the IAT score and the probability that the firm/recruiter invited the applicant with an Arab-Muslim sounding name for an interview. The results are identical in the two experiments and show that the probability to invite job applicants with names such as Mohammed or Ali decreases by five percentage points when the recruiter has at least a moderate negative implicit association toward Arab-Muslim men in Sweden.

On the other hand, no such stable, and statistically significant, correlations are found for the explicit measures and the probability to invite job applicants with Arab-Muslim sounding names. More important, even when being large in magnitude and statistically significant the inclusion of the explicit measures into the analysis does not affect the estimate of the (interacted) IAT score.

³¹ It should be mentioned that this is the only time it makes a difference if using the binary version instead of the standardized discrete version of the explicit measures. These results are available upon request.

What our results then indicate is that there are recruiters who implicitly discriminate, but who would not explicitly do so. The results present evidence of that recruiting behavior is being affected by implicit prejudice rather than by explicit discrimination being observed by a *new* and better empirical measure. This conclusion is also being guided by the results from a large number of studies showing a low correlation between explicit and implicit measures of attitudes and by the recent advances in neuroscience showing that explicit and implicit measures of attitudes/stereotypes are identified in distinct neural systems. While automatically activated attitudes and stereotypes predict the probability that employers will call an applicant with an Arab-Muslim sounding name less often for an interview, it seems to be unrelated to the processes of the explicit attitudes and stereotypes. This is not to say that we present evidence of that explicit discrimination does not exist, but rather that explicit and implicit discrimination coexist in society.

Since implicit discrimination may be unintentional and outside of the discriminator's awareness effective preventing policies/actions will differ compared to when preventing explicit discrimination. The obvious first stage is educational - in order to start reducing implicit prejudice human resource managers must be informed about the existence of the implicit bias when hiring, and how to reduce it. Also, if companies would like to use education programs to increase awareness of all sorts of work related prejudice among their employers, we argue that IAT offers a promising possibility to test whether these education programs have indeed had the desired effect. This claim rests on the assumption that the IAT is a better and more convincing indicator of attitudes and stereotypes than are explicit evaluation tools.

To reduce implicit prejudice in the screening process the firm should incorporate a review process that draws attention to task cues rather than to social cues. Given that hiring decisions

are made on more subtle grounds (e.g., ethnicity or race) than job related qualifications, it is important for companies to ensure that recruiters adopt more stringent and thorough evaluation criteria when assessing the suitability of job applicants. Demanding recruiters to rate the applicant on several pre-established job relevant dimensions is one way to reduce the impact of implicit discrimination on hiring decisions. Increasing accountability by requiring recruiters to motivate their hiring decisions could also contribute to a more objective evaluation process. Such general actions also have spill-over effects in that all types of implicit discrimination, not only the one for race or ethnicity, should be reduced.

In conclusion, our study highlights that much can be gained by combining research methods in social psychology with applied research in economics. Our research demonstrates the utility of the IAT when it comes to predicting ecologically valid behavior with “real-world” relevance for the labor market and the economy. Elucidating the role played by automatic processes in the hiring process offers new insights into labor market discrimination. Understanding the “forces” underlying discrimination is crucial in order to eventually be able to create a fair labor market.

References:

- Agerström, J., Carlsson, R., and Rooth, D. (2007) "Ethnicity and Obesity – Evidence of Implicit Work Performance Stereotypes in Sweden", IFAU WP 2007:20, IFAU, Uppsala, Sweden.
- Agerström, J., and Rooth, D. (in press) "Implicit Prejudice and Ethnic Minorities: Arab-Muslims in Sweden", *International Journal of Manpower*.
- Altonji, J., and Blank, R. (1999) Race and Gender in the Labor Market. In Ashenfelter O, Card D (eds) *Handbook of Labor Economics*, 3143–3259. Elsevier. Amsterdam.
- Åslund, O. and Rooth, D. (2005) "Shifts in Attitudes and Labor Market Discrimination: Swedish Experiences after 9-11", *Journal of Population Economics* 18 (4), 602-629.
- Bertrand, M., Chugh, D., and Mullainathan, S. (2005) "New Approaches to Discrimination: Implicit Discrimination", *American Economic Review*, 95 (2), 94-98.
- Bertrand M., and Mullainathan, S. (2004) "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination", *American Economic Review*, 94, 991-1013.
- Carlsson, M., and Rooth, D-O. (2007) "Evidence of Ethnic Discrimination in the Swedish Labor Market Using Experimental Data", *Labor Economics*, 14, 716-729.
- Carlsson, M., and Rooth, D-O. (2008) "Is it Your Foreign Name or Foreign Qualifications? An Experimental Study of Ethnic Discrimination in Hiring", IZA DP#3810, IZA, Bonn.
- Chaiken, S., and Trope, Y. (1999) *Dual-process theories in social psychology*, New York: Guilford Press.
- Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., and Banaji, M. R. (2004) "Separable neural components in the processing of Black and White Faces", *Psychological Science*, 15, 806-813.

