

Belot, Michèle; Schröder, Marina

Conference Paper

Remember Me? The Role of Gender and Racial Attributes in Memory

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2022: Big Data in Economics

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Belot, Michèle; Schröder, Marina (2022) : Remember Me? The Role of Gender and Racial Attributes in Memory, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2022: Big Data in Economics, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/264081>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Remember me? The role of gender and racial attributes in memory^{*}

Michèle Belot[†] and Marina Schröder[‡]

Abstract

Remembering people is at the core of many social and economic relationships. In this paper, we present evidence from two experiments showing systematic biases in the way we remember people. The first experiment examines memory in a real professional setting (Academia). Conference participants were asked to recall ‘who presented what’ a month after attending research conferences. The second experiment is a controlled computer-based version of the field: Participants are shown pictures of people (drawn from a picture database), matched with the title of a paper. In the second experiment, we exogenously vary the relative shares of women and non-white individuals. We find evidence that women and, to a lesser extent, ethnic minorities are more likely to be recalled in settings where they are in a minority. In contrast, they are more likely to be confused with each other when they are not a minority. These findings are in line with the theory of categorisation proposed by Fryer and Jackson (2008) and with a distinctiveness effect that has been identified in previous studies on memory (Slone et al., 2000, Meissner and Brigham, 2001). People with minority attributes appear to be categorised according to these attributes, and are “blended together”. In settings where there are few minorities, recall is enhanced. In settings where there are more of them, this leads to confusion. We conjecture that these biases in remembering could have important implications for the formation of professional networks.

Keywords: discrimination, gender, race, memory, experiment

JEL: C93, D83, J7

^{*} We thank seminar participants at the European University Institute, FAIR at NHH in Bergen, the University of Edinburgh, Cornell, Cologne, and University of Hannover and conference participants at the EALE and the annual meeting of the GfeW. We are especially thankful to Tommaso Battistoni for excellent research assistance.

[†] Department of Economics, Cornell University, 109 Tower Road, 404 Uris Hall, Ithaca, New York 14853-2501, mb2693@cornell.edu. ORCID: 0000-0002-6953-1200

[‡] Corresponding author; University of Hannover, Institute of Economic Policy, Königsworther Platz 1, 30167 Hannover, Germany, schroeder@wipol.uni-hannover.de, ORCID: 0000-0002-3265-6317

Introduction

Remembering people plays a key role in many social contexts and in the labour market in particular. Remembering is a necessary condition for forming social ties, and social ties are likely to matter for labour market careers and performance. It is well known for example that social networks are a major channel used to find jobs (Calvó-Armengol and Jackson, 2004). Thus, biases in remembering are a likely source for discrimination of minorities (Belot, 2015). One question that has received little attention so far is whether there are systematic biases in memory along attributes such as gender or race. When we meet someone and later try to recall who that person was, are we more likely to recall that person if he/she belongs to a minority group? Or is it the reverse?

This paper presents unique evidence from two experiments. The first experiment is conducted in a real professional setting context: Academia. We conducted an experiment with participants of two “plenary sessions-only” international conferences in Economics – the first was a conference organized at the University of California, San Diego (UCSD) on Deception, the second was a conference organized at the University of Edinburgh on Search and Matching. We sent an incentivised online questionnaire to all conference participants one month after each conference asking them to remember who presented what, i.e. to match pictures of presenters to titles of papers presented and to identify the presenters (name and institution) based on their picture. We study how recall accuracy depends on specific attributes (gender and race) conditionally on a range of measures of academic achievements, such as publications, work institutions and career path. The second experiment replicates key features of the field but is in a controlled environment. We match pictures of people (from a research database of pictures) to papers at random, and we vary exogenously whether or not women or non-whites are in minority.

Academia is an interesting environment because it shares many characteristics with other labour markets, such as the reliance on social networks for recruitment or activities relevant to promotions (e.g. recommendations) and concerns regarding the success of specific groups of the population such as women and non-whites. There is indeed ample evidence showing that women and non-whites fare less well in Academia compared to equally qualified peers who do not belong to these groups (Kahn, 1993, Blackaby and Frank, 2000, Blackaby et al., 2005, Carter et al., 1999). Notably, there is a belief that social networks may play an important role in explaining differentials according to gender and race; which led leading professional institutions to invest

specifically in improving the social networking opportunities of these sub groups.¹ One notable advantage of Academia relative to other labour markets is, as Blackaby et al. (2005) point out, that it has readily available measures of productivity that are comparable across institutions (such as number of publications, ranking of publications, etc.). That is, it is possible to compare researchers working in different institutions, which is usually very difficult to do in other labour markets.

How does the minority status of an individual effect whether or not an individual is remembered? Fryer and Jackson (2008) propose a theory of how memory may be affected by the minority traits of those who are to be remembered. They propose a model of memory based on limited cognitive resources. Their hypothesis is that we sort people into *categories*, defined as a vector of specific *attributes* (e.g. “white” & “woman”). People sorted into the same category are lumped together and cannot be distinguished from each other. One prediction is that categories may be finer for people for whom the expected value of interaction is high. In contrast, categories may be broader for people for whom the expected value of interaction is low. The expected value of interaction may be determined by an objective frequency of interactions in real life, or may be affected by previous experiences (Enke et al., 2020), stereotypes (Arrow, 1973) or preferences (Becker, 1971). This model of memory predicts systematic differences in the way we remember individuals from minority groups as compared to the majority. We know from research in psychology that race and gender are prime attributes encoded about others (Shephert et al., 1991; Valentine, 1992, Montepare and Opeyo, 2002, Tibbetts and Dale, 2007). That is, these attributes are likely to be used to form categories. These attributes are presumably playing a more important role in remembering individuals from minority groups than individuals from a majority. Thus, individuals with minority attributes are more likely to be confused with each other in settings where there is room for confusion (there are a number of individuals sharing these same minority attributes).

