Cognitive discrimination - a benchmark experimental study

Citation for published version:
Belot, M 2015, 'Cognitive discrimination - a benchmark experimental study', *Journal of Neuroscience, Psychology, and Economics*, vol. 8, no. 3, pp. 173-185. https://doi.org/10.1037/npe0000042

Digital Object Identifier (DOI):
10.1037/npe0000042

Published In:
Journal of Neuroscience, Psychology, and Economics

Cognitive Discrimination - A Benchmark Experimental Study

Michèle Belot

August 10, 2015

Abstract

This study provides experimental evidence for a cognitive mechanism underlying racial discrimination. Specifically, we study memory biases in the ability to remember people within and across races, in a context where there is no difference in the distribution of payoffs across racial groups. Participants see pictures of people - whom we call candidates - of White and East Asian origin, learn payoff-relevant information about them and then are asked to identify the candidates associated with the highest payoffs. I find that people are much better able to recall candidates with higher payoffs if they are of the same race. Candidates of the other race are more likely to be confused with each other. This leads to positive and negative discrimination at the same time: those at the bottom of the distribution benefit while those at the top lose out. These results suggest that cognitive biases could play a role in the nature of cross-racial relations, in particular for phenomena relying on repeated interactions and individual recognition, such as the formation and maintenance of social ties and the establishment of trust relationships.

Keywords: Statistical discrimination, Bounded Memory

JEL codes: J71, C91, D83

1 Introduction

Remembering people plays a key role in many social contexts. Most of our interactions involve people we have met before: acquaintances, professional relations, local merchants, etc. Keeping track of people - who they are and what they do - improves efficiency. It plays a key role in the formation and maintenance of social ties, in repeated interactions and in mechanisms relying on repeated interactions such as trust and cooperation. All involve the recall of another person's identity and past behaviour. In an increasingly racially diverse environment, one important question is: Are we better at remembering people from our own race? If biases exist in the ability to remember people, these could have important implications for the nature of cross-racial relations. It could provide, for example, a new perspective on phenomena such as *racial homophily* (McPherson et al. (2001), Fong and Isajiw (2000)), which is the tendency to form friendships and social ties with people from a similar race. And it could also provide a new perspective on racial discrimination.

One of the most prominent and widely used key to a person's identity is her face. In fact, neurological studies find that a large part of our brain is dedicated exactly to the highly complex task of face recognition.[1]. A robust empirical finding is the presence of an "own-race bias" in face recognition - the fact that people are generally better at rembering faces from their own race than other races (see Meissner & Brigham, 2001; Slone, Brigham, & Meissner, 2000 for reviews). This cognitive bias seems prevalent among all racial groups (Ng & Lindsay, 1994; Teitelbaum & Geiselman, 1997), although the evidence suggests the effect is most pronounced for Caucasians viewing members of racial minority groups (Meissner & Brigham, 2001).[2].

[1] See Duchaine (2008) for a review.

[2] Psychologists have repeatedly stressed the implications of such bias for eyewitness identification in the criminal justice system. Eyewitness misidentification appears to be *the* single greatest cause of wrongful convictions in the US, playing a role in more than 75% of convictions overturned through DNA testing. Source: "The Innoncence Project"; a national litigation and public policy organization dedicated to exonerating wrongfully convicted people through DNA testing (www.innoncenceproject.com)

The relevant question in the context of social interactions, however, goes beyond the issue of face recognition. The recall of a face is only useful to the extent that it is correctly mapped to the information relevant to the social interaction. To be specific, denote by x_i the information relevant to the social interaction with person i. x_i could for example capture how productive the person is, how enjoyable her company is or how trustworthy she has been. i is the identity and a face or name are markers uniquely mapping i to x_i. The existing experimental evidence in psychology focuses on face recall, that is, subjects are asked whether they recall having seen person i before. But the relevant question in many social interactions is whether people can correctly remember the mapping between identity and information. As we will see, in our sample, 56% of subjects report being able to remember faces but cannot recall the context. Thus, it is crucial to study the process of joint recall of identity and information in order to draw inferences relevant to a social context. The goal of this paper is to provide a first set of experimental results on this process and discuss a range of applications where such biases may play an important role.

Despite its obvious relevance for many economic and social interactions, there is surprisingly little work studying this type of biases. In Economics, Fryer and Jackson (2008) propose a model showing how own-race biases in memory could have discriminatory implications.[3] The mechanism is similar in nature to those present in contexts of imperfect information and noisy signals in the spirit of models of statistical discrimination.[4] In their model

[3] There is more prevalent theoretical work on the role of bounded memory in economic transactions. See Mullainathan (2002) and Mullhainathan et al. (2008) for recent contributions and review of the relevant literature.

[4] Beyond the seminal work by Arrow in 1973, a number of papers (Aigner and Cain (1977), Cornell and Welsh (1995), Lundberg and Startz (2007)) propose models where racial discrimination arises through biases in the technology used to *screen* people and

people are sorted into categories and each category has a prototype - a unique vector of attributes. People keep track of the variation in attributes *across* categories but not within a category. People sorted into the same category are blended together. They argue that minority groups may be sorted into coarser categories than majority members because they are less likely to be involved in frequently repeated interactions. This leads to positive discrimination for those at the bottom and negative discrimination for those at the top. In Psychology, the most relevant studies are those that study the recall of *associations* between faces and information ("*Who said What?*").[5] The seminal work in that area is Taylor et al.(1978), who study how participants recall the contents of interactions between people of mixed gender and race. They show that participants are more likely to misattribute statements of people of the same race than different races.

