

Do photos help or hinder field experiments of discrimination?

Judith Rich*

University of Portsmouth and IZA

Abstract

This paper assesses field experiments of labour and product markets that have attached photos to identify applicants (in the case of labour markets) or sellers/crowdfunders (in the case of product markets). The experiments seek to identify the contribution of attractiveness, race/ethnicity, skin colour, sexual orientation or religion to the behaviour of agents in markets. These experiments attach photos to CV to signal attractiveness, or the basis being tested such as race/ethnicity, skin colour or religion. Many experiments report significant findings for the impact of attractiveness or the identity revealed on positive callbacks to applicants. The issue considered here, however, is to what extent it is attractiveness or other perceived characteristics that may have had an impact on the behaviour recorded in the experiments. The results of the studies covered in this paper, to a lesser extent those of Weichselbaumer (2004) and Baert 2017, are compromised by including photos, with the possibility the responses received were influenced not only by the basis being tested such as attractiveness, race/ethnicity or religion but by some other characteristic unintended by the researcher but conveyed by the photo. There is evidence in experimental work of a range of characteristics that photos convey of individuals and their impact on labour and product market outcomes such as success in obtaining a positive response to job applications and success in obtaining funding to finance projects in the product market. Suggestions are made for future experiments: evaluation of photos for a range of characteristics; use of a 'no photo' application together with the photo applications; evaluation of responses for any bias from unobservable characteristics using Neumark (2012). This paper discusses for the first time three questions with some tentative answers. First we face introducing further unobservable characteristics by using photos. Second, we cannot fully control the experimental approach when using photos. Third, we are able to accurately evaluate the impact of the photos used on the response/probability of callback. Field experiments using photos need to ensure they do this for the range of factors that have been shown to affect judgments and therefore potentially influence call back response. However, the issue remains whether we have in fact identified all potential characteristics conveyed by the photos.

ACKNOWLEDGEMENTS

I would like to thank participants at the Conference "Hiring Discrimination: Measures, Moderators and Mechanisms", Växjö, Sweden August 2017 for discussion and helpful comments as well as the editor of the journal, two anonymous referees and Michael V. White.

- Department of Economics and Finance, University of Portsmouth, Richmond Building, Portsmouth St., Portsmouth, UK, P01 3DE; judy.rich@port.ac.uk.

Do photos help or hinder field experiments of discrimination?

INTRODUCTION

It will be useful to first consider the development of the correspondence testing method in field experiments as it provides some context for an evaluation of the use of photos in experiments of discrimination in markets.

The correspondence technique was developed specifically in response to the early criticism of the audit (in-person) approach that used people (trained testers, sometimes actors) to enquire as to job openings or housing availability. The in-situ method, created by Daniel in 1965, to test for the efficacy of UK anti-discrimination legislation at the time, used testers as one part of a wide-ranging investigation into differential treatment of Asians and West Indians (Daniel 1968). Daniel sent testers in pairs (one white British and one of the race/ethnicity being investigated) to enquire as to job vacancies, housing, car rental and hotel accommodation. The responses to the enquiries were recorded for the pairs of testers and assessed for any differences. Statistically significant lower callbacks were recorded for Asians and West Indians as compared to white British applicants. This was accepted as compelling evidence of discrimination on the basis of race and ethnicity, particularly in employment and housing, and was instrumental in the passing of legislative amendments to the UK Race Relations Act in 1968. This approach was critically assessed by Ward (1969) regarding the lack of control of the testers where the experimenter had no direct knowledge, or record, of the testers' actual behaviour in the test situation. It was possible that the motivation of the testers could influence their behaviour and therefore the responses to them (in either a positive or negative way). At the very least, it was not possible to demonstrate complete control in the enquiry so that the effect for the response being examined, such as race, could be satisfactorily isolated. Jowell and Prescott-Clarke (1970) responded by developing the now familiar correspondence testing approach which they asserted overcame any lack of control by removing any personal tester effects. The only feature of any applicant that differed in a pair of enquiries was the basis being tested for in their case race and immigrant background. In their written tests conducted in 1969, pairs of CVs were sent in applications for job vacancies, one white British, the other either Asian or West Indian. Callback responses for each applicant in the pair were recorded and analysed for any statistically significant differences. Subsequent to these developments, studies of

discrimination have been broadened to cover gender, age, sexual orientation, disability and obesity as well as religion, military service and mental illness, with the correspondence technique the more favoured approach for labour (and now housing) market studies. In particular, some researchers modified the technique by sending multiple CVs in application for jobs to test for discrimination against multiple ethnic groups or on multiple bases, at any one time in a country. Use of computer-generated CVs enabled researchers to randomize and then test aspects of the CV on callback rates. This paper cannot cover these many studies but details and evaluation of the experimental studies can be found in the surveys by Neumark (2018), Pager (2007), Quillian, Pager, Hexel and Midtbøen (2017), Riach and Rich (2002), Rich (2014) and Zschirnt and Ruedin (2016).

There was a subsequent criticism of experimental approaches by Heckmann and Siegelman (1993) and Heckmann (1998), which is now well known in the literature. Heckmann and Siegelman (1993) were critical of the ability of researchers using tester or written approaches to control for unobservable characteristics that could affect the employer's or landlord's assessment of the productivity of the applicant, regardless of whether they were made in person or in writing. Each applicant may have the same average productivity arising from the observed characteristic (say gender) and the unobserved characteristic. However the variation from the average for the unobserved characteristic could differ between the groups. Females, for example, may have a greater variation in productivity arising from this unobserved characteristic than males. If researchers are unable to control fully for productivity differences, there is a question as to whether the findings are an accurate reflection of discriminatory behaviour. In the case where researchers set a high standard/quality CV, there is a greater risk, to an employer, attached to choosing the minority applicant in that an individual with a much lower productivity could be chosen. If that is the case, a bias is introduced into the testing in favour of the majority applicant. An overestimation of discrimination against the minority may then arise in these tests. In the case where a low standard/quality CV is set, there is a lower risk attached to choosing the minority applicant (an individual with a higher productivity could be chosen). This would introduce a bias in favour of the minority applicant. An underestimation of discrimination against the minority group may arise in these tests.

An important technical innovation recently developed by Neumark (2012) enables researchers to evaluate whether responses have been affected by unobservable characteristics. However, the technique can only be applied if the original field experiment was designed to manipulate applicant details to differ on at least two productivity-related characteristics. Nevertheless, a recent re-evaluation by Neumark and Rich (forthcoming 2018) of field experiments for discrimination that enabled the impact of unobservable characteristics on callback rates to be isolated, raises concerns regarding experimental design, particularly of labour market tests. They found that just over half of the estimates of discrimination in the labor market tests either fall to near zero, become statistically insignificant, or change sign. They also found that in some cases it was the majority applicants who had a greater variation in productivity arising from the unobserved characteristic, indicating that researchers cannot presume which group could have the greater variance in unobservables.

