

Ethnic Discrimination in the Swiss Labour Market – Ethnic Hierarchies in Correspondence Test Results

Thèse présentée à la Faculté de Lettres et sciences humaines
Université de Neuchâtel

Par Eva Zschirnt

Pour l'obtention du grade de docteur en science humaines et sociales

Soutenu le 2 Novembre 2018

Directeur de thèse :

M. Gianni D'Amato, professeur, Université de Neuchâtel

Rapporteurs :

M. Giuliano Bonoli, professeur, IDHEAP, Université de Lausanne

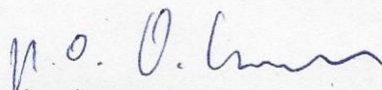
M. Arnfinn Haagenen Midtbøen (PhD), Senior Research Fellow at the Institute for
Social Research in Oslo, Norvège

Mme Judy Rich, Research Fellow at Portsmouth Business School, University of
Portsmouth, Royaume-Uni

IMPRIMATUR

La Faculté des lettres et sciences humaines de l'Université de Neuchâtel, sur les rapports de M. Gianni D'Amato, directeur de thèse, professeur, Université de Neuchâtel ; M. Giuliano Bonoli, professeur, IDHEAP, Université de Lausanne ; M. Arnfinn Haagenes Midtbøen (PhD), Senior Research Fellow at the Institute for Social Research in Oslo, Norvège ; Mme Judy Rich, Research Fellow at Portsmouth Business School, University of Portsmouth, Royaume-Uni autorise l'impression de la thèse présentée par Mme Eva Zschirnt en laissant à l'auteur la responsabilité des opinions énoncées.

Neuchâtel, le 2 novembre 2018



Le doyen
Pierre Alain Mariaux

Summary

As a part of the NCCR on the move's individual project on "Discrimination as an obstacle to social cohesion", this dissertation focuses on the discrimination of ethnic minorities in the Swiss labour market.

Since discrimination is no longer an overt and thus easily observable phenomenon, research on discrimination, and especially on discrimination in hiring decisions has become more challenging. In order to study and measure the extent of discrimination in hiring decisions, researchers have used diverse approaches, depending on the disciplines, mostly drawing on statistical analyses of observational data, behavioural research, attitude research, and victim research (Veenman 2010). While all of these approaches have their strength and weaknesses, a rise of field experiments and in particular correspondence testing can be observed in recent years. Since the late 1960s researchers have used field experiments in which two candidates with exchangeable qualifications that differ only in the characteristic to be measured apply for the same job. These experiments have been successfully used to quantify the extent of discrimination faced by minority applicants in an increasing number of (in particular) OECD countries.

This dissertation focuses on ethnic discrimination in hiring decisions in the German speaking part of the Swiss labour market and it uses a correspondence test as its core methodology. The main research question addressed in this project are:

- Do we find ethnic discrimination in the Swiss labour market?
- Do ethnic hierarchies exist in the Swiss labour market?
- Did discrimination in the Swiss labour market change of time compared to Fibbi et al. (2003)?
- How do the results from the Swiss labour market compare to other countries in which correspondence tests have been conducted?
- Does discrimination only occur in hiring decisions?

Next to contributing to the literature by providing data on ethnic discrimination in hiring decisions, this dissertation also contributes to the theoretical debate whether discrimination is due to taste or statistics. It shows that correspondence tests in German speaking labour markets have reported discrimination rates that are lower than the international average and offers specific characteristics of labour markets in German speaking countries as possible explanations. Compared to most correspondence tests on ethnic discrimination in hiring, it goes further than "just" presenting the classical descriptive results of a correspondence tests, i.e. whether applicants were invited for a job interview or not, but also qualitatively analyses the email responses that were received from potential employers.

Keywords: Ethnic Discrimination, Labour Market, Hiring, Correspondence Testing, Switzerland

Résumé

Dans le cadre du projet individuel "Discrimination as an obstacle to social cohesion" du NCCR on the move, cette thèse porte sur la discrimination des minorités ethniques sur le marché du travail en Suisse.

Étant donné que la discrimination n'est plus un phénomène ouvert et donc facilement observable, la recherche sur la discrimination, et en particulier sur la discrimination à l'embauche, est devenue plus difficile. Afin d'étudier et de mesurer l'ampleur de la discrimination dans les décisions d'embauche, les chercheurs ont utilisé diverses approches, selon les disciplines, s'appuyant principalement sur des analyses statistiques de données d'observation, de recherche comportementale, de recherche sur les attitudes et de recherche sur les victimes (Veenman 2010). Si toutes ces approches ont leurs forces et leurs faiblesses, on observe depuis quelques années une augmentation des expériences sur le terrain et en particulier des tests par correspondance. Depuis la fin des années 1960, les chercheurs ont eu recours à des expériences sur le terrain dans le cadre desquelles deux candidats ayant des qualifications échangeables qui ne diffèrent que par la caractéristique à mesurer postulent pour le même emploi. Ces expériences ont été utilisées avec succès pour quantifier l'ampleur de la discrimination à laquelle sont confrontés les candidats minoritaires dans un nombre croissant de pays, et en particulier des pays de l'OCDE.

Ce mémoire porte sur la discrimination ethnique dans les décisions d'embauche dans la partie germanophone du marché du travail en Suisse et utilise un test par correspondance comme méthodologie de base. Les principales questions de recherche abordées dans ce projet sont :

- Existe-t-il une discrimination ethnique sur le marché du travail en Suisse ?
- Existe-t-il des hiérarchies ethniques sur le marché du travail en Suisse ?
- La discrimination sur le marché du travail en Suisse a-t-elle changé dans le temps par rapport à Fibbi et al (2003) ?
- Comment les résultats du marché suisse du travail se comparent-ils à ceux d'autres pays dans lesquels des tests par correspondance ont été effectués ?
- La discrimination ne se produit-elle que dans les décisions d'embauche ?

En plus de contribuer à la littérature en fournissant des données sur la discrimination ethnique dans les décisions d'embauche, cette thèse contribue également au débat théorique sur la question de savoir si la discrimination est due au goût ou aux statistiques. La thèse montre que les tests de correspondance sur les marchés du travail germanophones ont révélé des taux de discrimination inférieurs à la moyenne internationale et offrent des caractéristiques spécifiques des marchés du travail des pays germanophones comme explications possibles. Par rapport à la plupart des tests par correspondance sur la discrimination ethnique à l'embauche, il va plus loin que la simple présentation des résultats descriptifs classiques d'un test par correspondance, à savoir si les candidats ont été invités ou non à un entretien d'embauche, mais analyse également qualitativement les réponses reçues par e-mail des employeurs potentiels.

Mots-clés : Discrimination ethnique, Marché du travail, Embauche, Tests par correspondance, Suisse

Table of Content

	List of Figures and Tables	3
	Acknowledgements	5
1	Introduction	7
2	Defining Discrimination	11
	2.1 Ethnic discrimination across disciplines	11
	2.2 Cumulative Discrimination	13
3	Explaining discrimination	15
	3.1 Taste-based Discrimination	15
	3.2 Statistical Discrimination	16
	3.3 Critical Discussion of economic theories on discrimination – why does discrimination still exist?	18
	3.4 Other explanatory approaches	19
	3.5 Conclusion	24
4	Measuring discrimination	27
	4.1 Statistical analysis	27
	4.2 Attitude Research	28
	4.3 Survey Experiments	30
	4.4 Field experiments	31
	4.5 Ethnographic Observations	31
	4.6 Victim Research: Perceived Discrimination and Legal Cases	32
	4.7 Conclusion	33
5	Methodology of field experiments	35
	5.1 Field experiments on discrimination in the labour market	36
	5.2 In-person audit studies	37
	5.3 Correspondence Test of Ethnic Discrimination in the Labour Market	38
	5.4 Criticism of in-person audit studies and correspondence tests	41
	5.5 Testing discrimination theories in field experiments	43
	5.6 Research Ethics Concerns	44
	5.7 Matched pair testing vs. single application	45
	5.8 Field experiments for research vs. field experiments for enforcement	45
	5.9 Conclusion	46

6	Context in Switzerland: Labour market inequalities between natives and immigrants	49
	6.1 Immigrants in Switzerland	49
	6.2 Public debate on immigration in Switzerland	50
	6.3 Immigrants in the Swiss labour market	52
	6.4 Evidence of labour market discrimination from previous studies	55
	6.5 Conclusion	63
7	A correspondence test in the Swiss labour market – the research design	65
	7.1 The Ethical Dimension	65
	7.2 Identifying groups for testing	67
	7.3 Signalling ethnicity: construction of names and portrayal of citizenship	69
	7.4 Identifying locations for testing	71
	7.5 Selecting occupations	72
	7.6 Sources for vacancies	73
	7.7 Constructing the applications	74
	7.8 Overview of the sample	80
8	Summary of the Papers	83
9	Conclusion	87
10	Abstracts of the Papers	99
	References	103

List of Figures and Tables

Figure 1:	Correspondence tests on ethnic and racial discrimination in OECD countries 1990-2017	p.39
Figure 2:	Unemployment rate based on ILO definition by nationality and gender	p.53
Figure 3:	Gross monthly wage in 2014 in CHF in the private sector	p.53
Table 1:	Warmth and competence of immigrant groups in Switzerland	p.63
Table 2:	Names used in the correspondence test	p.70
Table 3:	Characteristics of occupations chosen	p.73
Table 4:	Number of paired applications by nationality tested in the pair and occupation	p.80
Table 5:	Number of paired applications by gender, nationality tested in the pair and occupation	p.80
Table 6:	Number of paired applications in rural/urban areas by occupation	p.80
Table 7:	Number of paired applications by Canton and occupation	p.81
Table 8:	Number of paired applications in the private or public sector by occupation	p.81
Table 9:	Number of paired applications where customer contact was mentioned in the vacancy	p.81
Table 10:	Number of paired applications for full time or part time positions	p.82
Table 11:	Number of paired applications by contract duration	p.82
Table 12:	Number of paired applications by application method	p.82

Acknowledgements

During my dissertation project, I have been able to benefit greatly from the professional and personal support of colleagues, friends, and family, who deserve a big thank you!!

Since this dissertation project was embedded within the NCCR on the move that is funded by the Swiss National Science Foundation, I was able to benefit immediately from a ready-made interdisciplinary network of other PhD students, postdocs and project leaders. Being able to relate to the experience of other PhD students, benefiting from their experiences and feedback and exchanging ideas has been an invaluable asset, since writing a PhD can also be a very solitary endeavour. Some great working relationships and friendships have come out of this experience, and I am convinced that they will persist also after the PhD.

Some of these people deserve special mentioning. Starting with the rest of the NCCR IP 18, Gianni, Rosita and Marco. Although not officially part of this project, a very big thank you goes to Didier whose calm presence and support was invaluable during the last four years! Also, two people of the NCCR network office stood out in particular, Nicole, who was always open for our ideas and suggestions, and Gina, who made life so much easier and was always a friendly and welcoming face. I have also been able to benefit greatly from discussions with my NCCR colleagues in Lausanne, Daniel and Flavia, as well as Fabienne, who were always willing to exchange valuable feedback on work in progress.

On a more personal level, several colleagues and friends have made the last four years in Neuchâtel a great experience. Johanna deserves a special mentioning for all the dinners with her family and her amazing children that provided a wonderful distraction from academic life, the countless days spent at the lake swimming, cycling or inline-skating in the summer or hiking, cross country skiing or snowshoeing in the winter, and the private bike repair services. Steffi and Flo contributed immensely to making the time in Neuchâtel a great experience. Apart from teaching me to understand Swiss German, they also contributed a lot to the advancement of this thesis, either with helpful comments from a legal perspective or with apéros, dinners and excursion within Switzerland to provide a distraction from work. I also need to thank Annique for finally moving to Neuchâtel and always having an open ear in the final stage of the PhD write up. This was very much appreciated! Furthermore, I own a big thanks to Marco whose very open way and immediate invitations for apéros or after-work-drinks, barbecues or fondues, made me feel at home in Neuchâtel from the beginning. Jyothi and Jerry also contributed a lot to making me feel at home in Switzerland by making me discover Bern, the fun of swimming in the Aare and through Jerry's amazing food and wine. Especially in the last two years, Carolin, Melf and their little son have become good friends and apart from exploring the countryside around Neuchâtel and Biel, they have enabled me to keep up with the German ritual of watching *Tatort* together. I also lucked out with my office partner during the last nine months. It has been amazing to share the office with Verena and I expect it will be hard to find such a nice office partner again.