The question we ask here is how do people record and retrieve *payoff-relevant* information of the type "Individual i has productivity y". The fact that the information is payoff-relevant is important because it may affect how it is recorded in the first place. To the best of our knowledge, this is the first study providing evidence on how people memorise identities and payoff-relevant information. We know very little about how memory may operate in the presence of payoff-relevant information and whether cross-racial biases arise as well in this context.

This study is designed to serve as a benchmark and is conducted in a simple and neutral environment. Participants see people, learn information

evaluate their productivity. They conjecture that employers may be better at *evaluating* the productivity of workers from their own race than other races and this implies that the information is noisier for people of other races.

[5] I am very grateful to Oliver Curry for informing me about this literature

about them in a controlled manner and then make a decision requiring recall. I compare the accuracy of recall within and across races and provide a clean measure of the discriminatory implications of memory limitations. The focus of this study is on faces as identifiers because it is the most immediate visual clue that people use to identify others. Often a face is all we rely upon to identify someone. Local merchants typically know their customers only by face, not by name or any other identifier. Professional conferences, business social events, interactions between teachers and students, or between students themselves, are all examples of environments with repeated encounters and where facial recognition is an essential technology used to identify others.

The experiment proceeds as follows. Participants are presented with a sequence of 24 pictures of East Asian and white people - whom I call "candidates." Each candidate is associated with a value. In the second stage they are asked to select 8 people among the candidates they have seen. They see pictures again but they do not see the value. Their earnings depend directly on the values of the candidates they select. This setting shares similarities with a job recruitment process whereby, for example, an employer first sees a number of potential candidates and needs to decide whom to call for a second interview. Importantly, these values are exogenously assigned by a random technology and there is no difference in the distribution of values across racial groups.

We have two treatments. In a first treatment, participants see an equal number of candidates of each race. If there is an own-race bias in memory, we would expect to see less mistakes involving candidates from the same race as the decision maker. In a second treatment, participants only see two East-Asian candidates, a man and a woman. The idea is that race might enhance identification when there are few people of a particular race (in a crowd of white people, an East-Asian man will stand out). Race and gender have been found to be prime characteristics encoded about others (Montepare and Opeyo (2002)). It is a *distinctive* attribute - and distinctiveness has been found to enhance re-identification (Shepherd et al. (1991), Tibbetts and Dale (2007), Valentine (1992)).

We find a clear own-race bias in recall in the first treatment. This bias is most pronounced among white participants, who recall high-value candidates from their own race much more accurately than they recall high-value candidates from other races (66% of the white candidates with a value among the eight highest values are selected against 50% for East-Asian candidates). Second, participants select an equal number of candidates of each race - 4 on average. Thus, we find evidence of positive and negative racial discrimination at the same time. Higher ranked candidates of the other race are less likely to be selected, but the opposite is true for lower ranked candidates. In the second treatment I find that decisions involving East-Asian candidates are significantly more efficient than those involving an equal number of candidates of each race, and this is true both for white and East-Asian participants.

The study focuses on Caucasian White and East-Asian groups. In the UK, East-Asians are very distinct from other ethnic minorities in their socio-economic profile; it is a group that is perceived to do well and there is a widespread belief that they suffer little or no discrimination. According to the UK 2001 census, Chinese people are substantially more educated than other groups; with 33% of men and 29% of women holding a university degree, against 18% and 16% of British white men and women and 11% and 14% of Black Caribbean men and women. Yet East Asian groups remain heavily segregated. They are heavily concentrated in some sectors of the economy (e.g. fast-food industry), highly-qualified chinese people are more likely to be unemployed and the proportion of self-employment is substantially higher among Chinese than any other ethnic minority. These notable facts make it a particularly interesting group to study in the context of discrimination.

The rest of the paper is structured as follows. Section 2 presents the experimental design, Section 3 presents the data and summary statistics and Section 4 the results. Section 5 presents a calculation of the magnitude of losses incurred due to discrimination. I conclude in Section 6, by discussing relevant applications and new research questions.

2 Experimental Design

The experiment is structured in 3 stages. The first stage is a *viewing stage* where participants view an automated sequence of 24 pictures. Each picture is shown for 3 seconds and then the screen moves to the next picture.[6] Each picture appears together with its value. The second stage is a *selection stage* – participants see the same 24 candidates again, but without their associated values, and are asked to select 8 candidates that will enter a lottery (in a third stage). The lottery determines their final earnings. Importantly, pictures appear in a different sequence, and from a different angle. The sequence is not automated at this stage and participants can go back and forth between pictures for 3 minutes. In case of incomplete selection (less than 8 candidates selected by the subject), candidates are selected at random among the remaining non-selected candidates to complete the selection (it turns out that in the experiment all participants without exception se-

[6] The choice of number of pictures, time and mix of gender is in line with the common practice in psychology studies.

lected all 8 pictures themselves). Note that the sequence of presentation of candidates is randomized for each participant, and for each stage. The third stage is a *lottery*, whereby one of the candidates from the selection in stage 2 is picked at random.[7] The subject's earnings are equal to the value of the picture divided by 10, in British pounds.