The increasing use of the experimental approach, particularly correspondence testing, to cover numerous aspects of labour market hiring, has led to modifications and adaptations of the technique which allow for both the wider applicability and scrutiny of the approach. My concern in this paper is that designing the experiment to include personal features of an applicant though use of photos may defeat the purpose of correspondence testing and to actually be counter-productive, particularly in light of the Heckman-Siegelman critique. Specifically, the paper raises three questions: first, do we introduce further unobservable characteristics by using photos? second, can we fully control the experimental approach when using photos? Third, are we able to accurately evaluate the impact of the photos used on the response/probability of callback?

The focus here is the twelve experiments conducted between 2004 to 2017 that have used photos in their design of CVs: Arceo-Gomez and Campos-Vázquez (2014); Baert (2017); Galarza and Yamada (2014); Lopez Bóo, Rossi and Urzúa (2013); Maurer-Fazio and Lei (2014); Patacchini, Ragusa, and Zenou (2015); Rooth (2009); Ruffle and Shtudiner (2014); Weichselbaumer (2004; 2015; 2016a; 2016b). The paper first outlines the various ways the photos were constructed and used in the experiments. This is followed by a discussion of the experiments and the results of the studies. Then, using experimental research on the impact of two dimensions of social judgment, warmth and competence and the impact of facial perceptions on outcomes for projects listed on crowdfunding sites, other aspects of facial and

social judgment that have not been used by the researchers listed above are highlighted. These raise the possibility that the responses obtained were compromised by the use of photos.

THE PHOTOS USED IN THE EXPERIMENTS

Field experiments of discrimination in the labour market that attach photos of applicants to a CV when applying to job vacancies, do so for three reasons. The first is that the researchers want to specifically explore the impact of attractiveness on the probability of a positive response when applying for a job. This is irrespective of any necessary convention to supply a photo in the country where the testing occurred. The second is that the researchers want to specifically explore the impact of attractiveness on the probability of a positive response when applying for a job. Again, this is irrespective of any necessary convention to do so in the country where the testing occurred. The third is that photos of an applicant are an expected convention when applying for a job in the country where testing occurred, such as in Argentina, Austria, Germany and Mexico. These three reasons are not necessarily mutually exclusive as testing for race/ethnicity/phenotype can be conducted in a country where attaching a photo to a CV is customary.

The researchers then engaged in a process of evaluation of the photos. This is necessary to ensure that the photos convey the applicant characteristics being tested. If it is a study of the impact of attractiveness, then attractiveness needs to be rated and applicants with a range from unattractive to highly attractive need to be selected for inclusion. If it is a study of race/ethnicity or religion, then, together with other applicant characteristics that the researchers feel indicates productivity, attractiveness needs to be controlled to ensure average attractiveness of all applicants.

There are three sources for the initial photos used to create photo-profiles for the applications. They can be: purchased or obtained from an online source; obtained from undergraduates recruited for their participation; digitally constructed using agreed dimensions of facial attractiveness. The studies discussed in this paper have used a variety of these sources, sometimes in combination, such as buying the photos from an online site and then digitally manipulating them.

Rooth (2009) in a test of obesity in Sweden in 2006, Maurer-Fazio and Lei (2014) in a test of race/ethnicity in China in 2012, and Galarza and Yamada (2014) in a test for phenotype in Peru, early 2012, bought/obtained photos from an online site and then used digital enhancement to create the images wanted for the testing. Rooth (2009) obtained 100 photos from an online site which were then reduced by the group of researchers to 48 photos grouped as similar looking. The 48 were then rated for attractiveness by undergraduates, and this rating was then used to create pairs of photos rated the same. From these pairs, only one photo was chosen, which was considered by the evaluators as the better looking of the two, giving a final 7 photos. The 7 photos were then sent to a photo firm and manipulated to create an obese image of the individual. This enabled 7 pairs of applicants to be created, where a pair consisted of one photo of the individual as normal-weight (the original) and the other of the same individual as obese (the digitally-enhanced image). A matched pair of such applicants were then sent to an advertised job. After purchasing their photos for the experiment, Maurer-Fazio and Lei (2014) used an online survey where participants rated the photos. Those rated less attractive were then digitally altered (using psychological research on facial attractiveness) to enhance the difference between the attractive and unattractive applicants. They then created a matched pair of attractive/unattractive female applicants and attractive/unattractive male applicants to be used to apply for job vacancies to test for race/phenotype and sex. Galarza and Yamada (2014) downloaded photos from the internet and digitally modified these (by, for example, changing hair style, lips and nose) to obtain a sample of female indigenous/female white/male indigenous/male white that were standardized in terms of style of dress. An assessment of this sample found that, overall, whites were rated more attractive than indigenous individuals. A reduced group of 149 was then selected to exclude those rated very low or very high on attractiveness. The 149 photos selected for inclusion in their testing (74 Indigenous and 75 White) were rated on a scale 1 (unattractive) to 7 (highly attractive) by a panel of 7 professionals aged from 30 to late 50s. These photos were randomly assigned to the CVs that were created for the experiment, with four CVs sent to each job vacancy (female indigenous/female white/male indigenous/male white).

By contrast, Weichselbaumer (2004) and López-Bóo, Ross and Urzúa (2013) used only digitally-manipulated images. In a test for gender discrimination in Vienna over 1998-1999, Weichselbaumer (2004) manipulated photos to obtain images of three job applicants, a male,

a ‘masculine’ female and a ‘feminine’ female. The photos were rated by 119 undergraduates (who only saw one photo each) on five dimensions: attractiveness, competency, and three other dimensions, using the Bem Sex-role inventory (BSRI), covering femininity, masculinity and social desirability. These independent evaluations indicated that the photographs were effective in conveying the different female personality types (Weichselbaumer, 2004, pp. 170-171). CVs for three matched applicants, ‘feminine’ female/‘masculine’ female/male, with their identity signalled by the evaluated photos, were then created. López-Bóo, Ross and Urzúa (2013), in a test for attractiveness, used a professional designer to construct twenty-five photos using validated dimensions of attractiveness from a pool of one hundred (fifty men and fifty women), ranging from highly attractive to unattractive. To ensure the digitally manipulated photos conformed to the intended beauty norm, they asked undergraduates to rate the photos on a scale of 1 (unattractive) to 5 (highly attractive). This evaluation was consistent with rating by the validated measure. No other aspects of facial perception were evaluated by López-Bóo *et al.* (2013). The digitally-constructed photos that conformed to research on facial beauty measurements were then used in applications for jobs in Buenos Aires, Argentina, in 2010. While López-Bóo *et al.* (2013) created a pair of CV for attractive/unattractive applicants, they also used a ‘no photo’ CV as a control and so created sets of three applicants: attractive applicant/ unattractive applicant/ ‘no photo’ applicant.

