Does Racial and Ethnic Discrimination Vary Across Minority Groups? Evidence From Three Experiments*

Alison Booth
Economics Program
Research School of Social Sciences
Australian National University
alison.booth@anu.edu.au
http://econrsss.anu.edu.au/Staff/abooth/contact_ab.htm

Andrew Leigh
Economics Program
Research School of Social Sciences
Australian National University
andrew.leigh@anu.edu.au
http://econrsss.anu.edu.au/~aleigh/

Elena Varganova
Economics Program
Research School of Social Sciences
Australian National University
evarganova@yahoo.com

Abstract

We conducted several large-scale field experiments to measure labor market discrimination across different minority groups in Australia – a country where one quarter of the population was born overseas. To denote ethnicity, we used distinctively Anglo-Saxon, Indigenous, Italian, Chinese, and Middle Eastern names, and our goal was a comparison across multiple ethnic groups rather than focusing on a single minority as in most other studies. Our main experiment, an audit discrimination study, involved sending over 4000 fictional resumes to employers in response to job advertisements. In all cases, we applied for entry-level jobs and submitted a CV showing that the candidate had attended high school in Australia. We found economically and statistically significant differences in callback rates, suggesting that ethnic minority candidates would need to apply for more jobs in order to receive the same number of interviews. These differences vary systematically across groups, with Italians (a more established migrant group) suffering less discrimination than Chinese and Middle Easterners (who have typically arrived more recently). We also conducted two additional experiments to form a more nuanced picture of prejudice. These were a 'Return to Sender' experiment and an Implicit Association Test. The results from both experiments reveal societal prejudice against minority groups, although the ranking sometimes differs from that in the audit discrimination study.

JEL Codes: J71, C93
Keywords: discrimination, field experiments, employment

* We are grateful to Boyd Hunter, Gigi Foster, Steven Haider, and seminar participants at the Australian National University's Social and Political Theory Seminar, the Australian National University Centre for Aboriginal Economic Policy Research seminar, the Australasian Labour Econometrics Workshop, and Monash University for valuable comments. Iktimal Hage-Ali and Amy King put us in touch with Gabriella Hannah, who is quoted at the start of the paper. Pablo Mateos kindly allowed us to use a beta version of his *Onomap* software to impute ethnicity to the names of employers and letter recipients. The Implicit Association Test website was created by David Tulloh, and generously publicised by the authors of various blogs, including Larvartus Prodeo, Core Economics, Possum Comitatus, Andrew Norton, Ambit Gambit, Oz Politics, and Club Troppo. Mathias Sinning provided invaluable programming assistance and Susanne Schmidt outstanding research assistance. The background section of this paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) and is managed by the Melbourne Institute of Applied Economic and Social Research (MIAESR). The findings and views reported in this paper, however, are those of the authors and should not be attributed to either FaHCSIA or the MIAESR. We take very seriously the ethical issues surrounding this research. All three of our experiments received approval from the Australian National University's Human Research Ethics Committee. Our first two experiments involve some deception of participants – for a discussion on the ethics of deception in such field experiments, see Riach and Rich (2004).

> *"After completing TAFE in 2005 I applied for many junior positions where no experience in sales was needed – even though I had worked for two years as a junior sales clerk. I didn't receive any calls so I decided to legally change my name to Gabriella Hannah. I applied for the same jobs and got a call 30 minutes later."*
>
> *~ Gabriella Hannah, formerly Ragda Ali, Sydney*

I. Introduction

How should we measure racism and discrimination? Among economists, the most common approach has been to compare labor market outcomes across racial or ethnic groups. But this method may not provide an accurate answer. If an individual's race is correlated with some unobserved productive trait, then differences in economic outcomes will reflect more than just discrimination. Similarly, social researchers have often used surveys to measure the degree of racism in a society. But if respondents know the socially correct response, then this approach will also provide a biased estimate of true attitudes towards racial groups. When studying labor market outcomes, the problem arises from unobservable characteristics of racial minorities. When analyzing social attitudes, the problem stems from unobservable biases in the reporting of racial attitudes.

In both cases, field experiments can help solve the unobservables problem by creating a context in which all other factors except race are held constant. In a context where the subject is unaware that he or she is participating in an experiment – or in which it is difficult for the subject to provide a socially acceptable response – it is more likely that the outcome will provide an accurate measure of racism than with more traditional approaches.

In this paper, we present the results of three field experiments aimed at studying attitudes towards racial and ethnic minorities in Australia, a country whose immigration policy has been admired by other countries.[1] Unlike many field experiments, looking only at a single minority group, we take a broader focus: comparing attitudes to Anglo-Saxon Australians with attitudes to Indigenous Australians (the original inhabitants of the continent), Italian Australians (a relatively established migrant group), Chinese Australians (a more recent migrant group), and Middle Eastern Australians (another recent migrant group). By comparing across these groups, we hope to shed light on how the process of immigrant assimilation might change over time.

[1] For example, this points system has subsequently been taken up by other countries, including New Zealand and, from 2008, the UK.

With one in four residents born overseas, Australia is often regarded as something of a poster child for its ability to absorb new migrants into its social and economic fabric.[2] Skilled migrants are selected through a points system, which gives preference to applicants with high qualifications and workers in high-demand occupations.[3] Perhaps because of this, most research has found little discernable impact of migrants on the labor market conditions of Australian natives.

Yet recent events suggest that the Australian melting pot may not be so successful after all. In the late 1990s, Pauline Hanson's One Nation Party, with its policy of reducing Asian immigration to Australia, polled well in a number of federal and state elections. At the time of the 2000 Sydney Olympics, many journalists drew attention to the poor social indicators among Indigenous Australians. And in 2005, anti-Muslim riots on Sydney's Cronulla Beach drew international attention. As a series of reports have shown, some minority groups in Australia suffer extreme forms of persecution at work and in public places (see e.g. Poynting and Noble 2004; VicHealth 2007; Berman et al. 2008).

Our first experiment aims to estimate racial discrimination by employers. To do this, we conduct an audit discrimination study in which we randomly submit over 5000 fictional applications for entry-level jobs, varying only the name as an indicator of ethnicity. In terms of number of applications submitted, ours is one of the largest audit discrimination studies ever conducted. This allows us to look at multiple racial groups, and to see whether our effects differ by the gender of the fictitious applicant, the type of job advertised, and the city in which the job is located.

Our other two experiments seek to measure racial attitudes in the general population. In the second experiment, we send over 2000 letters to households randomly chosen from the telephone book. In place of the resident's name, the letters carry the same set of ethnically distinctive names as the resumes in the first experiment. Recipients can either put the letter in the trash (the low-cost option), or write "return to sender" on the envelope and mail it back. Here, we test whether householders are more or less likely to choose the low-cost option when the intended recipient has an Anglo-Saxon name than when the intended recipient has a minority name.

The third experiment is an Implicit Association Test. In this experiment, subjects are required to sort words into 'good' and 'bad' categories, and to sort names into 'Anglo' and

[2] The 2006 Census indicates that 28% of the foreign-born in Australia are from 'Anglo' countries, namely the UK, New Zealand, South Africa, USA, Ireland and Canada (listed in order of numerical importance).
[3] See Hatton (2005).

'non-Anglo' categories. In one stage of the experiment, the good words and Anglo names are grouped together, and the bad words and non-Anglo names are grouped together – thereby creating an 'implicit association' between good qualities and Anglo names, and between bad qualities and non-Anglo names. After testing how rapidly the subject can sort words and names in this context, the categories are reversed. Now, subjects must carry out the sorting task with an 'implicit association' between good words and non-Anglo names and between bad words and Anglo names. Unlike the first two experiments, subjects in the Implicit Association Test are aware that they are participating in an experiment designed to judge racial attitudes. However, the design of the test makes it extremely difficult to manipulate the results in order to achieve a socially desirable outcome.

Relative to other work on discrimination, our paper is novel in two respects. First, by comparing across multiple ethnic groups, we are able to learn more about the assimilation process than is possible with studies that focus on just one minority. Second, by using multiple experiments, we are able to form a more nuanced picture of discrimination – analyzing discrimination in the workplace, in the home, and via a subconscious test of attitudes.

The rest of the paper is structured as follows. In section 2, we present background information on the share of Australians falling into the four racial/ethnic categories studied in this paper, and review the available evidence on labor market outcomes and attitudinal surveys. In section 3, we present the results of our employment experiment, and compare our findings with those from other similar studies. In section 4, we briefly present the results from our return-to-sender and Implicit Association Test experiments. The final section concludes.

II. Background

We briefly outline the characteristics of the ethnic groups that are the focus of this study by reviewing the literature on their population share, employment outcomes, and levels of surveyed discrimination. Figure 1 depicts the share of Australian residents in each of the four ethnic minority groups, based upon data from the Australian census, which was conducted in 1901, 1911, 1921, 1933, 1947, 1954, and every five years from 1961 onwards. Until the 1960s, the share of Australians reporting their race as Indigenous was about 1 percent of the population. Since then, the share has risen steadily, and was over 2 percent in 2006. This change has been driven by two factors: higher fertility rates, and a growing willingness of respondents to self-identify as Indigenous.

For Italian, Chinese, and Middle Eastern Australians, our estimates are based upon country of birth (thereby ignoring second-generation immigrants). As the graph shows, Australia experienced a large influx of Italian migrants immediately after World War II. From the late-1970s, the share of Australians who are Italian-born has steadily declined. By contrast, immigration from China and the Middle East only began to expand in the 1970s and 1980s. By 2006, the share of Australians born in Italy, China, and the Middle East was about 1 percent each.

Since our experiments will focus on ethnicity rather than country of birth, a more appropriate comparator might be ancestry. However, the Australian census has not consistently asked respondents about their ancestry. Therefore it is only possible to look at recent data, and not to construct a time series of ancestry shares. We focus here on respondents' first answer to the ancestry question in the 2006 census (it was possible to give multiple ancestries). The ancestries that are relevant to our analysis are Italian (4%), Chinese (3%), and Arab (1%). By comparison, the most common ancestries are Australian (27%) and British (35%). It is not possible to distinguish Indigenous ancestry. While the country of birth figures suggest that Italians, Chinese, and Middle Easterners are about equally represented among first-generation migrants, the ancestry data indicate that Italians are substantially more numerous among second-generation (and higher generation) migrants.

Table 1 shows how these four minority groups perform in the Australian labor market. We estimate three outcome measures – participation, log annual hours, and log hourly wages – with the omitted group being Australian-born non-Indigenous respondents. For this analysis, we require a large dataset with good information on employment participation and hourly wages. Although the census samples are relatively large, earnings and hours are coded in bands, leading to very imprecise measures of hourly wages.[4] We therefore opt to use the 2001-06 Household, Income and Labour Dynamics in Australia survey (HILDA), pooling all six waves and clustering standard errors at the person level. The sample is restricted to those who are aged 21-64, with nonmissing information for all covariates.

Table 1 near here

Indigenous respondents are coded according to whether or not they self-identified as Aboriginal or Torres Strait Islander (HILDA respondents are not asked whether their parents are Indigenous). Respondents are coded as Italian, Chinese, or Middle Eastern if they – or

[4] An alternative approach would have been to simply look at unemployment rates, using data on country of birth from the August 2007 *Employee Earnings and Hours* Survey, and data on race from the August 2006 census. The unemployment rates by country of birth in 2007 were: born in Australia 4.0%, born in Italy 3.7%, born in China 7.2%, and born in North Africa/Middle East 9.5%. The unemployment rate by race in 2006 was 5.0% for non-Indigenous people, and 15.6% for Indigenous people.

either of their parents – were born in one of those countries/regions.[5] We exclude first-generation or second-generation migrants from other regions, so that the omitted group comprises respondents who were born in Australia and whose parents were both born in Australia. Across this particular sample, 3 percent of respondents are Indigenous, 5 percent are Italian, 3 percent are Chinese, and 3 percent are Middle Eastern.

In columns 1, 3, and 5, we include only a parsimonious set of controls – a survey year indicator, a gender indicator, and a quadratic in age. In this specification, most of the coefficients are negative, and there are four significant differences. In terms of employment, Indigenous respondents are 20 percentage points less likely to be employed, Chinese respondents are 9 percentage points less likely to be employed, and Middle Eastern respondents are 11 percent less likely to be employed. Conditional on being employed, Indigenous respondents work 19 percent fewer hours.

