



IFAU – INSTITUTE FOR
LABOUR MARKET POLICY
EVALUATION

Ethnicity and obesity: evidence of implicit work performance stereo- types in Sweden

Jens Agerström
Rickard Carlsson
Dan-Olof Rooth

WORKING PAPER 2007:20

The Institute for Labour Market Policy Evaluation (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 policies, studies of the functioning of the labour market, the labour market effects of educational policies 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.

IFAU also provides funding for research projects within its areas of interest. The deadline for applications is October 1 each year. Since the researchers at IFAU are mainly economists, researchers from other disciplines are encouraged to apply for funding.

IFAU is run by a Director-General. The institute has a scientific council, consisting of a chairman, the Director-General and five other members. Among other things, the scientific council proposes a decision for the allocation of research grants. A reference group including representatives for employer organizations and trade unions, as well as the ministries and authorities concerned is also connected to the institute.

Postal address: P O Box 513, 751 20 Uppsala

Visiting address: Kyrkogårdsgatan 6, Uppsala

Phone: +46 18 471 70 70

Fax: +46 18 471 70 71

ifau@ifau.uu.se

www.ifau.se

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.

Ethnicity and obesity: evidence of implicit work performance stereotypes in Sweden*

by

Jens Agerström*, Rickard Carlsson* and Dan-Olof Rooth*

August 31, 2007

Abstract

Using the Implicit Association Test, we investigate whether employers and students possess implicit and explicit negative attitudes and implicit performance stereotypes toward Arab-Muslim men relative to native Swedish men. We also examine if employers and students have implicit and explicit performance stereotypes toward obese individuals relative to people of normal weight. The results demonstrate that employers and students both implicitly and explicitly associate Arab-Muslim men with less work performance. Also, they have more implicit negative attitudes toward this ethnic group. Obese individuals are both implicitly and explicitly associated with less work performance compared with normal-weight individuals.

Keywords: Implicit, attitudes, stereotypes, discrimination, ethnicity, obesity
JEL-codes: J53, J64, J71

* We thank Magnus Carlsson, two anonymous referees and seminar participants at IFAU for valuable comments. A research grant from IFAU is gratefully acknowledged. Klara Johansson, Terese Johansson and Markus Mowitz provided excellent research assistance.

* Department of psychology, Lund University. Email: Jens.Agerstrom@psychology.lu.se

* Department of economics, Kalmar University. Email: Rickard.Carlsson@hik.se

* Corresponding author. Department of economics, Kalmar University. Email: Dan-Olof.Rooth@hik.se

Table of contents

1	Introduction	3
2	Prejudice: attitudes and stereotypes.....	5
2.1	Measuring attitudes and stereotypes.....	6
2.2	Measuring implicit attitudes and stereotypes: The implicit association test.....	9
3	The present research	12
3.1	Overview of the experiments.....	14
3.2	Experiment 1A.....	15
3.2.1	Method.....	15
3.2.2	Results and discussion	21
3.3	Experiment 1B.....	24
3.3.1	Method.....	24
3.3.2	Results and discussion	26
3.4	Experiment 1C.....	27
3.4.1	Method.....	27
3.4.2	Results and discussion	28
3.5	Experiment 2A.....	29
3.5.1	Method.....	29
3.5.2	Results and discussion	30
3.6	Experiment 2B.....	31
3.6.1	Method.....	31
3.6.2	Results and discussion	31
4	General discussion.....	32
	References.....	39
	Appendix: Development of the implicit measures.....	43

1 Introduction

The concept of *equal opportunity in employment* implies that job applicants and employees should be free to compete equally on the basis of merit and that they should not be discriminated against due to personal characteristics that are unrelated to job performance. Unfair employment opportunities, as manifested in unequal treatment of observable demographic groups, such as immigrants and females, is a topic that has attracted a lot of attention, both in the media and in academic research.

In research on discrimination in the labor market, economists have traditionally focused on taste-based and statistical discrimination theories. Taste-based discrimination occurs when an employer¹ discriminates against an individual because of the employer's explicit negative attitudes toward the social group which the individual is part of. Statistical discrimination on the other hand, is due to the employer having explicit stereotypes towards a social group, which include the beliefs that individuals belonging to this group are less suited for the job because of certain traits (e.g., low work performance). These types of discrimination have traditionally been assumed to operate on a conscious level, and consequently discrimination has been considered a deliberate act. However, current research on social cognition suggests that prejudice, as well as discrimination, may operate unconsciously. In other words, employers may unintentionally discriminate against individuals because of prejudice they are not aware of (Bertrand et al., 2005).

An important first step in understanding this *implicit discrimination* in the hiring process is to investigate if employers indeed possess *implicit* prejudice that could potentially be related to the discrimination² of groups that are in an unfavorable position in the labor market. Two groups that have been shown to be in an unfavorable position in the Swedish labor market are the Middle Eastern male³ minority and obese people.⁴

¹ Employer or recruiter or any other person responsible for the recruitment of staff. Henceforth we use the term "employer" to facilitate reading.

² Discrimination often refers to the juridical definition of illegal unequal treatment. However, while ethnicity is illegal to sort on when hiring people, obesity is not. To facilitate reading we interchangeably use the terms discrimination and unequal treatment.

³ This does not seem to be the case for women with the same ethnic background and we therefore focus only on men; see also the discussion in section 4. For obesity, no such distinction is made.

⁴ Henceforth, we mainly use the category "Arab-Muslim", which is included in the wider category "Middle Eastern".

Two field experiments conducted by Carlsson and Rooth (2007) and Rooth (2007a) have recently demonstrated that Swedish employers discriminate against people with male Middle Eastern sounding names and obese people when recruiting staff. In Carlsson & Rooth (2007), two fictitious job applications, one with a Swedish sounding name and one with an Middle Eastern sounding name, all other things being equal, were sent to employers for processing. It was found that applications with a Swedish name received fifty percent more callbacks for interview. Furthermore, in an analogous field experiment, Rooth (2007a) found that normal-weight applicants have a twenty percent higher chance of being called for an interview compared with obese applicants. Hence, ethnic minorities and obese individuals are both subjected to unequal treatment in the labor market. Also, in occupations where discrimination is high for one of the groups, discrimination is also high for the other. Taken together, this suggests that similar discriminating processes could be at work, affecting both target groups negatively.

There are several other sources indicating discrimination of these two groups. For instance, approximately forty percent of immigrants from Iran, which make up a large share of the Middle Eastern minority in Sweden, perceive that they have been discriminated against in the labor market within the last year (Lange, 2000). Furthermore, interviews with Swedish managers show that nine out of ten managers believe that employment decisions depend upon the applicant being obese and a major reason for this belief is that obese applicants are expected to be less productive at work (Dagens Nyheter, 2003).

International research has also shown that people tend to have implicit prejudice toward Arab-Muslims as well as obese people. For example, Nosek et al. (2007) have, in a mainly American sample, found that people hold more implicit and explicit negative attitudes toward Arab-Muslims, as compared to other people, and Schwartz et al, (2006) found that people tend to have more implicit and explicit negative attitudes as well as laziness stereotypes toward obese individuals relative to thin people. In Sweden, Ekehammar et al. (2003) have observed implicit and explicit prejudice toward a more general category of immigrants in a student sample.

Although research has been done internationally on implicit prejudice against Arab-Muslims⁵ and obese people, such studies have not been conducted on a sample of employers in a Swedish context, despite its apparent relevance.

⁵ Note that we focus on attitudes *and* stereotypes toward Arab-Muslim *males*, but only on stereotypes toward obese persons of *both sexes*.

Such research would have significant value, because if employers demonstrate a substantial bias in this regard, this could be related to actual discriminative behavior in the labor market. Consequently, the primary aim of this study is to investigate whether employers are implicitly prejudiced against the Arab-Muslim male minority⁶ and obese people. For this purpose, the Implicit Association Test (IAT, Greenwald et al., 1998; see section 2 for a description of the method) will be used. More specifically, we have developed an Arab-Muslim attitude IAT, an Arab-Muslim performance stereotype IAT, and an obesity performance stereotype IAT that allow us to examine to what extent employers have more negative attitudes and work performance stereotypes toward Arab-Muslims compared with Swedes, and more negative work performance stereotypes toward obese individuals compared with people of normal weight. These experiments were conducted in 2006 and 2007.

The remainder of this paper is outlined as follows. In Section 2, we present research on the role of implicit cognitive processes in social judgment. Furthermore, we discuss theories of implicit and explicit attitudes and stereotypes, and how these concepts can be measured. In section 3, we specify the aims of this study and present an overview of the experiments conducted. Then we present the methods used and the empirical data obtained in the five experiments conducted for this study. Finally, section 4 concludes the paper with a general discussion.

2 Prejudice: attitudes and stereotypes

In common language the word prejudice is most often used to describe unfavorable feelings toward and negative judgments of people based on their membership of certain social groups. Occasionally it is used to describe neutral or

⁶ Arab-Muslim male minority is a description of the group that we believed is best suited to describe the group that has been shown to be discriminated against, when male job applicants with Arab-Muslim/Middle Eastern sounding names such as "Mohammed" and "Hassan" are rejected (see Carlsson & Rooth, 2007). We realize that there are individuals with these names that do not consider themselves either Arabs or Muslims. However, these individuals may nevertheless be subject to the same level of discrimination, if they are perceived by others as belonging to that group, because of names or looks. This is perfectly plausible when employers recruit staff since the employer typically does not know all the details. Although the term Arab-Muslims is a simplification, we decided to use it in the present research. We are convinced that it is the most appropriate term for this particular group, given the methodological restrictions discussed in the appendix.

positive beliefs of social groups as well. However, in research the term prejudice typically refers to negative feelings, negative judgments and negative generalized beliefs of social groups. Thus, this definition includes both negative attitudes and negative stereotypes (Akrami, 2005). This is the definition used throughout this paper.

Greenwald & Banaji (1995) have reviewed the definitions of attitudes and stereotypes in the literature of the last century. Based on this, they define attitudes as “favorable or unfavorable dispositions toward social object, such as people, places and policies” (Greenwald & Banaji, 1995:7). Stereotypes, on the other hand, are defined as the generalized set of beliefs we have of certain groups of people. Typically the stereotype is a set of traits that is thought to describe the individuals belonging to the specific group. These traits do not have to be positive or negative and can sometimes be quite neutral. For example, a stereotype that says that Italian people love football could be seen as neutral to most people. Essentially, the attitude tells us whether we like or do not like a certain group. The stereotypes, on the other hand, can give us information about the specific beliefs.

The same traits can also be negative or positive depending on context (Greenwald & Banaji, 1995). For example, the stereotype of men could include the belief that they are generally aggressive, a trait that may be very positive in conjunction with certain sports activities such as boxing, but negative when it comes to child care. It is also common to have both positive and negative stereotypes toward the same group of people. For example, a traditional stereotype of women may include that they are good at housekeeping, emphatic, warm and caring, alongside with beliefs that they are weak and poor at math. Nevertheless, the overall existing attitude may be positive. This illustrates the sometimes rather complex relationship between coexisting attitudes and stereotypes.

2.1 Measuring attitudes and stereotypes

The traditional method of assessing prejudice is to ask the persons what beliefs they have (stereotypes) and what they think/feel (attitudes) about a social group. This is conventionally done in self-reports. Although these self-reports, henceforth referred to as explicit measures, may offer some information about the level of prejudice, there are substantial limitations with this method (Greenwald & Banaji, 1995). The most obvious limitation is caused by self-presentational strategies. It is almost a truism to say that people tend to not re-

port how they really feel about some groups of people and instead give the researcher a more socially desirable version of their attitudes and beliefs in order to convey a more positive persona (Akrami & Ekehammar, 2005). This problem can to some extent be reduced if the intention of the researcher is hidden and the questions are more subtle, as in the Modern Racism Scale (Akrami et al., 2000).