To contribute to the understanding of biased memory, we analyse whether there are systematic differences in recall accuracy depending on the race or the gender of the remembered individual. Are individuals from minority groups more likely to be confused with others who share the same minority attribute (race or gender) in settings where there is room for confusion?

¹ Prominent examples of such objectives are the professional network associations such as the “Black British Academics” or the “American Economic Association Committee on the Status of Women in the Economics Profession.”

The analysis of the field setting (where women and ethnic minorities are in a minority) shows that, controlling for the presenters' academic achievements, female presenters are more likely to be remembered than others: However, the female advantage is only present in the mapping of pictures of presenters to titles, not in recalling their name or institution. Participants may recall the paper was presented by a woman, but may not remember exactly who she was, in line with the broad categorisation hypothesis. The evidence regarding race in the field does not show such clear pattern.

As a second step, we exploit an important feature of the experimental design – the multiple choice format of a subset of questions. In these questions, the correct answer (presenter or title) must be identified among a set of alternatives. We use this feature to test if people are more likely to confuse people of the same race and gender if they belong to a minority group. We find evidence that this is the case: people are more likely to confuse women with each other than they are to confuse men with each other.

The field setting has the limitation that the minorities correspond to a limited number of people, and it could be that our results are driven by idiosyncratic characteristics of these people. We therefore also provide evidence from a complementary controlled on-line experiment replicating key features of the field setting: In a first stage of the controlled on-line experiment, participants see pictures of people matched (at random) with titles of papers in economics. In a second stage, participants are asked to either match a person to one of four titles, or one title to one of four people. The people appearing on pictures are drawn from a picture database and are not real economists. The advantage of the controlled setting is that we can use a larger battery of people, implement a random assignment of titles to people, guarantee that participants have no other information about these people than the information we give them, and we can contrast treatments where there is a balance of people of both gender and race, with treatments where females or non-whites are in minority.

The results from the controlled setting confirm the findings from the field: We find evidence of a positive recall bias for women when they are in minority. We also find that this positive bias completely disappears in the treatment where there is an equal proportion of men and women, as predicted. The pattern is less clear for non-white individuals.

As in the field setting, we also exploit the multiple choice format of questions to test for confusion. We find evidence that women and East Asians are more likely to be confused with

others who share the same gender or race, particularly in settings where they are not in minority, that is, in contexts where there is more room for confusion.

To our knowledge, we are the first to provide evidence on recall biases in a real professional context. We believe these findings have implications for our understanding of the structure of social networks and how they may impact success in real professional markets. In Academia for example, it is clear that recall plays an important role in network formation: Often we rely strongly on memory (who we remember meeting) to come up with names of potential candidates for seminars, workshops or even employment positions. The results of this study raise concerns about the practice of relying on recall in actions, such as seminar invitations, that may affect people's careers.

The rest of the paper is structured as follows. Section 1 presents the experimental design for the field setting and Section 2 presents the experimental evidence collected for the field setting. Section 3 presents the experimental design and evidence from the controlled setting. We conclude in Section 4.

1. Experimental design of Experiment 1 (real professional setting)

We conducted an on-line experiment following two “plenary sessions” economics conferences. The first conference was the “Deception, Incentives and Behavior” conference, which took place in San Diego in April 2012 at the Rady School of Management at UCSD (in the following, we will refer to this conference as the San Diego conference). The second conference was the annual European “Search and Matching Conference” (organised by the professional network SaM – “Search and Matching”), which took place in Edinburgh in May 2014 (in the following, we will refer to this conference as the Edinburgh conference). Both conferences were single-session conferences, i.e. all talks were plenary talks. The presentations lasted for 10 or 20 minutes in San Diego and for 30 minutes in Edinburgh. One important advantage of the single session setup is that participants do not self-select across sessions, which would complicate the interpretation of the results if that were a possibility. Most respondents attended all sessions. Nevertheless, it is possible that some participants did not attend some of the presentations, and it is important for the analysis not to confuse participants who cannot recall a presenter because they did not attend or because of imperfect memory. We will come back to this issue in the analysis.

We contacted all participants a month after each conference and asked them to participate in our experiment. We offered \$50 Amazon vouchers to the five best performers in the memory task

(that is, 10 vouchers in total for both data collections). These were distributed by e-mail (participants did not need to be identified to receive the voucher).

The questionnaire consisted of three memory tasks. The time to answer the questions was restricted in each one of these tasks (see on-line Appendix for screenshots), such that it would be very difficult for participants to check information on the internet.

The first two tasks consisted of matching pictures of presenters to the title of the paper they presented. The pictures were obtained from public websites (such as homepages).² In the first task, participants saw *one paper title* and *four pictures* of conference participants.³ The second task was the reverse of the first task. In the second task, participants saw only *one picture* of one presenter and *four titles* of papers. We chose to ask participants to recall the title. We could have also asked for alternative pieces of information such as presentation slides, or any other piece of information, the worry is that these pieces of information could contain identifying information. The advantage of a title is that it is a valuable piece of information, but it does not contain identifying information. Each of these tasks included six questions. Participants had 25 seconds to answer each of the questions. When selecting options for the first and second task, we made sure that we only asked one question about the same presenter. The first and second tasks are similar in nature. Participants know that one of the four options is the correct answer. In both cases, the options were randomly chosen from the pool of presenters and the pool of papers (from the same event). This feature is helpful because there will be a random variation in whether there are other people sharing attributes such as gender and race in the same choice set. Because minority groups are by definition small in size, most answer choice sets will have a majority of presenters from the majority groups and a minority of presenters from the minority groups. As a consequence, presenters from minority groups may be matched more accurately to the paper they presented, simply because of these distinctive characteristics. Indeed, if gender and race are prime attributes recorded in memory, participants may remember that the presenter had these attributes (even though she may not recall specifically who that presenter was). But we would also expect

² One concern that immediately arises is that there may be biases according to dimensions such as gender and race in how accurate the picture is. We are able to address this concern by asking independent raters to evaluate the similarity of the public picture with a picture taken in the conference room itself (for the Edinburgh conference). See on-line appendix for results with an additional control for the similarity of pictures.