Table 2 near here

In columns 2, 4, and 5, we include controls for years of actual labor market experience, years of education, and self-assessed English proficiency. In this specification, the coefficients tend to be closer to zero, and the only significant differences are for Indigenous respondents, who are 12 percent less likely to be employed, and work on average 15 percent fewer hours. However, the standard errors in Table 1 are sufficiently large that we cannot rule out modest levels of labor market discrimination, even controlling for observable productivity differences. Moreover, there are potentially important productivity differences that we are unable to observe, including school quality, interpersonal skills, and work ethics. To the extent that these are correlated with a respondent's race or ethnicity, they could help explain (or confound) estimates of labor market discrimination.

What is known about Australians' attitudes to these minority groups? In Table 2, we present results from four surveys, two of which ask about attitudes towards intermarriage, a third which asks about attitudes to immigration intakes, and a fourth survey that asks people of different races about their own personal experiences of racism. In drawing on these surveys, our aim is not to comprehensively catalog the evidence, but instead to provide a flavor of Australians' attitudes to various ethnic and racial groups in recent years.

In terms of outmarriage by a close relative, respondents were 3 to 4 times more concerned over marriage to someone of Asian or Indigenous descent than to a person of British

[5] We include Hong Kong and Taiwan as part of China. Countries defined as Middle Eastern are Algeria, Egypt, Libya, Morocco, Sudan, Bahrain, Iran, Iraq, Israel, Kuwait, Lebanon, Oman, Syria, and Turkey. Because of the way we code ethnicity, the categories are not mutually exclusive. Dropping respondents who are in more than one minority ethnic category makes no tangible difference to the results.

descent, and 4 to 6 times more concerned over marriage to a Muslim than over marriage to a Christian. This is consistent with attitudes to immigration, which suggest that respondents are more likely to support reducing immigration from the Middle East (38 percent) than from Asia (23 percent) or Europe (12 percent). We also analyze surveys on experiences of discrimination by different ethnic groups. Respondents from non-English-speaking backgrounds are approximately 1½ times as likely to have experienced discrimination in the past 12 months as long-time Australians. Self-reports of discrimination are similar among those who speak a major Asian language and those from a Middle Eastern background.

Finally, we tabulate outmarriage rates for three groups of second-generation migrants: those from Italy, China, and Lebanon or Turkey (from Birrell and Healy 2000). While over half of Italians outmarry, only about one third of Middle Easterners do. Though the share of Chinese outmarrying appears quite large, Birrell and Healy (2000, 40) caution that "when the movement of migrant spouses (that is those married overseas) is taken into account … the Asian experience may well have more in common with that of the Middle East pattern". In that sense, the figures on actual outmarriage rates seem reasonably consistent with surveys on attitudes to outmarriage.

III. The Audit Discrimination Experiment

Our first experiment is an audit discrimination study. The basic notion underlying such studies is that an unbiased estimate of the extent of hiring discrimination can be determined by conducting an experiment in which fake CVs, carrying ethnically or racially identifiable names, are sent to employers. By comparing the callback rates for different ethnic groups, the researcher can estimate the degree of racial or ethnic discrimination in a particular context.

According to a comprehensive review of the literature (Riach and Rich 2002), written audit discrimination studies were initially conducted by British sociologists in 1969 (Jowell and Prescott-Clarke 1970). Since then, researchers have applied the technique to Australia, France, the Netherlands, Sweden, and the United States. (Below, we compare our findings to those from previous studies.) Using written CVs, the audit discrimination technique has been used to measure discrimination on the basis of age, obesity, having a criminal record, facial attractiveness, and sexual orientation. As well as studies that use written applications, researchers have also trained pairs of actors to show up for job interviews, apply for rental housing, and negotiate to purchase used cars (for a recent survey, see Pager 2007).

While such audit discrimination studies using fake CVs have the advantage of providing unbiased estimates of the degree of discrimination in the hiring process, they can only observe the first stage of the employment process. In theory, the level of discrimination in the pre-interview stage could be negatively or positively correlated with discrimination in hiring decisions and wage offers. As Heckman (1998, 102) notes, "A well-designed audit study could uncover many individual firms that discriminate, while at the same time the marginal effect of discrimination on the wages of employed workers could be zero."[6]

During the six months from April 2007 to October 2007, we applied for over 5000 jobs using an online job-finding website. Such a large sample size provides sufficient statistical power to not only look at differences across five ethnic groups (Anglo-Saxon, Indigenous, Chinese, Italian, and Middle Eastern), but also to see whether such effects differed by gender, city, and job type. For example, we still have around 280 individuals per cell when looking at differences by ethnicity and city. However, our results are fragile once we go to three-level tabulations (e.g. ethnicity by job type by gender), so we do not show such results in our tabulations.

In selecting appropriate occupations for this study, we focused on jobs that did not require any post-school qualifications, and for which the application process was relatively straightforward (in order to ensure that we could complete a sufficient number of applications to have good statistical power). We also sought to apply for a mix of occupations, including those that involved face-to-face contact, and those that did not. Ultimately, we selected four occupations – waitstaff, data entry, customer service, and sales. Waitstaff jobs included positions at bistros, cafés, bars, restaurants, and hotels. Data entry positions – also known as document processing officers or technical records officers – included jobs working for an airline, a radio station, a bank, and a charity. Customer service jobs were a mix of telephone support and face-to-face positions (it was often difficult to distinguish these from the information available), and included staffing the front desk at a bowling alley, answering customer support calls at a private health insurance company, and staffing the front desk at a parking garage. Sales positions almost entirely involved in-person sales, and included jobs at a tiling store, a supermarket, an electrical goods store, and a pizzeria. Table 3 gives average wages and share female in these occupations, based on data from the Australian Bureau of Statistics' *Employee Earnings and Hours* survey, conducted in August 2007. The four jobs, more feminized than the non-managerial workforce as a whole, also have a slightly above-

[6] Heckman (1998) and Heckman and Siegelman (1993) present a number of additional critiques of the methodology used in audit studies. Since these primarily deal with studies that use actors, we do not address them here, but one response may be found in Pager (2007).

average share of employees from non-English-speaking backgrounds. Across the four jobs, workers are paid about three quarters of average wages.

Table 3 near here

To test for differences across localities, we applied for jobs in Australia's three largest cities: Sydney, Melbourne, and Brisbane. These cities differ in terms of their ethnic composition (with Sydney being the most ethnically diverse of the three) and the prevailing rate of unemployment at the time of our study (with Brisbane having the tightest labor market).

For each job category, we created four fictional CV templates that we used to apply for jobs. These were obtained from a broad Internet search for similar CVs, and tailored to the particular job. The CV template was augmented with the addition of an address (we selected four street-suburb combinations in middle-income neighborhoods, and randomized the street number between 1 and 20). Two sample CVs are depicted in Appendix Figures 1 and 2.

The ethnicity and race of the applicant was denoted by an ethnically distinguishable name, which appeared in large print at the top of the CV. For each ethnic/racial group, we identified five female first names, five male first names, and five last names, which were combined randomly to create the job applicant's name. Ideally, we would have obtained access to a large database of Australians, containing names and self-identified race/ethnicity. However, we were unable to locate a suitable public database, and sample surveys such as the HILDA survey (or Indigenous databases such as those held by the Australian Institute of Aboriginal and Torres Strait Islander Studies) turned down our requests to tabulate lists of common names. We therefore chose our Anglo-Saxon, Italian, Chinese, and Middle Eastern names by consulting the website www.behindthename.com, and our Indigenous names by consulting the indexes of various books listing Indigenous artists. The full list of names used in this study is provided in Appendix Table 1.

The job-finding website that we used had an online application process. For each advertised position, we submitted four applications, ensuring that each of the four applications was from a different ethnic group. Each application included a short covering letter, plus a fake CV. For each sex-race cell, we set up a separate phone line with an answering machine (all answering machines had a message left by a person with a regular Australian accent), plus an email address. Employers could invite the applicant back for an interview by either sending an email or making a telephone call.

Table 4 near here

Table 4 sets out the callback rates from the experiment. In Panel A, we show results pooling men and women. For Anglo-Saxon-sounding names, the mean callback rate was 35 percent.[7] However, names connoting the four minority groups received a lower callback rate, with Indigenous applicants obtaining an interview 26 percent of the time, Chinese 21 percent of the time, Italian 32 percent of the time, and Middle Eastern 22 percent of the time. For Indigenous, Chinese, and Middle Eastern applicants, the difference is highly statistically significant, but the Anglo vs. Italian difference is only statistically significant at the 10 percent level.

The middle column of Table 4 expresses the difference as a ratio. This is useful because it provides an intuitive metric for the level of discrimination in terms of the number of additional job applications that a minority applicant must submit to get the same number of callbacks as an Anglo applicant. These ratios indicate that, in order to get as many interviews as an Anglo applicant, an Indigenous person must submit 35 percent more applications, a Chinese person must submit 68 percent more applications, an Italian person must submit 12 percent more applications, and a Middle Eastern person 64 percent more applications.

Panels B and C separate the analysis into female and male applicants. This specification indicates that female Italian applicants are not discriminated against (relative to female Anglo applicants), but otherwise the minority groups all have significantly lower callback rates. Relative to Anglo applicants of the same sex, discrimination is generally worse for minority men than for minority women (the exception being those with Chinese-sounding names).[8] However, when we formally test whether racial discrimination differs by gender, we cannot reject the hypothesis that the level of discrimination is the same for men and women of the same ethnic group. In Booth and Leigh (2008), we explore gender differences in more detail and find that, overall, female candidates are more likely to receive a callback than male candidates (the differences are largest for waitstaff and data entry occupations).

One way to benchmark our results is to compare the number of additional applications that a minority candidate must submit in order to expect the same number of interviews. Another is to think about the kind of labor market that minority applicants face.[9] In effect, we

[7] We also tested for differences between Catholic and Protestant names, but found no mean difference between the two groups. Because Catholic respondents were identified both by name and by having a Catholic school on their CV, we were concerned that they might not make an appropriate control group for the purpose of focusing on ethnicity and race. We therefore dropped Catholic CVs from the sample for the current analysis.

[8] We are inclined not to make much of the larger effect for Chinese women, since many non-Chinese would probably have difficulty distinguishing between male and female Chinese first names.

[9] Another approach would be to benchmark the magnitude of our effects against the benefit of more education. However, returns to education did not differ systematically within jobs. We return to this issue below.

can ask the question: *what would the prevailing unemployment rate have to be for an Anglo person to face the same job-finding task as a member of a minority group?*

To answer this, we exploit the fact that the unemployment rate differs across time, and across the three cities in our experiment. Using only Anglo-Saxon respondents, we run a simple probit regression of whether a given respondent gets an interview on the prevailing unemployment rate in that month and city. The coefficient from this regression is -0.065 (standard error 0.033), suggesting that a 1-point increase in the unemployment rate reduces the probability of an Anglo-Saxon applicant getting an interview by 6.5 percent. On average, the prevailing unemployment rate during our analysis was 4.3 percent. However, when we combine the analysis in the previous paragraph with the results in Table 4, it suggests that:

- Indigenous applicants faced the same difficulties in obtaining an interview as an Anglo applicant when the unemployment rate was 5.6 percent;
- Chinese applicants faced the same difficulties in obtaining an interview as an Anglo applicant when the unemployment rate was 6.4 percent;
- Italian applicants faced the same difficulties in obtaining an interview as an Anglo applicant when the unemployment rate was 4.8 percent;
- Middle Eastern applicants faced the same difficulties in obtaining an interview as an Anglo applicant when the unemployment rate was 6.4 percent.

Next, we compare our results with those from similar audit studies conducted in other countries. A survey by Riach and Rich (2002), supplemented with a subsequent literature review, returned 15 comparable studies (including ours), covering 25 minority ethnic groups. The results are set out in full in Appendix Table 2, and graphed in Figure 2. The first comparison is with the earlier Australian audit discrimination estimates from Riach and Rich (1991), based on data collected in Melbourne between 1984 and 1988. In that study, the two minority groups were Greeks and Vietnamese. Although our study does not analyze either of those two groups, it is possible that discrimination involves regional stereotyping. To the extent that this is true, it is notable that we observe little change in the level of discrimination against migrants from Southern Europe (comparing Greeks in 1986 with Italians in 2007), but a substantial increase in discrimination against migrants from South East Asia (comparing Vietnamese in 1986 with Chinese in 2007).[10]

[10] However, if we restrict the 2007 sample to Melbourne applicants only, there is no apparent discrimination against Southern Europeans applying for jobs in Australia in 2007.