Apart from the self-presentational issues, researchers in the field of implicit social cognition (e.g. Greenwald & Banaji, 1995; Nosek, 2005) have proposed another limitation of these traditional explicit measures. The reason for people not reporting prejudice may not be due to the fact that they are untruthful to the researchers, or to themselves, but because some prejudice may exist outside of their awareness, which makes it impossible for them to report it. According to Greenwald and Banaji (1995), even if we are not aware of this implicit prejudice, it may still affect our behavior, perhaps even more so than explicit prejudice which operates consciously, and thus may be easier to control. Hence, discrimination in the labor market could possibly occur due to implicit attitudes and/or implicit stereotypes that people are unaware of, and thus unable to report even if motivated to do so truthfully.

With the distinction of explicit and implicit attitudes and stereotypes comes the question of the relationship between these constructs. As of yet, there is no general consensus on the matter. Today most researchers agree that there is some relation between the constructs. However, some view them as distinct, but related, constructs, while some researchers regard explicit and implicit prejudice as a single combined construct, arguing that the explicit measure of prejudice taps the same construct but “further down the line”, after controlled processing has been initiated (for reviews see Fazio & Olson 2003; Nosek & Smyth, 2007).

Regardless of what the case may be, it is clear that the results from implicit and explicit measures of attitude often differ substantially, and there is neurological evidence that implicit measures reflect more automatic than controlled reactions to social groups (Cunningham et al., 2004). For example, a study conducted by Phelps et al. (2000) found that participants’ performance on an implicit race measure was positively correlated with the activity in the amygdala, a structure in the brain that mediates fear responses and can trigger behaviors before people are consciously aware of this (LeDoux, 1994), indicating that those participants who displayed more implicit negative attitudes toward unfamiliar black faces also evidenced heightened levels of

amygdala activity. However, there were no association between the explicit measures and the activation of the amygdala.

The correlation between the results of explicit and implicit measures is naturally a part of this discussion. For a long period of time there seemed to exist little or no positive correlation between implicit attitudes and stereotypes and their explicit counterparts. This had researchers theorize that explicit and implicit attitudes and stereotypes are distinct and totally unrelated constructs. For example Devine (1989) argued that implicit stereotypes are activated automatically because of culturally shared knowledge, but that explicit prejudice reflects the individuals' personal thoughts and feelings. However, the recent explosion of research showing small, moderate and even strong correlations between explicit and implicit measures, have contributed to shift the question from whether implicit and explicit measures are related at all, to how, why, and under which conditions they are.

In one such study, Nosek (2005) found that the implicit-explicit consistency is contingent upon a number of factors, and one of the most important is self-presentation. Typically, for topics that are socially sensitive, the correlations between implicit and explicit measures are generally weaker. Nosek et al. (2007) find for a number of different topics that implicit and explicit stereotypes are on average only weakly correlated, while implicit and explicit attitudes toward social groups are on average moderately correlated.

Of greatest importance for the present study is the predictive validity of measures of stereotypes and attitudes. There are currently a great number of studies showing that both explicit and implicit measures can be successful in predicting behavior. In a meta-analysis conducted by Poehlman et al. (2005), it was demonstrated, however, that the explicit measures (self-reports) and implicit measures (in this study different types of IAT:s) are better predictors in different areas. In areas where self-presentational strategies are not an issue, the explicit measures will predict behavior better than the implicit measures. The strength of the implicit measures, on the other hand, is in the social sensitive areas where explicit measures often predict behavior only poorly or not at all.

The different predictive value of the explicit and the implicit measures does not only depend on how socially sensitive the topic is, but also which type of behavior it is supposed to predict. Typically, the ability for explicit measures to predict behavior weakens as the type of behavior becomes more automatic and harder to control. The implicit measures do not suffer from this decline in

predictive value. Consequently, implicit measures are good tools for predicting both highly conscious and controllable behavior, as well as behavior that is hard to control and outside awareness (Poehlman et al., 2005).

In summary, it is clear that in the social sensitive area of prejudice, implicit measures have qualities that make them the primary method of choice. Regardless of whether this has to do with the reduction in self-presentational bias, or whether implicit measures are tapping a construct that the participants are unaware of, the implicit measures offer researchers a new way to capture prejudice. Importantly, these measures offer a new way of examining the roots of labor market discrimination.

2.2 Measuring implicit attitudes and stereotypes: The implicit association test

In a review of the major implicit measures used in research, Fazio & Olson (2003) showed that even though there are currently a great number of different implicit measures, including small or large variations of the same measure, many are still in its infancy and not yet well suited for applied research. In applied research, the by far most widely used implicit measure is the Implicit Association Test. The Implicit Association Test (IAT) is a computer based test designed to measure individual differences in associations between concepts and attributes. It was first introduced by Greenwald et al. (1998) and has since become a widely used measure, particularly in areas where implicit attitudes toward and implicit stereotypes of social groups are studied.

In the IAT, the two keys, “d” and “k”, are mapped to categories on the left and right side of the screen, respectively. Using these keys, the participants classify the stimuli (words or images) appearing in the middle of the screen according to the category which they belong to, and for each and every response made by the participant to a presented stimulus, the computer will measure the response time in milliseconds. The category labels on top of the screen help the participant remember which key is mapped to which category. The following is a step by step description of the test based upon the original Black and White attitude-IAT presented by Greenwald et al. (1998).

First, the participant classifies the *target* stimuli, consisting of names such as “Latonya” or “Ebony” and “Katie” or “Peggy”, to the two *target* categories “Black” and “White”. In the next part of the IAT, the participant now instead classifies the *attribute* stimuli, which in this case are unpleasant and pleasant words. These include words such as “cancer” or “murder” and “happy” or

“joy”, which belong to the two *attribute* categories “unpleasant” and “pleasant”.

Then follows the critical part of the IAT, where there are now four categories but still only two keys for the user to respond with. Consequently, two categories will now share a response key. When two categories that are easily associated in terms of their nominal features (“White” + “Pleasant”/“Black” + “Unpleasant”) share the same response key, the participants generally classify the stimuli much faster and with fewer errors than when they are not associated, as is the case when “White” + “Unpleasant” share a response key and “Black” + “Pleasant” share the other. The key mapping that is easier, i.e., performed faster and with fewer errors, is called the *compatible* part and the other is called the *incompatible* part.

Later in the test the participant performs this critical part of combined tasks, with the key mapping of the target category in a reversed order. If the participant in the first critical part had “Black” + “Unpleasant” share a key (and consequently, “White” + “Pleasant” share the other), he or she will now instead have “Black” + “Pleasant” share the same key. Accordingly, “White” will share a key with “Unpleasant”.

Since the combined parts are done with both normal and reversed key mapping, it is possible to calculate differences in response times between these two critical parts of the test. This difference represents the relative association of the target concept and the attributes. In this case, if the part of the test with “Black” + “Unpleasant”/“White” + “Pleasant” generally has lower response times than “White” + “Unpleasant”/“Black” + “Pleasant”, it suggests that there is an automatic preference for Whites over Blacks, or, in other words, an implicit negative attitude toward blacks, as compared to whites. This difference in response latencies between the two critical parts is known as the IAT-effect.⁷

The design of the IAT is such that it can only assess relative difference in association. In the example above, the attitude toward Blacks is relative to the

⁷ The actual computation is somewhat more complicated. Since response times in computerized tests are highly individual, the average response time will vary greatly depending on a number of things, such as computer habits, cognitive skill, motivation, concentration and age. Having a generally higher response time will also result in higher variance. Simply calculating the mean difference in ms between the compatible and incompatible task will be highly confounded with the average latencies of the individual participants (Greenwald et al., 2003). Using large data sets Greenwald et al., (2003) evaluated a number of different scoring algorithms for the IAT and they concluded that the so called *D*-measure was the best choice in terms of several objective criteria. Consequently, the *D*-measure was used in the present study. How this computation was done is explained in detail in section 3.

attitude toward Whites. While the relative nature of the IAT can in some cases be a limitation, it is actually beneficial in the case of research on labor market discrimination, since it is in fact the relative difference in how the groups are perceived and treated that is of interest.

The target concepts in the IAT can easily be changed to address attitudes toward other target groups, such as Arab-Muslims or obese people. Also, the attributes can be changed to measure stereotypes rather than attitudes. For example, Intelligent/Unintelligent can be coupled with their respective target groups. In this regard, the IAT is a very flexible test which can measure associations between whichever concepts and attributes the researchers wish to investigate, using words, names, pictures, or photos as stimuli.

Apart from this flexibility, the IAT is also known for its high reliability, which is superior to other implicit measures. For example, in a meta-analysis of 50 studies internal reliabilities averaged .79 (Hofmann et al., 2005). In addition, the IAT typically has acceptable levels of test-retest interrelations (typically .60 or higher), whereas this varies greatly for other implicit measures, where figures close to zero are often reported (Fazio & Olson, 2003). However, the most compelling advantage of the IAT is an extensive body of research demonstrating correlations with explicit measures (Nosek, 2005; Hofmann et al, 2005), and good predictive validity with respect to behavior in various domains, e.g., judgment and decision making (Poehlman et al., 2005).

Other benefits of the IAT are that it is fairly easy to administer, takes little time to complete (typically 10 minutes), and the test itself does not involve deception as many other implicit measures often do (e.g., in subliminal priming tasks⁸). Also, compared to other implicit measures, the IAT can produce substantially larger statistical effects, allowing for studies with relatively few participants. Furthermore, it has been used extensively and successfully outside the laboratory, on the Internet, and within the *Project Implicit*⁹.

One important advantage for the present study is its high resistance to faking. In fact, participants typically cannot successfully fake their IAT-score even

⁸ In a subliminal priming task, the participants are typically flashed with words or images for a period of time that is too brief for them to consciously detect them. This subliminal prime, of, for example, a photo of an Arab-Muslim, may unconsciously affect their ability to classify words as good or bad. That is, a prejudiced person may have trouble classifying a positive word as such, if subliminally primed with a photo of an Arab-Muslim prior to the classification. In order for these test sessions to work properly, the participants have to be unaware of the researcher's intent.

⁹ *The Project Implicit* is the website that was developed by Greenwald and his associates to demonstrate the IAT to the public. Since 2000, over four million IAT sessions have been performed in this virtual online laboratory.

if explicitly instructed to by researchers. In order to fake their results the participants need prior experience with the IAT, as well as extensive knowledge of how the IAT works. (e.g., Kim, 2003; Steffens, 2004).

3 The present research

Bertrand et al. (2005) suggest that using the IAT to test for the existence of implicit negative attitudes among employers could be important in order to uncover the roots of overt discrimination. However, it is possible that implicit discrimination occurs as a result of negatively held implicit attitudes, biased implicit work performance stereotypes – or a combination of both. In order to investigate if employers possess implicit prejudice that may exert an influence on their likelihood to discriminate, both these possibilities should be considered.

Attitudes toward Arab-Muslims and obese individuals have been studied elsewhere, although not in a sample of Swedish employers (Nosek et al., 2007; Schwartz et al., 2006). However, there are to our knowledge no studies that have aimed to capture implicit stereotypes that pertain directly to work performance. The implicit laziness stereotype investigated by Schwartz et al. (2006) is close, but does only capture a small part of implicit stereotypes that could possibly influence employers' decisions when hiring staff. It would thus be fruitful to capture a more detailed implicit work performance stereotype that covers various aspects of low work performance.

In the present research, we investigate the degree to which employers and students have more implicit negative work performance stereotypes and more general negative attitudes toward Arab-Muslims and obese individuals compared with Swedes and people of normal weight.

All in all, we had to develop three IAT:s. The three IAT:s consisted of a performance stereotype IAT with the target labels “Muslim men” and “Swedish men”, a performance stereotype IAT with the target labels “Obese” and “Normal weight”, and an attitude IAT with the target labels “Muslim men” and “Swedish men”. It should be mentioned that the instructions given in conjunction with the Arab-Muslim IAT informed the participants to categorize names into two distinct labels: Arab-Muslim men in Sweden and native Swedish men. Furthermore, to simplify the information appearing on the screen only the labels “Muslim men” and “Swedish men” were used in the IAT procedure.

This makes us confident in that we have provided enough cues so that we measure the extent to which the participants' have implicit negative attitudes/stereotypes toward Arab-Muslim males relative to native Swedish males; see the appendix for a detailed description of how these IAT:s were constructed.

