

# Discrimination due to Outdated Decision Rules: A Field Study

-

## EXTENDED ABSTRACT

please do not cite or distribute without permission of the authors

Nina Stephan\*

University of Paderborn

April 15, 2016

### Abstract

This paper describes the problem of recruiters' decisions on hiring being based on heuristics. Even if simplifying rules of choice-making help the decision maker to avoid bad hires when information is incomplete and time is scarce, decision rules should at least be formed based on up to date information. However, it seems that recruiters in real life still make decisions based on simple rules that stem from their own time as young professionals. As an example for the failure of updating decision relevant information, this paper looks into discrimination against tattooed applicants in Germany. While in Germany tattoos would still be more widespread among certain groups of young people 30 years ago, they are almost equally spread among all socio-economic groups today. Consequently, it would no longer be rational to statistically discriminate against tattooed applicants, since they are on average equally qualified as people without a tattoo. The hypothesis is tested by running a field experiment in form of a correspondence study in the German labor market for vocational training jobs, where photographs are a must-have on any standard application. Results of a test run confirm that applicants with a visible tattoo on their application photograph, are indeed significantly less likely to be called back for an interview.

Keywords: discrimination, labor market, heuristics, correspondence study, field experiment

JEL-Codes:

---

\*nina.stephan@upb.de

# 1 Motivation

This paper shows that discrimination due to outdated decision-rules, applied by recruiters to make their choice over applicants, are an issue in the German labor market. The focus is not to distinguish between recruiters who actively discriminate since they have a taste against a certain applicant characteristic and those who discriminate since they assume certain applicant characteristics to be correlated with low ability. Rather, it focuses on the fact that recruiters use standardized decision rules when deciding to give applicants a chance, and that irrespective of the recruiter having a taste against a certain group of people or not, this poses a problem.

Imagine a recruiter while he is doing his job. To some extent, he will be interested in doing a good job by finding the most able applicant. At the same time, he knows that he will never be perfectly informed about all applicants and he will be constrained in time. On top of that, he may have a taste in favor of or against certain groups of people, for example because he likes to be surrounded by certain types of colleagues during work. No matter whether the recruiter is actually trying to do what is best for the company or also has an own agenda, he is limited when it comes to making the perfect decision. Information about the recruiter's ability will always be imperfect to some extent, and eventually the recruiter will infer from the information he has, that one applicant is suited better than another. A recruiter without an own agenda is likely to do this based on previous hiring experience and on perceived knowledge about statistical facts regarding some known applicant characteristics being correlated with unobserved characteristics that are relevant for ability. While at first glance this way of decision-making may seem totally reasonable, it can be problematic in real life. First of all, the position of a recruiter is usually only reached after a certain career, such that many recruiter positions are filled with rather old people, while most applicants are typically at the job entry level and thus belong to a younger generation. Older recruiters have more working experience than younger ones, and may not feel the need to still regularly investigate in statistics. They probably feel safe deciding for applicants with characteristics that they find in colleagues who have turned out to also do well at their jobs. And this does become a problem once applicants belong to another generation, where characteristics that used to go along with low ability are now going along with average or high ability, or vice versa. In that moment, decision rules are outdated, and the failure of updating them leads to a worse match between employer and employee, and to discrimination against a certain group of people. According to a review study by Riach and Rich (2002), discrimination simply remains an economic inefficiency, and thus a problem to be solved, as long as it means that recruiters do not choose the best man or woman for the job - no

matter what reason. Hence, this study aims at uncovering that discrimination as a consequence of outdated decision-rules is actually an issue on the German labor market.

Besides the obvious explicit statistical discrimination — where some might argue that on average at least it does help finding the better people for the job, or helps avoiding to hire the wrong person for a job, even if you may give up finding a raw diamond - previous literature has already tried to describe some more subtle forms of discrimination, called implicit discrimination (Neumark, 2011; Rooth, 2010). Implicit discrimination, as opposed to explicit discrimination, happens without the recruiter necessarily being aware of it, for example through —automatic association— as described by Rooth 2010. Some recruiters may also hold on to outdated heuristics, for example concerning people’s looks. Today, people in general dress less conservatively, extreme haircuts are no longer reserved for punks. However, a recruiter with a long history of experience may have had negative experiences with specific types of people and may still hold on to these beliefs, even though they are no longer making sense. The outer appearance of employees is one area, where a lot may have changed in the past decades in Germany. This is why the paper at hand will focus on the outer appearance of people as an example of an applicant characteristic that meant something else 30 years ago than it does today. I claim that recruiters do infer unobservable information from looking at applicants’ photographs, and that the effect of the first visual impression should not be underestimated. When investigating on discrimination based on outer appearance, the German labor market is an ideal place for a field experiment, as will be explained in detail below. The experiment will show that candidates with a modified application photograph, the modification consisting in a visible tattoo on the neck, have significantly lower chances of being invited for a job interview.