The biggest thanks goes to my family for all the support along the way, even in times where the goal was not always clear. Thank you for always being there – even long distance, but at least we already have quite a bit of experience with distances.

1. Introduction

Switzerland has one of the largest immigrant populations of all OECD countries. 27% of the working-age population in Switzerland are foreign born, similarly high rates are only reported in Australia and Luxembourg (OECD, 2012, p. 216). Furthermore, 7% of the Swiss population is considered as belonging to the second generation, leaving only about 60% of the population that is native Swiss and does not have a migration background (Bundesamt für Statistik, 2017b). However, it is not only the high percentage of immigrants and their offspring in the population that make Switzerland a case worth studying, but also the composition of this immigrant population. Most immigrants in Switzerland (almost 60%) were born in European Union countries, the most important countries being Italy, Germany, Portugal and France, followed by countries belonging to the Former Yugoslavia (Kosovo, Macedonia, Serbia, Croatia, and Bosnia) and Turkey (Bundesamt für Statistik, 2017b; OECD, 2012). While unemployment rates for foreigners are quite low in Switzerland in international comparison, there is still evidence that in particular the second generation faces discrimination in the Swiss labour market, e.g. when looking for apprenticeships or when transitioning from apprenticeships to their first positions.

Next to the quite unique composition of the Swiss immigrant working population, there are other, more structural reasons that make the study of discrimination in the Swiss labour market an interesting and relevant case study. The OECD emphasised three structural specificities for Switzerland: First, the decentralised nature of policy making as well as the linguistic heterogeneity between the different regions, second, the flexibility of the labour market, with its high share of female participation, high salaries, and low unemployment rates, and third, the high productivity and the strong international focus of the Swiss economy (OECD, 2012, p. 216). A fourth specificity, which has not been mentioned in the OECD report, but which is relevant for this research is the lack of a comprehensive anti-discrimination law in Switzerland, as it has been adopted in all EU countries. According to the MIPEX Results of 2015, Switzerland is “one of the very few countries without a comprehensive anti-discrimination law and equality body with legal standing; a sizeable number of potential victims are poorly protected against racial, ethnic, religious and nationality discrimination” (Huddleston, Bilgili, Joki, & Vankova, 2015b). Thus, the foreign born or the second generation in Switzerland not only encounter obstacles in the Swiss labour market and their social mobility, but also face problems when they try to enforce their rights when they encounter discrimination. Such

problems in access to the labour market and the judicial system pose a problem for a meritocratic society and can become a threat to social cohesion within the country.

Such a threat to social cohesion was already discussed by Mark Abrams in the introduction to the first field experiment on labour market discrimination that was conducted in the UK by Daniel (1968). Even in the late 1960s, Abrams warned that equal opportunities were an important element of a democratic society: “To deny this right to a minority of our fellow citizens amounts to a total rejection of the principles of democracy” (1968, p. 13). He cautioned readers that discrimination was a waste of manpower and that children of immigrants were probably not going to react as “docile when faced with the frustrations and humiliations of discrimination” as their newly arrived parents had reacted (p. 14). The fact that discrimination is still discussed as a threat to social cohesion in modern Western societies – 50 years after Adams wrote this introduction – shows that the issue is very persistent. Despite legislative progress in most countries in the field of anti-discrimination legislation – of which Switzerland is an exception as it does not have a comprehensive anti-discrimination law – discrimination still persists across societies and across the years as meta-analyses of field experiments on discrimination have shown (Quillian, Pager, Hexel, & Midtbøen, 2017; Zschirnt & Ruedin, 2016).

Studying ethnic discrimination in the Swiss labour market is not only of high relevance on a political and societal level, but it is also of high academic relevance and will contribute to the striving field of discrimination studies. Since the 1990s over 50 correspondence test have been conducted in OECD countries, several are currently ongoing, yet only one of them looked at ethnic discrimination in Switzerland (Fibbi, Kaya, & Piguët, 2003)¹. In contrast to most research conducted in other countries where candidates were trying to find a new position, this previous Swiss study focused on the transition from apprenticeships to first employment. This current research project therefore conducts a new field experiment on ethnic discrimination in the Swiss labour market that allows us to focus on the specificities of the Swiss case.

The main research questions addressed in this project will be:

- Do we find ethnic discrimination in the Swiss labour market?
- Do ethnic hierarchies exist in the Swiss labour market?

¹ A short correspondence test on unsolicited application in the German speaking part of Switzerland was included in a study by Diekmann, Jann, and Näf (2014). However, to ensure comparability the focus is laid on correspondence tests that are based on solicited applications for actual vacancies.

- Did discrimination in the Swiss labour market change of time compared to Fibbi et al. (2003)?
- How do the results from the Swiss labour market compare to other countries in which correspondence tests have been conducted?
- Does discrimination only occur in hiring decisions?

Next to contributing to the literature by providing data on ethnic discrimination in hiring decisions, this dissertation also contributes to the theoretical debate whether discrimination is due to taste or statistics. It shows that correspondence tests in German speaking labour markets have reported discrimination rates that are lower than the international average and offers specific characteristics of labour markets in German speaking countries as possible explanations. Compared to most correspondence tests on ethnic discrimination in hiring, it goes further than “just” presenting the classical descriptive results of a correspondence tests, i.e. whether applicants were invited for a job interview or not, but also qualitatively analyses the email responses that were received from potential employers. To my knowledge, this has not been done in previous correspondence tests on ethnic discrimination in hiring decisions.

The dissertation is part of a larger research project on “Discrimination as an obstacle to social cohesion” which is embedded in the framework of the NCCR on the move. The work presented here focuses on the results of a correspondence test that was conducted in the German-speaking part of the Swiss labour market between October 2017 and April 2018. It will later be complemented by the results from the French speaking area of Switzerland to provide an encompassing picture of discrimination in Swiss the labour market. The combined data will also be used to analyse if differences between the linguistic regions of Switzerland exist (German vs. French speaking Switzerland), if differences between sectors of the labour market can be observed, and whether there are factors that make enterprises more or less likely to discriminate against minority candidate (e.g. size, being a public or private enterprise, or whether an enterprise has a local, national or international orientation). This thesis consists of an introductory chapter, the five papers that emerged from the research project, and a conclusion. The introductory chapter is made up of three larger parts. Sections 2 and 3 first focus on defining and explaining labour market discrimination, which will lay the theoretical basis of the dissertation. Sections 4 and 5 then address the issue of measuring labour market discrimination. While Section 4 provides an overview of the methods used to study discrimination and their strengths and limitations, Section 5 focuses on the methodology chosen

for this research project: a correspondence test. The third part of the introductory chapter, Section 6 and 7 turn towards the Swiss context. Section 6 introduces the Swiss context by providing an overview of the labour market integration of immigrants in Switzerland and discussing previous research on discrimination in the Swiss labour market. Based on this information, Section 7 presents the research design for the correspondence test conducted on the Swiss labour market. The last two sections conclude the dissertation by providing a summary of the five papers which contain the research results and a final conclusion that summarizes the findings of the research project and puts them into a historical and international perspective.

2. Defining Discrimination

Before it is possible to study whether ethnic or racial discrimination occurs in the Swiss labour market, it is necessary to first define the concept of ethnic or racial discrimination². Focusing on the concept of “race”, Blank, Dabady, and Citro (2004, Chapter 2) discuss the challenge of providing a definition of race given that “there is little consensus what race actually means” (p.25). They briefly mention biological definitions of race, before focusing on race as a socially constructed concept. Referring to behavioural and social science work, they discuss that “race is a construct based on observable physical characteristics (e.g., skin color) that have acquired socially significant meaning” (p.27). In contrast to such physical characteristics “Cultural factors, such a language, religion, and nationality, have more often been used to refer to ethnicity – that is, groups of people who share a common cultural heritage, such as various European immigrant groups in the United States” (p.27). While the concept of race is still frequently used in the US, Canadians have started to use the term visible minorities, and the term race has almost disappeared from the European debate, with the exception of the UK. The European reluctance to use the term race becomes apparent in the EU Directive 2000/43/EC stating that “The European Union rejects theories which attempt to determine the existence of separate human races. The use of the term ‘racial origin’ in this Directive does not imply an acceptance of such theories” (Preamble (6)). Even if it might be hard to draw a line between the concepts of race and ethnicity, both are prohibited grounds of discrimination.

2.1 Ethnic discrimination across disciplines

The concept of discrimination is defined differently depending on the disciplinary background of the researcher. Economic definitions of discrimination in the labour market focus on the different treatment of identical individuals that differ only in an ascribed characteristic, such as gender or ethnicity, and where this characteristic does not have a direct effect on productivity. As Heckman formulated: “Discrimination is a causal effect defined by a hypothetical *ceteris paribus* conceptual experiment – varying race but keeping all else constant” (1998, p. 102). Discriminating on the basis of one characteristic thus diminishes economic efficiency although losses are hard to observe directly (Bendick, 2007). Heckman (1998) also emphasises the need to distinguish between discrimination by individual firms and discrimination on the market as

² This section extends the discussion of definitions included in Paper I of this dissertation. Since the concept of race is hardly used in Europe (except for the UK), the term ethnic discrimination will be used.

a whole, and points out that even if many individual firms behave discriminatory, labour market discrimination might still be minimal.

Legal scholars use definitions that are very similar to the definition of economists (at least in the case of direct discrimination) and also define discrimination rather narrowly. Yet by also accounting for cases of indirect discrimination and instances where discrimination is considered acceptable, the legal definition of discrimination is already slightly broader than the economic definition. Legal definitions of discrimination can be found in national legislation, in particular national anti-discrimination laws (if available), or in international law. Article 1 of the International Convention on the Elimination of All Forms of Racial Discrimination defines racial discrimination as

“any distinction, exclusion, restriction or preference based on race, colour, descent, or national or ethnic origin which has the purpose or effect of nullifying or impairing the recognition, enjoyment or exercise, on an equal footing, of human rights and fundamental freedoms in the political, economic, social, cultural or any other field of public life” (United Nations, 1965).

In the European Union, definitions of discrimination can be found in the two anti-discrimination directives that were adopted in 2000, the so-called “Race Directive” (Directive 2000/43/EC) and the “Employment Equality Directive” (Directive 2000/78/EC) (Council of the European Union, 2000a, 2000b). These directives had to be transposed into national law in all EU member states. Both directives provide a definition of the concept of discrimination in Article 2 (2), that differ only slightly in the scope of point 2 (b):

(a) *direct discrimination* shall be taken to occur where one person is treated less favourably than another is, has been or would be treated in a comparable situation on grounds of racial or ethnic origin;

(b) *indirect discrimination* shall be taken to occur where an apparently neutral provision, criterion or practice would put persons of a racial or ethnic origin at a particular disadvantage compared with other persons, unless that provision, criterion or practice is objectively justified by a legitimate aim and the means of achieving that aim are appropriate and necessary.” (Council of the European Union, 2000a, p. emphasis added).

In contrast to these two rather narrow definitions in law and economics, different definitions of discrimination exist among sociologists. While some sociologists define discrimination very broadly (similar to activists) to include all instances of inequality in their understanding of discrimination, others use very narrow definitions that encompass only instances where the intent was to harm the target by discriminatory actions (Quillian, 2006). A definition by the US National Research Council lies between the very narrow and the very broad definition. They

“use a social science definition of racial discrimination that includes two components: (1) *differential treatment on the basis of race* that disadvantages a racial group and (2) *treatment on the basis of inadequately justified factors other than race* that disadvantages a racial group (differential effect)” (Blank et al., 2004, p. 39, emphasis in the original).

Discrimination due to ethnicity or race, which corresponds to the first component of this definition, is prohibited in most countries. Instances that fall under the second part of the definition, the differential effect, “may or may not be considered discrimination under the law” (Blank et al., 2004, p. 39), depending on the national legal context and the reasoning behind the differential treatment. The concepts of differential treatment and differential effect follow a logic that is similar, but broader, than the legal concepts of disparate treatment and disparate impact discrimination in the US context.