A controlled environment presents two major advantages in this context. First, we can strictly limit the consequences of decisions to the decision-maker: the selection decision only affects the subject's payoff and the person who is the object of the decision (the candidate on the picture) does not incur any loss or gain by being selected or not. This rules out a role for other-regarding considerations, such as fairness or willingness to provide benefits to own-group members over others. Second, the payoffs associated with each candidate are fully controlled for and participants are fully informed about these individual payoffs and about the procedure of assignment of payoffs to candidates. This exogenous assignment means that the faces of candidates do not contain any payoff-relevant information. This is important given the recent evidence showing relationships between facial features (and in particular race) and inferences about personal characteristics such as competence and trustworthiness[8] . Here the only cognitive mechanism that can correctly map faces to values is memory.

Pictures of candidates

Pictures of candidates were drawn from a database provided by TAR-

[7]Note that if pictures were added to complete the selection in stage 2, these enter the lottery as well. No matter how many pictures were selected in stage 2 by the subject, there were always 8 pictures entering the lottery. In practice, all participants selected 8 pictures exactly.

[8]see Todorov et al. (2005), Eckel and Petrie (2008), Rule and Ambadie (2008), Todorov and Duchaine (2008) and Duarte et al. (2009).

RLAB [9] These pictures show only the face of the person. Pictures were selected according to a number of criteria to guarantee homogeneity in shooting conditions[10]. The database contains 11 East Asian men, 15 East Asian women, 35 Caucasian men and 44 Caucasian women.

For each subject, a set of 24 candidates was randomly chosen. Two pictures of each candidate from a different angle were randomly chosen (one used in the viewing stage and the other used in the selection stage). This ensures that the task involves face recognition rather than picture recognition and prevents participants from using other cues than the face itself to remember the person. The sequence of viewing is determined randomly for each participant, for both the viewing and selection stages.

A picture of a mixed race person has been chosen to illustrate the instructions (see instructions in the appendix).

The values of candidates

A unique value was pre-assigned randomly to each candidate. The values correspond to random draws from a discrete normal distribution truncated at 10 and 70, with mean 40 and standard deviation 15. This range has been chosen so that all values have 2 digit numbers. The choice of the normal distribution is motivated by the relatively low presence of extreme values in many real applications - extremes that are often those one would like to remember in taking decisions.

[9] Face-Place Face Database Project (http://www.face-place.org/); Copyright 2008, Michael J. Tarr. Funding provided by NSF award 0339122.

[10] All images were extracted from standard digital video (720x480), with the background removed and the faces scaled to be roughly equated in terms of size. The pictures were selected according to the following parameters: Race (East-Asian and Caucasian White); shave/stuble no make-up; no beards or mustache; no facial hair or visible make-up; no glasses; natural hair (no wig); neutral affect; orientations: 0°, 15° left, 15° right, 30° left, 30° right.

The instructions have been written carefully to inform participants in detail about the procedure of assignment of values to candidates (see Appendix). In particular, the discrete normal distribution was represented graphically with a *roulette wheel,* illustrating the differences of probabilities of occurence corresponding to each number from 10 to 70 by differences in areas. Participants are also informed that the procedure is repeated for each picture independently and there is a visual illustration of the procedure of assignment. The instructions do not mention race at any point.

Treatments

Two treatments are conducted:

Treatment 1 - "Equality treatment" - Choice sets are composed of an equal number of East-Asian and Caucasian candidates: 6 men and 6 women of each race.

Treatment 2 - "Majority-Minority treatment" Choice sets are composed of 22 Caucasian candidates (11 men and 11 women) and 2 East-Asian candidates (1 man and 1 woman).

3 Theoretical framework

This section outlines a simple model of bounded memory (inspired by Fryer and Jackson (2008)) and discuss the implications of alternative models of discrimination based on preferences or stereotypes. When we turn to the data analysis, we will then derive predictions from these models, predictions that we will take to the data.

3.1 A simple model of Bounded Memory

The simplest way to describe the decision problem is as follows. In a first stage, participants see a sequence of candidates with values x_i, $i = 1, ...24$. Candidates have a number of observable facial attributes y_i that she will observe again at the time of making the decision. The distribution of x is independent of y. Participants know the distribution of x (i.e. they know it is normally distributed with mean 40 and standard deviation 15) and they know that x is independent of y.

Decision makers form categories of people based on these attributes: for example, according to colour of the hair or eyes, attributes that she will observe again at the time of making the decision. Suppose there are K equal size categories for the own race and J equal size categories for the other race, with $J < K$. That is, categories are finer for candidates from one's onw race than for candidates for the other race. The decision maker cannot distinguish between people sorted into the same category.

In addition to keeping track of the observable attributes uniquely identifying each category, we assume that the decision maker keeps track of the expected value corresponding to each category. The observable attributes identifying the category constitute the "key" to retrieve payoff-relevant information. Thus, categories will have different expected values.