A further six studies used a third approach to obtain photos for the fictitious job applicants. Undergraduates were recruited to obtain suitable photos for use in tests which varied over attractiveness, race/ethnicity and religious discrimination by Arceo-Gomez and Campos-Vázquez (2014), Patacchini, Ragusa and Zenou (2015), Ruffle and Shtudiner (2014), Weichselbaumer (2016a, 2016b) and Baert (2017).

It is expected that job applicants attach a photo to their CV in Mexico. Arceo-Gomez and Campos-Vázquez (2014) exploited this convention to study the impact of phenotype on invitation to job interviews in 2011 to 2012. They did so by recruiting three men and three women representing the different phenotypes of white (European), Mestizo and Indigenous and taking photos of them (with their consent for use in the experiment).

Weichselbaumer (2016a) obtained photos of 12 selected undergraduates, matching the relevant ethnicity to be tested, recruiting participants via email. A photo session was held to

ensure consistency of presentation and the photos were rated by undergraduates on looks, intelligence, reliability and likeability. The photos were then put through an intensive process of digital manipulation to create photos that were equally matched on all the ratings of looks, intelligence, reliability and likeability for each identity. Weichselbaumer says of this process that she is confident that ‘the photos are carefully matched so not to introduce any bias’ (page 11 working paper). Weichselbaumer (2016b) testing in Germany on the basis of Muslim religion (date not specified) had no need for evaluation as only one photo of a female ‘model’ was used with three identities created: a German, a Turk and a Turkish muslim wearing a headscarf (but not strictly religious). Ruffle and Shtudiner (2014) tested for the impact of attractiveness in Israel over the period July 2008 to January 2010 by obtaining 300 usable photos from undergraduates and eliminating those (139) that were either at the extremes of attractive/plain or conveyed ethnic identity. The remaining 161 were judged for physical attractiveness, intelligence and ethnicity by a panel of eight persons (equal numbers of women and men) of varying age and profession. For both female and male photos, the mean attractiveness ratings were used to identify the four most attractive and the four plainest. This provided Ruffle and Shtudiner with a set of 8 attractive and 8 plain applicants. A final check for the robustness of the ratings of these 16 photos was done by conducting a further evaluation by undergraduates. This confirmed the judgement of the original panel. The 16 photos (8 attractive/8 plain) were then used for the creation of the CVs.

Baert (2017) wanted to investigate the impact of social media on callback rates given the evidence that employers/human resource managers are using Facebook to check/scrutinise applicants. He used photos and their evaluation from an earlier laboratory experiment assessing perceived attractiveness and personality traits on hiring decisions conducted by Baert and Decuypere (2014). The earlier study bought 22 photos of undergraduates from a picture database and used 159 people to evaluate them on attractiveness and five personality traits (agreeableness, conscientiousness, emotional stability, extroversion and openness). Baert (2017) then calculated a single index for personality and a rating for attractiveness. Four pictures were selected; high attractiveness/moderate conscientiousness (high other personality traits); moderate attractiveness/high conscientiousness (moderate other personality traits); moderate attractiveness/moderate all personality traits; low attractiveness/low all personality traits. These photos were then used on (fake) Facebook pages and CVs for applicants in the study.

Finally, Patacchini, Ragusa and Zenou (2015) tested for discrimination on the basis of attractiveness and sexual orientation in Italy in early 2012. They collected 89 photos of 20-35 year olds, though no details are provided of how they obtained them. A panel of 24 people then assessed the photos. The logit regression results reported in Table 3, page 1059, suggest that the average rating by the panel of the ‘beauty’ of the individual in the photo was used, enabling a test for the impact of a range of attractiveness, with a set of four applicants per job vacancy: attractive/unattractive homosexual and attractive/unattractive heterosexual.

In summary, there is a mix of sources for these photos. Some have manipulated photos to digitally create individuals over a range of attractiveness; others have recruited either undergraduates or models or the general public; some have used photos available from commercial websites. Most convened a panel of independent people to evaluate the photos on attractiveness and, in the case of Weichselbaumer and Baert (2017), to evaluate a range of personality traits and other characteristics. For attractiveness studies a range of attractiveness is used for applicants; for race/ethnicity (and religion) studies, all applicants are given the same average attractiveness.

RESULTS OF THE EXPERIMENTS

In conducting a field experiment of discrimination in labour markets, responses are collected and analysed for callback rate differences between the groups studied. A probit estimate is calculated to determine if there is a statistically significant different probability of receiving a positive response between the groups. Typically, the variables included in the estimates are taken from the generated CVs such as gender, age, qualifications, work history experience, hobbies, socio-economic background and skills (such as technical, computer or language). Other variables for inclusion in the probit estimates may have been collected such as job, recruiter and employer characteristics. When photos are attached to the CV, their impact is assessed in the probit estimates by including a variable which uses the average rating for each of the photos across the group of participants the researchers used in the evaluation process. The researchers typically have a group of participants rate the photos for attractiveness, often on a 5-point likert scale of 1 for very unattractive to 5 for very attractive. Cronbach’s alpha (or another effective reliability measure) is used to evaluate the internal consistency and reliability of the ratings. Tables 1 and 2 report the results for callback rates in labor market studies of, respectively, facial attractiveness and obesity and race/ethnicity and gender.

Details of the occupations tested, number of jobs applied to, and cities tested are also in these tables.

Results for field experiments of attractiveness

Research on the impact of attractiveness for offers to job interviews has shown callback rates to be affected by physical attractiveness. Maurer-Fazio and Lei (2014) and Rooth (2009) sent a matched pair of applicants, while Patacchini *et al.* (2015) sent four applicants to job vacancies, attaching evaluated/rated photos (either digitally enhanced or an actual individual) of the applicants. López-Bóo *et al.* (2013) created sets of three applicants for men and women, one attractive, one unattractive and one no photo, sending three applications to each job vacancy, while Ruffle and Shtudiner (2014) adopted the same idea of no photo condition but sent a matched pairs of applicants, always one applicant with no photo on CV, the other with either an attractive or an unattractive photo on their CV. A range of occupations were applied to: accountant, administrative assistant, banking, business sales assistant, computer professional, customer service, finance, industrial engineering, nurse, restaurant worker, secretary, shop sales assistant, software engineer, teacher pre-school.