- Devine, P. (1989) "Stereotypes and Prejudice: Their Automatic and Controlled Components", *Journal of Personality and Social Psychology*, 56(1), 5–18.
- De Houwer, J., Thomas, S., and Baeyens, F. (2001) "Associative Learning of Likes and Dislikes: A review of 25 Years of Research on Human Evaluative Conditioning". *Psychological Bulletin*, 127, 853–869.
- Eurobarometer (2006) Public Opinion in the European Union. Report from the European Commission, Brussels.
- European Social Survey (2003) Majorities' Attitudes Towards Minorities: Key Findings from the Eurobarometer and the European Social Survey. Report from the EUMC, Vienna.
- Fazio, R.H., Dunton, J.R., and Williams, C.J. (1995) "Variability in Automatic Activation of an Unobtrusive Measure of Racial Attitudes: A Bona Fide Pipeline?", *Journal of Personality and Social Psychology*, 69, 1013-1027.
- Fazio, R., and Olson, M. (2003) "Implicit Measures in Social Cognition Research: Their Meaning and Use", *Annual Review of Psychology*, 54, 297-327.
- Greenwald, A.G., and Banaji, M.R. (1995) "Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes", *Psychological Review*, 102, 4-27.
- Greenwald, A., McGhee, D., and Schwartz, J. (1998) "Measuring individual differences in implicit cognition: The Implicit Association Test", *Journal of Personality and Social Psychology*, 74, 1464–1480.
- Greenwald, A., Nosek, B., and Banaji, M. (2003) "Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm", *Journal of Personality and Social Psychology*, 85, 197-216.

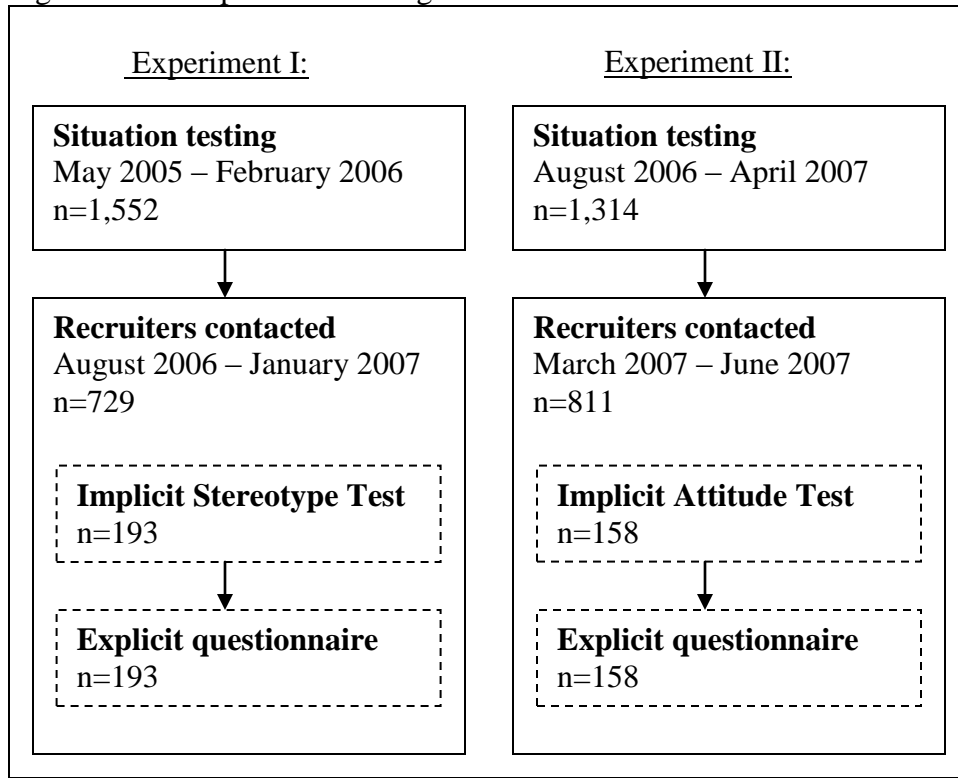
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E., and Banaji, M. R. (in press) “Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity”, *Journal of Personality and Social Psychology*.
- Hamermesh, D., and Biddle, J. (1994) “Beauty and the Labor Market”, *American Economic Review*, 84 (5), 1174-1194.
- Heckman, J., and Rubinstein, Y. (2001) “Importance of Noncognitive Skills: Lessons from the GED Testing Program”, *American Economic Review*, 91 (2), 145-149.
- Karpinski, A., and Hilton, J.L. (2001) “Attitudes and the Implicit Association Test”, *Journal of Personality and Social Psychology*, 81, 774–788.
- Kim, D. (2003) “Voluntary Controllability of the Implicit Association Test (IAT)”, *Social Psychology Quarterly*, 66, 83–96.
- Lane, K., Banaji, M., Nosek, B., and Greenwald, A (2007) “Understanding and Using the Implicit Association Test: IV. What We Know (So Far) About the Method”, in B. Wittenbrink and Schwarz, N. (Eds), *Implicit Measures of Attitudes: Procedures and controversies*, 59-102. New York: Guilford Press.
- Lange, A. (2000) *Diskriminering, integration och etniska relationer*, The Integration Board (Integrationsverket), Norrköping.
- McConnell, A. R., and Leibold, J. M. (2001) “Relations Among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes”, *Journal of Experimental social Psychology*, 37, 435-442.
- Nosek B., Greenwald, A., and Banaji, M. (2005) “Understanding and Using the Implicit Association Test: II. Method Variables and Construct Validity”, *Personality and Social Psychology Bulletin*, 31 (2), 166–180.