³The set of pictures mixed presenters and other attendants of the conference who presented a poster.

them to be more likely to be confused with someone sharing these attributes if confusion is possible (i.e. if there are more of them in the choice set).

Finally, the third task requires participants to provide the name and current academic institution of a presenter, shown on a picture. The answer here is free format (participants must write something and cannot choose between pre-specified answers) and allows us to evaluate the ability to recall the identity of a specific presenter. Here, remembering distinctive attributes is obviously not sufficient to provide a correct answer. Again, participants had 25 seconds to answer each question and were asked to identify six presenters.

After completing the three tasks, we asked respondents to answer a short survey. We asked about demographics of the respondents such as gender, age, field, research position, and race. We asked them to indicate which sessions they attended, and for the Edinburgh conference, we additionally asked respondents to indicate whether they knew the presenter before the conference. For the analysis, we will only consider responses to questions involving presenters in sessions that respondents report having attended. This is to minimize a potential selection bias and to ensure that the reason why participants do not recall a presenter is because of imperfect memory and not because they did not attend.

In both experiments, those participants who wanted to participate in the contest for an Amazon voucher were asked to provide an e-mail address. We had multiple versions of the questionnaire for each event (4 for each conference), varying the presenters involved. Our data includes questions about 37 of the 44 presenters from the San Diego conference⁴ and almost all presenters from the Edinburgh conference. Not all presenters are featured in each task though. The allocation of presenters to tasks was random.⁵

Participants earned points for each correct answer. They earned two points per correct answer for the first two tasks. For the third task, they earned one point for a correct name and one point for a correctly indicated affiliation.⁶ Vouchers were awarded to those respondents who achieved the highest number of overall points for the three tasks.

⁴ We could only include those with pictures on their public website and excluded one presenter from the San Diego conference because the presentation was not of an academic paper. Among the other presenters, we selected randomly.

⁵ As a consequence, it is possible that participants to the two conferences were asked to identify themselves. This should not bias the results in any specific way though, but could raise the level of accuracy slightly.

⁶ When assigning points and for the following analysis, we allowed for spelling mistakes and abbreviations.

Finally, we collected background information of presenters through the Internet. We recorded information about each presenter's current academic position, number of publications, number of top five journal publications, current research institution, and if applicable, time since completion of the PhD. We collected most of this information from personal webpages and CVs. We also recorded the ranking of the institution the presenter was affiliated to at the time of presentation, using the Tilburg Economics Schools and Research Ranking (<https://econtop.uvt.nl/>). We constructed a measure for race using the free on-line service karikos (<https://www.kairos.com/diversity-recognition>).⁷ The software identifies the racial background of a person based on a picture and indicates relative percentages of categories (e.g. 90% white, 5% Asian, 5% Hispanic). We categorized a presenter as non-white if the dominant category was not white).

2. Evidence from the field

The goal of the analysis is to establish whether there are systematic biases in memory according to gender and race, controlling for productivity variables. To test for systematic biases in remembering, we study the accuracy of memory conditioning on a large set of variables that are likely to correlate with “academic productivity”, such as the seniority of the presenter, the number of publications in top ranked peer reviewed journals and the rank of the current institution of the presenter. Of course, some of these variables are likely to correlate with how known a person is and on past interactions as well. These variables could also directly correlate with the quality of the presentation itself, which may make it more memorable. We will not be able to tease out between these alternative explanations. What matters though for our research question is whether there are systematic biases in memory along variables such as gender and race, conditioning on variables capturing productivity.

We first present summary statistics on presenters and respondents (2.1.). We then study how the presenters' characteristics relate to the accuracy of recall (2.2.) and test more specifically for the theory of categorisation (2.3.).

2.1. Summary statistics - Presenters and respondents

Table 1 presents summary statistics of the presenters included in the experimental study by conference and seminar series. There were 44 presenters in San Diego, of which 37 are included in the experiment, and all 25 presenters from the Edinburgh conference are included.

⁷ Unfortunately, the software is no longer available on-line.

The most notable differences between the events are in the percentages of women (about one third in the San Diego conference and 20% in the Edinburgh conference) and the percentages of economists (the San Diego conference was interdisciplinary, with a majority of economists, while the Edinburgh conference was dominated by economists).

Table 1 – Summary statistics of the presenters included in our study

Characteristics of presenters	San Diego	Edinburgh
Means (standard deviations)	April 2012	May 2014
# of presenters	37	25
% female	35%	20%
% non-white	11%	16%
% non-native English speakers	73%	64%
# of years since PhD	12.6 (9.8)	13.9 (11.9)
rank current institution	51.1 (38.1)	59.36 (37.54)
# top 5 publications in economics	1.6 (3.0)	2.3 (4.1)
# of publications	26.1 (27.3)	17.8 (22.5)
% economists	73%	96%

Table 2 presents summary statistics of the respondents, again split by event. There were 114 participants in the San Diego conference, among which 42 participated to the experiment; and 111 in Edinburgh, among which 46 participated.