Tables 1 and 2 indicate a greater variation in range of callback rates recorded for males compared to females: attractive females from 12.1% to 15.74%; unattractive females from 6.6% to 12.2%; attractive males from 6.2% to 19.7%; unattractive males from 3.8% to 15.8%. López-Bóo *et al.* (2013) and Maurer-Fazio and Lei (2014) found very similar premiums for attractive versus unattractive applicants, 36% for López-Bóo *et al.* and 32.5% for Maurer-Fazio and Lei, while Rooth (2009) found a lower premium of 22.4% although, as this is reported for obese versus average-weighted, the high correlation between attractiveness and obesity raises the question of whether the effects of weight and attractiveness were separated (Rooth 2009, p. 726). There is, however, no consistent pattern to a premium for attractiveness by gender, sexual orientation or occupation. Ruffle and Shtudiner (2014) found a beauty penalty for attractive females who received a statistically significant lower callback rate than the ‘no photo’ condition (as well as lower callback than plain females), whereas attractive men received a beauty premium with the highest callback rate which was statistically significantly higher than that for plain men (as well as higher than the ‘no photo’). Patacchini *et al.* (2015), on the other hand, found a beauty premium for females (callback rates of attractive 12.1%, unattractive 9.1%) but not for males (callback

rates of attractive 3.8%, unattractive 11.8%), with the beauty premium was lower for high-skilled as compared to low-skilled females. López-Bóo *et al.*'s (2013) assessment of callback rates across six occupational groups found a statistically significant beauty premium for females in one group of administrative support, covering receptionist, secretary and customer service jobs (López-Bóo *et al.*, 2013, Table 3, p. 172). Interestingly, while the Maurer-Fazio and Lei (2014) responses indicate that, in aggregate, attractive applicants received statistically significant higher callback rates than unattractive applicants, attractiveness was trumped by gender-stereotyping of jobs. The strength of stereotyping was so strong that attractive and unattractive females (males) received higher callback rates than attractive males (females) for administrative assistant/female-dominated (software engineer/male-dominated) jobs (Maurer-Fazio and Lei 2014, Table 6 p. 25). While Rooth's (2009) tests of discrimination on the basis of obesity in the Swedish cities of Gothenburg and Stockholm found statistically significant differences in callbacks (obese/average weighted female (31.9%/39.5%) and obese/average-weighted male (29.2%/35.3%) applicants), these varied across the seven occupations included in the tests. The statistically significant differences against obese men occurred in three job areas requiring customer contact (business, sales assistants and restaurant workers) while for obese females it was in the three occupations of accountant, pre-school teacher and restaurant worker. Ruffle and Shtudiner (2014) were able to compare responses by private companies (25% of the sample) as compared to employment agencies (75% of the sample). Over 90% of their sample of CVs were vetted by females. A compelling explanation for the job penalty was that female decision makers for private companies rejected the attractive female applicant on the basis of jealousy and envy (Ruffle and Shtudiner, 2014, page 21).

Baert (2017) in a test for the impact of social media applied to just over 100 online job vacancies for six months from late November 2013 in Belgium, specifically targeting entry level jobs for those with secondary education degrees in commerce and graduates with a master's degree in commercial services. Pairs of male applications were sent: one set where the photos (randomly assigned) were embedded in the CV; another where the photo was assigned to the (fake) Facebook profile; and another set of generated CVs with no photo. Baert (2017) recorded higher callback rates for more attractive applicants, with a significant difference in callback between those highest-rated on attractiveness compared to those

lowest-rated. No difference in callback rates was found for where the photo was placed: the callback rates were similar whether the photo was on Facebook or the CV.

Results for field experiments of race/ethnicity, religion and sexual orientation

Table 2 reports the results from the field experiments of race/ethnicity, religion and sexual orientation. Each of the studies sent a different number of applications/CVs in response to job vacancies and covered a different range of occupations, so it is difficult to provide an overview as in the previous studies of attractiveness where there was some commonality of approach.

Weichselbaumer's experiments covering race/ethnicity and religion discrimination in Austria and Germany found evidence of discrimination on each of the bases tested. Over the period December 2012 to August 2013, Weichselbaumer tested for discrimination against multiple ethnic/race groups in Austria (Weichselbaumer 2016a). Race/ethnicity was signalled by photos with CVs randomised across all necessary applicant details to ensure equivalence on productivity related characteristics. In particular, all applicants had been educated in schools, and were proficient in the language of the country of residence. A matched-pair of applications was sent. The jobs applied to ranged over entry-level jobs in accounting, IT, sales and service with either all male or all female sets of applicants. Table 1 provides full details of the race/ethnic groups investigated and the type of CVs constructed. Analysis of the responses found statistically significant lower callback rates for all migrant groups. In particular, the callback rate to Nigerians of 18.7% was half that received by Austrians and further they received lower callback rates across all four occupations. Higher callback rates were recorded for Vienna as compared to outside, which is consistent with the findings from housing market tests of ethnic/race discrimination in Sweden by Carlsson and Eriksson (2014). Carlsson and Eriksson found that ethnic discrimination in housing was concentrated largely in non-metropolitan areas and exercised by Swedish landlords. Attitude surveys confirmed higher levels of prejudicial attitudes in these areas and by that group.

In her earlier study of sex discrimination during 1998 and 1999 in Vienna, Weichselbaumer (2004) sent written applications to two sex-integrated occupations (computer programmer, accountant), one female dominated occupation (secretary) and one male-dominated occupation (network technician). She found no difference in treatment between feminine and

masculine females across the occupations. However both types of females were discriminated against in the male-dominated occupation of network technician, and males were discriminated against in the female-dominated occupation of secretary. Consistent with all other correspondence tests of sex discrimination, men applying to female-dominated occupations were discriminated against on average at three times the rate as women applying to male-dominated occupations. Net discrimination rates reported in Weichselbaumer (2004), Table 5, p. 173 indicate an average 7.2% bias against females in network technicians jobs and an average 20.5% against males in secretarial jobs.