- Nosek, B. and Hansen, J. (2008) "The associations in our heads belong to us: Searching for attitudes and knowledge in implicit evaluation", *Cognition and Emotion*, 22(4), 553-594.
- Nosek, B., and Smyth, F. (2007) "A Multitrait-Multimethod Validation of the Implicit Association Test: Implicit and Explicit Attitudes are Related But Distinct Constructs", *Experimental Psychology*, 54, 14-29.
- Pew (2007) "World Publics Welcome Global Trade – But Not Immigration: 47-Nation Pew Global Attitudes Survey", Report from the Pew Global Attitudes Project, Washington.
- Phelps, E.A. (2006) "Emotion and Cognition: Insights from Studies of the Human Amygdala", *Annual Review of Psychology*, 24, 27-53.
- Phelps, E.A., and Banaji, M.R. (2005) "Animal models of human attitudes: Integrations across behavioral, cognitive, and social neuroscience", In Cacioppo, J.T., Visser, P.S., and Pickett, C.L. (Eds.) *Social Neuroscience: People Thinking about Thinking People* (229-243). Cambridge, MA : MIT Press.
- Ranganath, K.A., Smith, C.T., and Nosek, B.A. (2006) "Distinguishing Automatic and Controlled Components of Attitudes from Direct and Indirect Measurement Methods", *Journal of Experimental Social Psychology*, 44, 386-396.
- Riach, P. A., and Rich, J. (2002), "Field experiments of Discrimination in the Market Place", *The Economic Journal*, 112, F480-F518.
- Rooth, D-O. (2008) "Automatically Activated Stereotypes and Differential Treatment Against the Obese in Hiring", IZA DP#3799, IZA, Bonn.
- Rooth, D-O., and Ekberg, J. (2003) "Unemployment and Earnings for Second Generation Immigrants in Sweden. Ethnic Background and Parent Composition", *Journal of Population Economics*, 16 (4), 787–814.

- Stanley, D, Phelps, E.A, and Banaji, M.R. (2008) “The Neural Basis of Implicit Attitudes”, *Current Directions in Psychological Science*, 17(2), 164-170.
- Steffens, M.C. (2004) “Is the Implicit Association Test Immune to Faking?”, *Experimental Psychology*, 51, 165–179.
- Uhlmann, E. L., and Cohen, G. (2007) “I Think, Therefore It’s True: Effects of Self-Perceived Objectivity on Hiring Discrimination”, *Organizational Behavior and Decision Processes*, 104, 207-223.
- Ziegert, J., and Hanges, P. (2005) “Employment Discrimination: The Role of Implicit Attitudes, Motivation, and a Climate for Racial Bias”, *Journal of Applied Psychology*, 90 (3), 553-562.

Figures:

Figure 1. The experimental design



Note: The two experiments were conducted in different time periods, but were otherwise similar in that first, the outcome variable was collected in the situation testing (hiring) experiment, while the explanatory variables from the IAT and the explicit measures were collected in the second stage. Not all contacted recruiters participated in the second stage of the experiment.

Figure 1a. The distribution of the non-standardized IAT score for implicit stereotypes against Arab-Muslim minority men relative to native Swedish men.