Figure 2 also provides an international benchmark for our results. For example, the level of discrimination against African-Americans in the United States in 2001 was higher than the level of discrimination against Indigenous Australians in 2007, but lower than the level of discrimination against Middle Eastern Australians in 2007. Compared with the UK, the level of discrimination against Chinese Australians in 2007 is higher than the discrimination against Asians in the UK in 1997. In comparison with Sweden, the level of discrimination against Middle Eastern Australians in 2007 appears similar to the level of discrimination against Arabic and Middle Easterners in Sweden in 2005-06.

Figure 2 near here

To what extent do levels of discrimination differ across the three cities in our experiment? In Table 5, we present results tabulated separately for Brisbane, Melbourne, and Sydney. In general, the patterns are quite similar. In each of the cities, discrimination is highest against Chinese and Middle Eastern applicants, followed by Indigenous applicants, followed by Italian applicants. However, the point estimates are suggestive of non-trivial differences. For example, if they are to get as many interviews as an applicant with an Anglo name, Chinese applicants must put in 57 percent more applications in Brisbane, but 92 percent more applications in Sydney. In addition, there is a statistically significant degree of discrimination against Italians in Brisbane, but no evidence of discrimination against Italians in Melbourne. To the extent that such differences exist, they could be due to the tightness of the labor market, the ethnic mix of the city, or differences in social norms. However, when we formally test the hypothesis that discrimination is equal across the three cities, we are unable to reject it for any of the three city-pair combinations. (Focusing on individual ethnicities, the only significant difference is the degree of discrimination against Italians in Brisbane and Melbourne.)

Table 5 near here

Next, we test whether the degree of ethnic/racial discrimination differs across the four job types in the survey – waitstaff, data entry, customer service, and sales. This is relevant because it helps to distinguish between discrimination that is motivated by customer discrimination, and discrimination that is motivated by employers or co-workers. If customer discrimination is the primary form of discrimination, then one should expect to see substantially more discrimination in jobs that involve the highest degree of interpersonal contact (waitstaff) than those involving no customer contact (data entry).

Across the four jobs, we observe the greatest amount of discrimination against minority applicants seeking waitstaff jobs. A Chinese and Middle Eastern person seeking a job as a waiter or waitress must submit fully twice as many applications in order to get as many interviews as an Anglo-Saxon applicant. However, there is only slightly less discrimination in data entry jobs, and a formal test cannot reject that the degree of discrimination is the same in both occupations. This suggests that relatively little of the discrimination observed can be attributed solely to customer-based discrimination.

Curiously, the one job in which the level of discrimination appears to be lower is customer service, in which there is no statistically significant discrimination against any of the minority ethnic groups. This is also the one occupation in which those with more education were significantly more likely to receive an interview (a pattern that did not hold in other occupations, as we discuss below). This suggests that there could potentially be less discrimination in higher-skill occupations than in the low-skill jobs analyzed here.

Table 6 near here

To what extent can minority applicants overcome discrimination with a better CV? To test this, we analyze whether callback rates are systematically different between high-quality and low-quality CVs. Recall that for each job, we sent four different CVs to each employer. The CVs differed in the type of experience that each applicant had, but the primary variation between CVs was in educational level. In anticipation of benchmarking our results against returns to education, we systematically varied the level of education of our applicants; assigning to the four CVs no post-school education, vocational training, a bachelor degree at a recently established ('brick') university, or a bachelor degree at an older ('sandstone') university. However, we did not observe consistent returns to education across the four job types. For waitstaff jobs, education appeared to be irrelevant; for data entry jobs, those with university degrees had significantly lower callback rates; for customer service jobs, those with any form of post-school education had higher callback rates; and for sales jobs, those with any post-school education had lower callback rates.

Because of this, our measure of CV quality is based not upon our own judgments about the CVs, but instead on employers' revealed preference. To be precise, we use the callback rate among Anglo-Saxon applicants, ranking employers' perceptions of the 'quality' of each CV. Within each job, we classify the two CVs with the highest callback rates as 'high-quality' CVs (with a mean callback rate among Anglo applicants of 42 percent), and the two with the lowest callback rates as 'low-quality' CVs (these had a mean callback rate among Anglo applicants of 28 percent).

In Table 7, we set out the extent of discrimination within each of these CV quality groups. For Indigenous and Italian applicants, the extent of discrimination does not vary much with CV quality (when measured as a ratio, it is worse with low-quality CVs; when measured as a difference, it is worse with high-quality CVs). However, there is a much starker pattern in the case of Chinese and Middle Eastern applicants, who clearly suffer much more discrimination when their CV is of high quality than when their CV is of low quality. The clearest case is for Middle Eastern applicants, who gain no apparent benefit from having a high-quality CV; despite the fact that Anglo applicants gain a 14 percentage point benefit (42 percent minus 28 percent) from the same increase in CV quality. A formal test easily rejects the null hypothesis that discrimination does not vary with CV quality.

Table 7 near here

Next, we test whether the level of discrimination varies systematically with employer characteristics. We do this in two ways. First, we match on the characteristics of the zipcode in which the employer is located, using data from the 2006 census. Although this has the advantage of precision, it suffers from the drawback that we cannot distinguish the channels through which neighborhood characteristics affect employer behavior. For example, employers in high-minority neighborhoods might themselves be non-Anglo, or they might have greater exposure to other minorities.

The results of this exercise are shown in Table 8 where the sample is the 2701 applicants for which we know the zipcode of the employer and the dependent variable the callback probability. In column 1, we interact the applicant's ethnicity with a measure of the share of respondents born overseas in the zipcode. The interaction coefficients are generally positive, suggesting that discrimination is lower when there are more migrants in a neighborhood. This interaction is significant (at the 10 percent level) for Middle Eastern applicants. However, the magnitude of the effect is quite small – suggesting that discrimination against Middle Eastern applicants is only wiped out when four-fifths of the zipcode is overseas-born.

In column 2, we interact the applicant's ethnicity with the share of people in the employer's zipcode that were born in that country. In column 3, we interact the applicant's ethnicity with the share of people in the employer's zipcode that have that ancestry. Although one main effect is significant (employers located in neighborhoods with more Chinese residents have higher callback rates), the interaction effects are insignificant (we do not observe any systematic relationship between applicants' ethnicity and the share of their ethnic group in the employer's neighborhood).

Table 8 near here

We next exploit the fact that for many jobs, we know the name of the contact person listed on the advertisement, the person who responded to one or more of our applicants, and sometimes both. Software known as OnoMap, developed by researchers at University College London, was used to impute the ethnicity of these individuals, providing a proxy measure of the ethnicity of the person who made the hiring decision. OnoMap assigns ethnicity based on first names and last names, exploiting large databases in which individuals' true names and ethnicities are known. For more details of the coding algorithm, see Mateos et al. (2007) and Mateos (2007).

The results of this exercise are shown in Table 9, in the form of probit regressions where the dependent variable is the callback rate. In the first three columns, we simply classify contact people and responding people as Anglo (i.e. with names in the OnoMap Celtic or English categories), or non-Anglo (i.e. with names in the following OnoMap groups: African, East Asian & Pacific, European, Greek, Hispanic, International, Jewish & Armenian, Muslim, Sikh, or South Asian). In columns 4-6, we classify employer names as being the same or different from the applicant's name (Italian applicants are matched to OnoMap's European and Greek groups, Chinese applicants are matched to OnoMap's East Asian & Pacific and South Asian groups, and Middle Eastern applicants are matched to OnoMap's Muslim group).[11]

We observe positive main effects for Chinese employers, who appear to have a higher callback rate. However, the only interaction effect we observe is for Italian employers, who appear to be significantly less likely to call back job candidates with Italian names. This is a surprising pattern, which suggests that a group with a relatively long history in Australia is actually less inclined to assist members of the same group.

Table 9 near here

IV. The Return to Sender Experiment and Implicit Association Test

In section 2, we noted survey evidence on attitudes to particular ethnic minorities in Australia. However, these surveys may suffer from 'social desirability bias', since respondents in racism surveys might be unwilling to admit to xenophobic views. Our second and third experiments therefore provide a direct assessment of attitudes in the general population towards the same

[11] Matching more narrowly – e.g. matching Italian applicants to OnoMap's Italian names, and Chinese applicants to OnoMap's Chinese names – makes little difference to the results.

four minority groups: Indigenous Australians, Italian Australians, Chinese Australians, and Middle Eastern Australians.

While others have previously administered an audit discrimination study with CVs, the Return to Sender experiment was developed specifically for this study. The experiment operates by incorrectly mailing thousands of letters to households that have been randomly chosen from the telephone book. Although the addresses are known to be correct, the names on the letters are fictional (we use the same set of names as in the previous experiment). Households therefore are faced with a choice: they can either take the low-cost option of putting the letter in the trash, or they can take the high-cost option of writing "return to sender" on the envelope, and mailing it back.

The closest previous experiment to the Return to Sender experiment is the 'Lost Letter' experiment (Milgram et al. 1965), in which a stamped addressed envelope is left in a public place as though its owner had dropped it on the way to a mailbox. The researcher then tests whether the finder posts the letter. Although our research was inspired by the Lost Letter approach, we believe that the Return to Sender experiment has advantages over its predecessor. One is that our experiment is less artificial. While mis-addressed mail is a regular occurrence in Australia, it is extremely rare to find an unposted letter lying on the street, sitting in a public telephone box, or placed under one's car windscreen wiper (three of the treatments in the original Lost Letter experiment). Another advantage is that the Return to Sender experiment is able to obtain a random sample of individuals listed in the telephone book, while the Lost Letter experiment is only able to obtain a random sample of people who use busy public areas during the day (who may not be representative of the general population).

Two features of the Australian postal system make it well-suited to the Return to Sender experiment. First, mis-addressed letters are typically delivered. Most Australian mailboxes do not display the name of the householder, and Australia Post will generally deliver a letter to an address regardless of the name on the envelope. Second, Australia Post does not pick up outgoing letters from household mailboxes. Instead, letters must be posted at post offices or kerbside letterboxes. This is an advantage for us because it raises the cost of returning letters, relative to countries in which outgoing mail is collected from households. For an analysis of the factors that predict letter return in general, see Leigh and Leigh (2008).

Appendix Figure 3 shows an example of a letter that was mailed out, while Appendix Figure 4 shows an example of a letter that was returned. To allow us to compare the results of this experiment with the previous one, we randomly assigned the same names to the letters that we used on CVs in the first experiment. Recipients' addresses were chosen randomly from the

2007 White Pages for Brisbane, Melbourne, and Sydney, and around 2500 letters were mailed in December 2007 and August 2008.[12] All letters contained an invitation to a child's birthday party, with an email address for RSVPs. We monitored this email address, and coded emails as returned letters for the purposes of this exercise (we received 8 emails in the first round, and 2 in the second round). We coded letters as not returned if they had failed to arrive 60 days after being mailed (only 14 letters were returned after this point, and our results are not particularly sensitive to shifting this cutoff point).

In both rounds, 23 percent of all those letters that were mailed out were returned bearing an official Australia Post 'return to sender' sticker. One possibility is that these letters were returned by postal officials without ever being delivered to a home. Another possibility is that they were returned by individuals, who took the incorrectly addressed letter to the counter of their local post office. Since we cannot distinguish empirically between these explanations, we show results both including and excluding letters that carried these official stickers.

Table 10 near here

Table 10 shows the results of this experiment. Whether we include or exclude letters with official stickers, we generally observe that letters with non-Anglo names are less likely to be returned than letters with Anglo names. Since the return rate for Anglo names is 53 percent (including all letters) or 38 percent (dropping those that might have been returned by the Post Office), the results suggest that 3-5 percent of individuals who would have returned a mis-addressed letter bearing an Anglo-Saxon name do not do so if the intended recipient has a Chinese, Italian, or Middle Eastern name. However, these differences are not statistically significant in either specification. Another result worth noting is that Indigenous names have the same return rate as Anglo names, while Italian names have a lower return rate (a different result from that observed in the job-finding experiment).