Arco-Gomez and Campos-Vázquez (2013) applied to over 1,000 online job ads for recent graduates in Mexico City, Mexico, over eight months from October 2011 to test for discrimination by phenotype - European (white), Mestizo and Indigenous. They sent 8 applications to each advertisement, usually of 4 men and 4 women: one application for each man and woman had no photo, while the other three for each of the men and women had a randomly attached picture covering the three phenotypes. The CVs controlled for all other aspects, such as education and experience. The jobs applied for covered vacancies for recent graduates with degrees in business administration, economics, engineering-computational systems, engineering-electronics and telecommunications, industrial engineering, public accounting. They found that European (white) and Mestizo phenotypes received higher callback rates than indigenous phenotypes or applicants with no photo. Table 2 shows this pattern for both women and men, with lower callback rates for all types of men compared to women. The callback rates for women and men respectively, were: European 17.1%/11.5%, Mestizo 15.8%/11.4%, Indigenous 13.8%/10.0% and no photo 12.8%/9.6%. Of interest to the correspondence testing approach is that the consistent lowest callback rates received for the 'no photo' applicants in a country where it is customary to attach a photo to a CV. Arco-Gomez and Campos-Vázquez (2013) also evaluated whether their responses had been affected by unobservable characteristics using Neumark technique. They found that the responses were robust when considered by phenotype or marital status but not by gender indicating the variance in unobservable characteristics was similar for all phenotypes and marital status, but dissimilar for females and males.

In their tests for phenotype in Lima conducted for six months from June 2011, Galarza and Yamada (2014) applied to approximately 1,200 entry-level positions in the professional,

technical and unskilled occupations. They sent 4 CVs to each vacancy using templates constructed from real CVs obtained from two employment websites. They used a database of 500 names of White (Spanish, Italian and French) and Indigenous (Quechuas) together with the photos attached to the CVs to convey phenotype. The results reported in Table 2 indicate the statistically significant higher callback rates to White as compared to Indigenous applicants, with this difference higher for females than males: White female 16.3%; Indigenous female 6.3%; White male 15.5%; Indigenous male 11.3%. (This pattern was repeated for all occupational groups, results which are not reported in the table). Probit estimates indicate a beauty premium as more attractive applicants received higher callback rates. Although Galarza and Yamada (2014) sound a note of caution for their results, due to the high correlation between physical appearance and attractiveness rating (as reported earlier), their results for phenotype cannot be distinguished from the results for the attractiveness rating. To do so would have required a study where the attractiveness of the photos was controlled to be of similar rating.

Comparing both studies of phenotype, much higher premiums for callbacks to Whites/Europeans as compared to Indigenous were found in Peru by Galarza and Yamada with callback premiums to Whites of 58.7% for females and 37.2% for males as compared to those found in Mexico City of 23.9% for females and 15.0% for males.

In their study of sexual orientation and attractiveness, Patacchini *et al.* (2015) put fake advertisements on job websites and collected CVs sent by real applicants. They created a bank of information from the CVs such as age, education, qualifications, work histories, etc. and then used a computer program to generate their CVs. As discussed earlier, they sent four CVs to each job vacancy in early 2012, covering a range of sales and clerical jobs. They found gay men were discriminated against but lesbians were not. The aggregate callback rate for gay men was 4.39% compared to 6.45% for heterosexual men. The aggregate callback rate for lesbians was 5.53% compared to 5.70% for heterosexual women. (The findings on attractiveness were presented in the previous section). Table 2 reports the callback rates for the aggregate responses only - details for the 7 occupations tested of administrative clerk, bookkeeper, call center, receptionist, sales clerk, secretary and shop assistant are available from Patacchini *et al.* (2015), Table 2, p. 1057 and for attractiveness, Figure 1, p. 1058.

ARE THERE CONFOUNDING INFLUENCES ON HIRING DECISIONS?

By including photos, the question that all these studies face is: are the responses received influenced only by the basis being tested for, such as attractiveness, phenotype, race/ethnicity, sexual orientation or religion? Alternatively, is there some other influencing characteristic unintended by the researcher, but conveyed in the photo? There is some evidence that suggests that may be the case.

Rooth (2009) in his study of obesity considered that attitudes and stereotypes may be affecting the reaction to the obese applicant. To investigate this, he recruited undergraduates in a laboratory study to assess the manipulated photos he had created of obese applicants. The students were asked to provide the first three attributes they ascribed to the photos. While the vast majority gave positive ones for the first three (such as happy/pleasant), just over half described them by weight (fat/plump) as well. This indicates other impressions besides productivity assessment may affect those making decisions on job applicants when photos are attached to a CV.

Further, work by Agerström, Björklund, Carlsson and Rooth (2012) indicates other characteristics of an individual that could be signaled in a personal profile are important to hiring decisions. They tested the impact of social judgments on warmth and competence, which have been found to be important by social psychologists, on the callback rate for job interviews. In a test of race/ethnic discrimination they applied to over 5,600 jobs on a Swedish online jobsite sending a pair of CVs, one from a Swede (in-group) and the other from an Arab (out-group). (It is not clear when the tests were conducted as this project was part of a larger project, Agerström *et al.*, p. 361). They applied to 12 occupational categories: accountants, business sales assistants, cleaners, computer professionals, construction workers, high school teachers, mechanics, motor vehicle drivers, nurses, primary school teachers, restaurant workers, shop sales assistants. The CVs were randomly generated using a computer program. They manipulated warmth and competence by attaching a letter of interest to the job application, creating combinations of high versus low warmth and high versus low competence. Those statements were evaluated by 84 students. This established that the intended characteristics were adequately conveyed and that there was an interaction between these characteristics, in that a higher perceived warmth was compensated with a lowered perceived competence. In aggregate, the tests recorded higher callback rates for

applicants signaling higher competence. Swedish applicants received higher callback rates than Arab applicants. Interaction effects found that this was due to a differential impact of being warmer and more competent for an Arab versus a Swede: “Only when the Arab applicants are both warmer and more competent than the Swedish applicants are the two ethnic groups invited for interview equally often” (Agerström *et al.* 2012, p. 364). There are two implications of this study for field experiments attaching photos to CVs in job applications. The first is that the characteristics of warmth and competence need to be considered when using photos on CV. The second is whether these characteristics are able to be controlled and evaluated accurately.

Research in product markets using photos in studies of peer-to-peer lending is also useful in reflecting on what characteristics photo images convey. Three research studies of crowdfunding sites are discussed here. Duarte, Siegel and Young (2012) studied the peer-to-peer lending site *prosper.com*. They investigated the impact of judgments about an individual’s trustworthiness, based on photographs of the individual, on success in obtaining loans. Duarte *et al.* had photos of borrowers on *prosper.com* rated by one group of evaluators for trustworthiness and by a separate group of evaluators for attractiveness and competence. They found that trustworthiness as judged by facial appearance predicted actual credit outcomes. The expected returns of borrowers perceived as more trustworthy were significantly higher as compared with the expected returns of borrowers rated as more attractive (Duarte *et al.*, 2012, p. 2481). In a study of discrimination, which also investigated *prosper.com*, Pope and Sydnor (2011) investigated the impact of photos that individual borrowers had provided on the probability of obtaining funding (where only 46% of individuals provided a photo). They employed undergraduates to rate the photos of individuals for a number of characteristics, actual and perceived, such as gender, race, age, happiness, weight and attractiveness (Pope and Sydnor, 2012, p. 64). Their estimates “imply that listings with a picture of an apparently black or African-American person are 3.2 percentage points less likely to get funded than an equivalent listing with a picture of a white person” (Pope and Sydnor, 2011, p.68). Listings with a picture of a woman were 1.1 percentage points more likely to get funding. Interestingly, attractiveness was not found to be a significant predictor of funding. A recent study by Rich, Cox, Nguyen and Morrison (2017) of the UK equity crowdfunding site ‘crowdcube’ used the photos of the crowdfunders along with all project details to analyse whether facial attractiveness, perceived trustworthiness and

perceived dominance were significant factors influencing the successful funding of the project. The photos of project crowdfunders were rated for all these features by the same participants in the evaluation. The average rating for each of the three facial characteristics were then used as the variables to capture attractiveness, trustworthiness and dominance. Of the three, only perceived dominance was found to be weakly significant in influencing the successful funding of the project.