Optimal strategy. In this framework, the optimal strategy consists in first choosing candidates sorted into the category corresponding to the highest expected value. If fewer than 8 candidates belong to this category, all candidates of this top category should be included. Then the decision maker should compare the expected value of the second highest category of the own race with the expected value of the highest category of the other race and select candidates from the category that has the highest expected value among the two. The procedure should be repeated up to the point where 8 candidates have been selected.

This simple model departs slightly from Fryer and Jackson insofar as we do not assume that values are used to form categories. They do not make a distinction between attributes that are observable or not observable at the time of making the decision. People observe a set of attributes (including values) and form categories based on these attributes. That model would imply that the decision-maker can distinguish between twins with different values if these values are sufficiently far apart that they are sorted into two different categories. Here it is more intuitive to assume that twins with different values - no matter how far apart these values are - cannot be distinguished by the decision maker.

3.2 Alternative models of discrimination

Here we briefly outline the implications of alternative models of discrimination in the context of the experiment.

A first alternative model of discrimination is one that would allow a role for preferences or stereotypes. Becker's (1961) model of taste-based discrimination conjectures that discrimination arises through preferences for interacting with people of one's own racial group. Preferences and tastes should have a limited role here by design since there is no "interaction" between the decision-maker and the person who is involved in the decision (the candidate on the picture). There is no other benefit from selecting a particular picture other than than the economic value attached to that picture. Also, decisions only affect the decision-maker and not the candidates, such that other-regarding considerations should not affect decisions. Yet participants come to the lab with a real world experience. Preferences may not be totally turned off and may affect decisions even in this environment. Preferences for people of similar race are equivalent to attributing a lower value to people of the other race. That is, a preference-based model would result in a distribution shift of the other-race productivity distribution.

A second alternative model of discrimination is a model of rational or statistical discrimination (Phelps, 1972 and Arrow, 1973), whereby discrimination arises through negative stereotypes regarding the other race. Stereotypes can affect decisions in environments where there is uncertainty about the value of a person. Here the environment rules out this uncertainty by design: Participants are perfectly informed about the values, and about the distributions of these values. But of course, they could bring stereotypes formed outside the lab into the lab. However, if memory is perfect, there is no reason for them to substitute these stereotypes to the perfect information they have about the candidates. It is possible though that stereotypes play a role in the presence of imperfect memory. Participants could use stereotypes to fill in memory gaps. Again, the implications of such negative stereotypes are identical to the implications of a preference bias. They also come down to attributing a lower value to people of the other race.

I now derive a number of predictions based on the models outlined above, predictions that I take to the data. The predictions relate to the probability of entering the selection and the probability of misallocation (i.e. mistakenly excluding or including a candidate from the selection).

3.3 Predictions

3.3.1 Probability of entering the selection

Let us start with the probability of entering the selection in the Treatment Equality. With perfect memory, we would expect on average 4 candidates of each race to enter the selection ($\frac{1}{3} \times 12$). With imperfect memory, the difference in probability of selection across races will depend on how fine the categories are for each of the races. For example, suppose participants form categories of 3 people for the own race and categories of 4 for the other race. They should then pick the three people from the highest own-race category, then the four people from the highest other-race category and pick one candidate at random from the second higher own race category. That means that she will have 4 candidates of each race on average. Another example is one where there would be categories of 4 people for the own race and categories of 6 for the other race. The optimal strategy would be to pick the 4 people from the highest own race category and then 4 people at random from the highest other race category, which will have on average a higher expected value than candidates in the second highest category of the own race. One could construct alternative examples, where the optimal composition holds more or fewer candidates of each race. The important point is that bounded memory does not necessarily imply that people should select a *lower* number of candidates of the other race.

In the Treatment Minority, we would expect that the candidates from the minority are more likely to be sorted uniquely into categories. Again, that does not mean they should be more or less likely to be selected.

This leads us to the following prediction:

Prediction 1 - The probability of selecting a candidate from one's own race is not necessarily higher than the probability of selecting a candidate from the other race.

This prediction is in contrast with the alternative models of discrimination, which both predict that the probability of selecting a candidate from one's own race should be *higher* than the probability of selecting a candidate from the other race. Note that the absence of discrimination (of any form) should lead to an equal number of white and East Asian candidates being selected.

3.3.2 Probability of misallocation

The bounded memory model has important predictions regarding the prediction of misallocation, which differ across treatments. We define misallocation as the incorrect selection of a candidate who does not belong to the Top 8 or as the incorrect exclusion of a candidate who does belong to the top 8. The predictions are as follows:

Prediction 2a - In the Treatment Equality, the probability of misallocation is higher for other race candidates than for candidates of the same race.

This is a direct implication of the categories being coaser for the other race candidates in comparison to candidates of the same race. In fact, since the probability of entering the selection should be identical for all candidates sorted into the same category, the coarser the categories the flatter the relationship between values and probability of being selected. If there is no memory and just one category, then the probability of selection is identical for all candidates and the probability of selecting a top 8 candidate is 1/3.

That is, we expect one third of the Top 8 candidates to enter the selection by chance. In the case of perfect memory, only the candidates belonging to the top 8 should be selected while those outside the top 8 should never be selected. In case of imperfect memory and coarse categories, then there is a positive probability that a candidate with a value belonging to the 8 eighest values is sorted into the category with low expected value, and similarly, there is a positive probability that a candidate who does not belong to the Top 8 is sorted into a category with high expected value. This is why the relationship becomes flatter.