Since implicit performance stereotypes have not been studied elsewhere it was important to compare this to the well established and frequently used attitude IAT.¹⁰ This would make it possible to establish a baseline to which the implicit performance stereotype IAT could be compared. Additionally, it was important to investigate whether the two measures seem to measure different constructs. However, due to the longer test-time, it was hard to motivate employers to perform two consecutive IAT:s.¹¹ Instead we had students perform both the performance stereotype IAT and the attitude IAT in conjunction with ethnicity. Since the main reason for this was to validate the two measures, rather than investigating the relationship of implicit stereotypes and attitudes of a certain population, using a student sample should not be a problem. It should be safe to assume that the measures will measure the same constructs regardless of whether the participants are students or employers. We also had students perform the obesity performance stereotype IAT. The inclusion of student samples, also allowed for a comparison with the employers' results, which is important to increase generalizability.

Although the primary focus is on the implicit measures, brief explicit measures of attitudes and work performance stereotypes were also included. The inclusion of explicit measures gave us a baseline to compare our implicit measure with, allowed us to analyze whether our explicit and implicit measures were correlated or not, and provided an opportunity to investigate whether the attitude and the stereotype IAT seem to measure distinct constructs. Indication of this would exist if the implicit performance stereotype IAT correlates more strongly with its corresponding explicit performance stereotype measure than

¹⁰ Note that other types of implicit stereotypes have been studied elsewhere (see Nosek et al., 2007)

¹¹ Even a compensation of 500 SEK (approximately 55 euros) could not motivate a sufficient number of employers to participate. All in all, one third of the employers asked to participate did so. Unfortunately, this could mean that only a certain subsample of employers chose to participate in the experiments, possibly limiting the generalizability of the results. However, even if this is the case, it does not seem unreasonable to assume that the employers who chose *not* to participate were those who felt they had something to hide and who are the most prejudiced. Thus, the IAT-effects obtained in the present research may be a somewhat conservative estimate of the existence of implicit prejudice among employers.

with the explicit attitude measure; and if the implicit attitude IAT correlates more strongly with its corresponding explicit attitude items than with the explicit work performance stereotype items.

3.1 Overview of the experiments

In Experiment 1A, 1B and 1C, we study implicit and explicit attitudes and performance stereotypes toward Arab-Muslim men. In Experiment 1A we study both implicit attitudes and performance stereotypes in a sample consisting of students. Experiment 1B examines implicit and explicit performance stereotypes in a sample of employers, and in Experiment 1C we examine implicit and explicit attitudes in another sample of employers. In Experiment 2A we investigate implicit and explicit performance stereotypes toward obese individuals in a student sample consisting of business students. Finally, Experiment 2B replicates this procedure on a sample consisting of employers.

The explicit measures used in the experiments were developed based on the measures used in Nosek et al. (2005), and included the feeling thermometer, the hiring preferences task, and the performance stereotype rating task.

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. The thermometer was only used in conjunction with ethnicity.

The *hiring preference task* asked the participants to choose which groups 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. For the ethnicity studies, 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 five alternatives in the obesity studies only differed in the two groups (Obese/ Normal weight) they contrasted.

Finally, 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. The obesity study used the same alternatives but compared “Obese” and “Normal weight” people instead.¹²

Table 1. Characteristics of the experiments

Experiment	IAT	Explicit measures
1A (students)	Arab-Muslim attitude Arab-Muslim performance stereotype	Feeling thermometer Performance stereotype rating
1B (employers)	Arab-Muslim performance stereotype	Feeling thermometer Performance stereotype rating Hiring preference
1C (employers)	Arab-Muslim attitude	Feeling thermometer Hiring preference
2A (students)	Obesity performance stereotype	Performance stereotype rating
2B (employers)	Obesity performance stereotype	Performance stereotype rating Hiring preference

3.2 Experiment 1A

The primary purpose of this experiment was to see if students have more implicit low-performance stereotypes of Arab-Muslims relative to Swedes and if they also have implicit negative attitudes of Arab-Muslims relative to Swedes. We also wanted to find out to what extent these implicit IAT measures correspond to how participants explicitly perceive this target group. As this was the first experiment, we also wanted to validate the two IAT:s and confirm that they measure the distinct, but not unrelated, constructs of attitudes and stereotypes, respectively. To do this we examined to what extent the implicit stereotypes and implicit attitudes are correlated, as well as to what degree the implicit attitudes and stereotypes are consistent with their explicit counterparts.

3.2.1 Method

3.2.1.1 Participants

The participants consisted of 87 students, of which 48 were enrolled in the *Human Resource Management Program*¹³ at Växjö University. The remaining

¹² Because of space restrictions, we do not list all the alternatives here. The full set of explicit measures can be provided if requested.

participants were a mix of teacher and psychology students. Mean age was 24.29 ($SD = 5.66$) years. The experiment was conducted in the autumn of 2006.

3.2.1.2 *Materials*

The explicit measures consisted of the feeling thermometer and the performance stereotype rating task. The implicit measures consisted of two IAT:s. The first was the Arab-Muslim attitude IAT constructed to assess whether participants more easily associate Arab-Muslim men with negative attributes and Swedish men with positive attributes, than they associate Arab-Muslim men with positive attributes and Swedish men with negative attributes.¹⁴

The second IAT was the Arab-Muslim performance stereotype IAT constructed to assess whether participants more easily associate Arab-Muslim men with low performing attributes and Swedish men with high performing attributes, than they associate Arab-Muslim men with high performing attributes and Swedish men with low performing attributes.

The two IAT:s used the same target labels comprising “Muslim men” and “Swedish men”, with the associated stimuli that included Arab-Muslim sounding and Swedish sounding names such as “Said” and “Lars”. The attitude IAT also included the attribute labels “positive” and “negative” with the associated stimuli that consisted of words such as “joy” and “pain”. The stereotype IAT, on the other hand, used the attribute labels “high-performing” and “low-performing”, with the associated stimuli that consisted of words such as “efficient” and “lazy”. All category labels and stimuli were presented in Swedish.¹⁵ An overview of the category labels and the corresponding stimuli can be found in *Table 2*. A detailed description of the selection of category labels and the creation of the stimuli, including an evaluation by a student sample, can be found in *Appendix A*.

¹³ A majority of the students were thus enrolled in a programme that educates future recruiters.

¹⁴ Both the target labels and stimuli were limited to include Arab-Muslim men only. See General discussion (in section 4) for a discussion of this.

¹⁵ The category labels and stimuli presented here are direct translations from Swedish, and thus may sound a little awkward. Rest assured, they sounded much better in Swedish. The stimuli words in Swedish, as they appeared in the experiments, can be found in *Appendix A*

Table 2. Labels and stimuli used in the Arab-Muslim IAT:s

Category labels	Stimuli
Target labels (used in both IAT:s) Muslim men Swedish men	Ali, Ameer, Mohammed, Said, Reza, Hassan Erik, Lars, Karl, Johan, Anders, Per
Attribute labels in the attitude IAT positive negative	happiness, love, joy , sun, laughter, wonderful murder, pain , misery, death, evil, hurt
Attribute labels in the stereotype IAT high-performing low-performing	effective, productive, hardworking, ambitious, goal-oriented ineffective, incompetent, slow, lazy initiative-lacking

3.2.1.3 Apparatus

The experimental session, consisting of the two IAT:s and a questionnaire, was conducted in a computer laboratory, using an online web version of Inquisit (2006).¹⁶ Inquisit includes an automatic check for compatibility and adequate system performance, and it was thus ensured that all computers met the specific requirements. All trials were recorded locally on the participant's computer using DirectX technology to ensure reliable millisecond accuracy. The data were then sent through a Secured Sockets Layer connection to a secured web-site, which ensured maximum confidentiality.

3.2.1.4 Design

The two IAT:s followed the design suggested by Nosek et al. (2005). Each IAT contained seven blocks of classifying tasks. The first block consisted of 20 trials where the participant learned the target categories by classifying the target stimuli into the two target categories, constantly reminded of which belongs to which key by the category labels at the top of the screen. In this case the target categories were "Muslim men" and "Swedish men" and the stimuli were Arab-Muslim sounding and Swedish sounding male names.

¹⁶ See www.millisecond.com for more information about the software. Also, an independent study of the timing accuracy of Inquisit has been conducted by De Clercq et al. (2003).

The second block had the participant learn the attribute categories by classifying the attribute stimuli for another 20 trials. In the Arab-Muslim attitude IAT the attribute labels were “positive” and “negative” and the stimuli were all words that were very positive or very negative (e.g. happy, misery). In the Arab-Muslim performance stereotype IAT the attribute labels were “high performing” and “low performing” and the attribute stimuli consisted of words such as “efficient” and “lazy”.

The third block was a pairing practice block where the participant sorted both the target *and* the attribute stimuli for 24¹⁷ trials into the four different categories – still only using two keys. This means that one of the attribute categories shared the key with one of the target categories, and that the two remaining categories (one attribute and one target) shared the other key. In the fourth block the participant continued to sort both target and attribute stimuli into the four different categories, but now for 40 trials.

The fifth block involved relearning how to sort the target stimuli again, but now with a reversed key mapping. Those who had the category “Muslim men” mapped to “d” now had it mapped to “k”. Consequently, “Swedish Men” were mapped to “d”. The fifth block consisted of 40 trials to relearn this reversed key mapping.

The sixth block consisted of a pairing practice block where the participant sorted both target and attribute stimuli to the four categories for 24 trials. But since the key mapping of the target stimuli now was reversed, the participants who in block three and four had “Muslim men” + “positive”/”Swedish men” + “Negative” now instead had the combination “Muslim men” + “negative”/”Swedish men” + “Positive”. In the seventh and last block the participant continued to sort both target stimuli (with reversed key mapping) and attribute stimuli into the four categories, but this time for 40 trials

During the whole test, on any trial when the participants made an error, a large red X was presented on the screen, slightly below the stimulus, and the participant had to make the right key choice to proceed. The response times reflected the time from that the stimuli were presented until participants chose the correct key response. The participants were repeatedly reminded that they had to sort the stimuli fast, but without making too many errors.

¹⁷ 24 trials were used instead of the traditional 20 trials, purely due to limitations in the software’s code. However, there are no indications that these additional trials would in any way be harmful to the procedure.

3.2.1.5 Procedure

The participants were asked to participate in a study on attitudes toward immigrant employees in exchange for a reward of two lunch coupons (approximately 10 euros). Once seated in front of the computer screen in the laboratory, they received written instructions necessary to complete the IAT. They then performed the two IAT:s, followed by a brief questionnaire, where they completed the explicit measures and provided demographic data. That the explicit measures were always performed afterwards was not considered a problem since a large amount of studies typically have found little or no effect of the order of the IAT and the same types of explicit measure we employed (Lane et al., 2007).¹⁸

Since an extensive body of research has demonstrated that there are two common confounding variables that might be relevant for the procedure of this experiment, these had to be minimized (Nosek et al., 2005). First, having the participants perform two IAT:s after each other might underestimate the IAT score on the second IAT. Although it has been demonstrated that effects pertaining to the attitude IAT get weaker (but only marginally so) if one IAT has been completed before another, nothing is known of how large this effect would be on our newly created performance stereotype IAT. Consequently, the safest option was to administrate it first, followed by the attitude-IAT.

Second, the order of which the compatible and incompatible part is administrated may exert an effect on the IAT score. Typically, if the incompatible part is presented first the effect is slightly lower, than if the compatible part is presented first. This is most likely due to difficulties pertaining to the relearning of the reversed key mapping. The response times will thus be somewhat longer in the part presented last (Nosek et al., 2005). However, the impact of pairing order appears to have an influence exclusively on the magnitude of the IAT effects. It does not appear to affect the reliability, relations with self-report, or increase the vulnerability to extraneous influences. It is therefore not possible to say that one of the two estimates is a better estimate of the underlying construct than the other (Nosek et al., 2005). Since this order-effect is inherent to the design of the IAT, we had to decide whether we would administrate the incompatible part first (which will provide a lower estimate), or the compatible part first (which will provide a higher estimate). Because our purpose was to validate the IAT:s as a useful tool to be used to test for implicit

¹⁸ The explicit measures we employ are adaptations of measures that have been used in conjunction with thousands of IAT sessions of various types in the Project Implicit.

discrimination, we chose to use the lower estimate.¹⁹ The reasoning behind this was that if the IAT:s yield effects when using the lower estimate, effects would also very likely be obtained if we were to use the higher estimate. Consequently, all participants completed the exact same test and no counterbalancing was made. This was necessary to enable a satisfactory correlation of the two IAT:s.