The research on crowdfunding sites raises a note of caution for discrimination experiments in labour markets that include photos. There is evidence in experimental work of a range of characteristics that photos convey of individuals and their impact on labour and product market outcomes, such as success in a positive response to job applications and in obtaining funding to finance projects. These photos need to be judged for a number of facial characteristics, not just attractiveness, as well as considering using the personality index, as in Baert (2017).

CONCLUDING REMARKS

The discussion and reflection here on the use of photos in field experiments provides some tentative answers to three questions posed at the outset of the paper. First, is there a problem of introducing further unobservable characteristics by using photos? The research by Agerström *et al.* (2012), Duarte *et al.* (2012), Pope and Sydnor (2011) and Rich *et al.* (2017) together with Baert (2017) and Weichselbaumer (2004) indicates that a range of characteristics, other than attractiveness, can affect the assessment of an individual via a photo. Attaching photos to CVs in application for jobs then introduces further unobservable characteristics. Second, can we fully control the experimental approach when using photos? Given the range of characteristics of an individual introduced by use of a photo, the cautious answer is that we cannot fully control the experiment for the impact of the basis being tested, such as race/ethnicity or sexual orientation, as we cannot fully control for all productivity-related characteristics. Thirdly, are we able to accurately evaluate the impact of the photos used on the response/probability of callback? As we can accurately evaluate the characteristics identified of attractiveness, personality traits, warmth, competence, trustworthiness, dominance and sex stereotyping, the tentative answer to this question is yes. So field experiments using photos can accurately evaluate photos and need to ensure they do this for the range of factors that have been shown to affect judgments and therefore

potentially influence call back response. However, the issue remains whether we have in fact identified all potential characteristics conveyed by the photos. The results of the studies covered in this paper, although to a lesser extent those of Weichselbaumer (2004) and Baert (2017), are likely compromised by including photos, with the possibility that the responses received were influenced by some characteristic unintended by the researcher but conveyed by the photo.

Given this, the following suggestions can be made for researchers conducting future experiments, particularly given that there are a variety of reasons for using photos. If researchers test for the impact of attractiveness on market outcomes (which could be, amongst others, job interviews, job offers, property rental viewings, funding opportunities), careful evaluation of photos is necessary, controlling for the impact of other characteristics, such as warmth, competence, trustworthiness, dominance and a range of personality traits, that are made in facial judgments. On the other hand, researchers may need to attach photos to enquiries because it is the convention in the country, although their study is concerned to test for discrimination on the basis of race/ethnicity, religion, gender, sexual orientation etc., in which case they need to isolate the impact from the photos by ensuring they are standardized to an average evaluation of the various facial characteristics. They could, of course, include photos covering a range of attractiveness etc. but this would necessarily complicate the findings on any differential treatment found. Analogously, researchers testing for phenotype and signaling this by using an image of the applicant, need to evaluate the photos of the various phenotypes for facial characteristics and, again, decide whether or not to standardize to an average evaluation. The ‘no photo’ condition used by Arco-Gomez and Campos-Vázquez (2013), Baert (2017), López-Bóo *et al.* (2013) and Ruffle and Shtudiner (2014) could be incorporated in the application design to provide a comparison for the callback rates of all applicants. Finally, all studies should evaluate their responses for any bias that may arise from unobservable characteristics (as Arco-Gomez and Campos-Vázquez 2013 did), in particular any bias introduced by photos in the design of the experiment by applying the Neumark technique to the probit estimates of the likelihood of a positive response to the applicants.

Table 1: Results for Callback Rates in Labour Market Studies - Facial Attractiveness and Obesity

Country and Study	Basis of preference test	Occupation	Applicant group	Callback Rate ¹ %
<i>China</i>				
Maurer-Fazio and Lei (2014) Conducted 2012 Beijing Chengdu Nanjing Shanghai Shenzhen Wuhan Applied to 12,096 jobs	Attractiveness /Gender	All occupations	Attractive Female	14.27
			Unattractive Female	10.10
			Attractive Male	12.27
			Unattractive Male	9.92
		Accountant	Aggregate Attractive	13.27
			Aggregate Unattractive	10.01
			Attractive Female	10.52
			Unattractive Female	7.80
		Admin assistant	Attractive Male	8.33
			Unattractive Male	6.61
			Attractive Female	12.76
			Unattractive Female	6.61
		Sales representative	Attractive Male	6.22
			Unattractive Male	4.63
			Attractive Female	18.06
			Unattractive Female	13.82
		Software engineer	Attractive Male	16.01
			Unattractive Male	12.63
			Attractive Female	15.74
			Unattractive Female	12.17
<i>Israel</i> Ruffle and Shtudiner (2014) Conducted July 2008- January 2010 Applied to 2,656 jobs	Attractiveness	All occupations	Attractive Female	12.8
			Plain Female	13.6
			No photo Female	16.6
			Attractive Male	19.7
		Accounts manager	Plain Male	9.2
			No photo Male	13.7
			Attractive Female	22.0
			Plain Female	16.0
			No photo Female	18.0
			Attractive Male	16.0
Plain Male	5.0			
No photo Male	8.0			

1. Difference in callback rate to the majority applicant for like CV: significant at the 0.05*; 0.01 **; 0.001***

Table 1: Results for Callback Rates in Labour Market Studies - Facial Attractiveness and Obesity continued