For example, it could be that categories for the own race candidates include 4 people, while those for the other race candidates include 6. In that case, it would indeed make sense to pick an equal number of each race, but we would expect more mistakes for the other race candidates than for the own race candidates.

In contrast in the Minority Treatment, the probability that minority candidates are allocated to a unique category is much higher, since there are only two candidates from the minority race and they are of both gender. Since gender and race are prime attributes recorded in memory, they are likely to be uniquely allocated to a category. Thus, the prediction for the Minority Treatment is the following:

Prediction 2b In the Treatment Minority, the probability of misallocation is higher for candidates of the majority than for candidates of the minority.

The Treatment Minority is useful because we expect it will *reduce* the probability of a misallocation of East Asian candidates for both types of participants (East Asians and Caucasians).

In other words, we expect that top 8 East Asian candidates would be less likely to enter the selection made by Caucasian Participants in the Equality treatment and more likely to enter their selection in the Minority treatment. Such reversal is not predicted by the alternative models of discrimination.

In contrast, the two alternative models of discrimination would predict that candidates of the other race should be less likely to be selected, irrespectively of the treatment and of their value.

Finally, if there was not discrimination at all, we would not expect any systematic differences in the probability of misallocation across races.

4 Data

The experiment was conducted at the laboratory of the Nuffield Centre for Experimental Social Sciences (Oxford) in September and November of 2009 and March 2010. The study was conducted with ethical approval of the Ethics Comittee of the Centre of Experimental Social Sciences at Oxford.

116 participants were recruited: 61 Caucasians and 55 East Asians. Invitations were sent by e-mail to participants in the pool with East Asian and British last names, without mentioning race or ethnicity[11]. Sessions were also relatively small in size (maximum 15 participants at a time) and the split across race was not equal for each session. The experiment lasted for 30 minutes in total, including a post-experimental questionnaire asking information about ethnicity, age, occupation, country of birth, age of arrival in the UK and a self-assessment of ability to remember faces of people in general. Participants received a £4 show-up fee and an additional payment depending on the performance in the memory task, bringing the total

[11] Participants were recruited using ORSEE (Greiner, 2004). The invitation asked for participants between the ages of 18 and 30. We have excluded participants above 50 years old from the analysis (n = 3).

payment to £8.60 on average.

Tables 1 and 2 present summary statistics of the participants. We had around 30 participants on average from each race in each treatment. The average age is around 25, and we had about 50% women among Caucasian White participants, and around 40% among East Asians.[12]

In terms of average score, the mean value of the selected candidates is 48 (S.E. .47), significantly higher than the benchmark for chance (40, $P = .000$). The benchmark is 33% for random selection and 100% for perfect memory. Overall, participants did significantly better than chance, with a proportion of correctly allocated candidates equal to 61% ($P = .000$)[13].

We now turn to the main analysis of the selection decision.

5 Results

Let us start with the first prediction - the probability of selection in the Treatment Equality. We find that on average Caucasian whites and East Asians select an equal number of Caucasian white candidates (4.1 for Caucasians and 4 for East Asians ($P = 0.93$) in the Treatment Equality[14]. We also find no systematic difference in Treatment Minority. Caucasian Whites and East-Asians select the same number of white candidates (7.33) and the same number of East-Asian candidates (0.66) on average ($P = .57$).

These results do not fit the alternative models of discrimination, while they are consistent with the model of bounded memory. Of course, the

[12] It is worth pointing out in addition that none of the East-Asian participants was born in the UK and the average time spent in the UK is 1.15 years.

[13] Two-sided Mann-Whitney tests with the individual proportions as independent observations.

[14] All reported tests correspond to post-estimation F-tests based on linear regression estimates pooling all choices made by the 116 participants and clustering the standard errors by faces.

model of bounded memory would have been consistent with any pattern, so the test for the bounded memory model is not strong.

We turn to the second set of predictions, regarding the probability of misallocation. I first present results for the Treatment Equality. Figure 1 shows the average proportion of top 8 candidates selected by racial groups of participants and candidates. The selection shows a substantial own-race bias[15]. Conditionally on being in the top 8, the average probability that an East Asian candidate enters the selection of white participants is .50 against .66 for white candidates ($P = .001$). The opposite trend is observed for East-Asian participants: the probability of entering the selection conditionally on being in the top 8 is .65 for East-Asian candidates and .56 for white candidates, but the difference is not statistically significant ($P = .25$).

Next, I estimate a probit model for the selection decision (Table 3). Again, we find that, conditional on being in the top 8, there is no significant difference in the probability of entering the selection according to race. We also find that the probability of entering the selection conditionally on being in the top 8 is higher on average, but for white participants, we find that East Asian candidates from the top 8 are *less* likely to be selected. Since they select an equal number of candidates of each race, these results imply that positive and negative discrimination occur at the same time: high-ranked candidates from the other race are less likely to enter the selection, but this benefits lower-ranked candidates (that do not belong to the top 8). Again, this prediction is in line with the model of bounded memory, but not with the alternative models of discrimination.