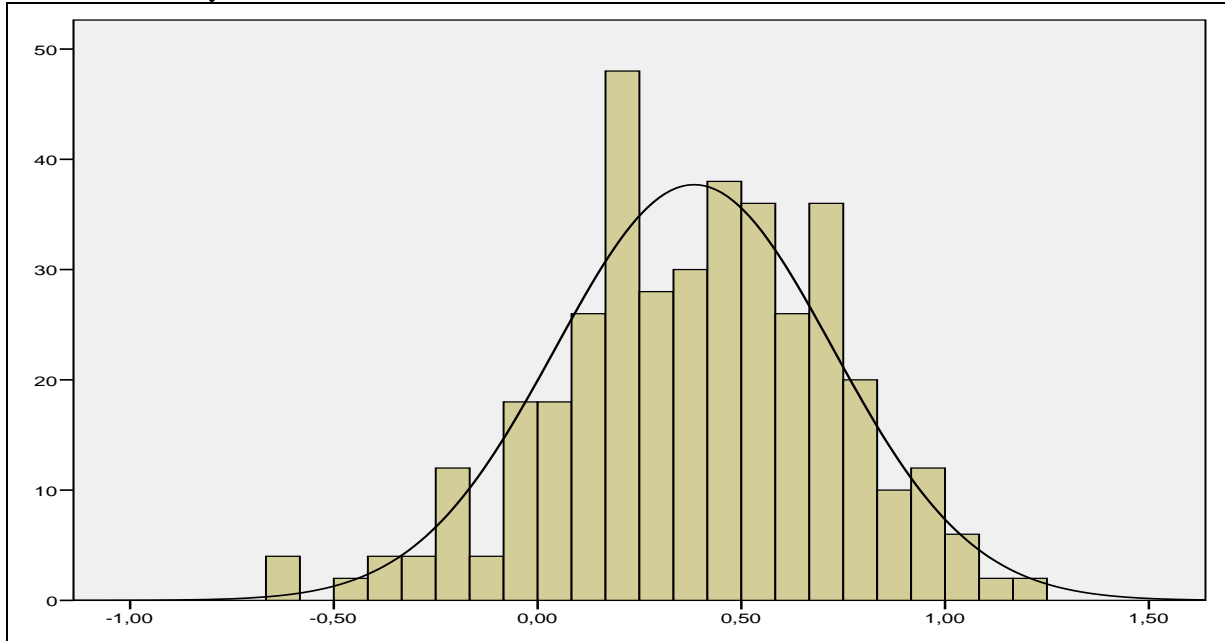
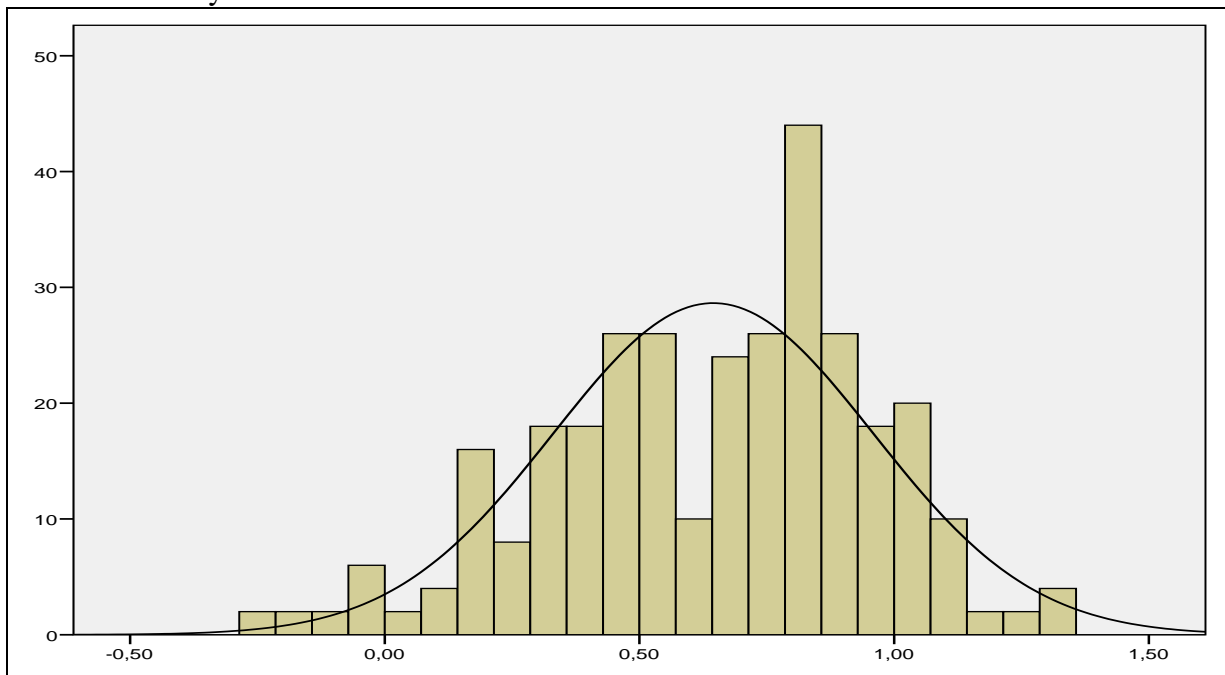


Figure 1b. The distribution of the non-standardized IAT score for implicit attitudes against Arab-Muslim minority men relative to native Swedish men.



Tables:

Table 1. Participation in the two experiments.

| Called for interview Different outcomes | Correspondence testing | IAT: Invited | Participated (share %) |
|--|---------------------------|-----------------|---------------------------|
| Experiment I: Implicit stereotypes | | | |
| Only Swedish applicant | 217 | 184 | 56 (30) |
| Only Arab-Muslim applicant | 66 | 59 | 18 (31) |
| Both | 239 | 192 | 62 (32) |
| Neither invited | 1,030 | 294 | 57 (19) |
| Total | 1,552 | 729 | 193 (26) |
| Experiment II: Implicit attitudes | | | |
| Only Swedish applicant | 265 | 230 | 43 (19) |
| Only Arab-Muslim applicant | 41 | 38 | 20 (53) |
| Both | 280 | 241 | 53 (22) |
| Neither invited | 728 | 302 | 42 (14) |
| Total | 1,314 | 811 | 158 (19) |

Note: From the field experiment we included all recruiters/companies from the original data (1,552 and 1,314, respectively) that were found in Statistics Sweden’s firm register. The reason why all firms were not invited to participate is because they have not been located in the register (due for example to close downs). Since only a fraction (29%) of the group “neither invited” were invited to participate in taking the IAT we need to use weights in the analysis.