Comparing these definitions by Blank et al. (2004), with the legal definitions of direct and indirect discrimination in EU law, it becomes apparent, that the concept of direct discrimination in EU law is very similar to their notion of differential treatment on the basis of race, while the EU’s concept of indirect discrimination is very similar to their understanding of differential effects. As this section has shown, the definitions used in the three disciplines discussed here all look at the differences in treatment between persons due to a characteristic that cannot be influenced, e.g. a person’s ethnic background. They do, however, differ in the narrowness of their understanding of discrimination.

2.2 Cumulative Discrimination

While the definitions provided above usually look at one particular instance or case of discriminatory treatment, sociologists have emphasised that being subjected to discriminatory behaviour is rarely a singular occurrence in the life of ethnic or racial minorities but rather a

continuous experience. Looking only at the domain of employment, Pager, Bonikowski, and Western (2009) point out that

“In reality, discrimination may occur at multiple decision points across the employment relationship. In this way, even relatively small episodes of discrimination – when experienced at multiple intervals or across multiple contexts – can have substantial effects on aggregate outcomes.” (p. 778)

The idea that discrimination can occur in multiple stages or areas of life is discussed in detail by Blank et al. (2004), who devote an entire chapter of their book to the concept of cumulative discrimination. Their main argument is that even if only small effects of discrimination are found in studies at one certain point in time, repeated exposure to small discriminatory effects across or within domains can over time lead to substantial differences. They argue that discrimination can cumulate across generations, within domains, or across domains. Furthermore, they emphasise that discrimination in early decisive situations can have increasing long-term effect. Using the example of the labour market, the authors elaborate how discrimination can occur at different points. Discrimination can take place in the educational system and impact the chances of entering the labour market or further education, it can occur in the transition from apprenticeships to the labour market, in the attempts to change positions, or in the level of wages, the status of the position, in performance evaluations, or termination decisions. Furthermore, discrimination in the labour market could also have consequences across domains, e.g. on the housing situation. Blank et al. (2004) further point out that tracing the effects of cumulative discrimination is even harder than to measure discrimination that only happens in one situation.

Similarly, Barbara Reskin develops her argument that different domains in which discrimination occurs are interconnect and can reinforce each other. She labels this “system of race-linked disparities” (2012, p. 17) as *über* discrimination and criticises that even though most sociologists are aware that discrimination is not an isolated occurrence, this awareness is not visible in quantitative research on discrimination.

3. Explaining discrimination

In order to set the theoretical foundation for this dissertation, the most commonly discussed theories on discrimination are addressed in this section. The literature offers several, sometimes competing, explanations on why discrimination occurs in the labour market. In the field of economics the classic and ongoing debate focuses on whether discrimination is due to taste (Becker, 1957) or statistics (Aigner & Cain, 1977; Arrow, 1973; Phelps, 1972). As Guryan and Charles (2013) have pointed out, this debate, which started in the 1950s, has recently been rekindled following the methodological advances in the study of labour market discrimination. Next to these economic theories, in particular sociologists and social psychologists have offered alternative explanations for discriminatory behaviour (e.g. focusing on ethnic hierarchies or the existence of stereotypes, prejudices or biases).

3.1 Taste-based Discrimination

Taste-based discrimination theory, sometimes also described as “preference-based discrimination”, developed by Becker (1957) in his seminal book *The Economics of Discrimination*, focuses on interethnic attitudes and is based on the notion that

“If an individual has a ‘taste for discrimination’, he must act *as if* he were willing to pay something either directly or in the form of a reduced income, to be associated with some persons, instead of others. When actual discrimination occurs, he must, in fact, either pay or forfeit income for this privilege. This simple way of looking at the matter gets at the essence of prejudice and discrimination” (Becker, 1957, p. 14).

Becker develops this theory further by addressing discrimination by three possible actors: employers, employees, and customers, always assuming that the other two that are not the focus of the analysis do not discriminate against minority candidates:

“When an employer discriminates against employees, he acts as if he incurs non-pecuniary, psychic costs of production by employing them; when an employee discriminates against fellow employees or employers, he acts as if he incurs non-pecuniary, psychic costs by working with them; when a customer discriminates against products, he acts if he incurs non-pecuniary, psychic costs of consumption by consuming them.” (p.153).

Employers are thus willing to pay a higher cost (e.g. a higher salary) to a worker belonging to their favoured social group in order to avoid working with members from a disliked group, or put more drastically “monetary costs of discriminatory behaviour are offset by nonmonetary benefits for being spared from interaction with despised groups” (Keuschnigg & Wolbring, 2016, p. 182). Employers are thus not seen as purely profit-maximising actors (Arrow, 1998).

Following Becker, in situations of perfect competition, discriminatory employers should eventually be driven out of the market, because these employers are willing to pay higher wages to candidates belonging to their preferred groups. Using data from an audit study in New York and information on firm survival six years onwards, Pager (2016) was able to show that firms that had been found to act discriminatory in their hiring decisions were less likely to be still in business. While these results are in line with taste-based discrimination theory, she cautions readers to make the link too quickly, as the firm failure could also be due to poor decision making in other areas. She concludes that “Discrimination may or may not be a direct cause for business failure, but it seems to be reliable indicator of failure to come” (Pager, 2016, p. 855).

3.2 Statistical Discrimination

Statistical discrimination theory was developed by Phelps (1972), Arrow (1973), and Aigner and Cain (1977) as a response to Becker’s theory of taste-based discrimination. It builds on the assumption that employers make hiring decision based on scarce information about job applicants. Phelps argues that

“the employer who seeks to maximize expected profit will discriminate against blacks or women if he believes them to be less qualified, reliable, long-term, etc. on the average than whites and men, respectively, and if the cost of gaining information about the individual applicants is excessive. Skin color or sex is taken as a proxy for relevant data not sampled” (1972, p. 659).

It is therefore not only employers with a clear taste preference who discriminate, but also those who are perceived as liberal and have no distaste to hire candidates belonging to a minority group, since they try to use group signals to compensate for the scarcity of information in the hiring process. These assumptions about behaviour due to an ascribed characteristic can be based on own experiences with members of the group concerned, reported experiences of others, or from other sources such as media reports or statistics (Keuschnigg & Wolbring,

2016). Hence, employers are seen as “profit maximizing actors who are confronted with the uncertainties accompanying selection decision” (Thijssen, 2016, p. 5) and who try to avoid the costs of hiring the wrong person or the cost of gathering more information about a candidate.

In the long run, however, employers should gather more information about employees of certain groups, thus making them less likely to (having to) resort to statistical discrimination in their hiring decisions (Flinn, 2015). Beliefs that have been proven wrong by experience should therefore not survive (Arrow, 1998).

Several researchers have tried to address the question whether discrimination is due to distaste or statistics in the design of their field experiments, however, results are inconclusive. The above-mentioned study by Pager (2016) whether firms that discriminate are more likely to go out of business seems to point towards the theory of taste based discrimination. Weichselbaumer (2016b) leans towards taste-based discrimination to explain the discrimination rates found in her Austrian study, too, arguing that the amount of application material provided in Austria makes it unlikely that employers have to resort to mechanisms of statistical discrimination. While Kaas and Manger (2012) found that providing additional information, in this case a reference letter, reduced discrimination and thus made a case in point for statistical discrimination theory, this did not hold true in the work of Bertrand and Mullainathan (2004), where a higher quality CV was only beneficial for the white candidates, but not the Afro-Americans. The studies by Oreopoulos (2011) and M. Carlsson (2010) also do not confirm assumptions from statistical theory, according to which candidates listing additional language skills or those belonging to the second generation should be preferred. As far as I am aware there is only one study that tried to address both taste-based and statistical discrimination that was conducted by Baert and De Pauw (2014) in Belgium. They used a vignette study, carried out with a population of students, to examine statements on attitudes related to taste-based and statistical discrimination as well as intentions to invite potential candidates for an interview or hire them. However, their sample of students is not representative of employers³, the vignette study design only allows them to study intended behaviour, and it is unclear if people with high levels of prejudice on the taste-based discrimination questions also rated non-native applicants more negatively in the questions relating to statistical discrimination.

³ In a study comparing vignette and conjoint survey experiments, Hainmueller, Hangartner, and Yamamoto (2015) found the convenient student sample to perform poorly compared to other more representative populations and study designs.

3.3 Critical Discussion of economic theories on discrimination – Why does discrimination still exist?

As the previous discussion has already mentioned, according to both taste-based and statistical discrimination theory, discrimination should not be a long-term phenomenon and should disappear completely over time in a competitive market.

Looking specifically at taste-based discrimination theory as developed by Becker (1957), Darity and Mason (1998) argue that “Standard neoclassical competitive models are forced by their own assumptions to the conclusion that discrimination only can be temporary” (p. 81). According to Becker’s taste-based discrimination theory, discriminatory employers should eventually be run out of business because of their non-competitive behaviour and their willingness to incur costs to avoid working with undesired groups. Thus, in a competitive market environment, non-discriminatory firms will have a comparative advantage. Bergmann (1989) criticised that Becker’s theory of taste-based discrimination is only deductive and that its applicability to real situations

“depends on three assumptions that may or may not be true in any particular time or place: (1) that there are large numbers of people who are willing and able to openly violate social customs, which they themselves support and enjoy, for purposes of making money, (2) that violating customs does not entail costs that cancel out the advantage of cheap wages, and (3) that competition is intense enough to put out business those who refrain from violating customs” (p. 50).

Another point of criticism concerning Becker’s theory is that he does not discuss where the interethnic attitudes or the “taste for discrimination” that employers have towards a minority group come from and if these attitudes are prone to change over time (Keuschnigg & Wolbring, 2016). Here, work by social psychologists could provide good additional explanations on how prejudices are established.

Statistical discrimination theory also predicts that – in the long run – discrimination should disappear. While group signals may – at the first contact with a new group – be used as a proxy for missing information about a specific group, the need to resort to such proxies should diminish over time and experiences made by employers should mitigate the need to resort to discrimination based on group signals (e.g. Arrow, 1998; Flinn, 2015). As Darity and Mason

(1998) point out “employers should *learn* that their beliefs are mistaken” (p. 83, emphasis in the original) if group differences are not found in real life, or, if these group differences were found to be real “employers are likely to find methods of predicting the future performances of potential employees with sufficient accuracy that there is no need to use the additional ‘signal’ of race or gender” (ibid.).

Despite these theoretical predictions, however, discrimination in the labour market persists. Numerous field experiments have documented continuous discrimination and meta-analyses of these audit and correspondence studies have shown almost no change in the level of discrimination of minority candidates over time (Quillian et al., 2017; Zschirnt & Ruedin, 2016). While correspondence tests usually focus only on one country at a certain point in time, results are strikingly similar, even if the external conditions, such as the economic situation or the legal anti-discrimination frameworks, might differ considerably (for rare comparative discussions of selected anti-discrimination regimes see e.g. Givens and Case (2014) or Mercat-Bruns (2016); while the meta-analysis by Zschirnt and Ruedin (2016) also addresses the economic context by using unemployment rates and GDP growth rates). Neither economic theory discussed above is able to explain these persistently high levels of discrimination. While numerous researchers have tried to design field experiments to test elements of taste or statistical discrimination theory, they still fail to provide convincing evidence if discrimination is based on taste or statistical assumptions. This focus on taste based or statistical discrimination as a driving force, and the debate whether discrimination should disappear in a comparative market, is also criticised by Darity and Mason (1998) who recommend looking at the mechanisms that “permit or encourage” (p. 82) discriminatory hiring behaviour. These could e.g. be studies on the legal anti-discrimination framework in different countries, which, as mentioned above, are still very rare, or studies focusing on the emergence of stereotypes and prejudices which can lay the foundation of discriminatory behaviour.

3.4 Other explanatory approaches

Next to these two prominent economic theories, other possible explanations have been brought forward that try to explain the reasons for discriminatory treatment. Some of them are briefly mentioned here, while this list is certainly not exhaustive.