[15] All reported tests correspond to post-estimation F-tests based on linear regression estimates pooling all choices made by the 116 participants and clustering the standard errors by faces.

East Asian participants do not exhibit an own race bias. That is, we find no evidence of an own race bias in the probability of selection or in the probability of misallocation. In that respect, it is worth pointing out though that East-Asian participants are living in the UK, and therefore might be a selected sample with respect to the ability to distinguish white candidates.

We now turn to the second treatment - Treatment Minority. Here the composition of facial stimuli only includes 2 East-Asian candidates, a man and a woman. The hypothesis tested here is whether race and gender can enhance identification when they are scarce attributes.Table 4 reports probit estimates testing for treatment effects for white and East Asian participants separately. The results show a substantial and significant improvement in the probability of entering the selection for top 8 East Asian candidates. The treatment effect is comparable in magnitude for both white and East-Asian participants.

These results are in line with the predictions of the bounded memory model, but not with the predictions of the two alternative models of discrimination. We see no significant differences in the probability of entering the selection across race in any of the treatments, and we find that the probability that a top 8 candidate from the other enters the selection made by White Caucasians is lower in the Treatment Equality and higher in the Treatment Minority. This reversal is not consistent with the alternative models of discrimination.

6 A measure of discrimination

Since we have detailed information about the values, we can calculate more precisely the discriminatory implications of cognitive limitations in the treatment *Equality*. That is, even though the candidates do not correspond to people actually incurring the consequences of the decisions, we can calculate gains and losses for each racial group of candidates. Specifically, we can distinguish between three types of outcomes:

(A) candidates with a value *within* the top 8 and *included* in the selection

(B) candidates with a value *lower* than the top 8 and *included* in the selection

(C) candidates with a value *within* the top 8 and *excluded* from the selection

I calculate the total value associated with each of these three categories and the total value corresponding to the optimal selection (sum of the values associated with candidates within the top 8) for each racial group of participants and candidates. The ratio between the first sum (corresponding to the actual selection) and the second (corresponding to the optimal selection) provides an indication of the relative gains and losses for each racial group of candidates due to the decisions of each racial group of participants. Then I calculate for each racial group the ratio of the sum of values of candidates included in the selection (whether they belong to the top 8 or not) and the sum of values associated with the candidates included in the top 8. This gives a measure of the overall losses and gains incurred by each racial group.

The results are presented in Table 5. Not surprisingly, the losses associated with exclusions of top 8 candidates are larger than the gains associated with inclusions of candidates outside the top 8. As a group, the white candidates realize 94% of their potential when they are in the choice sets of white participants, while they suffer a 23% loss when they are in the choice sets of East-Asians. The relative losses are comparable in magnitude for white and East-Asian candidates that are in the choice sets of East-Asian participants (20% and 8% respectively). Thus, overall, the cognitive biases identified here lead to an overall *worse treatment* on average of other racial groups, but this effect is particularly pronounced for East-Asian candidates in the choice sets of white participants.

7 Discussion and conclusion

This study highlights cognitive limitations in the recall of identities and payoff-relevant information and shows that these limitations are more pronounced across races than within race. These limitations lead to unequal treatment across races: negative discrimination at the top of the value distribution and positive discrimination for those lower in the distribution. This experiment is a first attempt to shed light on possible implications of cognitive limitations in identification for social interactions. As discussed in the introduction, these biases may have salient implications for the structure of cross-racial relations. Like a language, technologies used for individual recognition and identification enhance efficiency and seem to be developed very early on. In this context, understanding how these technologies develop and how they shape social relations could deepen our understanding of racial integration.

The current study provides a benchmark and aims at drawing attention to the phenomenon and its relevance for economic interactions. We think of three classes of applications where the phenomenon is likely to be relevant. A first class of applications relates to the literature on network formation. Any social tie starts with a first interaction and then requires re-identification. It is well-known that social networks are heavily biased towards one's own race - a phenonomenon known as *racial homophily* (McPherson et al. (2001), Fong and Isajiw (2000)). The literature in sociology and economics has mainly focused on two possible explanations for this phenomenon: one is *preferences* (relationships within race are valued more than across race). The other is the frequency of meeting or *opportunities* (people may be more likely to meet and interact with people of similar race) - see Currarini et al. (2009) for a model of network homogamy taking these two forces into account.[16] Limitations in the ability to remember people and the resulting uncertainty in the value of continuing a relationship have not been considered at all in this literature.

The second class of applications regards reputation-based mechanisms, relying on the repeated character of interactions and memory, such as trust and reciprocal altruism (Axelrod, 1981). These mechanisms are often very demanding in the information-structure required to sustain cooperation - in terms of players' identities and records of past behaviour. Basu et al. (2009) show that facilitating recordkeeping with an external support (e.g. allowing participants to take notes) enhances cooperation. Cognitive limitations in cross-race re-identification may therefore play a role in hampering cooperation across races and possibly compromising the use of these mechanisms to enforce cooperative behavior across races.