Next, we seek to understand discrimination via implicit attitudes. For this purpose, we use a computer-based Implicit Association Test (IAT) developed by psychologists to measure individuals' implicit attitudes, that is, the attitudes that a respondent might hold without being explicitly aware of them. In this test, a double categorization task measures the strength of the association between names from particular backgrounds and concepts of good and bad. Thus the IAT indirectly measures the participant's strength of implicit association between concepts. To do this, it relies on the participant's speed of response to various stimuli to be described below. Our procedure closely follows that of Greenwald et al. (1998), surveyed at length in

[12] In total, 3000 letters were mailed, but for consistency with the above exercise, we drop Catholic names from the analysis.

Nosek et al. (2006). For a discussion of the IAT from an economics perspective, see Bertrand et al. (2005). As we realized after embarking on this part of our study, the combination of an audit discrimination study with an IAT follows Rooth (2007), who persuaded a sample of recruiters (to whom he had previously sent fake CVs) to take the IAT.

In our experiment we measured participants' implicit attitudes towards four racial groups relative to the base group of Anglo-Saxon. The participants were individuals who self-located the test through blogs, our university website, or advertisements placed on Google.[13] When each test-taker logged in to take the test, they were randomly assigned to a particular ethnic group (Chinese, Indigenous Australian, Italian, Middle Eastern) to compare with a baseline of Anglo-Saxon. Participants were also asked, at the end of the IAT, to complete a brief online questionnaire eliciting explicit attitudes to different ethnic groups and also demographic details. Individuals whose IP address was non-Australian were dropped from our analysis, as our goal was to complement the two prior experiments with a comparison of Australians' implicit and explicit attitudes towards different races.

The design of the test is as follows. The IAT consists of seven parts, called 'blocks'. In each block, two categories appear, one in the top left hand corner of the screen and the other in the top right hand corner. Then a series of stimuli appears in the centre of screen and each stimulus needs to be sorted into the correct category. In the first set of tasks, the test-taker categorizes names appearing in the screen-centre into the appropriate ethnic group. For example, a person who was randomly assigned to the Chinese test might be asked to sort sequentially the following stimuli: Jennifer Adams (Anglo), Ping Chen (Chinese), Ming Lee (Chinese), Andrew Quinn (Anglo), and so on. To indicate the appropriate side of the screen to which the stimulus is to be sorted, the respondent hits either a key on the left or on the right of the keyboard. If an error is made, the participant has to correct the error before proceeding. For more details on the structure of the IAT, see Appendix A.

In the combined versions, the stimuli within a block are paired according to the stereotype; e.g. an Anglo-Saxon name with a good adjective in one corner, and a Chinese name with a bad adjective in the other corner. In the incompatible version, categories are paired

[13] Most participants came to the study via blogs, which advertised that the exercise was being done for the purposes of research on racial attitudes in Australia, and that participants would be doing an IAT. Among the blogs that publicized the study were Core Economics (economics.com.au), Possum Comitatus (possumcomitatus.wordpress.com), Andrew Norton (andrewnorton.info), Ambit Gambit (ambit-gambit.nationalforum.com.au), Oz Politics (www.ozpolitics.info), and Club Troppo (clubtroppo.com.au). We are grateful to the authors of these blogs for publicizing the website. The Google advertisement used the following keywords: Survey, survey research, survey example, survey design, attitude survey, psychological testing, psychological survey, implicit attitudes, self-perception, Chinese, Middle East, Indigenous, Aborigine, Australian society.

counter to the stereotype (Chinese with good, Anglo-Saxon with bad). Any implicit bias against non-Anglo ethnic groups would show up as a response time differential. In other words, IATs assume that sorting will be faster when the two concepts more closely fit the stereotype as compared to the situation when they do not.

The key outcome measure from an IAT is a 'D-measure', which compares response speeds when the stimuli are paired according to the stereotype with response speeds when the stimuli are paired counter to the stereotype. To be precise, the D-measure is defined as the difference in average response times between the two blocks, divided by the pooled standard deviation of response time in those two blocks. For example, a D-measure of 0 would denote that the typical respondent coded at the same rate in the stereotypical and counter-stereotypical setup, while a D-measure of 0.1 denotes that the typical respondent was one tenth of a standard deviation slower in coding in the counter-stereotypical setup than in the stereotypical setup.

Another way of interpreting this is that since the standard deviation in response times was about 6/10$^{\text{ths}}$ of a second, a D-measure of 0.1 would mean that the typical respondent took 6/100$^{\text{ths}}$ of a second longer to complete each coding exercise in the counter-stereotypical setup (e.g. Chinese+good and Anglo+bad) than in the stereotypical setup (e.g. Chinese+bad and Anglo+good). The typical time taken to complete each coding exercise was about 1 second.

The distribution of our D-measures is plotted in Figure 3. As can be seen from this chart, a non-trivial share of respondents had negative D-scores (indicating that they favored the minority group). In the case of the IAT tests comparing Anglo names with Chinese and Middle Eastern names, the D-measure distributions are bimodal. However, for all four minority groups, between 55 and 61 percent of respondents had positive D-scores. One approach commonly used in the IAT literature is to measure the share of respondents with a D-measure above 0.15. For each of the four minority groups, between 47 and 49 percent of respondents had scores above 0.15.

Figure 3 near here

At the end of the IAT, each participant was guided through an online questionnaire. From this we extracted background information about demographic attributes and racial attitudes. A major concern in the IAT literature has been that respondents to online surveys may be unrepresentative of the broader population. We address this concern in a new and unique manner: by obtaining both demographic and attitudinal data. This allows us to ensure that – although our results are obtained from a self-selected sample – we are able to re-weight them to be more representative of the general population. The creation of our population weights and prejudice weights is described in Appendices B and C.

Table 11 presents the means for our IAT scores, reporting the D-measure for each group relative to those of Anglo-Saxon ethnicity (the larger the number, the more implicit prejudice against the minority group). We present our results in three columns: unweighted (column 1), weighted so that the sample is representative of population demographics (column 2), and weighted so that the stated prejudice of the sample is representative of the general population (column 3). We exclude respondents with a non-Australian IP address, but our sample includes both Australian-born and non-Australian-born individuals (results are similar if we restrict the sample to those born in Australia).

Table 11 near here

In the unweighted results (column 1), the most negative D-measure is for Indigenous names, with a value of 0.108. This indicates that the typical respondent took $1/10^{th}$ of a standard deviation longer (about 0.06 seconds longer) to complete the exercise in the counter-stereotypical setup (Indigenous+good and Anglo+bad) than in the stereotypical setup (Indigenous+bad and Anglo+good). For the other three minority groups – Chinese, Italians, and Middle Easterners – the D-measure is around 0.07. In each case, we can reject the null hypothesis that the unweighted D-measure is equal to zero (column 4).

Since our sample is comprised of Internet users who read blogs or carry out particular Google searches, it is useful to see how our results are affected if we adjust its composition to match either the age-gender-education-birthplace composition of the general population (column 2) or the self-expressed racial prejudice of the general population (column 3). In these cases, we observe quite different patterns. Population-weighting increases the D-scores for the exercises in which Indigenous, Italian, and Middle Eastern names were rated, while decreasing the D-scores for Chinese names. Prejudice-weighting increases the D-scores for Chinese and Middle Eastern names, while decreasing the D-scores for Indigenous and Italian names.

Our preferred specification is that which uses prejudice weights, since we believe that this is most likely to match the sample to the general population. However, while the ordering of D-scores in column 3 approximately accords with the results from the audit discrimination experiment, it is clear that this result is quite sensitive to the weighting procedure. Interestingly, although our employment discrimination study uncovered very low levels of prejudice against individuals with Italian names in the labor market, the IAT results suggest that Australians still retain some implicit prejudice against those with Italian names.

Together, the results from the Return to Sender and IAT experiments suggest that, across our four minority groups, labor market discrimination is only weakly correlated with observed levels of prejudice in the general population. This is consistent with racial and ethnic prejudice having multiple facets: for example, individuals may be quite willing to hire or work alongside a member of a particular minority group, but reluctant to befriend that person.[14] Further experimental work on the complex facets of racial prejudice would be valuable in understanding these dynamics.

V. Discussion and Conclusion

The most common approach to estimating discrimination is through the use of surveys. However, such an approach may potentially provide biased estimates of the true extent of discrimination. For example, if earnings surveys do not contain good measures of productive characteristics such as school quality, and these characteristics are systematically correlated with both race and earnings, then their omission may bias estimates of labor market discrimination. Similarly, in the case of attitudinal surveys, there is a risk that survey respondents may proffer the socially acceptable answer rather than their actual belief.

To address these concerns, we conducted three large-scale field experiments. In our first experiment, which involved sending fake CVs to employers, we are able to obtain an experimental measure of the relationship between job callbacks and the racial soundingness of the applicant's name. In our second experiment, which involved sending mis-addressed letters to randomly selected homes, we are able to obtain an experimental estimate of the degree to which householders are willing to take some extra effort to return a letter addressed to someone with a name that connotes a particular racial or ethnic group. And in our third experiment, we measure the degree of unconscious prejudice through an online Implicit Association Test, which we then re-weight using a national population survey.

Of the three experiments, the results from the hiring experiment are the most precise. There, we find clear evidence of discrimination, with Chinese and Middle Easterners both having to submit at least 50% more applications in order to receive the same number of callbacks as Anglo candidates. Indigenous applicants also suffer a statistically significant level of discrimination, though the effects are smaller (for example, Indigenous applicants in Australia appear to fare a little better than African-Americans in the US job market). We observe virtually no discrimination against Italian applicants. To the extent that we can

[14] Note that this criticism could well be directed towards the explicit prejudice scale that we included in our IAT in order to create our prejudice weights (see Appendix C). However, in including these questions, we were constrained by the need to use the same questions that had previously been used in a survey of the general population.

compare our results with earlier evidence for Australia, our results do not suggest that ethnic and racial discrimination fell from 1986 to 2007.

The results from the Return to Sender and IAT experiments provide evidence of prejudice within the general population towards people in particular ethnic and racial groups. The Return to Sender experiment suggests that around 3-5 percent of individuals who would have returned a mis-addressed letter bearing an Anglo-Saxon name do not do so if the intended recipient has a Chinese, Italian, or Middle Eastern name (though these effects are not statistically significant). The IAT experiment indicates a modest level of implicit prejudice against members of each of the four minority groups. In our preferred approach (weighting the sample according to self-expressed racial prejudice), the ranking is similar to the hiring experiment, but is sensitive to the weights used. Together, the Return to Sender and IAT experiments provide suggestive evidence that racial and ethnic prejudice is not unidimensional, and that attitudes about which groups 'make good workers' may not mirror social attitudes towards these groups.

Naturally, the use of field experiments to measure discrimination has its own limitations. For example, the way in which race and ethnicity is denoted may not necessarily be representative of the general population. In our experiments, we use names that were chosen on the basis that we judged them to be representative of the various ethnic groups. This allows us to conduct experiments in which we only vary the names, but it has the limitation that our results will not necessarily generalize to individuals of the same ethnicity, but with an Anglicized name.

Another limitation of the experimental approach is that we are only able to observe narrow slices of behavior. In the case of hiring, our first experiment provides a precise estimate on the callback stage, but we are unable to speak to discrimination at the interview stage, nor on the job. Similarly, our Return to Sender and IAT experiments are precise measures of two narrow behaviors: the decision to return a letter, and reflex response speeds in classifying words. However, generalizing from these results requires us to make the assumption that they are correlated with other behaviors that are more difficult to measure, such as altruism towards strangers, cooperativeness towards co-workers, and friendliness in social settings.

References

Berman, G. and the Victorian Equal Opportunity & Human Rights Commission. 2008. *Harnessing diversity: addressing racial and religious discrimination in employment. A collaborative project between the Victorian Multicultural Commission and the Victorian Equal Opportunity & Human Rights Commission*. Melbourne: Victorian Equal Opportunity & Human Rights Commission.

Bertrand, M. and Mullainathan, S. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal: A Field Experiment on Labor Market Discrimination", *American Economic Review*, 94(4): 991-1013.

Bertrand, M., Chugh, D. and Mullainathan, S. 2005. "Implicit Discrimination", *American Economic Review*, 95(2): 94-98.

Birrell, B. and Healy, E. 2000. "Out-Marriage and the Survival of Ethnic Communities in Australia", *People and Place*, 8(3): 37-46.