Country and Study	Basis of preference test	Occupation	Applicant group	Callback Rate ¹ %
<i>Israel</i>				
Ruffle and Shtudiner (2014) continued	Attractiveness	Banking	Attractive Female	16.0
			Plain Female	15.0
			No photo Female	14.0
			Attractive Male	13.0
			Plain Male	13.0
			No photo Male	12.0
		Budgeting	Attractive Female	2.0
			Plain Female	2.0
			No photo Female	12.0
			Attractive Male	14.0
			Plain Male	2.0
			No photo Male	7.0
		Chartered accountant	Attractive Female	6.0
			Plain Female	6.0
			No photo Female	6.8
			Attractive Male	0.0
			Plain Male	6.0
			No photo Male	6.0
		Customer service	Attractive Female	14.0
			Plain Female	23.0
			No photo Female	21.0
			Attractive Male	31.0
			Plain Male	13.0
			No photo Male	21.0
		Finance	Attractive Female	9.0
			Plain Female	6.0
			No photo Female	7.0
			Attractive Male	4.0
			Plain Male	6.0
			No photo Male	8.0
		Industrial engineering	Attractive Female	9.0
			Plain Female	14.0
			No photo Female	27.0
			Attractive Male	27.0
			Plain Male	9.0
			No photo Male	16.0
Programming	Attractive Female	11.0		
	Plain Female	7.0		
	No photo Female	14.0		
	Attractive Male	36.0		
	Plain Male	11.0		
	No photo Male	14.0		

1. Difference in callback rate to the majority applicant for like CV: significant at the 0.05*; 0.01 **; 0.001***

Table 1: Results for Callback Rates in Labour Market Studies - Facial Attractiveness and Obesity continued

Country and Study	Basis of preference test	Occupation	Applicant group	Callback Rate ¹ %			
<i>Israel</i>							
Ruffle and Shtudiner (2014) continued	Attractiveness	Sales-junior	Attractive Female	13.0			
			Plain Female	16.0			
			No photo Female	18.0			
			Attractive Male	32.0			
			Plain Male	16.0			
			No photo Male	23.0			
		Sales-senior	Attractive Female	10.0			
			Plain Female	11.0			
			No photo Female	19.0			
			Attractive Male	13.0			
			Plain Male	7.0			
			No photo Male	10.0			
			<i>Sweden</i>				
			Rooth (2009) Conducted 2006 Gothenburg, Stockholm Applied to 985 jobs	Obesity	All occupations	Obese Male	29.2
Average-weighted Male	35.3						
Accountant	Obese Male	5.7					
	Average-weighted Male	7.6					
Business sales assist	Obese Male	37.5					
	Average-weighted Male	51.3					
Computer professional	Obese Male	22.2					
	Average-weighted Male	30.6					
Nurse	Obese Male	62.0					
	Average-weighted Male	60.8					
Restaurant worker	Obese Male	15.1					
	Average-weighted Male	23.7					
Shop sales assistant	Obese Male	9.9					
	Average-weighted Male	14.4					
Teacher pre-school	Obese Male	52.0					
	Average-weighted Male	58.7					
Obesity	All occupations	Obese Female			31.9		
		Average-weighted Female		39.5			
	Accountant	Obese Female		16.0			
		Average-weighted Female		30.0			
	Business sales assist	Obese Female		33.6			
		Average-weighted Female		39.8			
	Computer professional	Obese Female		29.7			
		Average-weighted Female		21.6			
	Nurse	Obese Female		51.1			
		Average-weighted Female		46.7			
	Restaurant worker	Obese Female		16.9			
		Average-weighted Female		39.4			
	Shop sales assistant	Obese Female		6.6			
		Average-weighted Female		9.1			
	Teacher pre-school	Obese Female		47.7			
		Average-weighted Female		56.0			

1. Difference in callback rate to the majority applicant for like CV: significant at the 0.05*; 0.01 **; 0.001***

Table 2: Results for Callback Rates in Labour Market Studies – Gender, Race/Ethnicity and Phenotype

Country and Study	Basis of preference test	Occupation	Applicant group	Callback Rate ¹ %
<i>Austria</i>				
Weichelsbaumer (2004) Conducted 1998-1999 Applied to 462 jobs Vienna	Gender	Accountant	Male	73.0
			Masculine female	63.0
			Feminine female	58.0
		Computer programmer	Male	81.0
			Masculine female	85.0
			Feminine female	81.0
			Male	40.0
			Masculine female	40.0
			Feminine female	43.0
		Network technician	Male	20.0
			Masculine female	46.0
			Feminine female	44.0
Weichelsbaumer (2016a) Conducted December 2012- August 2013 Applied to 1071 jobs all over Austria	Race/ethnicity Migrant	Aggregate ²	Austrian	37.0
			Serbian	28.2***
			Turkish	25.3***
			Chinese	27.1***
			Nigerian	18.7***
<i>Germany</i>				
Weichelsbaumer (2016b) Date of tests not specified, conducted for 1 year Applied to 1,474 jobs Berlin, Cologne, Dresden, Frankfurt, Hamburg, Munich, Stuttgart	Turkish ethnicity Religion/Muslim	Aggregate	German	18.8
			Turkish	13.5**
			Turkish with headscarf	4.2***
		Accountant	German	29.0
			Turkish	19.0
			Turkish with headscarf	5.8***
		Chief accountant	German	27.3
			Turkish	19.0
			Turkish with headscarf	3.6***
		Secretary	German	12.8
			Turkish	9.6
			Turkish with headscarf	3.7***
<i>Italy</i>				
Patacchini, Ragusa and Zenou (2015) Conducted January-February 2012 Applied to 531 jobs Rome, Milan	Sexual orientation/ attractiveness	Aggregate/Sexual orientation	Lesbian	5.5
			Heterosexual female	5.7
			Gay male	4.4**
			Heterosexual male	6.5

1. Difference in callback rate to the majority applicant for like CV: significant at the 0.05**; 0.01 ***

2. Callback rates by occupation in Weichelsbaumer 2016b, Table 4, page 250

Table 2: Results for Callback Rates in Labour Market Studies - Gender, Race/Ethnicity and Phenotype continued

Country and Study	Basis of preference test	Occupation	Applicant group	Callback Rate ¹ %
<i>Italy</i>				
Patacchini, Ragusa and Zenou (2015) continued	Sexual orientation/ attractiveness	Aggregate/ Attractiveness	Attractive Female	12.1
			Average Female	11.9
			Unattractive Female	9.1
			Attractive Male	11.8
			Average Male	11.2
		Aggregate/Sexual orientation/ Attractiveness	Unattractive Male	3.8
			Lesbian -Attractive	14.6
			Lesbian - Average	11.7
			Lesbian – Unattractive	8.4
			Hetero female -Attractive	10.0
			Hetero female - Average	12.0
			Hetero female – Unattractive	9.6
			Gay male -Attractive	6.1
			Gay male - Average	9.7
			Gay male – Unattractive	11.1
Hetero male -Attractive	9.7			
Hetero male - Average	12.7			
Hetero male - Unattractive	12.3			
<i>Mexico</i>				
Arceo-Gomez and Campos-Vázquez (2014) Conducted October 2011- May 2012 Applied to approx 1,020 jobs Metropolitan area of Mexico City	Phenotype	Aggregate	European female	17.1
			Mestizo female	15.8
			Indigenous female	13.8**
			No photo female	12.8**
			European male	11.5
			Mestizo male	11.4
			Indigenous male	10.0**
			No photo male	9.6**
			<i>Peru</i>	
Galarza and Yamada (2014) Conducted June –December 2011 Applied to 1,205 jobs Lima	Phenotype	Aggregate ³	White female	16.3
			Indigenous female	6.3***
			White male	15.5
			Indigenous male	11.3***