To ensure anonymity, the information we collected about respondents is more limited and coarse. About a third of our respondents are female and only a quarter above 40 years old. Except for the fact that all respondents are economists in the Edinburgh conference, against three-quarters at the San Diego conference, there are no large differences in the respondents' characteristics across the two conferences.⁸

Table 3 presents Summary statistics regarding the performance in each task. Overall, we find that people are much better able to map faces to paper titles than recalling names or institutions. Also, the mean accuracy rate in mapping faces and titles is around 65% on average for the two conferences. The accuracy is lower for task three (recalling name and institution of a person) at around 40%.

⁸ Since we collected information on the respondents as well, we are able to study to what extent their characteristics correlate with accuracy of recall (see Table A1 in the on-line Appendix). We do not find evidence of any significant variable affecting accuracy of recall, except for the respondent being an economist.

Table 2 – Summary statistics of the respondents

Characteristics of respondents	San Diego	Edinburgh
Means (standard deviations)		
# of respondents	41	46
% female	36%	30%
% tenured	33%	33%
% non-white	24%	20%
% economist	74%	100%
% over 40 years old	26%	26%

Table 3 – Summary statistics of performance in each task

	San Diego	Edinburgh
Tasks 1 and 2 (picture and title)	.57 (.50)	.65 (.48)
Task 3 (picture and name/institution)	.43 (.50)	.38 (.49)

Accuracy rate: Mean values with standard deviations in parenthesis.

Selection

One important issue we wish to address head on is selection. We have two potential sources of selection: One is in responding the survey, the second is in attending the presentation or not.

Regarding the first source of selection, in all events, we have more than a third of attendees responding to our survey, and the invitation e-mail did not give any hint as to what the survey would involve, except that it was related to the event of interest (conference attended). This response rate is in fact quite high relative to typical response rates in social sciences surveys. Cook et al. (2007) report that one should expect between a 25% and 30% response rate from an e-mail survey when no follow-up takes place. Of course, it would have been helpful to be able to compare characteristics of respondents and attendees, but we do not have the information on attendees. As is usually the case with surveys, we only have information on those who actually responded.

The second issue is key for the interpretation of the results. People could fail to recall someone either because of imperfect memory (what we are interested in) or because they did not attend the presentation. If participants are less likely to attend sessions of presenters with certain

characteristics (minorities for example), we would have a selection bias if we would categorize these people in the same category as those who cannot recall the presenters with these characteristics conditionally on attending. The ideal setup for our research question is one where either attendance is random or attendance is compulsory for everyone. As already mentioned, we chose these particular events precisely because we expected little selection into attendance to take place. In contrast to larger conferences, the norm in these smaller one-session events is very much that everyone attends all sessions, although of course in practice some people do not attend some of the sessions.

Table 4 presents summary statistics about self-reported attendance at both conferences. We have a mean attendance of 85% for all sessions at the San Diego conference (with a median of 86%) and a mean attendance of 82% at the Edinburgh conference (with a median of 85%). These high attendance rates should reassure that self-selection into attendance is not a major issue. Further, we find no evidence that attendance is different for minorities (women and non-whites) (see Table A2 in the on-line Appendix). To minimize any chance of bias, we will limit the analysis to those who claim they attended the session and we will check whether attendance is correlated with presenter characteristics.

Table 4 – Summary statistics on self-reported attendance at each session

	San Diego	Edinburgh
Average	0.85	0.82
Median	0.86	0.85
Max	1.00	1.00
Standard deviation	0.11	0.09
Min	0.57	0.70

2.2. Presenter characteristics and accuracy of recall

We pool the data from both conferences for the analysis, and we also pool the data from the first two tasks, as they both involve matching pictures of people to titles of papers.⁹ In Tables 5 and 6, we present the results of a two-way error component linear probability model, allowing for presenter and respondents' random effects.¹⁰ The dependent variable is a dummy for correct

⁹ We replicate the analysis for each conference separately (see Table A3 in the on-line Appendix). We should however take the results of this analysis with caution because the number of presenters falling into the different categories of interest (female and non-white) is small when considering each conference separately. Overall, the effects differ somewhat across events, but they are not inconsistent with each other.

¹⁰ The model is estimated in STATA 12.0 using the command “xtmixed”.

answer. Note that we only consider answers of respondents who also attended the corresponding session.

Table 5 reports the results related to the task of mapping pictures of presenters to titles of the paper they presented (multiple choice type questions). Column (1) shows the estimates of a model conditioning on gender and race. Column (2) conditions in addition on characteristics correlated with the productivity or expertise of the presenter. We control for rank of current institution, number of publications, number of top 5 publications in economics, a dummy for being an economist, number of years since PhD completion, and a dummy for presenter being a native English speaker (we conjecture that being a native English speaker may be correlated with the quality of the presentation). Column (3) additionally conditions on characteristics of social proximity. We have two main variables of social proximity: *Same field* (which is a dummy equal to 1 if both the presenter and respondent are economists; or both are psychologists, and equal to 0 otherwise) and *same gender* (which is a dummy equal to 1 if both the presenter and respondent are of the same gender, and equal to 0 otherwise).