The concept of ethnic penalties refers to differences observed between the majority and ethnic minorities, e.g. in labour force participation, unemployment, occupation or earnings, that persist after controlling for numerous characteristics, such as age, education, work experience or country of origin. Across many western societies immigrants and their offspring face disadvantages in their labour market outcomes compared to the white majority groups (e.g. Heath & Cheung, 2007; Heath, Rothon, & Kilpi, 2008; OECD, 2010, 2013). In the work of Heath and Cheung (2006) the term “ethnic penalty” is used to encompass all reasons that can explain the disadvantage of ethnic minority groups compared to the white majority group when qualifications are similar (p.19). They explicitly point out that the concept of ethnic penalties is “a broader concept than that of discrimination although discrimination is likely to be a major component of the ethnic penalty” (p. 19), yet other explanatory factors cannot be ruled out. Next to discrimination, Berthoud (2000) lists five other possible theoretical issues that are “commonly raised in discussions of the problems facing minorities in the labour market: migration, expectation, alienation, family formation [and] the structure of the economy” (p. 390). These factors may interact and affect certain immigrant groups differently.

So far, ethnic penalties have been documented across (western) countries and immigrant groups and/or generations, yet, as Midtbøen (2015a) pointed out “there is considerable variation in the magnitude and scope of ethnic penalties” (p. 188). While the statistical calculation of ethnic penalties provides important information about the disadvantages that members of the ethnic minority face in comparison to the majority group, it can only demonstrate that such disadvantages exist, but not explain why they occur. Discrimination in the labour market is only one possible explanation. Field experiments of discrimination in the labour market have documented persistent discrimination across countries and labour market contexts. Thus, while the statistical analysis of labour market disadvantages of ethnic minority candidates shows that they face ethnic penalties, it is not clear how much of this is attributable to labour market discrimination or what other factors might influence the extent of ethnic penalties encountered in a specific national context (e.g. Midtbøen, 2015a). A further, and more fundamental criticism of the “ethnic penalty” approach is discussed in detail by Modood and Khattab (2016), namely that the concept of ethnicity is problematic in itself, and that, if accepted as a concept, ethnicity by itself is not able to explain behaviour.

In contrast, departing from the position that immigrants and natives differ in important characteristics which will influence their labour market outcomes, human capital theory, focuses on the often lower human capital of members of the minority group to explain their lower position on the labour market compared to the majority. As Helland and Støren (2006) summarize, “human capital theory (see e.g. Becker, 1964, or Mincer, 1958) assumes that an increase in education leads to an increase in productivity, which in turn leads to increased income” (p.343). Yet, as Bursell (2012) points out, assessing the human capital of migrants becomes more complex, since they might lack some human capital that is country specific to the host country. This includes language barriers, lower levels of education, hard to transfer qualifications, or a lack of country specific human capital (Andriessen, Dagevos, & Iedema, 2008; Blommaert, Coenders, & van Tubergen, 2014; Gaddis, 2014; Midtbøen, 2015b). The longer immigrants are living in their new host country and the better they know the language and the host country’s institutions, the lower the gap in labour market outcomes due to a lack of human capital should become. However, statistical analysis on the position of immigrants as well as their children, shows that the second generation still faces labour market disadvantages. While employers place a greater value on qualifications and experiences obtained in the host country (Bursell, 2012), a finding that has also emerged in correspondence tests (Dechief & Oreopoulos, 2012; Oreopoulos, 2011), children of immigrants that received all their qualifications in the host country still encounter problems on the labour market. Although the concept of human capital should cause ethnic differences in labour market outcomes to diminish and disappear over time, these differences still persist.

As Bertrand and Duflo (2016) point out next to the two prominent economic theories of discrimination and the also economic focused approaches of ethnic penalties or human capital theories, another big strand of relevant literature on the causes of discrimination has been developed by psychologists “on a largely parallel track” (p.4). This work often focuses on topics such as stereotyping, prejudice, and their link to discrimination. According to Fiske (1998), “stereotyping is taken as the most cognitive component, prejudice as the most affective component, and discrimination as the most behavioural component of category-based reactions” (p.357). Categorization is thus an important element of social psychology works relating to discussion of (amongst others) social dominance, social distance and ethnic hierarchies, stereotype content models, or unconscious stereotypes, biases or implicit attitudes on discrimination.

One explanation for discriminatory behaviour is offered by the concepts of social distance and social dominance. These approaches argue that people prefer to be surrounded by others that are similar to themselves. According to the theory of social distance, employers choose candidates that are as close and as culturally similar to them as possible (Bursell, 2007). Taking the notion of social distance one step further, social dominance theory (Sidanius & Pratto, 1999) argues that those in more advantageous positions will try to maintain their advantages, which can be e.g. cultural, socioeconomic, political, or ideological (Blommaert, 2013; Sidanius, Pratto, Van Laar, & Levin, 2004). The theory includes both “individual and structural factors that contribute to various forms of group based oppression” (Sidanius et al., 2004, p. 846) and seeks to explain why all kinds of human societies organize around hierarchies based on group membership. People who score high on a scale measuring social dominance orientation “indicate agreement by respondents that some groups are just more worthy than others, that group hierarchy is inevitable and good, and that dominance is necessary” (Blank et al., 2004, p. 179).

Such strong in-group preferences, stereotypes against members of the outgroups and social dominance orientation lead to the creation of hierarchies between groups, e.g. by ethnic background or gender, or a combination of the two, which is often discussed in the Dutch context (Andriessen, Nievers, Dagevos, & Faulk, 2012; Andriessen, Nievers, Faulk, & Dagevos, 2010). Hagendoorn (1993) points out that ethnic hierarchies are very similar within majority groups, regardless of whether they hold positive or negative attitudes towards minorities and he lists several factors that can explain the formation of ethnic hierarchies. The first factor are stereotypes, which in a multi-ethnic context will be used to emphasise the negative differences between an out-group to the in-group. The more important these differences are, the greater will be the distance between the in-group and the out-group and “this means that the process of differentiation unavoidably entails a rank-ordering” (p.36) and thus leads to ethnic hierarchies. Secondly, he discusses how the dominant primary ethnic group relates to either (secondary) immigrant groups in ethnically rather homogenous societies or other groups vying for primacy (using the example of South Africa) and how this translates into ethnic hierarchies, when non-primary groups try to improve their position. Next, Hagendoorn looks at the socio-economic advancement of members of the ethnic minority groups. He argues that “status differentiation within ethnic groups will not necessarily lead to the mitigation of prejudice” (p.47), but might lead to the emergence of new stereotypes.

The existence of ethnic hierarchies has also been observed in correspondence tests on ethnic and racial discrimination in the labour market. As Zschirnt and Ruedin (2016) have shown looking at the most commonly studied ethnic groups in correspondence tests, a clear ethnic hierarchy exists between these groups. While Turkish candidates experience the lowest discrimination among these groups, Chinese and the group of Indian, Pakistani and Bangladeshi candidates face more discrimination, and candidates with an Arab or Middle Eastern background experience the most discrimination. Similar findings on a national scale are also reported in studies using multiple minority groups, e.g. Weichselbaumer (2016b) in Austria, Booth, Leigh, and Varganova (2012) in Australia or Fibbi et al. (2003) in Switzerland.

Another approach trying to explain how people categorize outgroups is the stereotype content model developed by Fiske, Cuddy, Glick, and Xu (2002). It “posits stereotyping along perceived warmth and competence” (T. L. Lee & Fiske, 2006, p. 753). T. L. Lee and Fiske (2006) also apply the stereotype content model to the perception of immigrant groups in the host country and argue that the perception of a certain group is strongly related to the groups immigration history. Yet, they furthermore point out, that stereotypes not only depend on the ethnic, racial, or national origin of an immigrant group “but also socio-economic status cross-cuts which cross-cuts the former” (p. 755). This model has also been applied to the Swedish (Agerström, Björklund, Carlsson, & Rooth, 2012) or Swiss (Binggeli, Krings, & Sczesny, 2014a; Krings, Johnston, Binggeli, & Maggiori, 2014) labour market and it is shown that portraying more warmth and competence increases the chances to be invited for a job interview in the case of Sweden and that groups which are categorized as high in competence but lacking warmth and which are perceived as direct competition are more likely to experience subtle discrimination in the case of Switzerland.

The strand of work focusing on unconscious stereotypes and implicit attitudes also addresses the topic of interracial interactions and discrimination from a social psychology perspective. It is argued that even decision makers who do not voice or avoid explicit stereotypes or prejudice can make decisions based on unconscious beliefs. As e.g. Dechief and Oreopoulos (2012) discuss, unconscious stereotypes and implicit associations can lead to erroneous statistical discrimination, thus referring back to the above-mentioned statistical discrimination theory that is often used by economists and sociologists. To predict the link between implicit attitudes and discrimination, researchers have increasingly used Implicit Association Tests (IATs), which measure the association of two target concepts with a certain attribute (Greenwald, McGhee, &

Schwartz, 1998). Results from correspondence tests on labour market discrimination have been combined with IAT test results from employers to study if IAT test results can predict discriminatory hiring decision (e.g. Bertrand, Chugh, & Mullainathan, 2005; Rooth, 2010). However, more recent work on IATs as a predictor for discriminatory behaviour has been more sceptical (R. Carlsson & Agerström, 2016; Frederick L Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013; Frederick L.; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2015). R. Carlsson and Agerström (2016) for example caution readers of their meta-analysis on IAT testing against using IAT results in practical applications based on the assumption that IATs can predict discrimination, but they still acknowledge that IATs can be a useful research tool in other contexts.

These alternative approaches to address discrimination and the emergence of stereotypes and prejudice are just the tip of the iceberg and serves as a point of orientation in this dissertation. Social psychology offers much more very detailed work on these topics, the discussion of which could be a dissertation of its own.

3.5 Conclusion

As this chapter has shown, economists as well as many sociologists researching discrimination often only refer to the question whether discrimination is taste-based or statistical. Yet, these two theories fail to explain why discrimination still persists instead of disappearing or at least diminishing, which both theories predict long term. Neither theory explains where assumptions about a group or certain attitudes actually come from, but rather take them and their continuous existences for granted. This is where a link to other explanatory approaches developed by sociologists and social psychologists can become very helpful, as they also allow to address contextual factors. Most discrimination research (e.g. field experiments) are normally conducted in one specific context, which leaves a comparison of external factors, e.g. the legal anti-discrimination regime, mostly out of the discussion. It would be interesting to see more comparative researchers on such factors, to see how external factors might influence discriminatory behaviour. Combining the quantitative focused economic and sociological approaches to measuring discrimination with more qualitative work from sociologists and work by social psychologists and legal scholars could probably contribute a lot to the understanding

why discrimination occurs. That discrimination is persistently occurring has been well documented in field experiments (Quillian et al., 2017; Zschirnt & Ruedin, 2016).

This dissertation will also make use of the theories of taste-based and statistical discrimination theory. In particular the meta-analysis presented in Paper II tries to address which factors might point towards taste-based or statistical discrimination. Furthermore, Paper IV which draws on the results of the correspondence test conducted in the Swiss German labour market, will discuss if the large amount of application material provided in the German speaking context might reduce discrimination according to statistical discrimination theory. Furthermore, it will examine whether ethnic hierarchies exist in the Swiss labour market. During the preparation phase discussed below, social psychology literature (e.g. on the stereotype content model applied to Switzerland) and attitude research were extensively consulted to prepare the experiment.

4. Measuring discrimination

In order to study ethnic discrimination in hiring decisions in the Swiss labour market, a suitable methodology had to be selected. Labour market discrimination has been studied using numerous methodologies, ranging from economic analysis of observational data to study differences in wages or employment rates between groups, analyses of reported legal cases of discrimination, studies of perceived discrimination with victims of discrimination, surveys and experiments on attitudes towards foreigners to experimental research, such as field experiments. Bovenkerk (1992), Veenman (2010), the OECD (2013), and Wrench (2016) provide good overviews over the various methods that have been used to study discrimination. While each of these overviews categorises methodologies slightly differently, they overlap in many instances. It has been acknowledged that no methodology on its own is able to provide a complete picture of labour market discrimination, which is notoriously hard to measure, and every method has different strengths and limitations, depending on the specific research question.