3.2.1.6 Scoring the IAT

The data were analyzed using the new scoring algorithm suggested by Greenwald et al. (2003), who created the new scoring algorithm in an effort to maximize the implicit-explicit correlations and minimize extraneous influences on the IAT. All trials from the practice and test blocks were used. Trials with 10,000 ms and above were removed. Participants with 10 percent of the responses falling below 300ms are conventionally removed. Yet in this study no participant fell into this category.²⁰ After this treatment, a mean value of the response times was computed for all four blocks. One pooled standard deviation for the practice blocks (3 and 6) and one pooled standard deviation for the test blocks (4 and 7) were computed. Then a difference in latency among the practice blocks (3 – 6) and a difference in latency among the test blocks (4 – 7) were computed.

The difference in latency of the practice blocks was then divided by the pooled standard deviation of the practice blocks, and the difference in latency of the test blocks was divided by the pooled standard deviation of the test blocks. This produced two separate so called *D*-measures, one for the practice blocks and one for the test blocks. These two *D*-values were then averaged to produce a *D*-measure which uses data from both the practice and the test trials. This *D*-measure is the individual's IAT-score. It was calculated so that a positive value reflected a implicit negative attitude toward Arab-Muslim men relative to Swedish men (in the attitude IAT) and an implicit association that Arab-Muslim men are more low-performing relative to Swedish men (in the stereotype-IAT). A negative *D*-value suggests the opposite.

¹⁹ Another possibility would have been to counterbalance the order of administration of the IAT:s. However, we wanted all participants to follow the exact same procedure in order to reduce unwanted variance with respect to the IAT outcome variable. This enables more clean-cut correlations with the explicit measures.

²⁰ The purpose of this lower tail treatment is to remove participants who did not perform the test correctly, e.g., pressing the keys as fast as possible on a random basis. While this may be more likely to occur in a public Internet study, in a controlled experiment this should be quite rare.

The *D*-measure is a measure of an individuals' effect size on the IAT, and it resembles Cohen's *d*, but its strength should be interpreted as roughly twice that of Cohen's *d* (Greenwald et al, 2003), with a *D*-value of 0 suggesting no association at all. In reporting the results we will present our results in the same manner as in Nosek et al. (2007), showing the share of the sample that evidenced at least a slight *D*-score of .15. This is not to imply that there is something theoretically specific about this bound, rather it is a conventionally used and conservative bound where we can be confident that there is a substantial effect on the IAT-measure.

3.2.2 Results and discussion

3.2.2.1 Arab-Muslim stereotype IAT

The response times were higher in the incompatible condition ("Muslim men" + "High Performance"/"Swedish men" + "Low Performance") than in the compatible ("Muslim men" + "Low Performance"/"Swedish men" + "High Performance"), mean difference = 155 ms; $D = .40$, $SD = .37$, $t(86) = 8.54$, $p < .001$. This shows that participants more easily associate Arab-Muslim men with low-performing attributes and Swedish men with high-performing attributes, than they associate Arab-Muslim men with high-performing attributes and Swedish men with low-performing attributes. In fact 66 (76%) of the participant showed at least a slight positive effect. Seven individuals (8%) showed the reverse pattern (i.e., a negative IAT score).

3.2.2.2 Arab-Muslim attitude IAT

Also in this test, the response times were higher in the incompatible condition ("Muslim men" + "Positive"/"Swedish men" + "Negative") than in the compatible ("Muslim men" + "Negative"/"Swedish men" + "Positive"), mean difference = 177 ms; $D = 0.50$, $SD = 0.31$, $t(86) = 11.91$, $p < .001$. This demonstrates that participants more easily associate Arab-Muslim men with negative attributes and Swedish men with positive attributes, than they associate Arab-Muslim men with positive attributes and Swedish men with negative attributes. In fact, a large majority of 72 (83%) showed at least a slight IAT-effect, while only two of the participants evidenced the reverse pattern.

3.2.2.3 The relationship between the stereotype and the attitude IAT

Participants' performance of the Arab-Muslim attitude IAT and the Arab-Muslim performance stereotype IAT was highly related, $r = .52$, $p < .001$.

The distribution of the participants' *D*-score pertaining to the two IAT:s is illustrated in the scatter plot below. As can be seen in *Figure 1*, a large majority of 57 (66%) of the participants show at least a slight effect on both IAT:s, which reveals that they in general have stronger implicit negative attitudes and implicit low-performance stereotypes toward Arab-Muslim men relative to Swedish men.

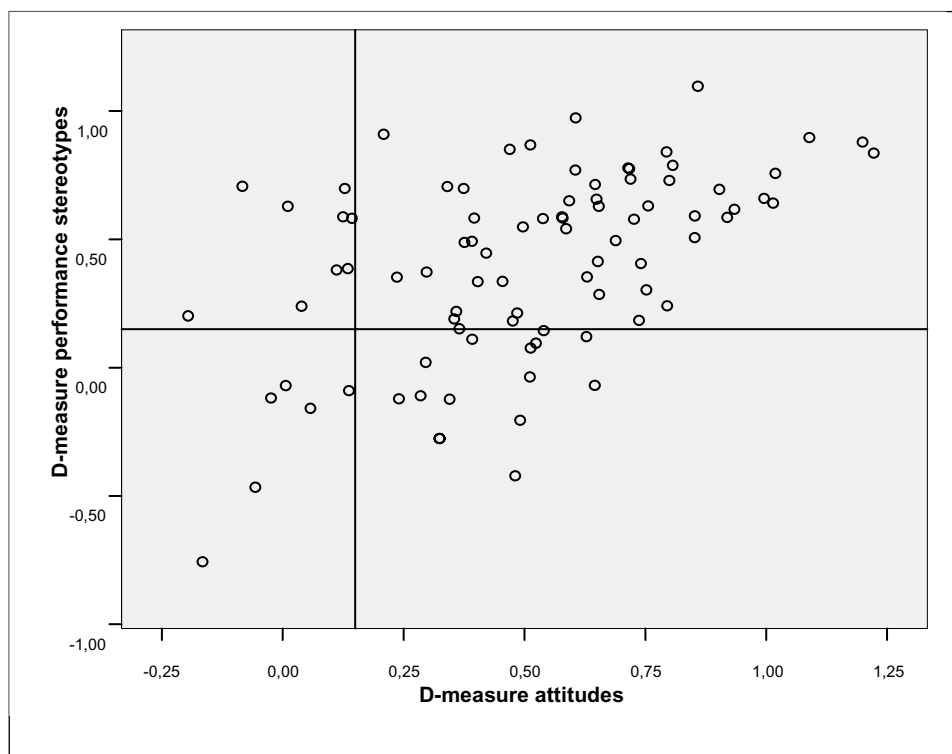


Figure 1. Scatter plot showing the positive relationship between students' scores (D-measures) on the two IAT:s (attitude IAT and work performance stereotype IAT)

3.2.2.4 *Explicit measures*

In order to test for relative differences in explicit feelings toward Arab-Muslim men and Swedish men, a paired sample t-test was carried out. The participants had less explicit positive feelings toward Arab-Muslim minority men ($M = 5.18$, $SD = 1.53$) than toward native Swedish men ($M = 7.53$, $SD = 1.55$), $t(86) = 11.24$, $p < .001$. A total of 68 individuals (78%) explicitly expressed more negative feelings toward Arab-Muslim men than Swedish men.

In response to the performance stereotype rating task, the participants stated that they consider Swedish men to be more productive at work than Arab-Muslim men ($M = .22$, $SD = .60$), $t(86) = 3.52$, $p < .001$. A total of 20 individuals (23%) indicated that Swedish men perform at least slightly better at work than Arab-Muslim men.

3.2.2.5 *The relationship between implicit and explicit attitudes*

The feeling thermometer showed a moderate correlation with the Arab-Muslim attitude IAT score, $r(87) = .40$, $p < .001$; see *Table 3*. In other words, participants with more implicit negative attitudes toward Arab-Muslim men relative to Swedish men also explicitly stated that they had more negative feelings toward Arab-Muslim men than Swedish men. The Arab-Muslim attitude IAT score did not significantly correlate with the explicit measure pertaining to the performance stereotype ratings ($p > .10$).

Table 3. Correlation matrix (Pearson's r) showing the relationship between the Arab-Muslim attitude IAT score, Arab-Muslim stereotype IAT score, and the explicit measures. Note: * $p < .10$ ** $p < .05$, *** $p < .01$.

Variables	1	2	3	4
(1) Attitude IAT	1	.52***	.40***	.17
(2) Stereotype IAT	.52***	1	.17	.21*
(3) Feeling thermometer	.40***	.17	1	.19
(4) Performance stereotype rating	.17	.21*	.19*	1

3.2.2.6 *The relationship of implicit and explicit performance stereotypes*

The performance stereotype rating task showed a significant correlation with the Arab-Muslim stereotype IAT score, $r(87) = .21$, $p = .06$. Participants who implicitly associated Arab-Muslim men with lower performance to some extent also explicitly declared that Arab-Muslim minority men are more low-performing than native Swedish men. The Arab-Muslim stereotype IAT score did not significantly correlate with the feeling thermometer ($p > .10$).

3.2.2.7 *Summary of the experiment*

The results from this experiment demonstrate that the present sample has more negative attitudes toward Arab-Muslim men than toward Swedish men, and possesses the stereotype that Arab-Muslim men are less performing than Swedish men. This was found to be true both at an implicit and an explicit

level, but a larger share of the participants have implicit negative attitudes and performance stereotypes. Even though we find an implicit-explicit correlation, for attitudes as well as stereotypes, it is clear that a large number of students report that they do not have negative feelings or low-performance stereotypes toward Arab-Muslims when their IAT results indicate otherwise.

As for the validation of the IAT:s, implicit attitudes were related to explicit attitudes, and implicit performance stereotypes were related to explicit performance stereotypes. At the same time implicit attitudes were not significantly related to explicit performance stereotypes, nor were implicit stereotypes significantly related to the explicit attitudes. These results suggest that the two IAT:s measure different constructs. However, that these correlations were not significant cannot fully prove that the two IAT:s measure different constructs. Higher power due to a larger sample might have yielded small but significant correlations. Nevertheless, the difference in strengths of the correlations with the corresponding explicit measures for the two different IAT:s indicates that they measure different constructs. Furthermore, the correlations of the two IAT:s suggest that the constructs measured are to a substantial extent related, which is to be expected since feelings toward a target group could influence how we evaluate its performance and the other way around. To conclude, the results of this experiment suggest that both the Arab-Muslim attitude IAT and the Arab-Muslim stereotype IAT can be useful tools for measuring implicit stereotypes and attitudes toward this particular group.

3.3 Experiment 1B

The main purpose of this experiment was to extend the results from Experiment 1A to a sample of employers in order to assess whether this target group implicitly holds more low-performance stereotypes of Arab-Muslim men compared to Swedish men. As in Experiment 1A, we wanted to examine to what extent the implicit stereotype IAT score corresponds to the explicit self-reports.

3.3.1 Method

3.3.1.1 Participants

The sample consisted of 193 employers (99 males, 94 females) responsible for the recruitment of staff at various companies situated in Stockholm and Gothenburg, Sweden. Mean age was 44.92 ($SD = 11.01$) years. The employers

were randomly selected from job advertisements pertaining to twelve different occupations²¹ which appeared in the database provided by the Swedish Labor Market Board (AMS) during a time period stretching from May, 2005 to February, 2006. The experiment was conducted in the autumn of 2006.²² Only employers who worked for a company with more than 10 employees were contacted.