Table 2. Correlation matrix showing the relationship between the Arab-Muslim performance stereotype IAT score/attitude IAT score and the standardized explicit measures. 193 and 158 cases, respectively.

| Variables | IAT | Feeling thermometer | Hiring preference | Performance stereotype |
|---|-------|------------------------|----------------------|---------------------------|
| Experiment I: Implicit stereotypes | | | | |
| Stereotype IAT | 1 | 0.10 | 0.23* | 0.05 |
| Feeling thermometer | 0.10 | 1 | 0.45* | 0.35* |
| Hiring preference | 0.23* | 0.45* | 1 | 0.21* |
| Performance stereotype | 0.05 | 0.35* | 0.21* | 1 |
| Experiment II: Implicit attitudes | | | | |
| Attitude IAT | 1 | 0.16* | 0.06 | |
| Feeling thermometer | 0.16* | 1 | 0.48* | |
| Hiring preference | 0.06 | 0.48* | 1 | |

Note: * $p < 0.05$. Non-weighted data. Almost identical results emerge when using the binary versions of the explicit measures and when using weights correcting for the sampling scheme.

Table 3a. Experiment I. The correlation between the callback rate for interview and the implicit stereotype and explicit attitude and stereotype measures. Percentage points.

| | Model: | | | | | |
|-----------------------------------|----------------------|----------------------|----------------------|--------------------|--------------------|--------------------|
| | A | B | C | D | E | F |
| Arab (job applicant) | -0.104*** (0.026) | -0.105*** (0.026) | -0.105*** (0.025) | - | - | - |
| Implicit measures: | | | | | | |
| IAT | 0.022 (0.034) | - | 0.018 (0.034) | 0.008 (0.032) | - | -0.007 (0.032) |
| IAT*Arab | -0.051** (0.023) | - | -0.044** (0.024) | -0.049* (0.027) | - | -0.044 (0.028) |
| Explicit measures: | | | | | | |
| Feeling thermometer | - | 0.077** (0.034) | 0.076** (0.034) | - | 0.089** (0.041) | 0.090** (0.040) |
| Feeling thermometer*Arab | - | -0.041 (0.030) | -0.039 (0.029) | - | -0.045 (0.036) | -0.038 (0.036) |
| Hiring preference | - | -0.003 (0.034) | -0.008 (0.035) | - | 0.002 (0.040) | 0.003 (0.041) |
| Hiring preference*Arab | - | -0.003 (0.025) | 0.009 (0.027) | - | -0.003 (0.024) | 0.009 (0.027) |
| Performance stereotype | - | -0.040 (0.040) | -0.041 (0.040) | - | -0.020 (0.041) | -0.020 (0.041) |
| Performance stereotype*Arab | - | -0.034 (0.028) | -0.031 (0.028) | - | -0.030 (0.041) | -0.032 (0.041) |
| Occupation fixed effects | No | No | No | Yes | Yes | Yes |
| Occupation fixed effects *Arab | No | No | No | Yes | Yes | Yes |
| No of cases | 386 | 386 | 386 | 386 | 386 | 386 |

Notes: (***), (**) and (*) indicate the one, five and ten percent level of significance, respectively. Each column shows the estimates from a regression of the callback rate on the set of variables given in the rows of the first column. The estimates are marginal effects reported after the dprobit command in STATA 9. Probability weights, taking both sampling design and non-participation into account, are used in all models. In columns D through F the estimate of the Arab indicator variable is discarded since it now corresponds to only one of the included occupations.

Table 3b. Experiment II. The correlation between the callback rate for interview and the implicit attitude and explicit attitude and stereotype measures. Percentage points.

| | Model: | | | | | |
|-----------------------------------|----------------------|----------------------|----------------------|-------------------|-------------------|--------------------|
| | A | B | C | D | E | F |
| Arab (job applicant) | -0.173*** (0.037) | -0.173*** (0.036) | -0.175*** (0.037) | - | - | - |
| Implicit measures: | | | | | | |
| IAT | 0.016 (0.044) | - | 0.016 (0.044) | 0.052 (0.042) | - | 0.057 (0.041) |
| IAT*Arab | -0.052* (0.030) | - | -0.058** (0.030) | -0.055 (0.037) | - | -0.066* (0.035) |
| Explicit measures: | | | | | | |
| Feeling thermometer | - | 0.003 (0.049) | 0.000 (0.050) | - | 0.001 (0.049) | -0.009 (0.050) |
| Feeling thermometer*Arab | - | 0.041 (0.041) | 0.053 (0.043) | - | 0.073 (0.049) | 0.085* (0.050) |
| Hiring preference | - | 0.024 (0.051) | 0.026 (0.051) | - | 0.051 (0.051) | 0.060 (0.052) |
| Hiring preference*Arab | - | -0.065 (0.050) | -0.070 (0.050) | - | -0.083 (0.055) | -0.092* (0.056) |
| Occupation fixed effects | No | No | No | Yes | Yes | Yes |
| Occupation fixed effects *Arab | No | No | No | Yes | Yes | Yes |
| No of cases | 316 | 316 | 316 | 316 | 316 | 316 |