A third class of applications is the use of recommendation, which plays a large role in school and employment selection systems. The issue of recall of information beyond recorded information - typically the main added value of a recommendation - is likely to have salient implications in such settings. University lecturers are very familiar with the difficulties of writing a recom-

[16] Interestingly, the homophilic feature of networks seems to appear very early on. Mollica et al. (2003) study friendship ties among MBA newcomers and find that friendship networks are already segregated 6 weeks after the beginning of the academic year.

mendation letter for students from large classes. Cognitive biases in memory could have implications for hiring and placement decisions and constitute an alternative mechanism for racial discrimination.

These three examples of applications have been extensively studied in economics. I propose here a novel angle and argue that cognitive mechanisms may play a role in these phenomena and explain possible racial biases associated with them.

There are many ways forward from here. An important first extension is to study how these biases vary across ethnic groups and social environments.

Second, the experiment conducted here is in a non-strategic setting. One obvious extension is to study the implications of such biases in the context of repeated strategic interactions - for example, in a trust game. One could investigate whether these biases lead to differences in the sustainability of cooperation within and across races. Moreover, in a strategic setting, subjective beliefs about other players' abilities to remember may play a role as well.[17] Why bother being cooperative if the other players will not remember?

Also, the study focuses on faces as identity markers. Faces are an obvious starting point because they play a key role in many social interactions. A face is generally unique to an individual and cannot be altered or copied easily (Gambetta and Bacharach (2001)). But we use other technologies to identify people and natural extensions to this study would be to investigate whether similar biases arise with other identification technologies - names being the other obvious candidate. Like faces, names typically have a cultural imprint, such that similar biases may arise. One could even investigate whether multiple identifiers (faces and names, for example) aggravate the

[17] This point was kindly brought to my attention by Robin Cubitt

biases or not.

Finally, the environment I consider is very simple. The payoff-relevant information is captured by a single value. It would be interesting to study recall in a more complex environment, where the relevant information takes a more complex structure. Also, the bridge between the experimental task and real world applications is long. One could ask whether these biases subsist in settings where people interact with each other for a longer period and may also have access to external support to memory.

In my opinion, these questions deserve attention and careful analysis, in the laboratory and also possibly in the field. The main challenge of field experiments on this phenomenon is to find reliable measures of payoffs and possibly even measures of perceptions of these payoffs.

These challenges are left for future research.

References

Aigner, D.J. and G.G. Cain (1977), Statistical theories of discrimination in labor markets, *Industrial and Labor Relations Review* 30 (2), 175-187.

Axelrod, R. and W.D. Hamilton (1981), The Evolution of cooperation, *Science* 211, 1390–96.

Arrow, K. (1973), The Theory of discrimination, in *Discrimination in Labor Markets*, O. Ashenfelter and A. Rees, eds. Princeton, NJ: Princeton University Press.

Basu, S., J. Dickhaut, G. Hecht, K. Towry and G. Waymire (2009) Record-keeping alters economic history by promoting reciprocity, *Proceedings of the National Academy of Science* 106(4), 1009-14.

Becker, G. S. (1961), The economics of discrimination, 2nd Ed. Chicago: University of Chicago Press.

Cornell, B. and I. Welch (1996), Culture, information, and screening discrimination, *The Journal of Political Economy* 104(3) 542-571

Currarini, S., M.O. Jackson & P. Pin (2009), An economic model of friendship: Homophily, minorities, and segregation, *Econometrica*, 77(4), 1003-1045.

Duarte, J., S. Siegel and L.A. Young (2009), Trust and Credit. Available at SSRN: http://ssrn.com/abstract=1343275

Duchaine, B. (2008). Review of Evolutionary Cognitive Neuroscience, *Quarterly Review of Biology* 83-100.

Duchaine,B., Yovel,G. (2008). Face Recognition. in Basbaum,A., R. Hoy, A. Kaneko, G. Shepherd and G. Westheimer (ed.) *The Senses: A Comprehensive Reference*. Amsterdam: Elsevier, 1st edition, 329-358. ISBN: 0-12-639482-2.

Eckel, C. and R. Petrie (2010), Face Value, *American Economic Review* forthcoming.

Fong E and W.W. Isajiw (2000), Determinants of friendship choices in multiethnic society, *Sociological Forum* 15, 249–269.

Fryer, R. and M.O. Jackson (2008), A Categorical Model of Cognition and Biased Decision Making, *The B.E. Journal of Theoretical Economics* (Contributions) 8 (1), Article 6.

Gambetta, D. and M. Bacharach (2001), Trust in signs, in K. Cook (ed.) *Trust and Society*, New York: Russell Sage Foundation, 148-184.

Golby A.J., Gabrieli J.D., Chiao J.Y., Eberhardt J.L. (2001), Differential responses in the fusiform region to same-race and other-race faces. Nature Neuroscience 4, 845–850.

Greiner, B. (2004), The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics. University of Cologne, *Working Paper Series in Economics* 10.

Hanley, J.R. and G. Cohen (2008), Memory for people: Faces, names and voices in Cohen, Gillian and Conway, Martin A., *Memory in the real world*, Psychology Press, UK

Kurzban, R., J. Tooby and L. Cosmides (2001). Can race be erased?: Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences*, 98(26), 15387-15392

Levin D.T. (1996), Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22, 1364–1382.

Lavrakas P.J., Buri J.R., Mayzner M.S. (1976). A perspective on the recognition of other race faces. *Perception & Psychophysics* 20:475–481.