3. Callback rates by occupation in Galarza and Yamada 2014, Table 1, page 87

References

- Agerström, J., Björklund, F., Carlsson, R. and Rooth, D-O. (2012). “Warm and competent Hassan = cold and incompetent Eric: A harsh equation of real-life hiring discrimination”. *Basic and Applied Social Psychology*, Vol. 34, No. 4, pp.359-366
- Arceo-Gomez, E. and Campos-Vázquez, R. (2014). “Race and marriage in the labor market: A discrimination correspondence study in a developing country”. *American Economic Review*, Vol. 104, pp. 376-380.
- Baert, S. (2017). “Facebook profile picture appearance affects recruiters’ first hiring decisions”. *New Media & Society*, pp. 1-20.
- Baert, S. and Decuyper, L. (2014). “Better sexy than flexy? A lab experiment assessing the impact of perceived attractiveness and personality traits on hiring decisions”. *Applied Economics Letters*, Vol. 21, No. 9, pp. 596-601.
- Carlsson, M. and Eriksson, S. (2014). “Discrimination in the rental market for apartments”. *Journal of Housing Economics*, Vol. 23, pp. 41-54.
- Daniel, W. (1968). *Racial Discrimination in England*. Middlesex: Penguin Books.
- Duarte, J., Siegel, S. and Young, L. (2012). “Trust and credit: The role of appearance in peer-to-peer lending”. *The Review of Financial Studies*, Vol. 25, No. 8, pp. 2455-2483.
- Galarza, F. and Yamada, G. (2014). “Labor market discrimination in Lima, Peru: Evidence from a field experiment”. *World Development*, Vol. 58, June, pp. 83–94.
- Heckman, J. (1998). “Detecting discrimination”. *Journal of Economic Perspectives*, Vol. 12, pp. 101-116.
- Heckman, J. and Siegelman, P. (1993). “The Urban Institute audit studies: Their methods and findings”., in Fix, M. and Struyk, R. (Eds.), *Clear and Convincing Evidence: Measurement of Discrimination in America*, Washington, D.C.: The Urban Institute

Press, pp. 187-258.

Jowell, R., and Prescott-Clarke, P. (1970). "Racial discrimination and white-collar workers in Britain". *Race*, Vol. 11, No. 4, pp. 397-417.

Lopez Bóo, F., Rossi, M. and Urzúa, S. (2013). "The labor market return to an attractive face: Evidence from a field experiment". *Economics Letters*, Vol. 118, pp. 170–172.

Maurer-Fazio, M. and Lei, L. (2014). ""As rare as a panda": How facial attractiveness, gender and occupation affect callbacks at Chinese firms". *IZA Discussion Paper* No. 8605.

Neumark, D. (2018). "Experimental research on labor market discrimination". Forthcoming *Journal of Economic Literature*.

Neumark, D. (2012). "Detecting discrimination in audit and correspondence studies". *Journal of Human Resources*, Vol. 47, No. 4, pp. 1128-157.

Neumark, D. and Rich, J. (2018). "Do field experiments on labor and housing markets overstate discrimination?: A re-examination of the evidence". Forthcoming *Industrial and Labor Relations Review*.

Patacchini, E., Ragusa, G., and Zenou, Y. (2015). "Unexplored dimensions of discrimination in Europe: homosexuality and physical appearance". *Journal of Population Economics*, Vol. 28, pp. 1045-1073.

Pager, D. (2007). "The use of field experiments for studies of employment discrimination: Contributions, critiques and directions for the future". *The Annals of the American Academy of Political and Social Sciences*, Vol. 609(1), pp. 104-133.

Pope, D. and Sydnor, J. (2011). "What's in a picture: Evidence of discrimination from prosper.com". *Journal of Human Resources*, Vol. 46, No. 1, pp. 53-92.

- Quillian, L., Pager, D., Hexel, O. and Midtbøen, A. (2017). “Meta-analysis of field experiments shows no change in racial discrimination in hiring over time”. PNAS Early Edition. www.pnas.org/cgi/doi/10.1073/pnas.1706255114
- Rich, J. (2014). “What do field experiments of discrimination in markets tell us? A meta analysis of studies conducted since 2000”. IZA Discussion Paper No. 8584. Bonn, Germany: IZA.
- Riach, P. A. and Rich, J. (2002). “Field experiments of discrimination in the market place”. *Economic Journal*, Vol. 112(483), pp. F480-F518.
- Rich, J., Cox, J., Nguyen, T. and Morrison, E. (2017). “Do attractiveness or perceived trustworthiness influence the crowd?”. (working paper).
- Rooth, D-O. (2009). “Obesity, attractiveness, and differential treatment in hiring: A field experiment”. *Journal of Human Resources*, Vol. 44, No. 3, pp. 710-735.
- Ruffle, B. and Shtudiner, Z. (2014). “Are good-looking people more employable?”. *Management Science*, Vol. 61, pp. 1760-1776.
- Ward, R. (1969). “A note on the testing of discrimination”. *Race*, Vol. 11, pp. 218–33.
- Weichselbaumer, D. (2016a). “Discrimination against migrant job applicants in Austria: An experimental study”. *German Economic Review*, Vol. 18, No. 2, pp. 237-265.
- Weichselbaumer, D. (2016b). “Discrimination against female migrants wearing headscarves”. IZA Discussion Paper Series 10217.
- Weichselbaumer, D. (2015). “Testing for discrimination against lesbians of different marital status: A field experiment”. *Industrial Relations*, Vol. 54, pp. 131–161.

Weichselbaumer, D. (2004). "Is it sex or personality? The impact of sex-stereotypes on discrimination in applicant selection". *Eastern Economic Journal*, Vol. 30, pp. 159-186

Zschirnt, E. and Ruedin, D. (2016). "Ethnic discrimination in hiring decisions: A meta-analysis of correspondence tests 1990-2015". *Journal of Ethnic and Migration Studies*, Vol. 42, No. 7, pp. 1115-34.