3.3.1.2 *Materials, apparatus, design and the scoring of the IAT*

In the present experiment the participants only completed the Arab-Muslim performance stereotype IAT. The design and the scoring of the IAT were the same as in Experiment 1A. As for the explicit measures, these consisted of the feeling thermometer, the hiring preference rating task, and the performance stereotype rating task. The only difference in apparatus was that the participants used their own computers at home, or at work, since they were not taken to a laboratory. Since the software checks for compatibility before running and uses the highly standardized DirectX input, running the test on own computers should be very similar to the computers in a laboratory. Internet connection quality, screen size, resolution and refresh rate etc. should not affect the results. Furthermore the design of the software is such that background running programs cannot distract the participant.

3.3.1.3 *Procedure*

The procedure was identical to that of Experiment 1A except for the following. Since a compensation of 500 SEK (approximately 55 euros) proved not to be enough to motivate the employers to perform two consecutive IAT:s, which the students had done in Experiment 1A, we were forced to limit their participation to a shorter procedure. We thus let the employers perform an exact replica of the first half of the test used in Experiment 1A. As a result, we effectively reduced test time by 50%, while still allowing for a direct comparison with the results from Experiment 1A. Consequently, the employers performed a performance stereotype IAT with the incompatible part first, followed by a brief questionnaire containing the explicit measurements.

²¹ See Carlsson & Rooth (2007) for details.

²² The experiments with employers as participants, e.g., 1B, 1C and 2B, were more time consuming since we had to locate each employer one by one. On average, each experiment took six months to complete. In contrast, the experiments with students as participants were completed in only one day.

We contacted the employers by phone and invited them to participate in a study on inclusion and exclusion processes related to the recruitment of staff. We then also explicitly stated that we had observed the hiring process of the advertised vacancy, and therefore we only invited the person responsible for the hiring to participate. After having properly performed the test, the participants sent us a confirmation mail with informed consent.

3.3.2 Results and discussion

For the employers the response times were higher in the incompatible condition (“Muslim men” + “High Performance”/”Swedish men” + “Low Performance”) than in the compatible (“Muslim men” + “Low Performance”/”Swedish men” + “High Performance”), mean difference = 193 ms; $D = .38$, $SD = .34$, $t(192) = 11.16$, $p < .001$. This shows that the employers more easily associate Arab-Muslim men with low-performing attributes and Swedish men with high-performing attributes than they associate Arab-Muslim men with high-performing attributes and Swedish men with low performing attributes. A total of 150 (78%) of the employers exhibited at least a slight positive D-value, indicating the existence of a stronger low-performance stereotype toward Arab-Muslim men relative to Swedish men. A total of thirteen participants (7%) showed the reverse pattern.

As for the explicit measures, a total of 95 (49%) of the employers explicitly stated in conjunction with the feeling thermometer that they have less positive feelings toward Arab-Muslim men ($M = 5.43$, $SD = 1.69$) than toward Swedish men ($M = 6.61$, $SD = 1.63$), $t(192) = 8.64$, $p < .001$. In the hiring preference rating task, a total of 103 employers (53%) explicitly stated that they slightly or strongly prefer Swedish men to Arab-Muslim men when hiring staff ($M = .63$, $SD = .69$), $t(192) = 12.76$, $p < .001$. In response to the performance stereotype rating task, 23 of the participants (12 %) explicitly stated that Swedish men are slightly or much more high-performing at work than Arab-Muslim men ($M = .12$, $SD = .50$), $t(192) = 3.49$, $p < .001$.

The Arab-Muslim stereotype IAT showed no significant correlations with the feeling thermometer, which is consistent with the results found in Experiment 1A. However, unlike in Experiment 1A we found no significant correlation between the stereotype IAT and the performance stereotype rating task.²³

²³ This could possibly be related to issues with the explicit measures. See the General Discussion for details on this matter (section 4).

The explicit hiring preference ratings did, however, show a significant correlation with the Arab-Muslim stereotype IAT: $r = .23$ ($p < .001$).

In summary, it is clear that the employers have implicit low-performance stereotypes of Arab-Muslim men as compared to Swedish men, and that the effect is almost identical to that of the student sample in Experiment 1A. Another interesting finding was that the performance stereotype IAT showed a correlation with whom the employers said they preferred when recruiting staff. Although not conclusive, since it cannot be assumed that the employers' reported hiring preference is an accurate reflection of their actual real-life decisions, this suggests that the implicit performance stereotypes used in the present IAT could be related to actual discriminative behavior in the context of staff recruitment.

3.4 Experiment 1C

In Experiment 1B we investigated implicit stereotypes toward Arab-Muslims, in a sample of employers. Since it was not possible for this sample to also complete an Arab-Muslim attitude IAT in a single session, we performed an additional experiment on a new group of employers in order to test for implicit negative attitudes.

3.4.1 Method

3.4.1.1 Participants

The sample consisted of 158 employers (80 males, 78 females) responsible for the recruitment of staff at various companies situated in Stockholm and Gothenburg, Sweden. Mean age was 43.41 ($SD = 10.95$) years. The employers were randomly selected from job advertisements pertaining to twelve different occupations²⁴ that appeared in the database provided by the Swedish Labor Market Board (AMS) during a time period stretching from August, 2006 to February, 2007. The experiment was conducted in the spring of 2007. Only employers who worked for a company with more than 10 employees were contacted.

²⁴ See Carlsson & Rooth (2007) for details.

3.4.1.2 Materials, apparatus, design, procedure and scoring of the IAT

Except for replacing the stereotype IAT with the attitude IAT, this experiment only differed in that the explicit stereotype rating was omitted²⁵.

3.4.2 Results and discussion

The response times were higher in the incompatible condition (“Muslim men” + “Positive”/“Swedish men” + “Negative”) than in the compatible (“Muslim men” + “Negative”/“Swedish men” + “Positive”), mean difference = 350 ms; $D = .64$, $SD = .31$, $t(157) = 18.20$, $p < .001$. This demonstrates that the employers more easily associate Arab-Muslim men with negative attributes and Swedish men with positive attributes, than they associate Arab-Muslim men with positive attributes and Swedish men with negative attributes. 148 (94%) of the employers showed at least a slight positive IAT-effect, while two employers showed the reverse pattern.

As for the explicit measures, a total of 83 (53%) of the employers explicitly stated in conjunction with the feeling thermometer that they have less positive feelings toward Arab-Muslim men ($M = 5.28$, $SD = 1.72$) than toward Swedish men ($M = 6.73$, $SD = 1.78$), $t(157) = 7.99$, $p < .001$. In the hiring preference rating task, a total of 52 employers (33%) explicitly stated that they slightly or strongly prefer native Swedish men to Arab-Muslim men when hiring staff ($M = .37$, $SD = .66$), $t(157) = 6.98$, $p < .001$. Furthermore, the Arab-Muslim attitude IAT showed a significant correlation with the feeling thermometer $r = .16$, $p = .05$. No significant correlation between the Arab-Muslim attitude IAT and the hiring preference rating could be found ($p > .10$).

The results show that this sample of employers has more implicit negative attitudes toward Arab-Muslim men relative to Swedish men. As in Experiment 1B, this sample also has more explicit negative attitudes toward this group. Furthermore, it seems as if this sample of employers has a somewhat stronger attitude IAT-effect than the students in Experiment 1A ($D = .64$ versus $D = .50$). However, it should be mentioned that this could be a result of the slightly reduced IAT-effect the students should have, since they had performed the performance stereotype IAT immediately before taking the attitude IAT. Moreover, the correlations between implicit and explicit attitudes were weaker in this experiment compared to those obtained in Experiment 1A. This weaker

²⁵ Motivating employers to take the test was very difficult, and we therefore tried to keep the time it took to complete the test session minimal. Omitting everything that was not of absolute priority was thus important to reduce test-time.

correlation is not surprising since the motivation to appear unprejudiced is likely to be stronger in this sample and self-presentational bias could obscure the actual relationships of the underlying constructs.

3.5 Experiment 2A

The primary purpose of this experiment was to investigate if business students implicitly hold the stereotype that obese individuals are more low-performing than normal weight individuals. In addition, we wanted to analyze to what degree their implicit IAT scores correspond to their explicit self-reports.

3.5.1 Method

3.5.1.1 Participants

The participants consisted of 88 (30 males, 58 females) business students from Kalmar university. The experiment was conducted in the autumn of 2006. Mean age was 25.22 ($SD = 5.73$) years.

3.5.1.2 Materials

The obesity stereotype IAT was identical to the Arab-Muslim performance stereotype IAT used in Experiment 1A and 1B except for the target labels, which now were the Swedish equivalents of “Obese” and “Normal weight”, and the target stimuli. The target stimuli consisted of pictures of average looking women and men, half of which had been professionally manipulated to look obese. More specifically, two pictures of the same individual were created that differed only with respect to whether they were obese or of normal weight. Thus each picture appeared in both an obese version and a normal weight version. A total of 12 stimuli pictures were used. They were based on the original photographs of three males and three females. An overview of the category labels and corresponding stimuli used can be seen in *Table 4*; see also Appendix A for a detailed description of how the target stimuli, i.e. the photos, were constructed. The explicit measure consisted only of the performance stereotype rating task.

Table 4. The category labels and stimuli used in the obesity performance stereotype IAT

Category labels	Stimuli
Target labels Obese Normal weight	Photos of obese people Photos of people of normal weight
Attribute labels High-performing Low-performing	effective, productive, hardworking, ambitious, goal-oriented ineffective, incompetent, slow, lazy initiative-lacking

3.5.1.3 Apparatus, design, procedure and the scoring of the IAT

The apparatus, design, procedure and the scoring of the IAT were the same as in Experiment 1A.

3.5.2 Results and discussion

In the obesity stereotype IAT, the response times were higher in the incompatible condition (“Obese” + “High Performance”/“Normal weight” + “Low Performance”) than in the compatible condition (“Obese” + “Low Performance”/“Normal weight” + “High Performance”), mean difference = 291 ms; $D = .61$, $SD = .35$, $t(87) = 11.80$, $p < .001$. This shows that participants more easily associate obese individuals with low performing attributes and normal weight individuals with high performing attributes than they associate obese individuals with high performing attributes and normal weight individuals with low performing attributes. Moreover, 76 (86%) of the participants showed at least a slight positive effect, while only one of the participants showed the reverse pattern.

In response to the explicit performance stereotype rating task, the participants stated that they consider people of normal weight to be slightly or much more high-performing at work than obese people ($M = .43$, $SD = .58$), $t(87) = 6.95$. 34 individuals (39%) indicated that they believe that obese people perform at least slightly worse at work. Surprisingly, no statistically significant correlation was found between this measure and the IAT score, $r(87) = -.012$, $p > .10$, suggesting that the participants’ implicit performance stereotypes of obese people did not correspond to their explicit performance stereotype ratings concerning these individuals.

Taken together, the results from this experiment show a large performance stereotype IAT effect. In this experiment, however, the participants' stereotype IAT score did not correspond to their explicit performance stereotype ratings.²⁶

3.6 Experiment 2B

Experiment 2A demonstrated the existence of more implicit low-performance stereotypes toward obese individuals relative to normal weight individuals in a student sample. The primary purpose of this experiment is to determine whether these findings can be generalized to a sample of employers.

3.6.1 Method

3.6.1.1 Participants

Participants were 166 employers (84 males, 82 females). Mean age was 43.91 ($SD = 10.45$) years. The employers were randomly selected from job advertisements pertaining to seven different occupations²⁷ which appeared in the database provided by the Swedish Labor Market Board (AMS) during a time period stretching from January to August, 2006, in Stockholm and Gothenburg. The experiment was conducted in the autumn of 2006 (see also footnote 21). Only employers who worked for a company with more than 10 staff were contacted.

3.6.1.2 Materials, apparatus, design, procedure and the scoring of the IAT

The participants completed the obesity performance stereotype IAT. The explicit measures consisted of the hiring preference task and the performance stereotype rating task. The apparatus, design, and the scoring of the IAT were the same as in Experiment 2A. As in Experiment 1B, the employers were contacted by phone and offered 100-500SEK for their participation.