Table 5 - Probability of correct mapping between face and title

	(1)	(2)	(3)
female	0.082 (0.056)	0.119** (0.059)	0.141** (0.058)
non white	0.074 (0.075)	0.103 (0.075)	0.107 (0.073)
non-native English speaker		0.056 (0.059)	0.053 (0.057)
# of years since PhD		0.005 (0.004)	0.004 (0.004)
rank current institution		-0.000 (0.001)	-0.000 (0.001)
# of top 5 publications in economics		0.016* (0.009)	0.015* (0.008)
# of publications		-0.000 (0.001)	-0.000 (0.001)
economist		0.018 (0.078)	-0.167* (0.088)
same field			0.250*** (0.061)
same gender			0.038 (0.034)
Constant	0.607*** (0.034)	0.464*** (0.092)	0.398*** (0.092)
Observations	892	850	850
Number of groups	1	1	1

Two-way error component linear probability model, allowing for presenter and respondents' random effects. The dependent variable is a dummy for correct answer. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

We find that women are significantly more likely to be remembered accurately, by around 14 percentage points, once we control for academic achievements and social proximity (Column (3)). We do find an effect of similar magnitude for race but it is not statistically significant. It is notable that the gender dummy becomes larger and significant when adding controls for productivity. This suggests that women are on average less established than men in our sample are and it is therefore unlikely that unmeasured productivity-related variables would explain the female positive coefficient.

Academic achievements do matter as well, but not to a large extent. All else equal, one needs eight top 5 publications in economics to achieve a similar improvement in recall as being female. This is perhaps surprising, as one would expect well-established researchers to be more likely to be remembered for several reasons: The value of social ties may be higher, they are likely to give higher quality presentations, and they are more likely to have been previously encountered. But, it appears that these factors do not matter much for recall accuracy. Being in the same field, on the other hand, is a strong and significant predictor of recall accuracy. On top of that, all else equal, economist presenters are less likely to be remembered. This is consistent with a value of standing out – economist respondents (who are in majority) can better distinguish between presenters who are psychologists in comparison to psychologist respondents, who have a harder time distinguishing between the large fraction of presenters who are economists.

Table 6 shows the results for the task consisting of providing the names of presenters and of their institutions (Task 3). Columns (1) to (3) relate to the naming of presenters and Columns (4) to (6) relate to the naming of institutions. Thus, the task here is to recall who a specific person is, and remembering of distinct attributes alone is not sufficient to get a correct answer. As in the previous Table, Column (1) includes gender and race dummies; Column (2) controls for productivity and expertise, and Column (3) controls for social proximity variables.

Here we find slight but non-significant evidence of a minority bias in favour of non-whites. Measures of academic performance and establishment do matter somewhat, but their effects are very small in magnitude. We find that social proximity matters greatly and in a similar way as for mapping between faces and titles. Of course, one obvious explanation for the effects of social proximity and of academic achievements is related to the fact that the respondent is more likely to know the presenter if they are in the same field and if the presenter is well established. This is a compelling story, but one we should be very cautious with, as knowing someone is not exogenous and is likely to be correlated with factors affecting how memorable someone is. There is an obvious circularity in the fact that respondents can only claim to know someone if they remember that person. Nevertheless, we collected information on whether participants claimed to know the presenter for the second of the two conferences (Edinburgh). We find that knowing the presenter is a strong predictor of accuracy of recall in all tasks. The effects of being female or non-white are similar in magnitude but not statistically significant in Edinburgh (whether one controls for knowing the presenter or not). The size of the coefficients remains similar to those excluding the dummy variable for “knowing the presenter” (see Table A4 in the on-line Appendix for detailed results).

Table 6 - Probability of correctly remembering of the name and the institution

	correct answer name			correct answer institution		
	(1)	(2)	(3)	(4)	(5)	(6)
female	-0.097 (0.086)	-0.015 (0.061)	0.022 (0.066)	-0.030 (0.078)	0.034 (0.063)	0.062 (0.066)
non-white	-0.019 (0.096)	0.107 (0.069)	0.100 (0.073)	0.020 (0.087)	0.084 (0.071)	0.078 (0.073)
non-native English speaker		-0.100* (0.060)	-0.092 (0.063)		-0.077 (0.062)	-0.069 (0.064)
# of years since PhD		0.006 (0.004)	0.006 (0.004)		0.000 (0.004)	-0.001 (0.004)
rank current institution		-0.001* (0.001)	-0.001 (0.001)		-0.002*** (0.001)	-0.002** (0.001)
# of top 5 publications in economics		0.015 (0.010)	0.016 (0.011)		0.016 (0.011)	0.016 (0.011)
# of publications		0.002 (0.002)	0.003* (0.002)		0.002 (0.002)	0.002 (0.002)
economist		0.048 (0.081)	-0.234** (0.109)		0.096 (0.084)	-0.163 (0.111)
same field			0.344*** (0.082)			0.314*** (0.084)
same gender			0.071 (0.045)			0.047 (0.046)
Constant	0.463*** (0.050)	0.360*** (0.091)	0.263*** (0.101)	0.427*** (0.047)	0.412*** (0.095)	0.338*** (0.102)
Observations	429	429	429	429	429	429
Number of groups	1	1	1	1	1	1

Two-way error component linear probability model, allowing for presenter and respondents' random effects. The dependent variable is a dummy for correct answer.

Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Turning to the naming of institutions (Columns (4)-(6)), we do not find evidence of significant biases in the accuracy of recall according to gender and race; although the point estimate of the latter is large and positive. People are better able to remember the institution of the presenter if the institution is highly ranked, although again the magnitude of the effect remains modest. We find evidence of a social proximity effect, similar to the one we found for the task of naming the presenter.

2.3. Probability of confusing people sharing similar attributes

Our findings so far show that women are more likely to be recalled accurately, in a setting where they are in minority. However, women are only significantly more likely to be recalled accurately in the multiple-choice type questions, where due to the minority status of women, remembering the gender may be sufficient in order to correctly identify a female presenter. In a next step, we evaluate whether minorities are more likely to be confused with those who share these minority attributes. This question directly relates to the theory of categorisation.