Notes: (***), (**) and (*) indicate the one, five and ten percent level of significance, respectively. Each column shows the estimates from a regression of the callback rate on the set of variables given in the rows of the first column. The estimates are marginal effects reported after the dprobit command in STATA 9. Probability weights, taking both sampling design and non-participation into account, are used in all models. In columns D through F the estimate of the Arab indicator variable is discarded since it now corresponds to only one of the included occupations.

Appendix:

Table A1a. Experiment I. Aggregated results for the correspondence testing.

| | Jobs No. | Neither Invited No. | Both invited No. | Only Swedish- sounding name invited No. | Only Arabic- sounding name invited No. | Callback rates | | |
|--|-------------|---------------------------|------------------------|--|---|---|--|------------------------------------|
| | | | | | | Swedish- sounding name (3 + 4)/(1) | Arabic- sounding name (3 + 5)/(1) | Difference (Swedish- Arabic) |
| | (1) | (2) | (3) | (4) | (5) | (3 + 4)/(1) | (3 + 5)/(1) | (Swedish- Arabic) |
| Computer professionals | 106 | 71 | 9 | 14 | 12 | 0.22 | 0.20 | 0.02 |
| Teachers (math and science) ^a | 42 | 16 | 17 | 7 | 2 | 0.57 | 0.45 | 0.12 |
| Business sales assistants | 278 | 164 | 57 | 39 | 18 | 0.35 | 0.27 | 0.08 |
| Preschool Teachers | 184 | 64 | 76 | 36 | 8 | 0.61 | 0.46 | 0.15 |
| Accountants | 186 | 155 | 10 | 14 | 7 | 0.13 | 0.09 | 0.04 |
| Nurses | 150 | 95 | 30 | 20 | 5 | 0.33 | 0.23 | 0.10 |
| Teachers - upper secondary school | 64 | 41 | 10 | 11 | 2 | 0.33 | 0.19 | 0.14 |
| Teachers (language) ^a | 60 | 26 | 9 | 19 | 6 | 0.47 | 0.25 | 0.22 |
| Construction workers | 64 | 44 | 7 | 12 | 1 | 0.30 | 0.12 | 0.18 |
| Restaurant workers | 140 | 128 | 3 | 8 | 1 | 0.08 | 0.03 | 0.05 |
| Motor-vehicle drivers | 78 | 59 | 6 | 13 | 0 | 0.24 | 0.08 | 0.16 |
| Shop sales assistants | 200 | 167 | 5 | 24 | 4 | 0.14 | 0.04 | 0.10 |
| Total | 1,552 | 1,030 | 239 | 217 | 66 | 0.29 | 0.20 | 0.09 |

Source: Table from Carlsson and Rooth (2007). Notes: (a) Upper level of compulsory school.

Table A1b. Experiment II. Aggregated results for the correspondence testing

| | Jobs No. | Neither Invited No. | Both invited No. | Only Swedish- sounding name invited No. | Only Arabic- sounding name invited No. | Callback rates | | |
|--|-------------|---------------------------|------------------------|--|---|---|--|------------------------------------|
| | | | | | | Swedish- sounding name (3 + 4)/(1) | Arabic- sounding name (3 + 5)/(1) | Relative (Swedish)/ (Arabic) |
| | (1) | (2) | (3) | (4) | (5) | | | |
| Computer professionals | 78 | 27 | 29 | 20 | 2 | 0.63 | 0.40 | 0.23 |
| Teachers (math and science) ^a | 30 | 11 | 10 | 9 | 0 | 0.63 | 0.33 | 0.30 |
| Business sales assistants | 258 | 148 | 51 | 50 | 9 | 0.39 | 0.23 | 0.16 |
| Preschool Teachers | 186 | 66 | 84 | 28 | 8 | 0.60 | 0.49 | 0.11 |
| Accountants | 138 | 89 | 20 | 26 | 3 | 0.34 | 0.17 | 0.17 |
| Nurses | 150 | 73 | 43 | 30 | 4 | 0.48 | 0.31 | 0.17 |
| Teachers - upper secondary school | 48 | 24 | 11 | 10 | 3 | 0.44 | 0.29 | 0.15 |
| Construction workers | 48 | 26 | 6 | 15 | 1 | 0.44 | 0.15 | 0.29 |
| Restaurant workers | 150 | 101 | 10 | 32 | 7 | 0.28 | 0.11 | 0.17 |
| Motor-vehicle drivers | 78 | 43 | 13 | 20 | 2 | 0.42 | 0.19 | 0.23 |
| Shop sales assistants | 150 | 120 | 3 | 25 | 2 | 0.18 | 0.03 | 0.15 |
| Total | 1,314 | 728 | 280 | 265 | 41 | 0.41 | 0.24 | 0.17 |

Source: Table from Carlsson and Rooth (2008). Notes: (a) Upper level of compulsory school.