Booth, A. and Leigh, A. 2008. *Do Employers Discriminate by Gender? A Field Experiment in Female-Dominated Occupations*. Canberra: Australian National University, mimeo.

Bovenkerk, F., Kilborne, B., Raveau, F. and Smith, D. 1979. "Comparative aspects of research on discrimination against non-white citizens in Great Britain, France and the Netherlands", in J. Berting, F. Geyer and R. Jurkovich (eds.), *Problems in International Comparative Research in the Social Sciences*, Oxford: Pergamon Press, pp. 105-122.

Bovenkerk, F., Gras, M. and Ramsoedh, D. 1995. *Discrimination against migrant workers and ethnic minoritiesin access to employment in the Netherlands*. International Migration Papers 4, Geneva: ILO.

Brown, C. and Gay, P. 1985. *Racial Discrimination 17 Years After the Act*. London: Policy Studies Institute.

Bursell, M. 2007. *What's in a name? A field experiment test for the existence of ethnic discrimination in the hiring process*. Working Paper 2007:7, Stockholm: The Stockholm University Linnaeus Center for Integration Studies.

Carlsson, M. and Rooth, D.-O. 2007. Evidence of ethnic discrimination in the Swedish labor market using experimental data, *Labour Economics*, 14(4): 716-729.

Dunn, K. 2003. *Racism in Australia: findings of a survey on racist attitudes and experiences of racism*. National Europe Centre Paper No. 77, Sydney: The University of New South Wales.

Esmail, A. and Everington, S. 1993. "Racial discrimination against doctors from ethnic minorities", *British Medical Journal*, 306: 691-692.

Esmail, A. and Everington, S. 1997. "Asian doctors are still being discriminated against", *British Medical Journal*, 314: 1619.

Firth, M. 1981. "Racial discrimination in the British labour market", *Industrial and Labor Relations Review*, 34: 265-272.

Forrest J. and Dunn, K. 2007. *Strangers in our midst? Intolerance and discrimination toward minority cultural groups in Victoria*. Report to VicHealth.

Goldberg, A., Mourinho, D. and Kulke, U. 1996. *Labour market discrimination against foreign workers in Germany*. International Migration Papers 7, Geneva: ILO. Available at http://www.ilo.org/public/english/protection/migrant/equality/discrimination/evidence.htm

Greenwald, A.G., McGhee, D.E. and Schwartz, J.L.K. 1998. "Measuring individual differences in implicit cognition: The implicit association test", *Journal of Personality and Social Psychology*, 74(6): 1464-1480.

Greenwald, A.G, Nosek, B.A., and Banaji, M.R. 2003. "Understanding and using the Implicit Association Test: I. An improved scoring algorithm", *Journal of Personality and Social Psychology*, 85: 197-216.

Hatton, T.J. 2005. "Explaining Trends in UK Immigration", *Journal of Population Economics*, 18(4), 719-740.

Heckman, J. 1998. "Detecting discrimination", *Journal of Economic Perspectives*, 12(2): 101-116.

Heckman, J. and Siegelman, P. 1993. "The Urban Institute audit studies: Their methods and findings", in Fix, M. and Struyk, R.J. (eds.), *Clear and convincing evidence: Measurement of discrimination in America*. Washington, DC: Urban Institute Press, pp. 187-258.

Hubbuck, J. and Carter, S. 1980. *Half a Chance? A Report on Job Discrimination against Young Blacks in Nottingham*. London: Commission for Racial Equality.

Issues Deliberation Australia. 2007. "The Voice Of Australians: A National Random Sample Survey Of The Australian Population", in *Australia Deliberates: Muslims And Non-Muslims In Australi*a. IDA, pp. 88-104.

Jowell, R. and Prescott-Clarke, P. 1970. "Racial discrimination and white-collar workers in Britain", *Race*, 11: 397-417.

Leigh, G. and Leigh, A. 2008. *The Return to Sender Experiment*. Canberra: Australian National University, mimeo.

Markus, A. and Dharmalingam, A. 2008. *Mapping Social Cohesion: The 2007 Scanlon Foundation Surveys*. Melbourne: Monash Institute for the Study of Global Movements.

Mateos, P., Webber, R. and Longley, P. 2007. *The Cultural, Ethnic and Linguistic Classification of Populations and Neighbourhoods using Personal Names*. CASA Working Paper 116, London: Centre for Advanced Spatial Analysis, University College London.

Mateos, P. 2007. "A review of name-based ethnicity classification methods and their potential in population studies", *Population Space and Place*, 13(4): 243-263.

McIntosh, N. and Smith, D. 1974. *The Extent of Racial Discrimination*. Political and Economic Planning Broadsheet no. 547, London: Political and Economic Planning.

Milgram, S., Mann, L. and Harter, S. 1965. "The Lost-Letter Technique: A Tool of Social Research", *Public Opinion Quarterly*, 29(3): 437-438.

Nosek, B.A., Greenwald, A.G. and Banaji, M.R. 2006. "The implicit association test at age 7: A methodological and conceptual review", in Bargh, J.A. (ed.), *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes*. Psychology Press, pp. 265-292.

Pager, D. 2007. "The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future", *The ANNALS of the American Academy of Political and Social Science*, 609: 104-133.

Poynting, S. and Noble, G. 2004. *Living with Racism: The experience and reporting by Arab and Muslim Australians of discrimination, abuse and violence since 11 September 2001. Report to The Human Rights and Equal Opportunity Commission.* Sydney: University of Western Sydney.

Riach, P. and Rich, J. 1991. "Testing for racial discrimination in the labour market", *Cambridge Journal of Economics*, 15: 239-256.

Riach, P. and Rich, J. 2002. "Field Experiments of Discrimination in the Market Place", *Economic Journal,* 112(483): F480-F518.

Riach, P. and Rich, J. 2004. "Deceptive Field Experiments of Discrimination: Are they Ethical?", *Kyklos*, 57(3): 457-470.

Rooth, D.-O. 2007. *Implicit Discrimination in Hiring: Real World Evidence*. IZA DP No. 2764, Bonn: IZA.

VicHealth 2007. *More than tolerance: Embracing diversity for health: Discrimination affecting migrant and refugee communities in Victoria, its health consequences, community attitudes and solutions – A summary report.* Melbourne: Victorian Health Promotion Foundation.

Figures

Figure 1

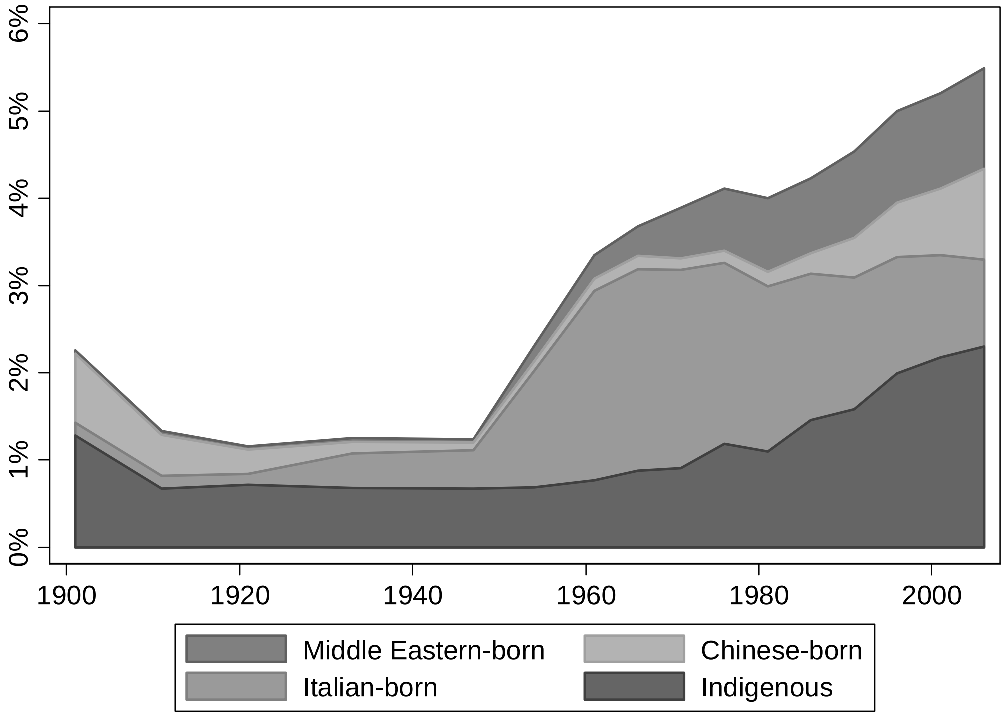
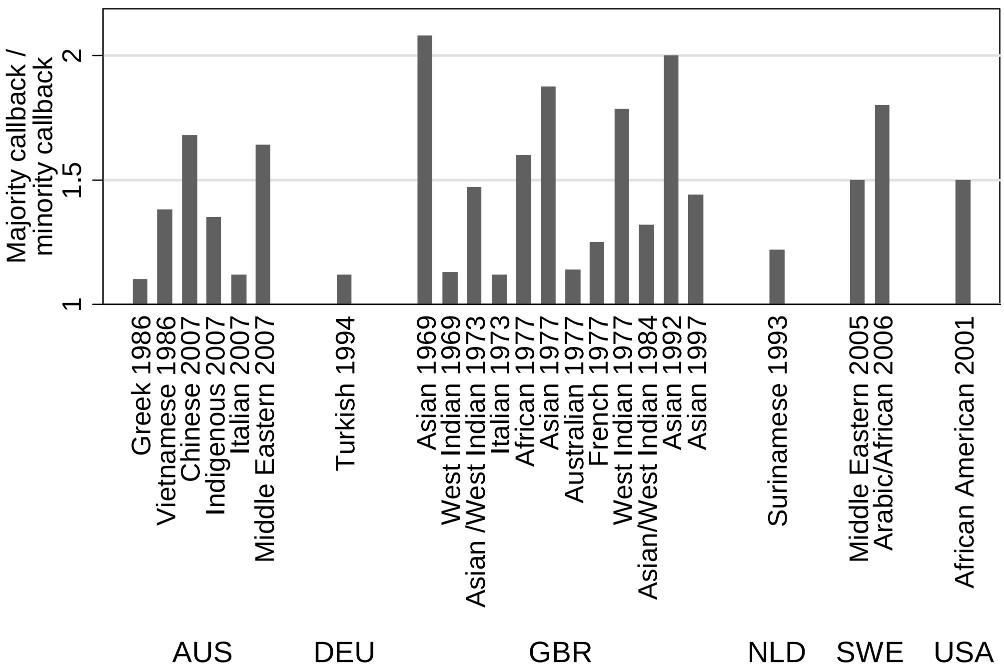
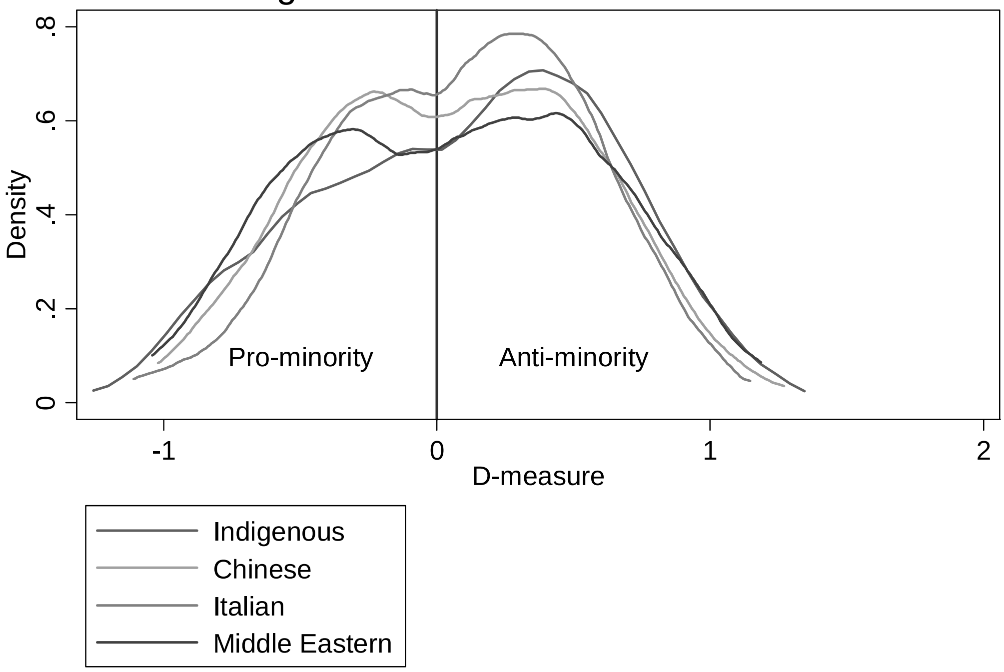


Figure 2



Note: Chart omits Bovenkerk et al. (1979), who found a ratio of 3.47 for Antillian job applicants in France in 1976.