For this analysis, we use data from tasks 1 and 2. In those two tasks, the respondent is asked to pick between four pictures of presenters and correctly match it to a presented paper (task 1) or to pick between four titles and correctly match it to a picture of a presenter (task 2). Because in each task, we chose the other three options at random, presenters from minority groups will often be shown with other presenters who are from majority groups. But, the design also allows us to study whether they are more likely to be confused with someone who has the same gender or race.

We test whether minority groups are more likely to be confused when there are several of them as possible answer options. To do this, we estimate a conditional logit model in the spirit of McFadden (1973), where the dependent variable is a dummy indicating whether or not an option is chosen, and the independent variables are attributes of the various options. The advantage of the conditional logit model is that it only uses variation in attributes within each choice set (i.e. between the four options) to estimate the effects of these attributes on the probability of being chosen.

We also control for a variable called “same session”, which is a dummy equal to one if the presenter shown as one of the four options presented in the same session as the presenter corresponding to the correct answer. We can consider “same session” as a proxy for similarity in research topic, as the sessions in the Edinburgh conference were organised according to topics, in the San Diego conference all talks were on one topic though (deception).

The results of the conditional logit model are reported in Table 7. We find that respondents are significantly more likely to choose the correct option compared to the other options, but we do not find evidence for confusion (i.e. choosing another option of the same gender and race instead of the correct option) overall. The theory of categorisation proposed by Fryer and Jackson (2008) would predict that confusion should be more likely to occur for minorities and/or for people who have attributes associated with a lower expected value of interaction.

In Table 7, Columns (2) to (5), we estimate the conditional logit for different subsamples, corresponding to the gender and race of the speaker associated with the correct answer. We find that men are not more likely to be confused with other men. Women, in contrast, are significantly more likely to be confused with other women. Interestingly, race also seems to play an important role when it comes to remembering women, while this is not the case when remembering men. For non-whites, we find evidence that gender is used as an attribute (more confusion is likely with a person from the same gender). Again, we do not find similar effects when it comes to remembering the majority group, i.e., white individuals. Our results are in line with the theory of categorisation discussed earlier, whereby broader categories are formed, for women in particular.

One drawback of the field is that minority groups are by definition in minority, so our results could be driven by idiosyncratic characteristics of those with minority attributes; and we have limited scope to test for the theory of categorisation. Ideally, we would like to have a larger set of people with minority attributes and vary exogenously the relative sizes of groups with minority attributes. As discussed earlier, we predict that minorities may be more likely to be confused with each other when there is more room for confusion, that is, when their share is relatively large. We conduct a controlled on-line experiment that will allow us to address the caveats of the field setting.

Table 7 - Conditional logit for probability of correct mapping between face and title – distinguishing by the attributes of the correct answer

	(1) All	(2) Male presenter	(3) Female presenter	(4) White presenter	(5) Non-white presenter
Correct option	1.820*** (0.223)	1.862*** (0.232)	1.668* (0.874)	1.759*** (0.228)	2.215* (1.267)
Option same gender	0.025 (0.131)	-0.168 (0.152)	0.539* (0.285)	-0.085 (0.135)	1.317** (0.518)
Option same race	0.228 (0.149)	-0.110 (0.172)	1.165*** (0.309)	0.101 (0.159)	1.073 (0.665)
Presented in same session	0.584*** (0.217)	0.605*** (0.227)	0.342 (0.829)	0.694*** (0.224)	-0.977 (1.090)
N. obs.	4,196	3,132	1,064	3,788	408

The dependent variable is a dummy indicating whether an option was chosen.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

3. Experiment 2: Controlled setting

The second experiment is also conducted on-line but is not based on a real setting. The experiment replicates key features of the field setting. We matched pictures of people drawn from a picture database¹¹ to titles of NBER working papers (drawn at random from the NBER paper series in 2019). The picture database has pictures from individuals of different gender and race; we used those from the categories “Caucasian white” and “Asians”. In the latter, we selected pictures of East Asians and excluded pictures of South Asians.

In a first stage, participants see 12 pictures of people randomly matched with titles of papers. In a second stage and as in the first experiment, participants are asked to either match a person with a title (choosing between four possible titles), or to match a title with a person (choosing between four possible people). Four questions were asked in total, one with the correct answer being a white man, a second a white woman, a third a non-white man and a fourth a non-white woman (the order was randomized). The set of possible answers was chosen at random. Two of the three incorrect answers corresponded to titles or people shown in the first stage, one was a completely new person or title. The race and gender of the new option was with equal chances a white man, white woman, non-white man or non-white woman. This randomization ensures we have a variation in the characteristics of the choice set and allows us to examine again, whether people just remember attributes (gender, race) of the correct answer or are able to remember who the specific person was.

We implemented four treatments as summarized in Table 8. We recruited 387 participants from the subject pool of the experimental laboratory at the University of Cologne. The summary statistics of these participants are shown in Table 9 below. In all treatments, we have a majority of white and female respondents. The gender composition is however more balanced than the racial composition.

¹¹ Stimulus images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University, <http://www.tarrlab.org/>. Funding provided by NSF award 0339122. We used a subset of pictures (we removed those where the person had uncommon accessories such as a hat or a piercing), consisting of 75 pictures of Caucasian whites and 29 East Asians. To avoid the task to be about picture recognition rather than person recognition, we use pictures with different profile orientation (front facing or face turned left or right by 15 to 45 degrees).