Table A2. Explicit measures and their binary counterparts. 193 and 158 cases, respectively.

| Explicit measure | Values | | | | | |
|---|--------|-----|-----|----|-----|----|
| | <0 | 0 | 1 | 2 | 3-4 | >4 |
| Experiment I: Implicit stereotypes | | | | | | |
| <i>Feeling thermometer:</i> | 6 | 92 | 38 | 24 | 19 | 14 |
| Binary version | - | 98 | 95 | - | - | - |
| <i>Hiring preference:</i> | 2 | 88 | 82 | 21 | - | - |
| Binary version | - | 90 | 103 | - | - | - |
| <i>Performance stereotype:</i> | 6 | 164 | 16 | 7 | - | - |
| Binary version | - | 170 | 23 | - | - | - |
| Experiment II: Implicit attitudes | | | | | | |
| <i>Feeling thermometer:</i> | 3 | 72 | 25 | 24 | 17 | 17 |
| Binary version | - | 75 | 83 | - | - | - |
| <i>Hiring preference:</i> | 2 | 104 | 42 | 10 | - | - |
| Binary version | - | 106 | 52 | - | - | - |

Note: Non-weighted data.

Table A3. Experiment I and II. The correlation between the callback rate for interview and the implicit stereotype and explicit attitude and stereotype measures. Percentage points. Unweighted data and weighted data based only on the sampling design.

| | Experiment I: Implicit stereotypes | | | | Experiment II: Implicit attitudes | | | |
|---------------------------------|------------------------------------|-------------------|----------------------|--------------------|-----------------------------------|---------------------|---------------------|---------------------|
| | Model: C | F | Model: C | F | Model: C | F | Model: C | F |
| Arab (job applicant) | -0.224*** (0.048) | - | -0.138*** (0.031) | - | -0.147*** (0.049) | - | 0.128*** (0.038) | - |
| Implicit measures: | | | | | | | | |
| IAT | -0.008 (0.038) | -0.030 (0.040) | -0.008 (0.037) | -0.029 (0.039) | 0.049 (0.041) | 0.064 (0.042) | 0.045 (0.041) | 0.073 (0.042) |
| IAT*Arab | -0.038 (0.043) | -0.042 (0.055) | -0.029 (0.028) | -0.041 (0.037) | -0.090** (0.045) | -0.106** (0.049) | -0.068** (0.034) | -0.087** (0.038) |
| Explicit measures: | | | | | | | | |
| Feeling thermometer | 0.036 (0.043) | 0.098* (0.051) | 0.065 (0.041) | 0.104** (0.046) | 0.010 (0.048) | 0.027 (0.050) | 0.015 (0.048) | 0.022 (0.048) |
| Feeling thermometer* Arab | 0.009 (0.051) | 0.010 (0.068) | -0.015 (0.032) | -0.012 (0.042) | 0.029 (0.057) | 0.042 (0.063) | 0.028 (0.045) | 0.036 (0.052) |
| Hiring preference | 0.010 (0.042) | 0.014 (0.047) | -0.006 (0.038) | 0.004 (0.046) | 0.010 (0.047) | 0.050 (0.054) | -0.006 (0.045) | 0.043 (0.049) |
| Hiring preference*Arab | -0.028 (0.050) | -0.053 (0.060) | -0.014 (0.032) | -0.029 (0.038) | -0.084 (0.074) | -0.131 (0.081) | -0.059 (0.058) | -0.101 (0.064) |
| Performance stereotype | 0.019 (0.056) | -0.011 (0.058) | -0.024 (0.052) | -0.034 (0.048) | - | - | - | - |
| Performance stereotype* Arab | -0.099 (0.068) | -0.124 (0.085) | -0.047 (0.038) | -0.056 (0.054) | - | - | - | - |
| Occupation fixed effects | No | Yes | No | Yes | No | Yes | No | Yes |
| Occupation fixed effects *Arab | No | Yes | No | Yes | No | Yes | No | Yes |
| Weighted by design | No | No | Yes | Yes | No | No | Yes | Yes |
| No of cases | 386 | 386 | 386 | 386 | 316 | 316 | 316 | 316 |

Notes: (***), (**) and (*) indicate the one, five and ten percent level of significance, respectively. Each column shows the estimates from a regression of the callback rate on the set of variables given in the rows of the first column. The estimates are marginal effects reported after the dprobit command in STATA 9. The non-weighted analysis is found in column 2, 3, 6 and 7, while the probability weighted analysis, taking only the sampling design into account, is found in column 4, 5, 8 and 9.