Figure 3: Distribution of IAT Results



Tables

Table 1: Observed Labor Market Differences by Race/Ethnicity

	[1]	[2]	[3]	[4]	[5]	[6]
Dependent variable:	**Employed**		**Log annual hours**		**Log hourly wage**	
Indigenous (self-identified)	-0.200***	-0.122***	-0.185***	-0.152**	-0.042	0.023
	[0.035]	[0.032]	[0.060]	[0.060]	[0.033]	[0.028]
Italian (by birth or parentage)	-0.018	0.001	-0.011	-0.007	-0.025	-0.007
	[0.026]	[0.024]	[0.029]	[0.031]	[0.023]	[0.021]
Chinese (by birth or parentage)	-0.086**	0.014	-0.079	-0.057	0.020	0.053
	[0.038]	[0.034]	[0.052]	[0.057]	[0.058]	[0.049]
Middle Eastern (by birth or parentage)	-0.108***	-0.029	-0.08	-0.058	0.056	0.023
	[0.039]	[0.030]	[0.066]	[0.063]	[0.035]	[0.032]
Control for education, experience, and English proficiency?	No	Yes	No	Yes	No	Yes
Person-year observations	61530	61530	22895	22895	22895	22895
Individuals	8368	8368	6387	6387	6387	6387
R^2 or Pseudo R^2	0.04	0.25	0.11	0.16	0.04	0.18

Source: HILDA survey, waves 1-6. Robust standard errors, clustered at the individual level, in brackets. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels respectively. All regressions control for survey year indicators, a quadratic in age, and a gender dummy. Employment results are marginal effects from a probit model, while results for annual hours and hourly wages are OLS coefficients. Experience is actual labor market experience, education is years of education, and English proficiency is measured by indicators for the four options on a self-assessed scale (very well, well, not very well, not at all). Those who do not speak a language other than English are assumed to speak English very well. Sample is all respondents aged 21-64 in columns 1-2, and employed respondents aged 21-64 in columns 3-6.

Table 2: Attitudinal Evidence		
Group	**Rate**	**Ratio to Majority group**
Concern about outmarriage by race/ethnicity (NSW/Qld sample, Oct-Dec 2001)		
British	7.9%	N/A
Italian	12.3%	1.5
Asian	27.6%	3.5
Indigenous	29.1%	3.7
Concern about outmarriage by race/ethnicity (Vic sample, 2006)		
British	8.3%	N/A
Italian	9.7%	1.2
Asian	19.9%	2.4
Indigenous	25.1%	3.0
Concern about outmarriage by religion (NSW/Qld sample, Oct-Dec 2001)		
Christian	8.9%	N/A
Muslim	53.4%	6.0
Concern about outmarriage by religion (Vic sample, 2006)		
Christian	11.1%	N/A
Muslim	43.9%	4.0
Think immigration from this region should be reduced (2007)		
Europe	12.0%	N/A
Asia	23.0%	N/A
Middle East	38.0%	N/A
Have experienced discrimination in last 12 months, by respondent ethnicity (Vic sample, 2007)		
Long-time Australian	10.3%	N/A
Non-English-speaking background	16.3%	1.6
First language Cantonese, Mandarin, or Vietnamese	16.0%	1.6
Middle East background	15.1%	1.5
Outmarriage rate among second generation (1996-98)		
Italy	55.9%	N/A
China	71.4%	N/A
Lebanon and Turkey	34.6%	N/A

Source: NSW/Qld outmarriage attitudes from Dunn (2003), and Vic outmarriage attitudes from Forrest and Dunn (2007). In both surveys, the question asks "In your opinion how concerned would you feel if one of your close relatives were to marry a person of [group name]" (Not at all, Slightly, Somewhat, Very, Extremely, Don't know). We drop 'Don't know', and code as concerned all respondents who give any response except 'Not at all'. Immigration responses from Issues Deliberation Australia (2007). Discrimination experience from Markus and Dharmalingam (2008). Outmarriage rate among second-generation immigrants is the share of Australian-born brides and grooms whose parents were born in a given country, and who do not marry a first-generation or second-generation migrant from that country. The outmarriage rates are from Birrell and Healy (2000, Table 1), averaging figures for brides and grooms.

Table 3: Characteristics of the Jobs			
	Wage	**Share Female**	**Share NESB**
Waitstaff	$18.90	80%	17%
Data entry	$19.10	85%	15%
Customer service	$21.60	68%	17%
Sales	$18.50	69%	16%
All full-time non-managerial	*$26.00*	*46%*	*15%*

Source: Wage and share female from Australian Bureau of Statistics, *Employee Earnings and Hours* survey, August 2007. Share NESB from HILDA, pooling waves 1-6. NESB denotes respondents who were born in a non-English-speaking country. Since we only have access to the 2-digit occupation code, we classify the four occupations using ISCO-88 codes 51, 41, 42, and 52 respectively.

Table 4: Callback rates by soundingness of name and applicant gender				
	Callback rate	Ratio (Anglo-Saxon rate / Minority rate)	Difference (Anglo-Saxon rate − Minority rate)	P-value on difference
Panel A: Male and Female Applicants				
Anglo-Saxon (N=837)	35%	NA	NA	NA
Indigenous (N=848)	26%	1.35	0.09	0.0000
Chinese (N=845)	21%	1.68	0.14	0.0000
Italian (N=835)	32%	1.12	0.04	0.0940
Middle Eastern (N=845)	22%	1.64	0.14	0.0000
Panel B: Female Applicants				
Anglo-Saxon (N=422)	38%	NA	NA	NA
Indigenous (N=442)	31%	1.23	0.07	0.0311
Chinese (N=374)	21%	1.82	0.01	0.0000
Italian (N=410)	37%	1.03	0.01	0.7858
Middle Eastern (N=434)	25%	1.52	0.13	0.0001
Panel C: Male Applicants				
Anglo-Saxon (N=403)	33%	NA	NA	NA
Indigenous (N=426)	22%	1.51	0.11	0.0003
Chinese (N=403)	22%	1.54	0.12	0.0002
Italian (N=461)	28%	1.21	0.06	0.0686
Middle Eastern (N=435)	19%	1.76	0.14	0.0000
Does ethnic discrimination differ by applicant gender?		$Chi^2(4)$=6.68 P-value=0.15		

Note: To test whether ethnic discrimination differs significantly by applicant gender, we run the probit regression

$$\text{Interview}(0,1) = \alpha + \beta I^{Female} + \gamma \mathbf{I}^{\mathbf{Ethnicity}} + \lambda(I^{Female} \times \mathbf{I}^{\mathbf{Ethnicity}}) + \varepsilon$$

The dependent variable is a dummy for receiving an interview, while I^{Female} and $I^{Ethnicity}$ are, respectively, indicators for being female and being in each of the four minority ethnic categories. The Chi^2 test above is a test for the joint significance of the four λ coefficients.

Table 5: Callback rates by soundingness of name and city

	Callback rate	Ratio (Anglo-Saxon rate / Minority rate)	Difference (Anglo-Saxon rate − Minority rate)	P-value on difference
Panel A: Brisbane				
Anglo-Saxon (N=269)	42%	NA	NA	NA
Indigenous (N=281)	30%	1.41	0.12	0.0030
Chinese (N=283)	27%	1.57	0.15	0.0002
Italian (N=286)	33%	1.28	0.09	0.0261
Middle Eastern (N=280)	28%	1.51	0.14	0.0005
Panel B: Melbourne				
Anglo-Saxon (N=282)	27%	NA	NA	NA
Indigenous (N=272)	18%	1.48	0.09	0.0154
Chinese (N=278)	17%	1.61	0.10	0.0039
Italian (N=282)	29%	0.93	-0.02	0.5722
Middle Eastern (N=284)	16%	1.64	0.10	0.0026
Panel C: Sydney				
Anglo-Saxon (N=286)	38%	NA	NA	NA
Indigenous (N=295)	31%	1.25	0.08	0.0537
Chinese (N=284)	20%	1.92	0.18	0.0000
Italian (N=267)	34%	1.14	0.05	0.2450
Middle Eastern (N=281)	21%	1.80	0.17	0.0000
Does ethnic discrimination differ by city?	**Sydney vs. Melbourne** $Chi^2(4)=4.59$ P-value=0.33	**Sydney vs. Brisbane** $Chi^2(4)=4.47$ P-value=0.35	**Brisbane vs. Melbourne** $Chi^2(4)=5.00$ P-value=0.29	

Note: To test whether ethnic discrimination differs significantly by city, we run the probit regression

$$\text{Interview}(0,1) = \alpha + \beta I^{City} + \boldsymbol{\gamma} \mathbf{I}^{\mathbf{Ethnicity}} + \boldsymbol{\lambda}(I^{City} \times \mathbf{I}^{\mathbf{Ethnicity}}) + \varepsilon$$

The dependent variable is a dummy for receiving an interview, while I^{City} and $I^{Ethnicity}$ are, respectively, indicators for being in a particular city and being in each of the four minority ethnic categories. The Chi^2 test above is a test for the joint significance of the four $\boldsymbol{\lambda}$ coefficients. We run this test three times, for each of the three city-pair combinations.

Table 6: Callback rates by soundingness of name and job type	Callback rate	Ratio (Anglo-Saxon rate / Minority rate)	Difference (Anglo-Saxon rate − Minority rate)	P-value on difference
Panel A: Waitstaff				
Anglo-Saxon (N=223)	50%	NA	NA	NA
Indigenous (N=215)	29%	1.70	0.20	0.0000
Chinese (N=200)	25%	1.99	0.25	0.0000
Italian (N=211)	39%	1.27	0.10	0.0288
Middle Eastern (N=214)	22%	2.27	0.28	0.0000
Panel B: Data Entry				
Anglo-Saxon (N=222)	34%	NA	NA	NA
Indigenous (N=209)	21%	1.60	0.13	0.0031
Chinese (N=199)	19%	1.82	0.15	0.0004
Italian (N=213)	29%	1.18	0.05	0.2472
Middle Eastern (N=207)	20%	1.71	0.14	0.0011
Panel C: Customer Service				
Anglo-Saxon (N=196)	26%	NA	NA	NA
Indigenous (N=215)	28%	0.91	-0.02	0.5836
Chinese (N=215)	23%	1.12	0.03	0.5196
Italian (N=201)	32%	0.79	-0.07	0.1337
Middle Eastern (N=220)	25%	1.02	0.01	0.9048
Panel D: Sales				
Anglo-Saxon (N=196)	31%	NA	NA	NA
Indigenous (N=209)	27%	1.16	0.04	0.3369
Chinese (N=231)	18%	1.71	0.13	0.0018
Italian (N=210)	26%	1.19	0.05	0.2717
Middle Eastern (N=204)	20%	1.59	0.12	0.0081
Does ethnic discrimination differ between waitstaff and data entry?		$Chi^2(4)=3.55$ P-value=0.47		

Note: To test whether ethnic discrimination differs significantly by job, we run the probit regression

$$Interview(0,1) = \alpha + \beta I^{Waitstaff} + \boldsymbol{\gamma} \mathbf{I}^{\mathbf{Ethnicity}} + \boldsymbol{\lambda}(I^{Waitstaff} \times \mathbf{I}^{\mathbf{Ethnicity}}) + \varepsilon$$

The dependent variable is a dummy for receiving an interview, while $I^{Waitstaff}$ and $I^{Ethnicity}$ are, respectively, indicators for applying for a waitstaff job and being in each of the four minority ethnic categories. The Chi^2 test above is a test for the joint significance of the four $\boldsymbol{\lambda}$ coefficients. We run this test with waitstaff and data entry positions only.