## **2 Experimental Design**

Typically, as described by Rich (2014) or Bertrand and Duflo (2016), field studies who aim at revealing a case of discrimination on the labor market, so called correspondence studies, create several application profiles with a set of random characteristics and try to match them into pairs of equally well qualified applicants but with different identities. On top of the identities being different, the treatment characteristic is manipulated in one out of two applications sent to the same company. Of course, as a large number of such applications is created, on average all candidates are not only equally well qualified but do also have the same average characteristics.

Comparing the groups of applicants with and without the treatment criterion does then reveal the effect of the treatment. In countries as the US this method works relatively well, as applications do not include too many details or official documents. That way, a relatively large number of different applications of equal quality can be generated randomly and each application is still consistent in itself.

This would not work with German applications. First, in Germany it is expected that any application is accompanied by a relatively large number of additional information regarding the candidate, including not only a CV with a photograph, but also school certificates (Kaas & Manger, 2012; Weichselbaumer, 2015). Applications without a photograph are often categorized as incomplete and candidates may even remain unconsidered. Moreover, creating many different identities would be much more costly and effortful as a result.

Second, critics of correspondence studies claim that relative to the gains through new insights on what is happening on labor markets, the costs such studies create on the side of the recruiting companies is too large. Keeping this critique in mind it makes sense to, if at all, only bother a company with as little fake applications as possible. Especially since a study of equal size as some US studies would be disproportionately more costly to the relatively smaller German economy.

Third and maybe most importantly, experiments are likely to be detected as two very similar applications get in at a similar point in time. As explained by Weichselbaumer (2015) German companies are sensitive for such kind of investigations and may, once they suspect an investigation for discrimination, adjust to a different way of dealing with applicants, for example by calling every single applicant and asking for an additional detail. Weichselbaumer therefore advises not to send several similar applications to one company.

Despite all of the above, the German labor market also has its advantages when it comes to running a correspondence study. On the one hand, a large amount of characteristics being specified in the application leaves relatively little variables open for the recruiters to speculate about, which rules out many alternative stories. On the other hand, as a photograph is always included in an application, manipulations of the outer appearance can be signaled directly.

As a result, the experiment at hand is designed in a different way than that used in most other studies. Instead of creating applications of equal quality but different identities the present study looks for pairs of two very similar job advertisements of different companies. Then, two applications of the same applicant are sent

to the two companies respectively. The only thing that will differ between the two applications will be the treatment modification on the photograph. That way, looking at each pair on its own we can already learn something about equal or unequal treatment. Of course, even if two job advertisements are matched on a large variety of characteristics, they are never the same.

However, on average the two groups of companies, those who received treatment and those who received non-treatment applications, will not differ from each other. Taking all cases together one can thus draw conclusions regarding discrimination based on outer appearance without running the risk of being detected. Even if a recruiter happens to learn that a person of the same name has also applied at a competitor's company, this would not raise his awareness, as real applicants do usually apply for a job at more than one company as well. It would only make the application appear more realistic.

## **2.1 Treatment**

The photographs on the applications are actual professional photographs of a real person, of which a manipulated second version for the treatment group is created using Photoshop. The photograph must avoid to signal that the applicant does not care enough to create a professional application, but at the same time the treatment cannot be too small, as it may otherwise be overlooked. Therefore, a tribal tattoo was chosen for the trial run photograph, which covers the entire neck of the candidate .

## **2.2 Job Advertisements**

As higher level jobs require even more extensive applications, this study focusses on entry-level jobs. In Germany, vocational-training jobs are suitable for this purpose out of several reasons. First, a relatively large number of very similar job advertisements for vocational trainings are published every month. This does not only raise the potential number of comparable observations but also makes it easier to find matching job advertisements.