4.1 Statistical analysis

All four of the above-mentioned overviews discuss the use of statistical analysis of the position of ethnic minorities in the labour market. These statistics often show differences in unemployment rates for natives and immigrants, differences in income or if immigrants are more likely to work in certain sectors or under certain working conditions compared to natives (e.g. Wrench, 2016). Controlling for variables such as (amongst others) the level of qualifications, years of education, job experience, age, or gender, differences in outcomes for natives and immigrants are argued to be due to unobservable differences. This can be done by using Oaxaca-Blinder decompositions (Blinder, 1973; Oaxaca, 1973) to analyse gaps in employment and wages. As Riach and Rich (2004a) point out, “Basically this technique interprets any wage differential that cannot be explained by productivity-determining characteristics such as education, length of employment, etc. as measuring the extent of discrimination” (p. 463), even though other plausible explanations might not have been observed in the model. Thus, it is criticised that “These methods can control for too little but they can also control for too much, and both can lead to incorrect inferences of discrimination” (Guryan & Charles, 2013, p. F419). The differences between natives and immigrants, sometimes called “ethnic penalty”, suggest that discrimination could be at play, yet, the “ethnic penalties identified in statistical research are only *indirect* indicators of the operation of

discrimination” (Wrench, 2016, p. 119, emphasis in the original). Well known examples of such statistical analysis are the work by Anthony Heath and Sin Yi Cheung (Heath & Cheung, 2006, 2007) on the unequal chances and ethnic penalties that ethnic minorities encounter in the labour market. While statistical analysis might give a first indication of discrimination “This method has many disadvantages, one of them being that [it] gives no conclusive proof of discrimination as long as all other possible relevant variables have not been identified” (Bovenkerk, 1992, p. 4). Another problem with this approach are omitted or poorly-measured variables which might influence the residual difference between natives and immigrants, language skills being one variable that is usually not available in data sets (OECD, 2013). As Guryan and Charles (2013) have pointed out, “Concerns about the limited ability of regression-based methods to isolate the portion of disparities in economic outcomes that might be due to discrimination led to search for alternative methods” (p. F420). These alternative methods are discussed in the remainder of this section.

4.2 Attitude Research

Attitude research provides valuable information about the attitudes towards immigration and foreigners in the country in general, and also specifically on attitudes towards foreigners in the labour market. There is a growing and well-developed literature in this field, with researchers using laboratory experiments, attitude surveys, or interviews with employers.

One method of researching attitudes are laboratory experiments or experimental research on ethnic discrimination in the labour market and Lane (2016) conducted a meta-analysis of studies of discrimination using laboratory experiments. While laboratory experimental approaches allow the researchers to control all the conditions except for the dependent variable and can provide information on the mechanism determining discriminatory behaviour, they also face limitations. Laboratory experiments face problems of external validity (influenced both by the realistic setting of the experiment and the representativeness of the participants chosen), of social desirability bias, of only assessing behaviour in an artificial situation but not in real life, and of not being able to assess the extent of discrimination in the labour market (Bovenkerk, 1992; Veenman, 2010).

Attitude surveys are one of the most frequently used methods in the studies of attitudes towards foreigners. They can take the form of interviews or surveys with employers, large scale surveys

with a certain target population or national or international surveys. Well-known examples of international attitude surveys are the Eurobarometer Survey, the European Social Survey, or the World Values survey. Furthermore, in many countries national surveys on attitudes exist as well. These attitude surveys allow the researchers to obtain a quite nuanced picture on the attitudes towards foreigners in general or in certain areas of life, such as the labour market, and the methodology has been well developed over time. Yet, as in all areas concerning attitude research, problems arise pertaining to social desirability bias, since respondents might figure out the goal of a survey or refrain from voicing less accepted opinions. Furthermore, unless researchers were involved in the survey design, they have to work with the questions that were asked, even if they do not perfectly correspondent to their research focus.

Bovenkerk (1992) also proposes interviews with managers to study the employment process. While he cautions that social desirability biases may be strong in interviews, he also remarks that employers have often been surprisingly open in discussion about discriminatory hiring decisions. I am aware of four studies where interviews with employers complemented field experiments in the labour market: Midtbøen (2014), who conducted interviews with participants from his correspondence test in Norway, Pager and Quillian (2005) who interviewed employers by phone about their hiring intentions, Schneider, Yemane, and Weinmann (2014), who conducted focus group discussions with employers to complement their correspondence test in Germany, and Oreopoulos (2011), who added interviews with employers that had not been part of his correspondence test, but commented on his findings on discrimination in the Canadian (Toronto) labour market.

Yet, results from attitude research – particularly asking people how they would behave in a certain situation – have to be treated with caution. Two examples on attitude surveys that also looked at the behaviour of participants at a different stage of the research, show this discrepancy between stated and actual behaviour. Already in the 1930s, LaPiere conducted research in US restaurants and hotels. After travelling with a Chinese couple and frequenting restaurants and hotels with them, he enquired several months later if these establishments would accept bookings from Chinese customers. While they were actually served in all but 1 out of 251 situations, only 2 persons of more than 256 questionnaire recipients were willing to host a Chinese guest (LaPiere, 1934). In a second example Pager and Quillian (2005) combined results from an audit study on hiring candidates with a criminal record with a telephone survey of the same employers. They found that employers who were more open to the idea of hiring a

candidate with a criminal record in the survey, were not more likely to invite this candidate for a job interview or to offer him a job. Additionally, while the telephone survey did not show differences by race, the results of the audit study showed significant differences depending on the race of the applicants. Both LaPiere (1934) and Pager and Quillian (2005) showed that the results obtained in surveys did not correspond to the actual behaviour of participants. Contrary to those findings, M. Carlsson and Rooth (2012) and M. Carlsson and Eriksson (2017) find a link between attitudes expressed in surveys and the discriminatory outcomes in their correspondence test on the Swedish labour and rental housing market respectively, when they combine their testing results with national attitude survey data. The conflicting results of these studies show that using attitudes expressed in surveys to predict discriminatory behaviour is difficult and should be treated with caution.

4.3 Survey experiments

A further way to study attitudes and stated preferences of participants that lies between attitude surveys and field experiments “which are often regarded as the methodological gold standard” (Protsch & Solga, 2017, p. 392) are vignette experiments, also known as (factorial) survey experiment. Rather than asking a single item question about hiring intentions or attitudes towards foreigners, in vignette experiments, real world situations are simulated, and participants are supposed to make a choice. In the field of hiring discrimination these questions usually focus on the intention to hire candidates or to invite them for a job interview (e.g. Auer, Bonoli, Fossati, & Liechti, 2018; Damelang & Abraham, 2016; Di Stasio, 2014; Humburg & Van der Velden, 2015; Protsch & Solga, 2017). Characteristics of the applicants are varied among several dimensions, in order to avoid participants guessing the intended topic of the research. Proponents of vignette analysis emphasise that “Compared with single-item questions, a situational description with varying dimensions leads to more subtle questioning; therefore, the responses are less likely to be influenced by social desirability bias” (Auspurg & Hinz, 2015, p. 4). Furthermore, they emphasise the possibility to consider “multiple theoretically important dimension simultaneously” (Protsch & Solga, 2017, p. 392). Finally, high internal and external validity are listed as advantages of using survey experiments, although the latter also depends on the target population (*ibid.*). However, vignette experiments only measure stated preferences, not real behaviour and do not provide information about the mechanism causing discriminatory behaviour (Veenman, 2010). Also, the choice of participants and the design of the vignette experiment can greatly influence the results, as

Hainmueller et al. (2015) have shown using Swiss data on naturalisation decisions. To my knowledge, there is only one example of research on ethnic discrimination in the labour market that combined a vignette experiment with students making hiring decisions with a correspondence test on the labour market (Larja et al., 2012; Liebkind, Larja, & Brylka, 2016).

4.4 Field experiments

Field experiments on discrimination in the labour market have become more and more popular as they enable researchers to observe “Laboratory like controlled conditions in quasi-experiments in real-world hiring situations” (Bendick & Nunes, 2012, p. 238). They are built on the notion that real world employers are presented with equally qualified fictitious candidates that differ only in the characteristic to be studied. Candidates apply for publicly advertised vacancies either in person or in writing. It is carefully recorded which applicants get invited for a job interview and the differences between majority candidates and minority candidates being invited is then attributed to discriminatory treatment. Field experiments have the advantage that they provide a direct measure of the actual extent of discrimination (Midtbøen & Rogstad, 2012). Since a written correspondence test is the methodology chosen for this dissertation, this methodology is discussed in detail in the next part of this dissertation.

4.5 Ethnographic Observations

Information about labour market discrimination can also be obtained through ethnographic observations. Veenman (2010) gives the example of directly observing job interviews although this is rarely possible to organize. However, in field experiments where applications are made by trained testers (so-called audit studies), researchers have been able to observe job interviews that took place either in-person or on the phone through the testers’ observations. In these field experiments employers were not aware of their participation in an experiment. Such in-person audit studies have particularly been done in US field experiments on labour market discrimination (e.g. Bendick, 1996; Bendick, Jackson, Reinoso, & Hodges, 1991; Bendick, Rodriguez, & Jayaraman, 2010; Ghumman & Ryan, 2013; Lodder, McFarland, & White, 2003; Pager et al., 2009). In Europe, the studies conducted under the ILO project have also included some observations on the interview process, particularly at the telephone stage (Arriijn, Feld, & Nayer, 1998; Attström, 2007; Bovenkerk, Gras, Ramsoedh, Dankoor, & Havelaar, 1995; Cediey & Foroni, 2008; de Prada, Actis, Pereda, & Molina, 1995; Goldberg, Mourinho, &

Kulke, 1995). As Paper V of this dissertation discusses, correspondence tests and particular the email correspondence with employers, also offer ways to directly observe differences in the treatment of candidates.

4.6 Victim Research: Perceived Discrimination and Legal Cases

All of the above-mentioned methods focus on the role of the alleged offender who acts discriminatory (i.e. the employer). Studies of perceived discrimination and to some extent studies of legal cases shift the focus to the victims of discrimination. Victim surveys focus on the experiences of the targets of discrimination and what they perceive as discriminatory treatment. Questions on perceived discrimination have been incorporated into larger surveys, e.g. the Eurobarometer asking about whether respondents had observed cases of discrimination, or there can be specific surveys dedicated to the subject of perceived discrimination (Larja et al., 2012). Such thematic surveys on perceived discrimination have been carried out e.g. in the Netherlands (Andriessen, Fernee, & Wittebrood, 2014) or Germany (Beigang, Fetz, Foroutan, Kalkum, & Otto, 2016) and, on an EU level in the EU-MIDIS survey, where the results of its second wave were published in 2017 (European Union Agency for Fundamental Rights, 2017). They enable researchers to not only address discrimination at the moment of hiring, but also in other areas of the work life, such as promotions or termination decisions (OECD, 2013). While victim surveys provide great information on experiences of discrimination since they report more instances than only cases that were reported to the police (Wrench, 2011), they also face numerous limitations. First of all, there is a problem of validity, since there is no clear and common understanding of what constitutes discrimination in these surveys, but victims report their subjective experiences of discriminatory treatment. It is therefore unclear whether studies of perceived discrimination over- or underestimate the extent of discrimination encountered by minorities (Larja et al., 2012; OECD, 2013). Second, it cannot be assumed that all ethnic minority groups have the same level of awareness and sensitivity about discriminatory treatment (OECD, 2013). Finally, sampling and contacting participants poses a challenge. As Larja et al. (2012) point out, in nationally representative samples, minorities often “drown in the masses” (p. 25).