3.6.2 Results and discussion

In the obesity performance stereotype IAT, the response times were higher in the incompatible condition (“Obese” + “High Performance”/“Normal weight” + “Low Performance”) than in the compatible condition (“Obese” + “Low Performance”/“Normal weight” + “High Performance”), mean difference

²⁶ This may be due to methodological issues with the explicit measure. See General discussion for details (section 4).

²⁷ See Rooth (2007a) for details.

= 558 ms; $D = .72$, $SD = .38$, $t(165) = 16.77$, $p < .001$. This indicates that these employers more easily associate obese individuals with low performing attributes and normal weight individuals with high performing attributes, than they associate obese individuals with high performing attributes and normal weight individuals with low performing attributes. Moreover, 150 (90%) of the participants showed at least a slight positive effect, while only three of the participants showed the reverse pattern..

In response to the explicit performance stereotype rating task (coded as -2 to +2), 41 (25%) participants stated that they consider normal weight people to be slightly or much more productive at work compared with obese people ($M = .28$, $SD = .53$), $t(165) = 6.93$, $p < .001$. Furthermore, the hiring preference rating task (coded as -2 to +2) showed that 97 (58%) of the participants slightly or strongly preferred normal weight people to obese people when recruiting staff ($M = .67$, $SD = .63$), $t(165) = 13.75$, $p < .001$. No statistically significant correlation was found between the implicit and explicit measures²⁸.

The results from this experiment closely resemble those from the student sample in Experiment 2A. It is clear that a great majority of the employers in this sample possesses stronger implicit low-performance stereotypes of obese people than of normal weight people, whereas relatively few (25%) explicitly state those stereotypes. However, a large number of the participants (58%) did indeed explicitly state that they had a *preference* for normal weight individuals when hiring staff.

In summary then, it is clear that a large majority of the employers have more implicit work related low-performance stereotypes of obese people relative to normal weight people, and that a majority explicitly states that they prefer normal weight people when hiring new staff..

4 General discussion

Although increasing attention has been directed at the unequal treatment of immigrants and other demographic groups in our society in the media as well as in research, little is known about what causes this discrimination. This study examines one such potential cause: implicit attitudes and performance stereo-

²⁸ This may be due to methodological issues with the explicit measure. See the General Discussion (section 4) for details.

types. While considerable research has already been conducted on implicit attitudes toward stigmatized social groups, this study is, to our knowledge, the first that investigates implicit attitudes toward Arab-Muslims and obese people in a sample of employers. It is also the first study to investigate whether employers (or students) implicitly associate Arab-Muslims and obese people with less work performance.

The main purpose of this study was to see if employers possess implicit prejudice to such a substantial degree that this could possibly explain part of the ongoing unequal treatment of Arab-Muslims and obese people in the labor market. The results (see *Table 5* for an overview) clearly show that implicit prejudice toward these groups exists as manifested in the strong IAT-effects and the large number of participants (employers and students) who associate the two groups more strongly with negative attributes pertaining to work performance (stereotypes) as well as with concepts of negative valence that are more global in nature (attitudes).

Taking these findings one step further, there is reason to believe that these automatic associations could play a significant role in applied contexts as well. Importantly, the employers' performance on the stereotype IAT was related to the explicit measures that intended to tap their overt preferences for Swedes over Arab-Muslims when hiring staff (Experiment 1B). Although we do not know whether the IAT is better at predicting actual discrimination than traditional explicit measures, it seems likely that part of the prejudice is disguised when the employers answer the explicit questions regarding this relatively sensitive matter (Greenwald & Banaji, 1995; Nosek, 2005). Moreover, since it could be assumed that the real hiring process may be more automatic in nature (as compared to when the employers report their preferences in a questionnaire) when gut-feelings may exert a prominent role, and stress may enter the equation etc, it may be that the IAT is a better predictor of real world discrimination than explicit questionnaires. The central idea behind this is that a test that measures automatic associations should be a better predictor of behavior that takes place in situations where deliberate and more time consuming cognitive processes are mitigated. Ongoing research (Rooth, 2007b) suggests that an implicit low performance stereotype acts as an important determinant of employers' decisions in the context of staff recruitment.

Since we studied participants' implicit and explicit cognitions toward two social target groups, it would be of interest to compare the results obtained in relation to each group. When doing this, we find that the effects on the

performance stereotype IAT is substantially higher in the obesity experiments than in the ethnicity experiments. This is not so surprising given previous research which shows that people tend to perceive obese individuals as lazy and slow. While we have no data on people’s attitudes toward obese individuals, previous research (e.g. Schwartz et al., 2006 found an average IAT-effect of $D = .48$) suggests that the effects could be similar to those that we found in the ethnicity experiments.

Table 5. A summary of the results of the experiments. All results are relative to the contrasting groups Swedish men and normal weight individuals.

	Arab-Muslims		Obese individuals	
	Students	Employers*	Students	Employers
Implicit measures:	.50	.64	-	-
Attitude IAT (D-measure)				
Share with at least a <i>slight</i> attitude IAT effect	.83	.94	-	-
Stereotype IAT (D-measure)	.40	.38	.61	.72
Share with at least a <i>slight</i> stereotype IAT effect	.76	.78	.86	.90
Explicit measures:				
<i>Feeling thermometer:</i> share with negative feelings	.78	.49	-	-
<i>Hiring preference:</i> share with preference of other group	-	.53	-	.58
<i>Performance stereotype rating:</i> Share with low-performance stereotypes	.23	.12	.39	.25

Note: The employers consist of two different samples. One took the Stereotype IAT and one took the attitude IAT. In this table we only present the results of the explicit measures from Experiment 1B, since that sample completed all of the explicit measures. See Experiment 1B and 1C for more details.

It is also interesting to note that compared to the employers, the students demonstrated very similar results on the implicit measures. This finding suggests that the implicit work performance stereotypes are as prevalent among students as among employers, which adds some generalizability to our results. When looking at the participants’ explicit statements, we find that both employers and students have more explicit negative attitudes toward Arab-Muslims relative to Swedish people, but that employers express this to a substantially less degree. Yet, half of the employers still report more negative

feelings toward Arab-Muslim men as compared to Swedish men. Additionally, a majority of the employers reports that they slightly or strongly prefer Swedish men and normal weight people over Arab-Muslim men and overweight people, respectively, when recruiting staff. On the other hand, the results of the explicit performance stereotypes are quite different, with few of the students and even fewer of the employers displaying explicit low-performance stereotypes toward Arab-Muslim men. Interestingly, the share that shows explicit low-performance stereotypes toward obese people relative to normal weight people, is substantially higher, but still only constitutes a minority.

As pointed out in the introduction, the most obvious limitation with explicit measures is that of self-presentational strategies. In addition, we may not be aware of how we perceive various social targets (Greenwald & Banaji, 1995). In relation to this, it is striking that 90% and 78% of the employers in our study show (relative to the contrasted group) implicit low-performance stereotypes toward obese people and Arab-Muslims, respectively, whereas only 25% and 12%, respectively, explicitly state that these two target groups are more low-performing. This disparity is less extreme with respect to explicit and implicit *attitudes* toward Arab-Muslims (see *Table 5*), but it is nonetheless clear that both the employers and the students display more implicit negative attitudes and stereotypes than they explicitly report.

While the results of the implicit measures were similar for the student and employer sample, the results differed substantially for the explicit measures. This could be because the employers had a stronger motivation to appear unprejudiced. Even though anonymity was guaranteed, the employers were invited to participate in the study because of their profession, and it is reasonable to assume they were more motivated to appear socially desirable.

However, some of the differences between the student and the employer sample could possibly be explained by other problems with explicit measures, apart from self-presentational issues. Spontaneous comments from the employers suggest that some of them found the *obesity performance stereotype rating* irrelevant since their ratings would depend upon job requirements. In addition, they tended to be unsure whether they should have interpreted the question in a more general manner, so that it refers to how they perceive this to be in Sweden as a whole, or in a more specific manner, so that it refers to how things are at their own work place. For the students, however, it should have been natural to answer “in general” since many of them do not yet have a permanent work place. However, we do not consider these issues as a large

threat to our study since it is the implicit measures that are of primary interest and these types of problems inevitably come with administering such brief explicit measures of this kind.²⁹

Although the explicit measures were primarily included as a baseline to compare the implicit measures with, they were also included as a way to further strengthen the validity of the implicit measures. The logic behind this is that even though implicit and explicit measures should not provide identical results, the underlying constructs should be related to some degree, and thus some level of correlation between the measures could be expected.³⁰

For both the students and the employers, the implicit and explicit attitudes were found to correlate significantly, further strengthening the validity of the Arab-Muslim attitude IAT. On the other hand, the relationship between the performance stereotype IAT and the corresponding explicit measures were somewhat mixed. Yet, in the student sample the implicit and explicit performance stereotypes of Arab-Muslims were found to be significantly correlated. Moreover, in the employer sample, the work performance stereotype IAT score was found to significantly correlate with *explicit hiring preference*, suggesting a possible relationship between implicit performance stereotype and decisions in hiring. However, no significant correlation was found between the work performance stereotype IAT score and the explicit *performance stereotypes* for the same employers in the Arab-Muslim experiment. In addition, in the obesity experiments, no significant correlations between the implicit and explicit measures of *performance stereotypes* were found in either the student or the employer sample.

The lack of significant correlations between implicit and explicit *performance stereotypes* in two of the experiments could possibly be due to the problems with the explicit measures mentioned above. Since implicit and explicit stereotypes are typically only weakly correlated, as opposed to attitudes that are often moderately correlated, (see Nosek et al., 2007), even small problems with the explicit measures could obscure an actual relationship. Because of this, we do not see the mixed results (regarding implicit-explicit

²⁹ It is possible that a more elaborate and time consuming measure of explicit performance stereotypes (e.g. a battery of questions) would have yielded different results, but that would not have been justified given the aims of this study.

³⁰ This should be regardless of whether the underlying constructs are the same, but the measures tap them before or after self-presentational processes are activated, or if implicit and explicit attitudes and stereotypes are distinct, but related, constructs. Note that this study does not seek to address the theoretical issue of how the constructs are related.

correspondence) as signs of problems with the performance stereotype IAT, but rather as inherent to the explicit measure pertaining to work performance stereotypes. Nevertheless, the correlations found in Experiment 1A strengthen the validity of both the attitude IAT and the performance stereotype IAT in conjunction with the study of ethnicity.

The last methodological issue that deserves attention has to do with the fact that we only included male names in the ethnicity IAT. The reason for this is that Arab-Muslim men is the category expected to be hurt the most by labor market discrimination in the Swedish labor market. Moreover, since previous research has studied discrimination specifically against Arab-Muslim men (Carlsson & Rooth, 2007), including Arab-Muslim men in the present study seems natural if one wants to be able to make a direct comparison between the present findings and those obtained in previous studies. Accordingly, we can only conclude that employers have substantial negative attitudes and stereotypes toward Arab-Muslim men and not Arab-Muslims in general.

This study has found evidence of strong implicit negative attitudes and implicit low-performance stereotypes of two social groups that are known to be in an unfavorable position in the labor market. As suggested earlier, the natural next step is to examine the relationship between implicit attitudes/stereotypes, and actual discriminative behavior. However, given that a relationship exists, it is still possible to believe that there are many employers who do not discriminate, despite that they possess both implicit and explicit prejudice toward a target group. In fact, Carlsson and Rooth (2007) and Rooth (2007a) find that only a minority actually discriminates against Arab-Muslims and obese people. This seems counterintuitive given the great majority of the employers who possess implicit negative attitude and/or a low-performance stereotype toward these groups. This is an encouraging finding, but it raises some new questions on the study of discrimination in the labor market. While it is of great interest to examine if people with stronger implicit negative attitudes or low-performance stereotypes toward certain groups discriminate more, it is also of great importance to understand why some individuals do not discriminate, even though they implicitly associate certain groups of people with negative or low-performing attributes, as well as explicitly state that they prefer one group over the other.