Table 8: Treatments – Experiment 2

	Distribution of pictures (Numbers in each treatment)				Number of subjects
	white men	white women	Asian men	Asian women	
1 Balanced	3	3	3	3	99
2 Female minority	5	1	5	1	97
3 Asian minority	5	5	1	1	96
4 Female and Asian minority	9	1	1	1	96

Table 9 – Summary statistics of respondents – Experiment 2¹²

	Balanced	Female Minority	Asian Minority	Female and Asian Minority
# of respondents	99	97	96	96
% female	59%	54%	55%	64%
% non-white	15%	16%	20%	24%
% east Asian	4%	6%	5%	11%

Table 10 reports the mean accuracy rates in recall, according to the attributes of the person corresponding to the correct answer and the treatment. These descriptive statistics point at a positive minority bias especially in the female and Asian minority treatments. Comparing the recall accuracy of different groups in the balanced treatment, we find no significant differences.

Table 10 – Mean rate of accuracy in recall across treatments

Race and gender of “correct answer”	Balanced	Female Minority	Asian Minority	Female and Asian Minority
white men	0.60 (0.49)	0.55 (0.50) ^{n.s.}	0.64 (0.48) ^{n.s.}	0.49 (0.50) ^{n.s.}
white women	0.66 (0.48)	0.71 (0.46) ^{n.s.}	0.70 (0.46) ^{n.s.}	0.78 (0.42)*
Asian men	0.60 (0.49)	0.54 (0.50) ^{n.s.}	0.64 (0.48) ^{n.s.}	0.73 (0.45)*
Asian women	0.57 (0.50)	0.72 (0.45)**	0.66 (0.48) ^{n.s.}	0.76 (0.43)***

Upper cases refer to results from a Fisher’s exact test comparing results to the Balanced setting.
Where *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, n.s. $p > 0.10$.

In Table 11, we present estimates of the probability of a correct answer, depending on the gender and race of the person corresponding to the correct answer. In line with our findings from the

¹² The distribution of respondent characteristics does not differ significantly across treatments except for one case: There are significantly more East Asian respondents in the “Female & Asian minority” treatment than in the Balanced treatment (p-value of Fisher test, $p = 0.062$). Robustness checks do not indicate that this imbalance drives any of the results.

field, we find evidence of more accurate recall of women. However, this difference arises only in the treatments where they are in minority. When women are a minority, they are 15 to 18 percentage points more likely to be correctly matched to the title of the paper they were matched with in the first stage. We provide evidence of a bias in favour of non-whites only in the treatment where both women and non-whites are in minority. Whether the respondent shares the same gender or race with the person corresponding to the correct answer does not matter except in the balanced treatment, where having the same gender enhances recall.

Table 11 - Probability of correct mapping between face and title (Experiment 2)

	Balanced		Female minority		Asian minority		Female & Asian minority	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	0.015 (0.044)	-0.011 (0.045)	0.175*** (0.045)	0.177*** (0.045)	0.042 (0.045)	0.039 (0.045)	0.164*** (0.043)	0.148*** (0.046)
East Asian	-0.045 (0.044)	-0.107 (0.079)	-0.000 (0.045)	0.033 (0.074)	-0.023 (0.045)	-0.058 (0.070)	0.108** (0.043)	0.156*** (0.058)
Same gender as resp.		0.130*** (0.045)		-0.021 (0.046)		0.033 (0.043)		0.055 (0.045)
Same race as resp.		-0.076 (0.082)		0.043 (0.075)		-0.047 (0.072)		0.076 (0.058)
Constant	0.619*** (0.044)	0.634*** (0.083)	0.541*** (0.043)	0.515*** (0.078)	0.647*** (0.045)	0.670*** (0.075)	0.554*** (0.042)	0.478*** (0.062)
Observations	396	396	388	388	384	384	384	384

Two-way error component linear probability model, allowing for picture and respondents' random effects. The dependent variable is a dummy for correct answer. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In Table 12, we distinguish specifically between settings in which women or East Asian are a minority and those in which they are not. For example, white women are in minority in two treatments ('female minority' and 'female & Asian minority') and are not a minority in two others ('balanced' and 'Asian minority'). We report the results for the relevant treatments in which the relevant group is not a minority in Panel A and those in which the relevant group is a minority in Panel B of Table 12. Since white men are never in the minority in any treatment, we report results only in Panel A.

We find evidence for differences in the way women and East Asians are remembered. However, we find such differences only when they are not in minority. For white men, we find no indication

that people are more likely to confuse them with other white men. The options that are either white or male are not significantly more likely to be chosen. However, when the correct answer involves a white woman or an East Asian man, we find evidence of reliance on gender and/or race to select an answer. People are more likely to select a wrong option of the same gender and race. This is true for white women, and this is true for East Asian men. The evidence is weaker for the case where the correct answer corresponds to an East Asian woman, but there is only one treatment where they are not in minority (balanced), and so the number of observations is smaller. The point estimates are however quite large and thus do not point to the absence of a categorisation effect.¹³

The analysis presented in Table 12 pools all the respondents, with no distinction according to their gender or race. To examine whether there are own-group biases in memory, we estimate the models in Table 12 separately for (1) white respondents only (our sample of non-white respondents is too small to analyse separately), (2) women only and (3) men only (see Tables A7, A8, and A9 in the on-line Appendix). We find stronger evidence for a categorisation according to gender if we restrict the sample to white respondents only. Regarding gender, we find evidence that the confusion effect is stronger for male respondents; we do not find significant evidence that female respondents confuse women with other women, while male respondents do.