Not included in Veenman’s (2010) classification are studies analysing reported legal cases of discrimination, yet these works also focus on the victims’ experiences. Such information is often provided by equality bodies or specific monitoring institutions. Looking at the example

of the US, the Equal Employment Opportunity Commission (EEOC) not only provides information on court cases concerning labour market discrimination, but is also authorised to make complaints on behalf of the victims of discrimination. These legal cases provide valuable information about concrete experiences of discriminatory treatment and of “the *nature* and *forms* of ethnic discrimination in employment” (Wrench, 2016, p. 119, emphasis in the original). Yet, focusing only on the number of formal complaints is likely to underestimate the extent of discrimination, as most victims of discrimination are not aware that they have experienced discriminatory treatment, do not start legal proceedings, do not know their legal rights and the legal procedures in these cases, or are discouraged by the low levels of compensation that can be obtained (OECD, 2013; Wrench, 2011). As the MIPEX⁴ indicators have shown, there are big differences between equality bodies across countries and their legal standing and independence to help victims of discrimination to take legal action. The number of complaints is low compared to how many people indicated having been a victim of discrimination in other studies (Huddleston, Bilgili, Joki, & Vankova, 2015a, pp. 15, 17). This low number of complaints is also confirmed in the EU-MIDIS II survey, according to which only 12% of the participants had reported the most recent experience of discrimination that they were subjected to because of their ethnic or migration background (European Union Agency for Fundamental Rights, 2017). Court cases are thus not able to provide information about the extent of discrimination in the labour market, but add valuable insights on the mechanisms of discrimination.

4.7 Conclusion

Researchers have used numerous methods to measure discrimination, all of which have their strength and limitations. While some are better suited to measure the extent of discrimination, other methods perform better when the mechanisms behind discriminatory treatment are the research focus. More recent research projects have therefore increasingly combined different methodologies to get a broader picture of labour market discrimination and to work around the limitations of using only one methodology. Triangulation of methods has become increasingly popular in international discrimination research.

⁴ MIPEX is the Migration Policy Index. It measures integration policies in 38 countries in eight policy areas: labour market mobility, education of children, political participation, family reunion, access to nationality, health, permanent residence and anti-discrimination.

As e.g. Protsch and Solga (2017) have pointed out, of the methodologies listed above, field experiments “are often regarded as the methodological gold standard” (p. 392) in discrimination research. Hence, they warrant a section of its own that specifically addresses the use of field experiments. Due to a lack of research on discrimination in the Swiss labour market, a correspondence test has been chosen as the methodology for this dissertation. While survey experiments and social psychological research about discrimination on the labour market exists for Switzerland and have gained traction in the last years, there is so far only one correspondence test of discrimination on the labour market for candidates transitioning from apprenticeships to their first jobs (Fibbi et al., 2003). The particular limitations and advantages of field experiments are discussed in the following section, before the last sections explicitly focus on the Swiss case and the literature on labour market discrimination in Switzerland, which has increased in recent years.

5. Methodology of field experiments

Given the choice of a field experiment (a correspondence test) for this research project, the methodology will now be discussed in great detail. Field experiments have become increasingly popular in economics, political science and sociology over the last decades, particular since the 1990s (Jackson & Cox, 2013; List, 2009), and a growing body of literature addresses the advantages and disadvantages of field experiment and how to design and report them (e.g. Field & Hole, 2002; Gerber & Green, 2012; Jackson & Cox, 2013; List, 2011; Teele, 2014b).

While laboratory experiments often face the criticism in how far results were affected by the setting in a laboratory, or the use of student samples, field experiments take place outside of the laboratory in the natural environment of the research subjects. Harrison and List (2004) distinguish between three types of field experiments, artefactual field experiments, framed field experiments and natural field experiments. Of those three, natural experiments in which participants are unaware of their participation in an experiment, combine “the most attractive elements of the method and naturally occurring data: randomization and realism” (List, 2011, p. 6). List even claims that

“Such experiments are a useful marriage between laboratory and naturally occurring data in that they represent a mixture of control and realism not usually achieved in the lab or with naturally occurring data” (List, 2009, p. 439).

Furthermore, List (2011) argues that the random inclusion of participants in contrast to self-selected participants increases the representativeness of the sample of participants.

Field experiments have been shown to be a valuable research tool to study causal effect in the social world, yet they also face criticism: they can be extremely expensive and/or difficult to implement (Gerber & Green, 2012), particularly in the case of natural field experiments they raise ethical concerns, because participants did not provide their voluntary and informed consent (Barrett & Carter, 2014; List, 2011; Teele, 2014a), and they are often difficult to replicate (Levitt & List, 2009). However, even if it is usually not possible to completely replicate the experiments leading to identical results, it is still possible to repeat the experiment and replicate the research design, which should lead to similar results if the external

circumstances remain similar. Finally and similar to the question of replicability, looking at the field of development economics, Banerjee and Duflo (2014) question the impact of environmental dependence and if similar results would be found in a different setting.

5.1 Field experiments on discrimination in the labour market

One of the many areas in which field experiments have been frequently used are studies of discrimination. Since discrimination is usually prohibited, it is hard to observe the phenomenon directly and other methodologies, as previously discussed, face limitations when it comes to measuring the extent of discrimination (Jackson & Cox, 2013). They have been conducted in several domains, most well-known in the employment or housing market and focused on different grounds of discrimination. While the majority of studies measured ethnic and racial or gender discrimination, experiments have also been conducted on the grounds of age, disability, sexual orientation, caste, religion, or obesity (for reviews see Baert, 2018; Bertrand & Duflo, 2016; Neumark, 2016; Riach & Rich, 2002; Rich, 2014b). Despite this vast range of applications, the following discussion focuses on field experiments on racial and ethnic discrimination in the labour market. Two existing meta-analyses on studies of ethnic and racial discrimination have shown that minority job candidates encounter discrimination in all countries studied (Zschirnt & Ruedin, 2016) and that discrimination rates remain stable over time (*ibid*; for the US in particular Quillian et al. (2017)).

Field experiments on labour market discrimination have also increased in the last decades, thus contributing to the overall growth of field experiments across disciplines. Since the late 1960s field experiments have been conducted to measure the extent of ethnic or racial discrimination in a given labour market (Gaddis, 2018a; Zschirnt, 2016) and the methodology has been advanced rapidly since the first testings. This development is the focus of the first paper of this dissertation. In the early 1990s, Bovenkerk (1992) established the research framework for an comparative research project on labour market discrimination that was conducted by the International Labour Office (ILO). It provided the first clear guidelines on how to plan and execute research projects on ethnic or racial discrimination in the labour market. More than 25 years later, Gaddis (2018b) published an edited book that discusses several recurring methodological questions regarding the design, implementation and analysis of field experiment on hiring discrimination in detail.

The field experiments that measure discrimination against ethnic or racial minorities present real employers that advertised open positions with fictitious but equally qualified candidates, which differ only in the characteristic of interest, in this case race or ethnicity. Responses from employers are carefully recorded and differences in treatment of the majority and minority applicants are then attributed to discrimination. It is possible to distinguish between two types of field experiments based on the way employers are contacted: in-person audit studies and written correspondence tests.

5.2 In-person audit studies

In-person audit studies have been mostly conducted in the US. Employers are contacted by matched pairs of testers, who apply for an open position in-person, either on the business premises or by phone. The individual testers are matched as carefully as possible on numerous characteristics (e.g. age, height, weight, and attractiveness) and receive extensive training on how to behave in the application process.

Conducting in-person audit studies has several advantages. First, ethnicity or race of applicants is easily signalled since applicants present themselves in person. Second, in-person audits are not limited to positions in which applications have to be made in writing, but also allow researchers to study discrimination in lower skilled positions, where applicants often apply in-person at the place of business. Third, by sending individual testers to the job interviews it is possible to not only collect data on the outcome of the application (i.e. job interview or rejection), but also to gather information on the way that applicants were treated during the application process (Pager, 2007). These personal observations can provide valuable qualitative data on instances of subtle discrimination, especially when experiences of testers are analysed side-by-side. Several US studies have included information on the length of phone calls or interviews, where the interview took place (in private or in the store?), if the interview was conducted by a manager or subordinate, e.g. a receptionist, the number of topics discussed during the interview, the differences in compensation, hours or shifts offered, if information about the job duties was offered without being asked, how politely applicants were treated (waiting time before the interview, did employers introduce themselves, did employers call candidates by (the right) name, did they shake their hands?), if additional vacancies were mentioned, or if applicants were steered towards other positions (e.g. Bendick, 1996; Bendick

et al., 1991; Bendick et al., 2010; Ghumman & Ryan, 2013; Lodder et al., 2003; Pager et al., 2009).

Critics of in particular in-person audit studies have been vocal in pointing out the problems of this methodology. First of all, matching and training the individual testers takes a lot of time, effort, and resources and conducting the experiment requires intensive supervision (Pager, 2007). Heckman and Siegelman (1993) and Heckman (1998) have been the most vocal in their criticism of the methodology. Their points focus on (1) the small number of tests being carried out for a limited sample of low-skilled occupations that do not allow for the generalisation of the results, (2) the presentation of results, which according to Heckman (1998) “is far less decisive than on the issue of market discrimination than it is claimed to be” (p. 101), and (3) the problem of unobservable characteristics among tester pairs which could distort results, that matching testers too closely could privilege rather small differences and “may be forced to privilege relatively minor characteristics simply out of necessity of breaking the tie” (Pager, 2007, p. 116). Heckman and Siegelman (1993) also pointed out the problem of experimenter effects, since “the experimenter is not simply a passive runner of subjects, but can actually influence the results” (p. 215) or the limitations of having a sample based only on publicly advertised positions. Summarising Heckman’s and Siegelman’s arguments, “different degrees of success in the hiring process should be attributed to the “failure by the researchers to match the testers on some subtle productivity-related characteristics” (Bendick & Nunes, 2012, p. 248).

5.3 Correspondence Test of Ethnic Discrimination in the Labour Market

Like in-person audit studies, correspondence tests present an employer with equally qualified fictitious applicants, yet in this case applications are made in writing, not in-person, and they only address the first stage of the hiring process, the question whether applicants are invited for a job interview or not. As soon as a candidate is invited for a job interview, the researchers quickly and politely decline the invitation to limit the burden on the employer. Looking only at correspondence tests, our meta-analysis (Paper 2 of this dissertation, Zschirnt and Ruedin (2016)), identified 43⁵ studies that had been conducted in OECD countries between 1990 and 2015, and another 9 studies meeting these criteria were published between 2016 and 2018

⁵ Since Akintola (2011) focuses on both Canada and Sweden, this study is treated as two separate studies in the meta-analysis.

(Baert, Albanese, du Gardein, Ovaere, & Stappers, 2017; Baert & Vujić, 2016; Bagley & Abubaker, 2017; Gunn Elisabeth Birkelund, Ugreninov, Chan, Midtbøen, & Rogstad, 2017); Darolia, Koedel, Martorell, Wilson, and Perez-Arce (2016); (Drydakis, 2017; Koopmans, Veit, & Yemane, 2018; Liebkind et al., 2016; Pierné, 2018)⁶. Since Georgia is not part of the OECD the correspondence test by Asali, Pignatti, and Skhirtladze (2017) is not included in this list or in Figure 1.

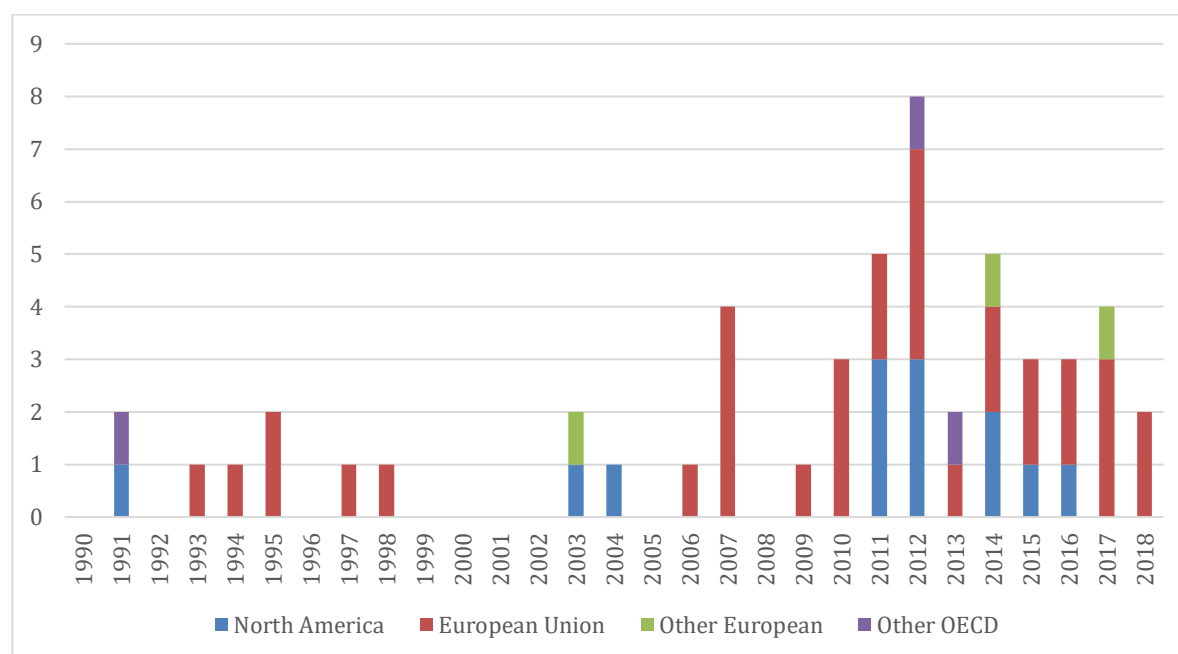


Figure 1: Correspondence tests on ethnic and racial discrimination in OECD countries 1990-2018

Creating matching candidates is easier in correspondence tests, as they only use written applications and do not have to match individual testers. Yet, by only focusing on written contact, correspondence tests are limited to the first stage of the application process, i.e. whether applicants are invited for a job interview or not. Once an invitation is received, it is quickly and politely declined by the researchers. Yet, looking at the ILO studies that tested discrimination in two stages of the hiring process, Riach and Rich (2002) and Rich (2014a) point out that more than 80 percent of discrimination occurred at the first hiring stage, thus denying ethnic minority candidates the chance to even present themselves in a job interview.