In conclusion, these findings could have important implications for the unequal treatment of Arab-Muslims and obese individuals (compared with Swedes and people of normal weight) in the current labor market in Sweden

and elsewhere. As noted earlier, ongoing research suggests that the performance stereotype IAT is capable of predicting actual discriminative behavior against Arab-Muslims that takes place in the hiring process (Rooth, 2007b). Furthermore, as our results suggest, people may not be aware of these automatic tendencies since they are far from always reflected in the explicit measures. Accordingly, the IAT could serve as an important tool for making prejudiced individuals in general, and employers working with recruiting staff, become aware of this cognitive bias. In the long run, we believe this to be an important step in creating a fair job market.

References

- Akrami N (2005) "Prejudice: The interplay of Personality, Cognition and Social Psychology", Doctoral dissertation, Uppsala University.
- Akrami N & B Ekehammar (2005) "The association between implicit and explicit prejudice: The moderating role of motivation to control prejudiced reactions", *Scandinavian Journal of Psychology*, vol 46 (4), pp. 361-366.
- Akrami N, B Ekehammar & T Araya (2000) "Classical and modern racial prejudice: A study of attitudes toward immigrants in Sweden", *European Journal of Social Psychology*, vol 30 (4), pp. 521–532.
- Bellezza F, A Greenwald & M Banaji (1986) "Words high and low in pleasantness as rated by male and female college students", *Behavior Research Methods, Instruments, & Computers*, vol 18 (3), pp. 299-303.
- Bertrand M, D Chugh & S Mullainathan (2005) "New Approaches to Discrimination: Implicit Discrimination", *American Economic Review*, vol 95 (2), pp. 94-98.
- Bluemke M & M Friese (2006) "Do features of stimuli influence IAT effects?", *Journal of Experimental Social Psychology*, vol 42 (2), pp. 163-176.
- Carlsson M & D Rooth (2007) "Evidence of Ethnic Discrimination in the Swedish Labor Market Using Experimental Data", *Labour Economics*, vol 14, pp.716-729.
- Cunningham W, M Johnson, C Raye, J Gatenby, J Gore & M Banaji (2004) "Separable neural components in the processing of Black and White faces", *Psychological Science*, vol 15 (12), pp. 806-813.
- Dagens Nyheter* (2003), "Företag nobbar överviktiga", published on 27th of February 2003.
- De Clercq A, G Crombez, A Buysse & H Roeyers (2003) "A simple and sensitive method to measure timing accuracy", *Behavior Research Methods, Instruments and Computers*, vol 35 (1), pp. 109-115.
- De Houwer J (2001) "A structural and process analysis of the Implicit Association Test", *Journal of Experimental Social Psychology*, vol 37 (6), pp. 443–451.

- Devine, P (1989) "Stereotypes and prejudice: Their automatic and controlled components", *Journal of Personality and Social Psychology*, vol 56 (1), pp. 5–18.
- Ekehammar B, N Akrami & T Araya (2003) "Gender differences in implicit prejudice", *Personality and Individual Differences*, vol 34 (8), pp. 1509-1523.
- Fazio R & M Olson (2003) "Implicit measures in social cognition research: Their meaning and use", *Annual Review of Psychology*, vol 54, pp. 297-327.
- Govan C & K Williams (2004) "Changing the affective valence of the stimulus items influences the IAT by re-defining the category labels", *Journal of Experimental Social Psychology*, vol 40 (3), pp. 357-365.
- Greenwald A & M Banaji (1995) "Implicit social cognition: Attitudes, self-esteem, and stereotypes", *Psychological Review*, vol 102 (1), pp. 4–27.
- Greenwald A, D McGhee & J Schwartz (1998) "Measuring individual differences in implicit cognition: The implicit association test", *Journal of Personality and Social Psychology*, vol 74 (6), pp. 1464-1480.
- Greenwald A, B Nosek & M Banaji (2003) "Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm", *Journal of Personality and Social Psychology*, vol 85 (2), pp. 197-216
- Hofmann W, B Gawronski, T Gschwendner, H Le & M Schmitt (2005) "A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures", *Personality and Social Psychology Bulletin*, 31, pp. 1369-1385.
- Inquisit 2.0.60616 [Computer software] (2006), Seattle, WA: Millisecond Software LLC.
- Kim, D (2003) "Voluntary controllability of the Implicit Association Test (IAT)", *Social Psychology Quarterly*, vol. 66 (1), pp. 83-96.
- Lane K, M Banaji, B Nosek & A Greenwald (2007) "Understanding and using the implicit association test: IV. What we know (So Far) about the Method", in Wittenbrink, B. & N Schwarz (Eds), *Implicit Measures of Attitudes: Procedures and controversies*, pp. 59-102, Guilford Press, New York
- Lange A (2000) "Diskriminering, integration och etniska relationer", Integrationsverket, Norrköping.

- LeDoux J (1994) "Emotion, memory, and the brain", *Scientific American*, vol. 270 (6), pp. 32-39.
- Mitchell J, B Nosek & M Banaji (2003), "Contextual Variations in Implicit Evaluation", *Journal of Experimental Psychology*, vol. 132 (3), pp. 455-469.
- Nosek B (2005) "Moderators of the relationship between implicit and explicit evaluation", *Journal of Experimental Psychology : General*, vol 134 (4), pp. 565-584.
- Nosek B, A Greenwald & M Banaji (2005) "Understanding and using the Implicit Association Test: II. Method variables and construct validity", *Personality and Social Psychology Bulletin*, vol 31 (2), pp. 166–180.
- Nosek B & F Smyth (2007) "A multitrait-multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs", *Experimental Psychology*, 54, pp. 14-29.
- Nosek B, F Smyth, J Hansen, T Devos, N Lindner, K Ranganath, C Smith, K Olson, D Chugh, A Greenwald, & M Banaji (2007) "Pervasiveness and correlates of implicit attitudes and stereotypes", *European Review of Social Psychology*, 18, pp. 36-88.
- Phelps E, K O'Conner, W Cunningham, E Funayama, J Gatenby, J Gore, & M Banaji (2000) "Performance on indirect measures of race evaluation predicts amygdala activation", *Journal of Cognitive Neuroscience*, vol 12 (5), pp. 729-738.
- Poehlman T, E Uhlmann, A Greenwald & M Banaji (2005) "Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity", Unpublished manuscript, Yale University.
- Rooth D (2007a) "Evidence of unequal treatment in hiring against obese applicants – a field experiment", IZA Discussion Paper #2775. Bonn.
- Rooth D (2007b) "Implicit discrimination in hiring: real world evidence", IZA Discussion Paper #2764. Bonn.
- Schwartz M, L Vartanian, B Nosek & K Brownell (2006) "The Influence of One's Own Body Weight on Implicit and Explicit Anti-fat Bias", *Obesity*, Vol 14(3), pp. 440-447.

Steffens M (2004) “Is the Implicit Association Test Immune to Faking?”, *Experimental Psychology*; vol 51 (3), pp. 165-179.

Steffens M & I Plewe (2001) “Items' cross-category associations as a confounding factor in the Implicit Association Test”, *Zeitschrift für Experimentelle Psychologie*, vol 48 (2), pp. 123-134.

Appendix: Development of the implicit measures

In order to be able to measure prejudice toward Arab-Muslims in Sweden and obese individuals, we needed to develop three IAT:s. The central idea behind an IAT is always the same, and to change the constructs being measured is done by changing the category and/or the attribute labels and their corresponding stimuli. The selection of these categories and stimuli is thus very important.

The first part of this supplementary study consists of a brief background on previous research on how the stimuli and labels work in the IAT, which dictated how we selected the stimuli for the present study. The second part consists of the development of the categories and stimuli for the implicit work performance stereotype IAT and implicit attitude IAT that were used with Arab-Muslims as the target group. The third and last part explains the development of the implicit work performance stereotype IAT that was used with obese people as the target group.

Labels and stimuli

The category labels and the stimuli used to represent the concepts at hand must be carefully chosen. According to Nosek et al. (2005), the most important consideration is to choose stimuli that are easily associated with only one of the four labels in the IAT. Also, one category of stimuli must be easily distinguished from the others. In this study, for the IAT:s, wherein all stimuli were names and words, the target stimuli (and labels) differed in color and font from the attribute stimuli (and labels) according to convention.

The IAT score is mostly derived from the automatic associations of the four different labels. De Houwer (2001) found that even extreme manipulations of the stimuli, (e.g. substituting Einstein with Hitler as foreigner stimuli), did not change the IAT score as long as the stimuli was representative of the labels. Furthermore, Mitchell et al. (2003) showed that the very same stimuli of black athletes and white politicians yielded different results depending on whether the labels were politicians and athletes or white and blacks. However, there are a number of other studies where manipulations of the stimuli have altered and even reversed the IAT score. Govan & Williams (2003) found that choosing target stimuli that confound with the attribute stimuli can greatly alter the IAT.

Using a set of stimuli of only “nice insects” (e.g. butterfly) and “nasty flowers” (e.g. poison ivy,) they found that the classical Insect/Flower-IAT yielded the opposite IAT-effect, without changing the labels. They also found similar results for a black/white attitude-IAT. This IAT used very negatively valenced white names (e.g., Adolf Hitler and Charles Manson) and very positive black names (e.g., Michael Jordan and Bill Cosby) which resulted in a substantial decrease in the automatic preference of white people over black people, as compared to a typical IAT with neutral names. Govan & Williams provide the explanation that when it is not clear how to sort the stimuli, the participants will recode the labels so that they will make more sense. For example, although the label clearly says “insects” the participants will interpret it as “nice insects”. These findings have received support from other studies as well (see Bluemke & Friese, 2006; Steffens & Plewe, 2001).

Nosek et al. (2005) emphasize that each stimuli category should clearly represent its corresponding label, without confounding it with any of the other labels. It is also important that the stimuli chosen do not only represent a subgroup of the concept of interest. For example, using only the sub sample of male gay stimuli in a “Homosexual-Straight”-IAT made the participants interpret the label “gay” as “gay men” and sort the stimuli accordingly. The results differed substantially from an IAT that used both homosexual male and homosexual female (lesbian) stimuli.

When the labels and the corresponding stimuli match and are without confounds the IAT-effect is mostly based on the labels, rather than the stimuli. This makes the choice of labels especially important. Consider, for example, researchers who wish to measure the relative attitudes towards White Americans and African-Americans. They will then have to choose photos or names that are clearly distinguishable from each other and good exemplars of the categories. But what labels should be chosen could prove difficult. For example, choosing labels that are clearly distinguishable, easy and short without any confounds, might mean the choice of labels that are offensive to some people (i.e. using “Black Americans” and “White Americans” instead of “African-Americans” and “European Americans”). However, Nosek et al. (2007) have shown that alternating between associated labels, such as those mentioned in the above example, only has minimal effects on the measure. This finding is also important as it allows for valid comparisons between studies that have used almost, but not entirely, identical labels.

Given the empirical evidence from the aforementioned studies, it can be concluded that a poor choice of labels and stimuli used in Implicit Association Tests can reduce, increase, or even reverse the effect. Accordingly, we were careful when choosing the category labels and stimuli. It was also necessary to create labels and stimuli in the Swedish language in order to adapt them to the present samples.

The category stimuli were chosen from Carlsson & Rooth (2007) and consisted of Arab-Muslim/Middle Eastern sounding names and Swedish sounding names. These names included: Erik, Lars, Karl, Johan, Anders, Per and Said, Ali, Ameer, Mohammed, Reza (Persian name), Hassan. The names had already been used in the above mentioned study, and we wanted to use the exact same stimuli in order to facilitate comparisons. Thus, the task at hand was to choose the appropriate category labels. We chose the labels “Swedish men” and “Muslim men”. Since the stimuli only consisted of male names it was necessary to specify this in the label. Moreover, while the category label “Muslim men” (on top of the screen) is not a perfectly correct one, it is short and easy to understand, and the risk of misinterpretations was considered to be minimal since the pretest information given to the participants made sure that by “Muslim men” we meant Arab-Muslim minority males living in Sweden. The critical reader may be concerned that we, by using the label “Muslim men”, have investigated attitudes/stereotypes toward religion rather than toward ethnic groups in Sweden. However, since the contrast pair is native “Swedish men”, and not “Christians”, we are confident that we have captured the relative implicit attitudes/stereotypes toward these two ethnic groups rather than different religions.