Table 7: Callback rates by soundingness of name and CV quality

	Callback rate	Ratio (Anglo-Saxon rate / Minority rate)	Difference (Anglo-Saxon rate − Minority rate)	P-value on difference
Panel A: Low-Quality CVs				
Anglo-Saxon (N=390)	28%	NA	NA	NA
Indigenous (N=460)	21%	1.34	0.07	0.0162
Chinese (N=451)	18%	1.54	0.10	0.0007
Italian (N=381)	24%	1.17	0.04	0.1980
Middle Eastern (N=413)	22%	1.30	0.06	0.035
Panel B: High-Quality CVs				
Anglo-Saxon (N=447)	42%	NA	NA	NA
Indigenous (N=388)	33%	1.28	0.09	0.0056
Chinese (N=394)	24%	1.73	0.18	0.0000
Italian (N=454)	38%	1.10	0.04	0.2261
Middle Eastern (N=432)	22%	1.93	0.20	0.0000
Does ethnic discrimination differ by CV quality?	$Chi^2(4)=18.75$ P-value=0.0009			

Note: To test whether ethnic discrimination differs significantly by CV quality, we run the probit regression

$$\text{Interview}(0,1) = \alpha + \beta I^{HighQualityCV} + \boldsymbol{\gamma} \mathbf{I}^{\mathbf{Ethnicity}} + \boldsymbol{\lambda}(I^{HighQualityCV} \times \mathbf{I}^{\mathbf{Ethnicity}}) + \varepsilon$$

The dependent variable is a dummy for receiving an interview, while $I^{HighQualityCV}$ and $I^{Ethnicity}$ are, respectively, indicators for having a high quality CV and being in each of the four minority ethnic categories. The Chi^2 test above is a test for the joint significance of the four $\boldsymbol{\lambda}$ coefficients.

Table 8: Applicant Ethnicity and Employer Neighborhood Characteristics

	[1] Overseas-born share	[2] Born in same country	[3] Same ancestry
Indigenous applicant	-0.167***	-0.090***	-0.090***
	[0.046]	[0.027]	[0.026]
Chinese applicant	-0.153***	-0.127***	-0.130***
	[0.048]	[0.028]	[0.028]
Italian applicant	-0.098*	-0.068**	-0.075**
	[0.052]	[0.028]	[0.035]
Middle Eastern applicant	-0.205***	-0.127***	-0.127***
	[0.042]	[0.024]	[0.025]
Indigenous addressee × Overseas born share	0.201		
	[0.140]		
Chinese addressee × Overseas born share	0.04		
	[0.142]		
Italian addressee × Overseas born share	0.137		
	[0.137]		
Middle Eastern addressee × Overseas born share	0.239*		
	[0.140]		
Overseas born share	0.005		
	[0.100]		
Indigenous addressee × Indigenous share		-1.578	-14.919
		[1.475]	[14.715]
Chinese addressee × Chinese share		-0.326	-0.165
		[0.379]	[0.270]
Italian addressee × Italian share		2.283	1.079
		[2.041]	[1.098]
Middle Eastern addressee × Middle Eastern share		-1.626	-0.585
		[1.717]	[0.762]
Indigenous share		0.328	1.11
		[0.409]	[0.809]
Chinese share		0.650***	0.441***
		[0.189]	[0.136]
Italian share		-1.879	-0.759
		[1.203]	[0.628]
Middle Eastern share		-0.219	-0.211
		[0.532]	[0.253]
Observations	2701	2701	2701
Pseudo R^2	0.07	0.07	0.07

Note: Table shows marginal effects from a probit model. Standard errors in brackets. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels respectively. All estimates include indicator variables for job type, city, and CV template. Share variables are the share born in a given country in column 2, and the share with a given ancestry in column 3.

Table 9: Applicant Ethnicity and Employer Ethnicity

	[1] Contact non-anglo	[2] Responder non-anglo	[3] Contact or responder non-anglo	[4] Contact same race	[5] Responder same race
Indigenous applicant	-0.111***	-0.132***	-0.111***	-0.106***	-0.126***
	[0.025]	[0.036]	[0.025]	[0.024]	[0.033]
Chinese applicant	-0.178***	-0.236***	-0.169***	-0.169***	-0.225***
	[0.023]	[0.034]	[0.024]	[0.021]	[0.031]
Italian applicant	-0.065**	-0.054	-0.055**	-0.063**	-0.059*
	[0.027]	[0.038]	[0.027]	[0.026]	[0.035]
Middle Eastern applicant	-0.145***	-0.231***	-0.160***	-0.146***	-0.218***
	[0.024]	[0.034]	[0.024]	[0.022]	[0.031]
Indigenous applicant × Non-Anglo employer	0.044	0.033	0.085		
	[0.084]	[0.087]	[0.074]		
Chinese applicant × Non-Anglo employer	0.048	0.024	0.065		
	[0.087]	[0.085]	[0.073]		
Italian applicant × Non-Anglo employer	0.081	-0.003	0.077		
	[0.088]	[0.086]	[0.074]		
Middle Eastern applicant × Non-Anglo employer	-0.019	0.068	0.079		
	[0.078]	[0.085]	[0.073]		
Non-Anglo employer	-0.001	0.021	0.016		
	[0.053]	[0.060]	[0.048]		
Chinese applicant × Chinese employer				0.14	0.055
				[0.102]	[0.101]
Italian applicant × Italian employer				-0.178**	-0.244*
				[0.086]	[0.131]
Middle Eastern applicant × Middle Eastern employer				-0.125	-0.078
				[0.141]	[0.219]
Chinese employer				0.157	0.209***
				[0.101]	[0.077]
Italian employer				0.02	-0.001
				[0.039]	[0.043]
Middle Eastern employer				0.041	0.021
				[0.090]	[0.099]
Observations	2335	2319	3313	2335	2319
Pseudo R^2	0.07	0.09	0.06	0.07	0.09

Note: Table shows marginal effects from a probit model. Standard errors in brackets. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels respectively. All estimates include indicator variables for job type, city, and CV template. Employer ethnicity is imputed using the name of the contact in the job advertisement in columns 1 and 4, the name of the person who responded to candidates in columns 2 and 5, and either of those people in column 3 (if either is non-Anglo, the employer is coded as non-Anglo).

Table 10: Letter return rates by soundingness of name

	Letter return rate	Ratio (Anglo-Saxon rate / Minority rate)	Difference (Anglo-Saxon rate − Minority rate)	P-value on difference
Panel A: All Letters				
Anglo-Saxon (N=499)	53%	NA	NA	NA
Indigenous (N=497)	53%	0.99	0.00	0.8963
Chinese (N=509)	48%	1.09	0.04	0.1652
Italian (N=504)	48%	1.09	0.04	0.1742
Middle Eastern (N=497)	49%	1.07	0.04	0.2543
Panel B: Excluding Letters with Post Office Stickers				
Anglo-Saxon (N=383)	38%	NA	NA	NA
Indigenous (N=377)	38%	0.99	0.00	0.9229
Chinese (N=389)	33%	1.17	0.05	0.1118
Italian (N=394)	34%	1.13	0.04	0.2049
Middle Eastern (N=387)	34%	1.11	0.04	0.2787

Table 11: Implicit Association Test Results
Larger scores on the D-measure correspond to greater implicit prejudice against the minority group

	[1] Unweighted	[2] Population weighted	[3] Prejudice weighted	[4] Test that unweighted mean equals zero (P-value)
Indigenous (N=203)	0.108	0.158	0.063	0.003
Chinese (N=216)	0.067	0.058	0.078	0.047
Italian (N=202)	0.076	0.126	0.063	0.018
Middle Eastern (N=184)	0.067	0.088	0.123	0.088

Note: Results are from an IAT comparing a given minority group with Anglo-Saxon. Results in column 2 are weighted so that the sample is representative of population demographics. Results in column 3 are weighted so that the stated prejudice of the sample is representative of the general population. Column 4 shows the p-value from a t-test against the hypothesis that the unweighted mean D-score for that group equals zero.

Appendices

Appendix Table 1: Ethnically Distinctive Names	
Anglo first names	Female: Jennifer, Lisa, Kimberly, Sarah, Amy Male: Martin, Andrew, Phillip, Adam, Brian
Anglo last names	Abbott, Adams, Johnson, Mitchell, Robinson
Middle Eastern first names	Female: Fatima, Lala, Nadine, Anan, Hiyam Male: Ahmed, Hassan, Bilal, Mahmoud, Rafik
Middle Eastern last names	Hariri, Baghdadi, Chikhani, Kassir, Gemayel
Indigenous first names	Female: Betty, Winnie, Daisy, Dorothy, Peggy Male: Bobby, Jimmy, Tommy, Wally, Ronnie
Indigenous last names	Japanangka, Tjungarrayi, Djukukul, Tipungwuti, Puruntatameri
Chinese first names	Female: Ping, Ming, Xiu, Ya, Nuying Male: Tai, Hong, Yin, Peng, Hu
Chinese last names	Chen, Lin, Huang, Lee, Chang
Italian first names	Female: Maria, Anna, Rosa, Angela, Giovanna Male: Giuseppe, Giovanni, Antonio, Mario, Luigi
Italian last names	Rosso, Ferrari, Bianchi, Romano, Galeotti

Appendix Table 2: Comparison with Other Audit Discrimination Studies

Study	Country	Year(s) of test	Minority	Ratio (majority callbacks divided by minority callbacks)
Riach and Rich (1991)	Australia	1984–88	Vietnamese	1.38
			Greek	1.10
Booth et al. (This study)	Australia	2007	Indigenous	1.35
			Chinese	1.68
			Italian	1.12
			Middle Eastern	1.64
Bovenkerk et al. (1979)	France	1976–77	Antillian	3.47
Goldberg et al. (1996)	Germany	1994	Turkish	1.12
Bovenkerk et al. (1995)	Netherlands	1993/4	Surinamese/m	1.27
			Surinamese/f	1.17
Carlsson and Rooth (2007)	Sweden	2005/6	Middle Eastern	1.50
Bursell (2007)	Sweden	2006/07	Arabic/African	1.80
Jowell and Prescott-Clarke (1970)	UK	1969	Asian	2.08
			West Indian	1.13
McIntosh and Smith (1974)	UK	1973	Asian /West Indian	1.47
Firth (1981)	UK	1977/8	Asian	1.95
			West Indian	1.76
			Australian	1.14
			French	1.25
			African	1.60
Hubbuck and Carter (1980)	UK	1977/9	Asian	1.80
			West Indian	1.81
			Italian	1.12
Brown and Gay (1985)	UK	1984/5	Asian/West Indian	1.32
Esmail and Everington (1993)	UK	1992	Asian	2.00
Esmail and Everington (1997)	UK	1997	Asian	1.44
Bertrand and Mullainathan (2004)	US	2001/02	African American	1.50

Note: All studies except Bertrand and Mullainathan (2004), Carlsson and Rooth (2007), Bursell (2007), Goldberg et al. (1996), and Booth et al. (this study) are summarized in Riach and Rich (2002). Note that Jowell and Prescott-Clarke (1970) changed not only the names but also the qualifications.

Appendix Figure 1

Matthew O'Brien

Personal Detail:

15 Boundary Rd
Mortdale, Sydney
Tel: 91149283
E-mail: obrienluck@gmail.com

Personal Profile:

World class customer service agent with excellent communication & inter-personal skills. Consistently achieve quality, service and financial metrics. Committed to team success - able to multitask, and meet deadlines. Flexible and detail oriented with a commitment to understanding procedures. Demonstrate integrity & compliance - confidentiality in handling correspondence & customer files and code of ethics.

Skills:

-Accurate and rapid typing.
-Proficient in Microsoft Office/Word & the Internet.
-Knowledge of Excel, Access, Lotus Notes, IDT, IP Agent and other software applications.
-Word processing skills
-Performing administrative duties for Senior Management Personnel.
-Banking, processing procedures, product and services

Work Experience:

2005 - 2007
Service Specialist
Medibank Private, Sydney

2001 to 2005
Customer Service Representative
Optus, Sydney

1997 to 2001
Administrative Assistant
National Bank, Sydney

Highlights of Qualifications:

•Inbound call center, customer service and banking experience.
•Work effectively in a changing environment refocusing efforts with a positive attitude.
•Not afraid to think outside the box.
•Ensure customer care goals are achieved efficiently and effectively.
•Provide excellent customer service to both clients and providers quickly and accurately.
•Communicate unpleasant or negative information in a tactful manner.
•Establish and maintain control of inbound calls using a well organized structure.
•Resolve complex or basic inbound calls using sound business judgment.
•Promoting the company's products and services as a benefit to the client.
•Provide ongoing and comprehensive communication.