⁶ Like Akintola (2011), Bagley and Abubaker (2017) also focused on two countries.

In correspondence tests researchers have complete control in constructing fictitious CVs that are matched for socio-economic background, work experience and education, family status, etc., and only differ in the characteristic to be studied. They are easier to administer than in-person audits and Lahey and Beasley (2009, 2018) discuss how the process of creating resumes can be simplified by relying on a resume creation software. The standardised resumes can then be sent to a great number of employers at a much lower cost compared to in-person audit studies. It allows to apply for a wider range of positions, not only those typically lower skilled positions in which applications in person are common. However, the higher the qualification level, the more challenging it becomes to construct resumes for the fictitious applicants that show the required qualifications.

In contrast to in-person audit studies, in which trained testers are chosen based on certain characteristics, e.g. their skin colour, written correspondence tests have to signal ethnicity or race differently. The most common is the use of distinct names – names that are perceived as black, Hispanic or white American or as belonging to the native group of the country or immigrant groups. Government statistics, such as census data, population registers or lists with baby names published per country are often used as an important source in the construction of names. If this wealth of data from official sources is not readily available, researchers have used websites listing common names for the respective ethnicities or have relied on the work from previous studies, especially the names used in Bertrand and Mullainathan (2004) (e.g. Dechief & Oreopoulos, 2012; Oreopoulos, 2011). In countries where it is common to attach a photograph to the CV, this can also be used to convey racial or ethnic clues, yet including photographs introduces a new set of unobservable variables and they need to be very carefully prepared and pretested (Rich, 2018; for an example on the use of photographs see Weichselbaumer, 2016b).

However, the use of names also raises the problem that the selected names might not only signal the race or ethnicity of a fictitious candidate, but that they introduce unintended and unobserved socio-economic connotations “thus confounding the effects of race and class” (Pager, 2007, p. 111). These socio-economic connotations have been discussed in the US context since the much-cited correspondence test by Bertrand and Mullainathan (2004) was published. Fryer and Levitt (2004) analyse government data on first names given to babies born in California between 1961 and 2000 for the “causes and consequences of distinctively black names”, showing changes in naming patterns over this time frame. They argue that “until the late 1970s

the choice of Black names was only weakly associated with socioeconomic status; in the 1980s and 1990s distinctively Black names have come to be increasingly associated with mothers who are young, poor, unmarried, and have low education” (p. 787). They also specifically discuss their findings in relation to results from audit studies using distinctively black names, but conclude “that carrying a black name is primarily a *consequence* rather than a cause of poverty and segregation” (p. 801, emphasis in the original). Recently the topic has garnered renewed attention with the work of Gaddis who focused on the perception of black first names, referring explicitly to the work of Bertrand and Mullainathan (2004) in his title, showing “that a number of characteristics of an individual name matter: gender, popularity, type of last name included, and the average level of education of mothers who commonly give that name, among others.” (Gaddis, 2017a, p. 485). He conducted a similar experiment on Hispanic names (Gaddis, 2017b), cautioning researchers to carefully pre-test names used in correspondence tests for these signals. While Gaddis has focused (mostly) on first names, Crabtree and Chykina (2018) built on his work to study the racial perception of last names and how these might change across geographical contexts.

Recently, more work on the methodology of in-person audit and correspondence testing has been published and the just released book by Gaddis (2018b) provides a wealth of information that will be extremely helpful for researchers that plan their first correspondence test as well as for those who are already more experienced. The contributions systematically address issues such as internal and external validity (Lahey & Beasley, 2018; also Ross, 2017), the sampling of research participants and the timing of the experiment (Lahey & Beasley, 2018), using single, paired or multiple applications (Lahey & Beasley, 2018; Vuolo, Uggen, & Lageson, 2018), calculation of the sample size (Lahey & Beasley, 2018; Vuolo, Uggen, & Lageson, 2016; Vuolo et al., 2018), or the use of emails in correspondence studies (Crabtree, 2018). Furthermore, a recent paper by Bonoli and Fossati (2018) analyses instances where minority candidates are preferred to majority candidates, which in “traditional” correspondence test discussion is hardly discussed.

5.4 Criticism of in-person audit studies and correspondence tests

Many of the major points criticised by Heckman and Siegelman in the case of in-person audit testing have become obsolete by changing to written correspondence tests. Yet, one important criticism also applies to correspondence studies, namely the problem of unobserved variance

between groups which can “cast serious doubt on the validity of the evidence from these studies” (Neumark, 2012, p. 1129).

David Neumark has focused on this question in his 2012 article “Detecting Discrimination in Audit and Correspondence studies” and developed a method to “recover an unbiased estimate of discrimination” (p. 1128), which is possible if the studies include variation in the design of the fictitious applications that affect productivity and hiring decisions. In a complex mathematical procedure, which is explained in detail in his article, Neumark makes use of the variations in fictitious applications to study how not only the characteristic to be studied in the testing, but also the quality of an application influences hiring decisions. He points out that “audit and correspondence studies can generate spurious evidence of discrimination” (Neumark, 2012, p. 1129), i.e. over or underestimate the extent of discrimination or can even fail to identify evidence of discrimination, usually caused by unobserved variables in the application process.

In his original design Neumark uses the data of Bertrand and Mullainathan (2004), who included variations in their applicant pool, dividing it into high and low quality resumes, using “criteria such as labor market experience, career profile, existence of gaps in employment, and skills listed” (p. 994) and adding more features to the high quality resume, such as e.g. “summer or while-at-school employment experience, volunteering experience, extra computer skills, certification degrees, foreign language skills, honors, or some military experience” (p. 994). Neumark argues that his test for unobserved variables shows, in the case of the Bertrand and Mullainathan data, that there is “stronger evidence of race discrimination that adversely affects blacks than is obtained when differences in the variances of the unobservables are ignored” (2012, p. 1149). Furthermore, he states that “the method proposed here can easily be implemented in any future correspondence (or audit) study. All that is needed is for the resumes or applicant to include some variation in characteristics that affect the probability of being hired. ... All that needs to be done is to intentionally create resumes of different quality” (p.1149). Both M. Carlsson, Fumarco, and Rooth (2014) and Neumark and Rich (2018) use the Neumark test to assess the robustness of findings of field experiments on discrimination. The former focus on Swedish studies and their results suggest that unobserved group differences may play an important role in the design of a correspondence test (M. Carlsson et al., 2014, p. 14), the

latter examine ten correspondence tests that were conducted either in the housing⁷ or labour⁸ market, that provide sufficient information to apply Neumark's method. Neumark and Rich (2018) report that the findings of housing market studies have proven to be quite robust to the corrections added by the Neumark test, yet only half of the labour market studies show robust results, while in the rest “estimates of discrimination fall to near zero, become statistically insignificant, or change sign” (p.1). Thus, they encourage researchers to include variation in the quality of resumes created for a field experiment to allow this control for unobserved variables.

5.5 Testing discrimination theories in field experiments

The research design for recent field experiments have often tried to address the question whether discrimination is due to taste-based or statistical discrimination and Thijssen (2016) provides an overview of studies that leaned towards either form of discrimination. Looking at taste-based discrimination first, he argues that the discrimination rates found in Austria by Weichselbaumer (2016b) could point towards taste-based discrimination, since the amount of application materials provided should mean that there is hardly any reason for employers to resort to statistical discrimination. Thijssen (2016) also categorises the field experiment by Jacquemet and Yannelis (2012) as one that leans towards taste-discrimination, since they test for a distaste of foreign names, so called “ethnic homophily”, in the Chicago labour market by using amongst other fictitious names which cannot be attributed to a certain ethnic group. He also includes studies on ethnic hierarchies among the fictitious candidates in this group (Booth et al., 2012; McGinnity & Lunn, 2011; Pager et al., 2009). While trying to find signals to explicitly test for taste-based discrimination is challenging, focusing on statistical discrimination theory is more straight-forward, since the underlying assumption of statistical discrimination theory is that employers resort to discrimination if they lack information about the applicant. Thus, varying the amount of information in an application should also influence the propensity of employers to discriminate. Examples of correspondence test that varied information or resume quality include among others Kaas and Manger (2012), Bertrand and Mullainathan (2004), Pager (2003), Pager et al. (2009), or Gunn Elisabeth Birkelund, Heggebø, and Rogstad (2017).

⁷ The housing market studies include: Ahmed, Andersson, and Hammarstedt (2010), Bosch, Carnero, and Farre (2010), M. Carlsson and Eriksson (2014), and Ewens, Tomlin, and Wang (2014)

⁸ The labour market studies are: Baert, Cockx, Gheyle, and Vandamme (2015), M. Carlsson and Rooth (2007), Drydakis (2014), H.-A. Lee and Khalid (2016), Oreopoulos (2011), and Bertrand and Mullainathan (2004)

In our meta-analysis we also tried to point out which findings would rather support taste-based or statistical discrimination theory (Zschirnt & Ruedin, 2016). We interpreted the finding of almost no differences in discrimination rates between the first and second generation, as pointing towards taste-based discrimination theory. Since the second generation is portrayed as holding local qualifications, employers should not have to resort to statistical discrimination to compensate a lack of information about the qualifications. Furthermore, as also argued in Thijssen (2016), the existence of ethnic hierarchies that we also documented in our meta-analyses also points towards the theory of taste-based discrimination. While Thijssen (2016) argues that the discrimination rates reported by Weichselbaumer (2016b) in the Austrian labour market support the theory of taste-based discrimination, we argue in our meta-analyses that the lower discrimination rates found in German speaking countries seem to confirm the assumptions of statistical discrimination theory, i.e. that more information about applicants will make employers less likely to resort to statistical discrimination theory. However, both lines of argumentation are sides of the same coin – on the one hand focusing on the discrimination rates that are measured, and on the other hand of the lower rates of discrimination compared to countries where such extensive application material is not the norm.

It is thus possible to design correspondence tests in a way to specifically tests certain assumptions of taste-based or statistical theories of discrimination. Findings of such studies could also have policy implications. If it is for example confirmed that more extensive application packages improve the chances of minority applicants, it would be comparatively easy to adopt such application procedures also in other countries.

5.6 Research Ethics Concerns

Next to considering the theoretical framework, another important methodological element of in the preparation of correspondence tests is the question of research ethics. Like the beginning of this section indicated, field experiments raise several research ethics concerns. Objections that are made regarding field experiments on discrimination in the labour market are (1) that correspondence testing infringes the principles of voluntary participation and informed consent, (2) that researchers are deceiving their research participants, and (3) that correspondence testing can have negative consequences for employers who unwillingly participated in the experiment. A detailed discussion of the research ethics concerns is the focus of Paper 3 of this dissertation. Like in the seminal article on research ethics by Riach and Rich (2004a), it is argued that

thorough preparation and careful execution of a correspondence test can mitigate most of these concerns.