Using a more correct label on the computer screen (e.g. “Arab-Muslim male minority in Sweden”) would provide too much information and might introduce confounds. Furthermore, by using names representative of Arab-Muslims in Sweden, and typically native Swedish male names, enough cues were provided to make sure that it was the associations pertaining to Arab-Muslims that were being assessed, and not only non-Swedish people in general, or native Swedish people that have converted to Islam.³¹ In other words, we believe that the target labels we chose are a good reflection of how people perceive such

³¹ We are also aware that there are many individuals who do not consider themselves either Arabs or Muslims, who may have the names used in this study. We do however, believe that most people in Sweden perceive these names as Arab-Muslim, and that some discrimination could be based on this perception.

names. Each contrast pair consisted of six “Arab-Muslim” male and six “native Swedish” male sounding names with their respective category labels.

The attitude-IAT required a contrast pair with positive and negative valence for the category labels. Exactly what labels are traditionally used in attitude-IAT:s usually differs. While “Good” vs. “Bad” (e.g. in Project Implicit), or “Pleasant” vs. “Unpleasant” (e.g. Greenwald et al., 1998) are often used, they do not translate well into Swedish without introducing unwanted connotations. The clearest and most easily understandable labels we could choose were the direct Swedish translation of “Positive” and “Negative”³², a contrast pair also recommended by Lane et al. (2007).

The attribute stimuli needed to be clearly associated with these two labels. Since the IAT was in Swedish, the Swedish equivalent of the words commonly used (see Greenwald et al., 1998) were chosen. A simple translation would entail the risk of introducing confounding factors and not maximizing the difference in valence between the positive and the negative stimuli. Therefore, we had 68 students evaluate candidate words, according to the procedure recommended by Bellezza et al. (1986). The students were shown a list of words that were thought to be positive, negative, or neutral and were then asked to rate how positively or negatively they perceived these words. The students rated the candidate words on a 5 point-scale ranging from 1 = “very unpleasant / negative” to 5 = “very pleasant / positive”. Twelve words were selected with half of them rated as highly positive and half of them rated as highly negative. This included the Swedish equivalent of words such as “death”, “pain”, “happy”, and “wonderful”. Further, it was ascertained that these words were general in character and did not have any direct semantic connotations with either of the two target groups (Arab-Muslims and Swedes). The student mean ratings of the 12 words are shown in *Table 6*.

Table 6. Positive and negative words as rated by students

Words rated as highly positive			
Lycka	(happiness)	m= 4.88	SD = .32
Kärlek	(love)	m= 4.80	SD = .47
Glädje	(joy)	m= 4.64	SD = .48
Sol	(sun)	m= 4.48	SD = .58
Skratt	(laughter)	m= 4.46	SD = .55
Underbar	(wonderful)	m= 4.42	SD = .58

³² ”Positiv” and ”Negativ” in Swedish.

Words rated as highly negative			
Mord	(murder)	m= 1.29	SD = .73
Misär	(misery)	m= 1.53	SD = .70
Död	(death)	m= 1.57	SD = .81
Ondskefull	(evil)	m= 1.58	SD = .65
Smärta	(pain)	m= 1.61	SD = .67
Sårad	(hurt)	m= 1.61	SD = .55

The work related performance stereotype IAT required another set of attribute labels and corresponding stimuli. The attribute labels chosen were the Swedish equivalent of “high-performing” and “low-performing”.³³ The labels of this attribute dimension of performance stereotypes were carefully selected to closely capture the desired concept. They were also evaluated by students along with the stimuli used to represent them. These words were evaluated by 113 students following the procedure above, but this time the students were instructed to rate the words on a 5-point scale according to what extent they are associated with performance, where 5 = high performance and 1 = low performance. Ten words were selected with half of them being rated as very high and half of them being rated as very low in performance. The words included, for example, the Swedish equivalents of “efficient” and “lazy”. The students’ mean ratings of the ten words used in this study are presented in *Table 7*. The final words were carefully examined in order to avoid introducing any confounds. We also made sure that they were easy to sort into their respective category.

Table 7. High-performing and low-performing words as rated by students

Words rated as highly associated with high performance			
Effektiv	(effective)	m= 4.56	SD = .59
Produktiv	(productive)	m= 4.56	SD = .68
Flitig	(hardworking)	m= 4.35	SD = .72
Ambitiös	(ambitious)	m= 4.27	SD = .65
Målmedveten	(goal-oriented)	m= 4.25	SD = .61
Words rated as highly associated with low performance			
Ineffektiv	(ineffective)	m= 1.39	SD = .54
Inkompetent	(incompetent)	m= 1.51	SD = .66
Slö	(dull)	m= 1.67	SD = .80
Initiativlös	(initiative-lacking)	m= 1.77	SD = .65
Trög	(slow)	m= 1.89	SD = .73

³³ ”Högpresterande” and ”Lågpresterande” in Swedish.

The development of category stimuli for the obesity IAT

For the performance stereotype IAT used in relation to obesity, we could use the same attribute stimuli and attribute labels as in the Arab-Muslim stereotype IAT. We did, however, need to create new category labels and stimuli. The category labels chosen were the Swedish equivalents to “normal weight individuals” and “obese individuals”.³⁴

It would not be appropriate to use words as category stimuli and we therefore created photos of obese and normal weight individuals. Since it is important that the photo set of obese people only differs with respect to weight when contrasted with the photo set of normal weight people, we had to manipulate the photos. For this purpose, we hired a professional photo editor, who manipulated the photos of normal weight people which he transformed into an obese version that was identical to the original photo in every respect except for the obesity factor. The obesity and normal weight versions of the same person are shown in *Figure 2*. It was necessary to use original pictures of normal weight people and manipulate these stimuli to become obese rather than the other way around, since it is much harder to remove weight, due to the fact that less of the bone structure is visible in photos of obese people and thus need to be simulated. These photos were also used as part of creating fake job applications in Rooth (2007). We are confident that our photos are of high enough quality to be used in an IAT without introducing any confounds, and that most people will find it easy to sort them into the appropriate category.

³⁴ ”Normalviktiga” and ”Överviktiga” in Swedish.



Figure 2. Photos of normal weight men and women manipulated to appear obese

Publication series published by the Institute for Labour Market Policy Evaluation (IFAU) – latest issues

Rapporter/Reports

- 2007:1** Lundin Daniela ”Subventionerade anställningar för unga – en uppföljning av allmänt anställningsstöd för 20–24-åringar”
- 2007:2** Lundin Daniela, Eva Mörk & Björn Öckert ”Maxtaxan inom barnomsorgen – påverkar den hur mycket föräldrar arbetar?”
- 2007:3** Bergemann Annette & Gerard van den Berg ”Effekterna av aktiv arbetsmarknadspolitik för kvinnor i Europa – en översikt”
- 2007:4** Junestav Malin ”Socialförsäkringssystemet och arbetsmarknaden – politiska idéer, sociala normer och institutionell förändring – en historik”
- 2007:5** Andersson Christian ”Lärartäthet, lärarkvalitet och arbetsmarknaden för lärare”
- 2007:6** Larsson Laura & Caroline Runeson ”Effekten av sänkt sjukpenning för arbetslösa”
- 2007:7** Stenberg Anders ”Hur påverkar gymnasialt komvux löneinkomster och vidare studier?”
- 2007:8** ”Forslund Anders & Kerstin Johansson ”Lediga jobb, arbetssökande och anställningar – den svenska matchningsfunktionen”
- 2007:9** Kennerberg Louise ”Hur förändras kvinnors och mäns arbetssituation när de får barn?”
- 2007:10** Nordin Martin ”Invandrades avkastning på utbildning i Sverige”
- 2007:11** Johansson Mats & Katarina Katz ”Underutnyttjad utbildning och lönegapet mellan kvinnor och män”
- 2007:12** Gartell Marie, Ann-Christin Jans & Helena Persson ”Utbildningens betydelse för flöden på arbetsmarknaden”
- 2007:13** Grönqvist Hans & Olof Åslund ”Familjestorlekens effekter på barns utbildning och arbetsliv”
- 2007:14** Lindqvist Linus ”Uppföljning av plusjobb”
- 2007:15** Sibbmark Kristina ”Avidentifierade jobbansökningar – erfarenheter från ett försök i Göteborgs stad”
- 2007:16** Hesselius Patrik & Malin Persson ”Incitamenteffekter och Försäkringskassans kostnader av kollektivavtalade sjukförsäkringar”
- 2007:17** Eriksson Stefan & Jonas Lagerström ”Diskriminering i anställningsprocessen: resultat från en Internetbaserad sökkanal”

- 2007:18** Eriksson Robert, Oskar Nordström Skans, Anna Sjögren & Olof Åslund "Ungdomars och invandrades inträde på arbetsmarknaden 1985–2003"
- 2007:19** Agerström Jens, Rickard Carlsson & Dan-Olof Rooth "Etnicitet och övervikt: implicita arbetsrelaterade fördomar i Sverige"

Working Papers

- 2007:1** de Luna Xavier & Per Johansson "Matching estimators for the effect of a treatment on survival times"
- 2007:2** Lundin Daniela, Eva Mörk & Björn Öckert "Do reduced child care prices make parents work more?"
- 2007:3** Bergemann Annette & Gerard van den Berg "Active labor market policy effects for women in Europe – a survey"
- 2007:4** Andersson Christian "Teacher density and student achievement in Swedish compulsory schools"
- 2007:5** Andersson Christian & Nina Waldenström "Teacher supply and the market for teachers"
- 2007:6** Andersson Christian & Nina Waldenström "Teacher certification and student achievement in Swedish compulsory schools"
- 2007:7** van den Berg Gerard, Maarten Lindeboom & Marta López "Inequality in individual mortality and economic conditions earlier in life"
- 2007:8** Larsson Laura & Caroline Runeson "Moral hazard among the sick and unemployed: evidence from a Swedish social insurance reform"
- 2007:9** Stenberg Anders "Does adult education at upper secondary level influence annual wage earnings?"
- 2007:10** van den Berg Gerard "An economic analysis of exclusion restrictions for instrumental variable estimation"
- 2007:11** Forslund Anders & Kerstin Johansson "Random and stock-flow models of labour market matching – Swedish evidence"
- 2007:12** Nordin Martin "Immigrants' return to schooling in Sweden"
- 2007:13** Johansson Mats & Katarina Katz "Wage differences between women and men in Sweden – the impact of skill mismatch"
- 2007:14** Gartell Marie, Ann-Christin Jans & Helena Persson "The importance of education for the reallocation of labor: evidence from Swedish linked employer-employee data 1986–2002"
- 2007:15** Åslund Olof & Hans Grönqvist "Family size and child outcomes: Is there really no trade-off?"

- 2007:16** Hesselius Patrik & Malin Persson “Incentive and spill-over effects of supplementary sickness compensation”
- 2007:17** Engström Per & Patrik Hesselius “The information method – theory and application”
- 2007:18** Engström Per, Patrik Hesselius & Malin Persson “Excess use of Temporary Parental Benefit”
- 2007:19** Eriksson Stefan & Jonas Lagerström “Detecting discrimination in the hiring process: evidence from an Internet-based search channel”
- 2007:20** Agerström Jens, Rickard Carlsson & Dan-Olof Rooth “Ethnicity and obesity: evidence of implicit work performance stereotypes in Sweden”

Dissertation Series

- 2006:1** Hägglund Pathric “Natural and classical experiments in Swedish labour market policy”
- 2006:2** Savvidou Eleni “Technology, human capital and labor demand”
- 2006:3** Söderström Martin “Evaluating institutional changes in education and wage policy”
- 2006:4** Lagerström Jonas “Discrimination, sickness absence, and labor market policy”
- 2006:5** Johansson Kerstin “ Empirical essays on labor-force participation, matching, and trade”