Second, vocational training jobs usually start around the first of February or first of October of a year, which again makes it easier to match job advertisements, but also makes it less likely that seasonal changes in macroeconomic factors change, which would bias the results if they were not controlled for in the model. Third,

the area of vocational training job markets, also called second order job markets, are currently largely debated in German speaking countries. It is therefore not only contributing to the investigation of discrimination but also contributing to a deeper understanding of such second order job markets. Lastly, applications to vocational training jobs are relatively easy to write and can be prepared by young student assistants, who are at a similar age as comparable true applicants. Format and style of the applications are based on brochures from the Bundesagentur für Arbeit, the federal employment agency. Student assistants are instructed to aim at rather giving a picture of an overqualified candidate, and not an underqualified one. This is done to avoid a large number of cases where both applicants of a pair of job advertisements are rejected, which at first sight appears to be a case of fair equal treatment. However, on a second sight it does not reveal anything about whether both were rejected for the same reason (which would lead us to believe in non-existence of discrimination), or whether they simply had a much better alternative candidate, but would have discriminated against the tattooed applicant otherwise. One does simply do not learn anything from these applications. While in the test run only data from a female candidate were collected, both male and female candidates will be represented in the applications sent out now. In order to appear realistic, only female applicants will apply to those jobs where in the past 85% or more of the applicants were female, and only male applicants will apply to those jobs where in the past 85% or more of the applicants were male. Most jobs will still be applied for by both genders. As it is widely known that men and women are in most cases discriminated against in different ways (Rich, 2014; Bertrand & Mullainathan, 2004) the results of this study are likely to differ between genders, and therefore one should be looking at both. It may be that beauty matters more in women, and as a result a tattoo on a beautiful woman's neck does not make a difference, but on a man's neck it does. On the other hand the results may reveal that men's looks, as opposed to women's, do not matter at all, and hence recruiters would not even notice a tattoo on their neck. In both cases a difference in callback rates of tattooed men and women is likely, but we have to wait for more results before either can be confirmed.

### **2.3 Timing and Regional Coverage**

While the test run of 2015 was only conducted in two states, Hamburg and Nordrhein-Westfalen, the applications in the main study are sent out in four different states (Hamburg, Niedersachsen, Nordrhein-Westfalen, and Bayern), starting in March 2016 and presumably until February 2017. Covering more than one region is necessary as Weichselbaumer's 2015 study showed that discrimination in the

labor market does differ a lot between states in Germany. Of course, covering these four states does not mean that all the country is represented, however, covering more territory would require creating more identities, as the typical applicant does not apply very far away from his current residential area, and it would thus not be realistic to let the same candidate apply for jobs all over the country. Of course, confirming the validity of results for other regions of the country would be a valuable addition to the work at hand.

Technically, applications can be sent out at any time as job ads appear between half a year and a few weeks before the start of the job. However, a pair of matching job ads must be found, which ideally appeared published around the same date and for which applications are expected by roughly the same date. Moreover, to increase the quality of the matching it would not make sense to match pairs of job ads and send out applications every day, as the best match for one job advertisement may appear just a day later. There is thus a trade-off between matching applications as often as possible and waiting for a longer time before making matches but to then make them match fully. The aim of this study is to find the best matches of every week, by looking for new job advertisements on a weekly basis, and also sending out job applications on a weekly basis.

## **2.4 Test Run**

In the trial run of the experiment conducted in 2015, 40 paired applications were sent out to paired job openings, where the only difference between a treatment-group application and a control-group application consisted in a modified application picture on the CV. As a photograph is a must-have on a German application, one can conclude that recruiters hope to draw some additional information on the qualification of the candidate from it, thus they do consider the photograph when they make their choice. Studies investigating the effect of handsomeness, as Hamermesh & Biddle (1994) and many studies afterwards confirm that looks make a difference. As mentioned before, the modification here consisted in a tribal tattoo photoshopped onto the neck of the applicant. Even though nowadays tattoos appear in all kinds of socioeconomic groups in Germany (Brähler & Hofmeister, 2009; Dillingh et Al., 2016), it is likely still considered a negative signal rather than a positive one regarding the candidate's qualification, since formerly prisoners or seamen were traditionally tattooed, and these groups may not be very qualified. The results of the test run confirm these thoughts, and verify the salience of the tattoo as a signal.

## **3 Results**

### **3.1 Test Run Results**

Of course, so far only first results from the test run can be described at this point in time. While the callback rate for applications with regular photographs was 45%, the callback rate for applications with modified photographs was only 15%. This is a statistically significant difference of 30 percentage points, which means that sending out the same amount of applications candidates with a regular photograph will be called back three times as often. The difference between the callback rates is much larger in the state of Hamburg than in the state of Nordrhein-Westfalen, which would confirm the necessity of looking at different regions to get an accurate picture of what is going on. Due to the high significance of the outcomes it seems justifiable to run this experiment at a larger scale to confirm the validity of the results. It is not unlikely that it will be confirmed that people with an unfavorable outer appearance will have less chances in the labor market despite being equally well qualified, as the results of the test run suggests.