Lundberg, S. and R. Startz (2007), Information and racial exclusion, *Journal of Population Economics* 20, 621–642

Maclin O.H. and R.S. Malpass (2003), The ambiguous-race face illusion. *Perception* 32, 249–252.

McPherson M, L. Smith-Lovin and J.M. Cook (2001). Birds of a feather: homophily in social networks. *Annual Review of Sociology* 27, 415–444.

Meissner C and J. Brigham (2001), Thirty years of investigating the own-race bias in memory for faces. *Psychology, Public Policy, and Law.*7, 3–35.

Mollica, k.A., B. Gray and L.K. Treviño (2003), Racial homophily and its persistence in newcomers'social Networks, *Organization Science*14 (2), 123-136.

Montepare J.M., Opeyo A. The relative salience of physiognomic cues in differentiating faces: A methodological tool, *Journal of Nonverbal Behavior* 26, 43–59.

Mullainathan, S. (2002), A memory-based model of bounded rationality, *The Quarterly Journal of Economics* 117(3), 735-774.

Mullainathan, S., J. Schwartzstein and A. Shleifer (2008), Coarse thinking and persuasion, *The Quarterly Journal of Economics* 123(2), 577-619, 05.

Ng W, Lindsay R. (1994) Cross-face recognition: Failure of the contact hypothesis. *Journal of Cross-Cultural Psychology* 25, 217–232.

Phelps, E. (1972), The statistical theory of racism and sexism, *The American Economic Review* 62(4), 659-661

Rhodes G, Brake S, Tan S, Taylor K. (1989), Expertise and configural coding in face recognition, *British Journal of Psychology* 80:313–331

Rule, N. O. and N. Ambady (2008). The face of success: Inferences from chief executive officers' appearance predict company profits. *Psychological Science*, 19, 109-111.

Shepherd, J.W., F. Gibling and H.D. Ellis (1991), The Effects of distinctiveness, presentation time and delay on face recognition, *European Journal of Cognitive Psychology 3,* 137-45.

Tanaka J., Kiefer M., Bukach C.M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study, *Cognition* 93:B1–B9

Taylor, S., S. Fiske, N. Etcoff, and A. Ruderman (1978) Categorization and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology* 36, 778-793.

Tibbetts, E. A. and J. Dale (2007), Individual recognition: It is good to be gifferent, *Trends in Ecology and Evolution* 22 (10), 529-37

Todorov, A., A.N. Mandisodza, A. Goren and C.C. Hall (2005), Inferences of competence from faces predict election outcomes. *Science*, 308, 1623–1626.

Todorov, A. and B. Duchaine. (2008). Reading trustworthiness in faces without recognizing faces. *Cognitive Neuropsychology* 25, 395-410.

Valentine, T. (1992), Towards an exemplar model of face processing The effects of race and distinctiveness, *The Quarterly journal of experimental psychology. A, Human experimental psychology* 44 (4) 671-703.

Table 1: Treatments and number of participants

	Treatment Equality	Treatment Minority
Caucasian White participants	32	29
East-Asian participants	28	27

Table 2 - Participants summary statistics

	Caucasian White participants		East-Asian participants	
	Equality	Minority	Equality	Minority
Age	25.3	24.7	25.6	25.3
	(5.4)	(6.2)	(4.4)	(5.4)
Share women	50%	50%	38%	41%
Mean score	47.6	50.3	48.2	45.4
	(5.4)	(12.1)	(4.6)	(11.5)

Table 3 - Probability of entering the selection - Treatment "Equality"

	White part.	EA part.
other race	-.065 (.044)	.008 (.052)
top 8	.486 (.048)***	.458 (.053)***
other race × top 8 (i)	-.192 (.054)***	-.085 (.069)
n.obs	768	672
n. participants	32	28
Pseudo R^2	.13	.14

Probit estimates - Marginal effects - Standard errors clustered by candidate

Table 4 - Probability of entering the selection - Treatment effects

	White subj.	EA subj.
Asian candidate	-.046 (.043)	.017 (.050)
Top 8	.485 (.048)***	.373 (.051)***
Asian candidate × Treatment minority	-.017 (.032)	.027 (.046)
Asian candidate × top 8	-.171 (.052)***	.103 (.082)
Top 8 × Treatment minority	.016 (.069)	-.042 (.062)
Asian candidate × top 8 × Treatment minority	.255 (.127)**	.334 (.149)**
n.obs	1464	1320
n. participants	61	55
Pseudo R^2	.16	.13

Probit estimates - Matginal effects - Standard errors clustered by candidate

Table 5 - Losses and gains due to cognitive limitations
Share in Potential Realization

	White participants (1)	East-Asian participants (2)	Z-test (1)=(2) P-value
Included - Top 8 (A)			
white top 8 selected	.69	.56	.07
East-Asian top 8 selected	.52	.65	.03
Included - Outside Top 8 (B)			
white not top 8 selected	.25	.21	.55
East-Asian not top 8 selected	.28	.27	.91
Excluded - Top 8 (C)			
white top 8 not selected	.31	.44	.07
East-Asian top 8 not selected	.48	.35	.03
Total (A)+(B) white candidates	.94	.77	
Total (A)+(B) East-Asian candidates	.80	.92	