Education:

Certificate in Paralegal Studies
Completed Year 12

Diploma of Business Administration

Reference:

Available on request

Appendix Figure 2

Dorothy Japanangka

6 Cavendish St
Stanmore, Sydney
Tel: 91149463
E-Mail: japanangka.f@gmail.com

Experience:

10/2004-03/2007
Food Prep Chef
University Union Court

Prepared food items for chef to use in recipes being served the following day. Helped in other areas of the kitchen and in dining area as needed.

7/2003-10/2004
Hungry Jacks, Sydney

Prepared buns and burger meat, chicken etc for use in burgers during business hours. Also washed dishes and assisted in garbage disposal.

8/2001 - 11/2002
Domino's Pizza, Sydney

Preparation of pizza toppings, pizzas and salad ingredients at various stages. Took telephone orders. General cleaning of the store. I was promoted to a management position which involved assisting with all operational tasks, scheduling, inventory, training of new employees, book keeping and cash handling.

Skills:

Food Preparation and Serving Related First-Line Supervisors/Managers of Food Preparation and Serving Workers.

Education:

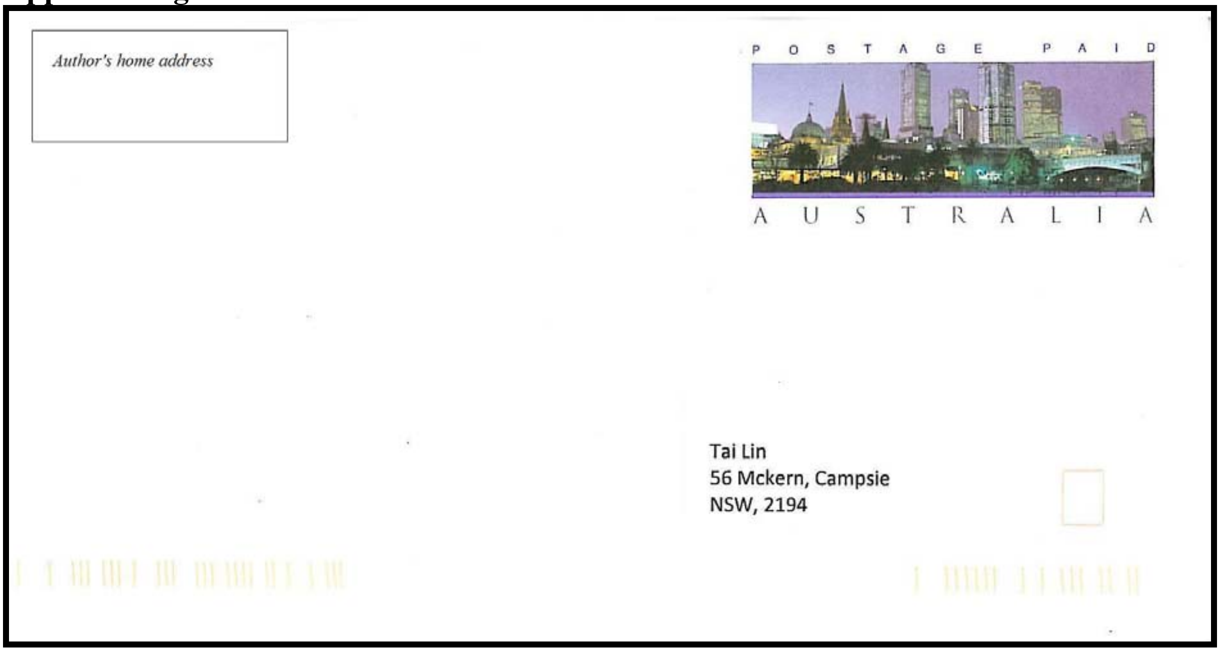
Finished year 12
Strathfield South College

Bachelor of Arts
University of Technology, Sydney

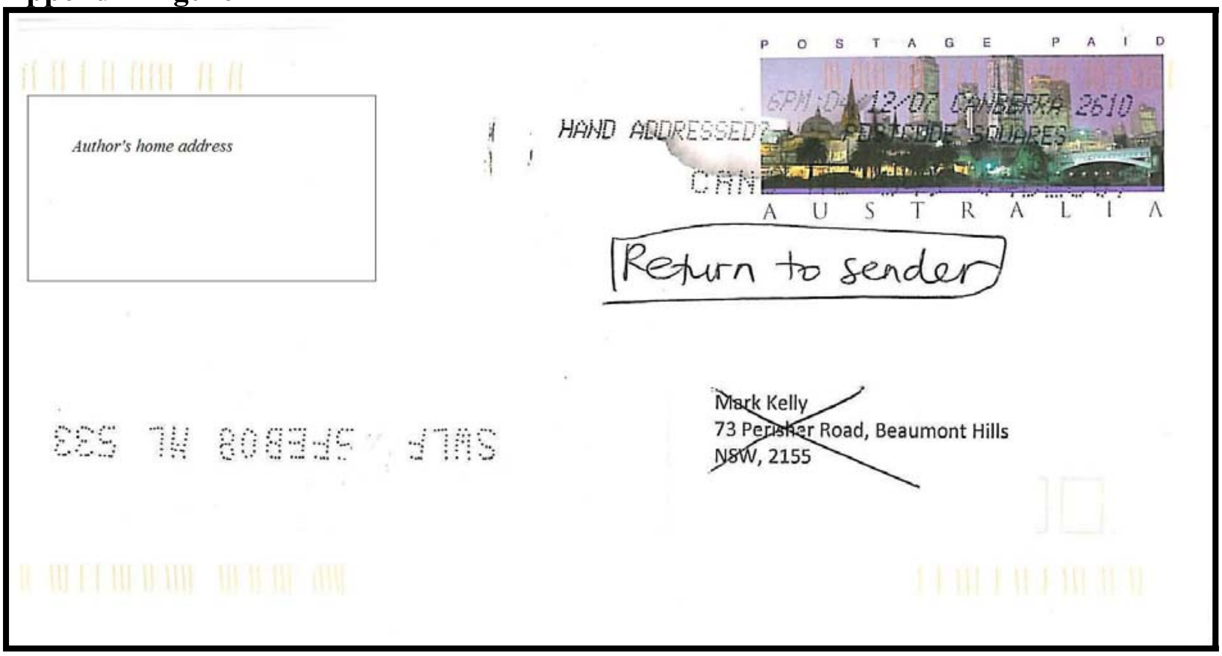
References:

Please let me know if you need references.

Appendix Figure 3



Appendix Figure 4



Appendix A: Details on the construction of the IAT

Sequence of Blocks in the IAT

Block	No. Trials (stimuli)	Name of Block	Task
B1	20	Ethnicity training	Categorize stimulus appearing in screen centre (either Anglo or particular ethnic-group name) to correct side of screen
B2	20	Good/bad training	Categorize stimulus appearing in screen centre (words relating to good/bad) to correct side of screen
B3	20	Combined training	Categorize sequence of combined names and words appearing in screen centre to correct side of screen
B4	40	First test	Categorize sequence of combined names and words appearing in screen centre to correct side of screen
B5	40	Sides of ethnicities now reversed before ethnicity retraining	Categorize names appearing in screen centre to correct side of screen
B6	20	Combined retraining	Categorize sequence of combined names and words appearing in screen centre to correct side of screen
B7	40	Second test	Categorize sequence of combined names and words appearing in screen centre to correct side of screen

Notes:

(i) Items assigned to left-key response or right-key response in each block were randomized across participants.

(ii) A trial is defined as the time from the appearance of a single stimulus to the correct categorization of that stimulus. If an error is made in that trial, the participant has to correct the error before proceeding.

The data on average response times obtained from the IAT are used to measure respondents' implicit attitudes towards people of particular ethnic backgrounds. For our analysis, we use a transformation of the data, known in the literature as the D-measure. This is defined as the difference in average response times between the two relevant blocks, divided by the standard deviation of response time in those two blocks. Following the 'improved scoring algorithm' for the IAT set out in Greenwald et al. (2003), we omitted trials taking longer than 10 seconds, and dropped subjects for whom more than 10 percent of trials took less than 0.3 seconds. Thus we computed our first D-measure, D1, as the difference between the means for each of B3 and B6, divided by the 'inclusive' standard deviation for those two blocks. Our second D-measure, D2, is computed as the difference between the means for each of B4 and B7, divided by the 'inclusive' standard deviation for those two blocks. Finally, we calculated the overall D-measure as D=(1/3)D1 + (2/3)D2, because of the different number of trials in each block.

Appendix B: Creation of population weights for the online IAT sample

We created two weights for each observation in the IAT: a ‘population weight’ (described in the following paragraphs) and a ‘prejudice weight’ (described in Appendix C).

Our population weights in the IAT are designed to ensure that our sample better fits the age-sex-education-birthplace distribution in the general population. We were unable to obtain a public-use tabulation that suited our purposes (the ABS’s Education & Work tables use different age bands to ours, and do not tabulate education in the same way). We considered using the population weights in the Australian Survey of Social Attitudes (AuSSA), but since they are only based on education, we decided not to pursue this (our IAT sample is skewed by age and sex as well).

Instead, we estimated the age-sex-education-birthplace distribution in the HILDA survey, incorporating HILDA’s own survey weights in our calculations. We used the latest wave of HILDA that was available to us at the time, being 2006. We divide the sample into 140 cells, based on two genders, six age categories (18-30, 31-40, 41-50, 51-60, 61-70, 71-80, over 80), five education categories (Less than year 12, Year 12, Trade/Apprenticeship, Certificate/Diploma, Bachelor Degree and Above), and two birthplace categories (born in Australia, born overseas). There are HILDA respondents in all 140 cells, but IAT respondents in only 74 cells. IAT respondents aged under 18, or with missing values for age, sex, or education, are assigned a population weight of 1.

Appendix C: Creation of 'prejudice weights' for the online IAT sample

One of our concerns with the online IAT is that individuals may self-select into doing an online survey of racial attitudes because they are more racially tolerant than the mainstream population. To correct for this, we included questions on the survey about self-reported levels of racial prejudice. These questions also appear on the 2007 Australian Survey of Social Attitudes, which allows us to benchmark our results against those in the AuSSA.

The two questions we use are:

How close are you prepared to be with Lebanese?

1. *Welcome as family member*
2. *Welcome as close friend*
3. *Have as next door neighbour*
4. *Welcome as work mates*
5. *Allow as Australian citizen*
6. *Have as visitor only*
7. *Keep out of Australia altogether*
8. *Don't know*

How close are you prepared to be with Aborigines?

1. *Welcome as family member*
2. *Welcome as close friend*
3. *Have as next door neighbour*
4. *Welcome as work mates*
5. *Allow as Australian citizen*
6. *Have as visitor only*
7. *Keep out of Australia altogether*
8. *Don't know*

We also asked a question about closeness with Italians in the IAT, but we do not use it for creating our prejudice weights.

In order to match the the AuSSA and the IAT, we create a 'prejudice index', which is simply the sum of responses to the two questions, for respondents who provided answers between 1 and 7 on both questions (i.e. we drop those who answered 'don't know' on either question). This prejudice index ranges from 2 (most tolerant) to 14 (least tolerant).

The distribution across this prejudice index in the two surveys is shown below.

Prejudice Index	Share in AuSSA	Share in IAT	Prejudice weight
2	17.54%	61.39%	0.29
3	2.01%	6.63%	0.30
4	16.02%	14.26%	1.12
5	3.65%	4.65%	0.78
6	11.84%	5.05%	2.35
7	6.24%	2.48%	2.52
8	11.37%	2.97%	3.83
9	6.66%	0.79%	8.41
10	9.92%	0.69%	14.31
11	6.43%	0.30%	21.65
12	5.35%	0.30%	18.01
13	1.47%	0.20%	7.42
14	1.50%	0.30%	5.04
	100.00%	100.00%	

Our prejudice weight variable is simply the ratio of the AuSSA share to the IAT share, and is shown in the final column above.

IAT respondents who failed to answer either of the closeness questions, or who answered 'don't know' to either of the prejudice questions, are given a prejudice weight of 1.

At an individual level, the correlation between the prejudice weight and the population weight is 0.11 (this is the same whether we include or exclude weights of 1).