Table A4. Implicit and explicit variable means by group.

| Called for interview | IAT | Explicit measures | | |
|---|--------------|---------------------|-------------------|------------------------|
| Different outcomes | | Feeling thermometer | Hiring preference | Performance stereotype |
| Experiment I: Implicit stereotypes | | | | |
| Only Swedish applicant | 0.18 (1.03) | 0.13 (1.11) | 0.15 (1.06) | 0.26 (1.24) |
| Only Arab-Muslim applicant | 0.09 (0.83) | 0.11 (0.98) | -0.03 (1.13) | 0.09 (0.78) |
| Both | -0.17 (0.91) | 0.01 (1.07) | -0.05 (0.85) | -0.15 (0.68) |
| Neither invited | -0.02 (1.10) | -0.17 (0.80) | -0.08 (1.06) | -0.11 (1.07) |
| Total | 0.00 (1.00) | 0.00 (1.00) | 0.00 (1.00) | 0.00 (1.00) |
| Experiment II: Implicit attitudes | | | | |
| Only Swedish applicant | 0.19 (0.85) | 0.00 (0.91) | 0.11 (1.10) | - |
| Only Arab-Muslim applicant | -0.29 (1.02) | -0.20 (0.61) | -0.33 (1.21) | - |
| Both | -0.01 (1.04) | 0.07 (1.12) | -0.04 (0.78) | - |
| Neither invited | -0.05 (1.06) | 0.01 (1.07) | 0.09 (1.00) | - |
| Total | 0.00 (1.00) | 0.00 (1.00) | 0.00 (1.00) | - |

Note: All variables are presented as standardized means of their continuous/discrete versions.

Appendix: Description of Application Construction

An ideal design for the hiring experiment would be to use exactly the same application, only varying the name of the applicant. Using this design, any differential treatment by employers would be due solely to the name manipulation. However, for obvious reasons such a strategy is not possible to implement empirically without employers being suspicious. Hence, we had to use a design using similar but different application texts, and alternate these randomly between applicants. In their study, Carlsson and Rooth found no systematic differences in callback rates between the different occupation specific applications being used (nor for the names used).

Construction of Job Applications

Job seekers can upload their CV on the webpage of the Swedish employment agency for employers to view. From this source, Carlsson and Rooth (2007) collected approximately 20 templates for each occupation, which were finally rewritten into only two application types per occupation (see below for the ones used in the *computer specialist* occupation, having been translated into English). The constructed application types consist of a quite general biography on the first page and a detailed CV (not shown) of education and work experience on the second page. The only difference between the two types of applications is the font and font size (Verdana 10 or Times New Roman 12) being used. The content in the two applications is the same and is formulated in a similar, but not identical, way. Colleagues at Kalmar University were asked to assess whether the applications were perceived as similar in qualifications within occupations, and applications were rewritten if not.

Application exemplars (translated into English): Computer specialists

Application A (in Verdana 10)

Hi,

My name is Karl Johansson and I am 30 years old. Previously I worked as a system designer at Telenor AB between 1998 and 2005 in an environment based on win2000/SQL Server. I then participated in three different projects and my work contained development, maintenance and every day problem solving. Development work was done in ASP, C++ and Visual Basic and we used the development platform .Net and MS SQL. In addition I have experience in HTML, XML, J2EE and JavaScript.

I enjoy working with development and problem solving. And I now hope that I will develop further at your company. To my personal characteristics one could add that I find it easy to work both on my own and in a group. I am a dynamic person that likes challenges. I really like my occupation which I think is mirrored in the work I do. I have a degree in computer engineering. I graduated with good grades from Stockholm university.

In my spare time I like to read and listen to music. Me and my wife also like to socialize with our friends during weekends.

I look forward to being invited for interview and I will then bring my good certificates and diplomas.

Best regards

Karl Johansson

Application B (in Times new roman 12)

Hi,

First of all I would like to introduce myself. My name is Mohammed Said, and I am 29 years old and live in Stockholm.

I previously worked at EssNet AB for about 7 years and my work tasks included designing, implementing as well as testing different financial applications. The programming was mainly done in a Java J2EE environment. A lot of the development was also done in C/C++ and Visual Basic. Sometimes I also participated in projects which were more web based and linked to data bases. This work gave me experience in ASP, JSP and MS SQL among other things.

To my skills one could also add a great work joy and that I do not want to finish up work only half done. I work independently and am good at taking own incentives. At my previous job I also learned how to cooperate and to listen to others and together create great results.

As regards my educational background I have an engineering degree in technical physics with a focus on computer science at Uppsala University.

Since I am very social person a great deal of my spare time is spent together with my friends and my wife. Traveling is also a great interest of mine. I hope I will get the opportunity to come for an interview and then get the chance of telling you more about myself and to show my good grades and credentials.

Best regards

Mohammed Said