After sending out applications everything is neatly protocolled, and an answer is waited for. Recruiters may answer via e-mail or phone-call. Both e-mail and online phone-numbers are created for this reason. In order to hold the time-costs of companies who are affected by the study as low as possible, a simple rejection of the callback is sent to the recruiters within the next two days. Phone-calls are answered by voice-mails and rejected in the same way.

### **3.2 Future Results**

While in the test run only data from a female candidate were collected, both male and female candidates will be represented in the applications sent out now. Moreover, a wider range of regions and companies will be covered, giving a more complete picture of what is going on. This will be described by looking at subsets of the bigger 2016 sample.

In order to find out whether the initial thoughts of this work and the analysis of the results are in line with the actual behavior of recruiters, a recruiter survey, similar to that conducted by Oreopoulos (2011) will be run. By asking a random subsample of recruiters, who will previously have participated in the experiment, to fill out a simple questionnaire, a closer look at what is actually going on during the decision process can be taken. The questionnaire will include asking the recruiters about preferable candidate characteristics and will also ask them to state to which

degree they support certain methods for choosing the best qualified candidate.

## **4 Conclusion**

Discrimination as a consequence of outdated decision rules is an issue in the German labor market. In other words, recruiters do read more from applicant photographs than what the picture actually reveals, and recruiters draw wrong conclusions regarding the ability of applicants from looking at them. Discrimination is thus posing a problem, even if it is only in the form of implicit discrimination. Consequently, both employees and recruiters could be better off if the latter were made aware of the heuristics or automatic associations which they base their decisions on. The best skilled employees would have higher chances to get hired and the companies would more likely end up with the actually most suitable candidate. In order to reveal this problem, the study at hand will show that a modification on an application photograph, which does not make a difference for the qualification of the candidate, does make a difference for the likelihood of being invited to a job interview. The expected findings confirm an inefficiency of the German labor market. It is the aim of the paper to point this problem out as a general one, rather than to point out an exact magnitude of discrimination. Of course, any conclusion drawn now is only based on a trial run of the planned experiment. However, the first observations do confirm the hypothesis that recruiters discriminate against applicants with a visible tattoo on their application photograph. A final analysis of the empirical evidence can only be expected after the data collection is completed. It will then be confirmed whether applicants with a manipulated photograph on their CVs do actually have lower callback rates than applicants with regular photographs, and thus, whether outer appearance matters, since recruiters' decision rules are actually outdated.

## 5 References

- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *The American Economic Review*, 94(4), 991-1013.
- Bertrand, M., & Duflo, E. (2016). Field Experiments on Discrimination. *NBER Working Paper*, 22014.
- Dillingh, R., Kooreman, P. & Potters, J. (2016). Tattoos, Life Style and the Labor Market. *IZA Discussion Paper*. 9675.
- Hamermesh, D. S., & Biddle, J. E. (1994). Beauty and the labor market. *The American Economic Review*, 84(5), 1174-1194.
- Kaas, L., & Manger, C. (2012). Ethnic discrimination in Germany's labour market: a field experiment. *German Economic Review*, 13(1), 1-20.
- Neumark, D. (2011). Detecting discrimination in audit and correspondence studies. *Journal of Human Resources*, 47(4), 1128-1157.
- Oreopoulos, P. (2011). Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes. *American Economic Journal: Economic Policy*, 3(4), 148-147.
- Riach, P. A., & Rich, J. (2002). Field Experiments of Discrimination in the Market Place. *The Economic Journal*, 112(483), F480-F518.
- Rich, J. (2014). What do field experiments of discrimination in markets tell us? A meta analysis of studies conducted since 2000. *IZA Discussion Paper*. 8584.
- Rooth, D. (2010). Automatic Associations and Discrimination in Hiring: Real World Evidence. *Labour Economics*, 17(3), 523-34.
- Weichselbaumer, D. (2015). Testing for discrimination against lesbians of different marital status: A field experiment. *Industrial Relations: A Journal of Economy and Society*, 54(1), 131-161.
- Brähler, E., & Hofmeister, D. P. D. (2009). Verbreitung von Tätowierungen, Piercing und Körperhaarentfernung in Deutschland. *Pressemitteilung der Universität Leipzig*.