¹³ One may argue that these results are driven by the fact that there are no similar individuals to confuse minorities with. To test for such an alternative interpretation, we conducted a similar analysis restricting the data to those questions where there was an answer option of the same gender and the same race as the correct option. Results are robust to this alternative approach. For details see A6 in the on-line Appendix

Table 12 - Conditional logit for mapping between face and title – by type of individual

Panel A - Treatments in which individual is not a minority				
	(1)	(2)	(3)	(4)
Gender correct answer	white men	white women	Asian men	Asian women
Race correct answer				
Treatments	1,2,3,4	1, 3	1, 2	1
Correct option	1.668*** (0.291)	2.053*** (0.419)	3.237*** (0.542)	1.878*** (0.513)
Option same gender	0.230 (0.253)	0.653* (0.393)	1.578*** (0.502)	0.664 (0.481)
Option same race	0.385 (0.258)	-0.180 (0.392)	0.951*** (0.358)	0.677 (0.465)
Observations	800	384	412	192
Panel B - Treatments in which individual is a minority				
	(1)	(2)	(3)	(4)
Gender correct answer	white men	white women	Asian men	Asian women
Ethnicity correct answer				
Treatments		2, 4	3, 4	2,3,4
Correct option		1.710*** (0.388)	1.578*** (0.345)	2.231*** (0.255)
Option same gender		-1.118 (0.757)	-0.567 (0.405)	0.189 (0.381)
Option same race		-0.188 (0.439)	0.562 (0.464)	0.450 (0.378)
Observations		380	340	532

The dependent variable is a dummy indicating whether an option was chosen.

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

4. Conclusion

This paper investigates the presence of systematic biases in memory in two different settings. First, we collected data in a real professional context, among researchers in Economics in the context of two international high profile conferences. We study how accurately conference participants can remember who presented what, a month after the conference. Specifically, conference participants were asked to map pictures of people to titles of papers presented and they are asked to provide the presenters' name and institution based on their picture. We complement the evidence from the field with evidence from a controlled experiment where the individuals are matched at random to titles of papers and where we vary exogenously the fractions of men/women and non-whites.

We find evidence that women and, to a lesser extent, racial minorities, are more likely to be recalled in settings when they are in minority; and are more likely to be confused with others who share the same attributes. That is, they seem to be lumped into broad categories according to gender and race. These findings are in line with the theory of categorisation proposed by Fryer and Jackson (2008) and with a distinctiveness effect that has been identified in previous studies on memory (Slone et al., 2000, Meissner and Brigham, 2001). People with minority attributes appear to be categorised according to these attributes, and are "blended together". In settings where there are few people with minorities, recall is enhanced. In settings where there are more of them, this leads to confusion.

Overall, we argue that biases in recall is an understudied but important research avenue; as such biases may have implications for career prospects, since social networks play a large role in many labour markets. At this stage, we do not know the implications of these biases for people's careers, but given the importance of recall in network formation, we conjecture these effects may not be small.

References

- Arrow, K., 1973. "The Theory of discrimination," in *Discrimination in Labor Markets*, O. Ashenfelter and A. Rees, eds. Princeton, NJ: Princeton University Press.
- Becker, G.S., 1971. *The Economics of Discrimination*. Chicago: University of Chicago Press
- Belot, M., 2015. "Cognitive Racial Discrimination: A Benchmark Experimental Study," *Journal of Neuroscience, Psychology and Economics*.
- Blackaby, D., Booth, A., and Frank, J., 2005. "Outside offers and the gender pay gap: Empirical evidence from the UK academic labour market," *Economic Journal* 115 (501), 81–107.
- Blackaby, D. and Frank, J., 2000. "Ethnic and Other Minority Representation in UK Academic Economics," *Economic Journal* 110, 293-311.
- Calvó-Armengol, A. and Jackson, M.O., 2004. "The Effects of Social Networks on Employment and Inequality," *American Economic Review* 94(3), 426-454.
- Carter, J., Fenton, S., and Modood, T., 1999. "Ethnicity and employment in higher education," London, Policy Studies Institute.
- Chen, D.L., Schonger, M., Wickens, C., 2016. oTree - An open-source platform for laboratory, online and field experiments. *Journal of Behavioral and Experimental Finance* 9, 88-97.
- Enke, B., Schwerter, F., & Zimmermann, F. (2020). Associative memory and belief formation (No. w26664). National Bureau of Economic Research.
- Fryer, R. and Jackson, M.O., 2008. "A Categorical Model of Cognition and Biased Decision Making," *B.E. Journal of Theoretical Economics (Contributions)* 8 (1), Article 6.
- Kahn, S., 1993. "Gender Differences in Academic Career Paths in Economists," *American Economic Review* 83 (2), 52-56.
- Meissner, C. and Brigham, J., 2001. "Thirty years of investigating the own-race bias in memory for faces," *Psychology, Public Policy, and Law* 7, 3-35.
- Montepare, J.M. and Opeyo, A., 2002. "The relative salience of physiognomic cues in differentiating faces: A methodological tool," *Journal of Nonverbal Behavior* 26, 43–59.
- Reuben, E., Sapienza, P., and Zingales, L., 2014. "How stereotypes impair women's careers in science," *Proceedings of the National Academy of Sciences*, 111(12), 4403-4408.
- Shepherd, J.W., Gibling, F., and Ellis, H.D., 1991. "The Effects of distinctiveness, presentation time and delay on face recognition," *European Journal of Cognitive Psychology* 3, 137-45.
- Slone, A. E., Brigham, J.C., and Meissner, A., 2000. "Social and Cognitive Factors Affecting the Own-Race Bias in Whites," *Basic and Applied Social Psychology* 22(2), 71-84.
- Tibbetts, E. A. and Dale, J., 2007. "Individual recognition: It is good to be different," in Tibbetts, E. A. and J. Dale 2007, *Trends in Ecology and Evolution*.
- Valentine, T., 1992. "Towards an exemplar model of face processing - The effects of race and distinctiveness," *Quarterly journal of experimental psychology* 44 (4), 671-703.