5.7 Matched pair testing vs. single applications

With the recent emergence of more research designs that only send one application per vacancy (e.g. Koopmans et al., 2018; Weichselbaumer, 2015, 2016a) instead of the “usual” carefully matched pairs or sets of applications, new questions arise in how far this change to the methodology can still measure labour market discrimination. While the risk of detection is certainly lower in experiments using only single applications, critics of this unpaired methodology have argued that these experiments can only measure preferential treatment in the labour market. To measure discrimination in hiring decisions, individual employers would have to be forced to choose between potential candidates (Riach & Rich, 2004b). It is therefore not possible to say if an employer who chooses not to invite the minority candidate does so because of the minority status or for any other reason. Recent contributions by Vuolo et al. (2016) and Vuolo et al. (2018) have focused on the statistical implications of choosing paired or unpaired research designs, yet a thorough discussion whether findings from single application research designs can still measure discrimination is still missing in the literature. Furthermore, the number of applications sent per employer can also have ethical implications if too large sets of applications are being submitted for one vacancy.

5.8 Field experiments for research vs. field experiments for enforcement

The field experiments discussed in this section have been conducted for the purpose of researching the extent of hiring discrimination in the labour market in question. However, the methodology has also been used for enforcement purposes to collect evidence of discrimination against employers. As Pager and Western (2012) or Cherry and Bendick (2018) emphasise, the audit methodology was initially designed to test the effectiveness of anti-discrimination laws. US law recognises the role of testing in the enforcement of these laws by giving legal standing to testers and NGOs that employ testing “to become plaintiffs in litigation based on testing evidence alone” (Bendick & Nunes, 2012, p. 255). Results from testings can be used as corroborative evidence in discrimination cases against employers. Testings that are explicitly conducted for enforcement purposes make one important adjustment in comparison to testing for research: Instead of sampling each employer only once, they conduct multiple audits with

the same employer, to document whether discriminatory decisions occur systematically across multiple application processes (Pager & Western, 2012).

While most discussions of testing for enforcement have focused on the US context, there are also countries in Europe that recognise the use of testing in anti-discrimination cases. Veronique van der Plancke (2007) and Calvès (2007) discusses the experiences made in Belgian and French courts respectively, while Rorive (2009) provides an overview on situation testing in 11 European countries (Belgium, Czech Republic, Denmark, Finland, France, Hungary, Latvia, The Netherlands, Slovakia, United Kingdom, and Sweden).

5.9 Conclusion

Field experiments on labour market discrimination have become an invaluable method to measure the extent of discrimination in hiring decisions and contribute to the societal debate on discrimination with these findings. As reviews with a more historical focus have shown, the methodology of field experiments has advanced considerably over time since Daniel (1968) published the first field experiment on ethnic and racial discrimination, sometimes directly in response to criticism of the methodology (Cherry & Bendick, 2018; Gaddis, 2018a; Zschirnt, 2016). The lists of countries and the contexts in which correspondence tests have been conducted has become much more diverse and research questions have become more targeted. Furthermore, there have been more studies making use of multi-method or interdisciplinary approaches which can offer new perspectives. Field experiments also increasingly record more variables that could influence hiring decisions and incorporate variations in the design of resumes to allow for further statistical analysis of findings using the Neumark test. Still, as discussed e.g. in Neumark (2012), field experiments also face limitations – most of which have been listed in the critique by Heckman and Siegelman. While most of these points have been alleviated in correspondence tests and by applying Neumark’s method to test for the robustness of findings of discrimination, some limitations still remain. One criticism that is often brought forward is that field experiments, and in particular correspondence tests, only measure discrimination at one very specific point in time. However, as mentioned above, the decision whether an applicant receives an invitation for a job interview is a crucial point in the application procedure. One debate that is currently gaining traction given the increasing number of these research designs, is in how far field experiments using only single applications per employer can measure labour market discrimination. While the issue of matched vs. unmatched

testing has been discussed in detail by Vuolo et al. (2018), the more fundamental question that discrimination can only be measured if employers are forced to make a choice between candidates is hardly discussed (for exceptions that briefly mention this question see Cherry & Bendick, 2018; Riach & Rich, 2004b).

6. Context in Switzerland: Labour market inequalities between natives and immigrants

While the previous two sections have focused on providing an overview of the methodologies used to study labour market discrimination in international research, with a particular focus on field experiments, the next two sections will address labour market inequalities between natives and immigrants in Switzerland. They will first provide an overview of the current state of knowledge on labour market discrimination in Switzerland obtained by using the previously discussed methodologies, before presenting the research design for a correspondence test for ethnic discrimination in hiring decisions in the Swiss labour market.

6.1 Immigrants in Switzerland

Switzerland has one of the highest shares of immigrants in its resident population among OECD countries, similar levels are only reported for Luxemburg and Australia (OECD, 2012). As one of the first European countries, Switzerland started a labour recruitment programme already in 1948 to attract Italian guest workers. A second wave of guest workers arrived in the 1970s and 1980s, coming mostly from former Yugoslavia and Portugal. Migration peaked again in the 1990s, but at that time due to humanitarian migration, with refugees from the former Yugoslavia constituting the biggest group. While Switzerland is not part of the European Union, the gradual establishment of free movement of persons within the European Economic Area led to a fourth wave of migration. Since 2002 migration from EU countries has increased considerably, with the majority of immigrants being highly qualified, coming from high-income countries, and being already able to speak one of the national languages upon arrival (OECD, 2012).

Looking at data provided by the Statistical Office for 2016 shows that of the approximately 8.33 million Swiss residents 2.05 million hold foreign nationalities. The biggest share of immigrants is made up by people from European countries (1.7 million, 20.8% of the Swiss resident population), which distinguishes Switzerland from other countries in Europe. Immigrants from the countries belonging to the EU-28 or the European Free Trade Area (EFTA) account for 66.5% of the immigrant population with Italy (15.2%) being the first country of origin, followed closely by Germany (14.7%) and Portugal (13.1%). French immigrants make up 6.0% of the immigrant population, before Kosovars (5.2%), Spaniards

(4.0%), Serbians (3.5%) and Turks (3.4%). Looking at the share of immigrants from non-European countries, 6.7% of foreigners in Switzerland come from Asia, 4.6% from Africa and 3.8% from America (Bundesamt für Statistik, 2017c, plus own calculations). In 2016, the greatest number of foreigners lived in the Cantons of Zurich (almost 400.000), Vaud (almost 264.000), Geneva (almost 200.000), and Bern and Aargau (both almost 165.000). In ten cantons the share of foreigners amongst the resident population was above 24% (Bundesamt für Statistik, 2018b).

Among the share of foreigners aged 15 and above, the majority are first generation migrants. Due to the restrictive Swiss naturalization policy, in 2016, 35% of the second generation are still legally considered as foreigners, while two thirds of the second generation were considered as Swiss. Of those considered Swiss, 78% obtained the Swiss passport through naturalisation and 22% received it upon birth (Bundesamt für Statistik, 2017b).

6.2 Public debate on immigration in Switzerland

Internationally, Switzerland is often portrayed as a tolerant society, that has been able to “incorporate different religions and languages without destroying their cultural identities” (Freitag, Vatter, & Mueller, 2015, p. 1). Direct democracy has been pointed out as one important feature that makes the inclusion of different minorities in the territory possible. However, due to the restrictive Swiss nationality laws with their strong focus on *ius sanguini*, only approximately 5 million of the 8 million inhabitants of Switzerland are able to participate in the direct democracy e.g. by voting in popular initiatives or referenda (Boulila, 2018). As Freitag et al. (2015) discuss, scepticism towards immigration has increased and it has been possible to observe “a cultural shift in Switzerland from the tolerant consensus democracy to a nation characterized by increasing animosity towards immigrants” (p.2).

Immigration has traditionally been a highly-debated issue in the Swiss society and xenophobia and fears of “*Überfremdung*” (foreign domination) have led to a rising number of popular initiatives and referenda since the 1960s/1970s with most of these votes trying to restrict immigration and mobility and the access to asylum (Arrighi, 2018). As Boulila (2018) pointed out in a recently published article on race and racial denial in Switzerland, several of these “successful referendums and popular initiatives, accompanied by racist campaigns, have been identified to violate international law” (p.1) and have raised international concern about racism

in Switzerland. Prominent examples are the popular initiative banning the building of minarets (accepted in 2009) or the Initiative against Mass Immigration (accepted in 2014). Both in the media as well as in political campaigns immigrants have frequently been portrayed as a threat to Swiss society. One of the most notorious examples has been a poster showing cartoon white sheep on a Swiss flag kicking a black sheep of the Swiss flag in the campaign to automatically deport foreign nationals that had committed a crime. Similar threatening posters were used in a campaign against the proposed facilitation of access of citizenship for the second generation in 2004, showing the strong culturalist discourse employed in these debates (Wessendorf, 2008). These initiatives have not only been directed at immigrants coming from third countries, but in the case of the Initiative against Mass Immigration, which was adopted in February 2014, has been targeted towards highly skilled and well qualified immigration from EU countries. Since 2002 immigration from EU countries has increased considerably under the EU free movement provisions. As Freitag et al. (2015) show, the adoption of this initiative is “the latest earthquake to have shaken Swiss politics” (p.2), by voting to stop the free movement of persons and to reintroduce a quota system. This vote and its conflict with international laws (i.e. treaties signed with the European Union) had serious repercussions and caused Switzerland to be temporarily excluded from the EU’s research funding schemes of Horizon 2020. After long negotiations trying to combine the results of this vote with international obligations, Switzerland introduced an “*Inländervorrang*” (preference of residents in the country), according to which job candidates that are Swiss or already residents in Switzerland have to be given preference above candidates from abroad.

In the last decades, the political right has been able to exploit the topics of immigration and asylum to further their populist agenda and campaigns frequently focus on security and immigration issues. Thus, the SVP/UDC has become the biggest political party in the Swiss parliament, winning almost 30% of the votes in the last Federal Election in 2015. This “climate of hostility and threat” (Boulila, 2018, p. 5) created by the SVP/UDC has also been problematized by the UN Committee on the Elimination of Racial Discrimination that expressed its concerns with “the discriminatory effects of the initiatives and referendums put forward by the far right since 2007” (ibid.) and criticised the lack of federal legislation providing a clear definition of racial discrimination and clearly prohibiting racial discrimination and making it an offence under criminal and civil law. The Committee specifically referred to the above-mentioned initiative against the construction of minarets (2009), the initiative on the “expulsion of foreign criminals (2010) and the initiative “against mass immigration” (2014)

and expressed the concern about the racist stereotypes found in the media and political posters and the lack of persecution in such instances of racism (CERD, 2014).

However, as the electoral success of right wing parties across Europe has shown, such anti-immigration discourse has not been particular to Switzerland, but can also be observed in many other European countries (e.g. Boulila, 2018; Wessendorf, 2008). What is, however, peculiar in the case of Switzerland are the possibilities of referenda and popular initiatives in the direct democratic system, the high share of foreigners that are excluded from participating in these direct democratic instruments, and the good integration of immigrants in the labour market.

6.3 Immigrants in the Swiss Labour Market

The Swiss labour market has been very stable since the turn of the millennium, despite facing several severe challenges, such as “the Great Recession in 2008, a massive inflow of foreign workers, and a sharp appreciation of its currency” (Lalive & Lehmann, 2017, p. 10). Compared to other OECD countries and despite the portrayal as immigration as a threat in the public discourse, the outcome of immigrant integration is favourable in Switzerland, with immigrant men and women having higher employment rates than in most other OECD countries (Liebig, Kohls, & Krause, 2012). Yet, looking only at the Swiss labour market, immigrants face higher rates of unemployment compared to natives. Overall, unemployment in Switzerland is low, in the last quarter of 2017 the unemployment rate was at 4.7%, below the OECD average of 5.5% (OECD, 2018c). While the native-born unemployment rate was 3.2% in 2016, both for men and women, it was more than twice as high for the foreign-born population, for which the unemployment rate in 2016 was 8.2% (7.6% for foreign born men and 8.9% for foreign born women) (OECD, 2018a, 2018b).

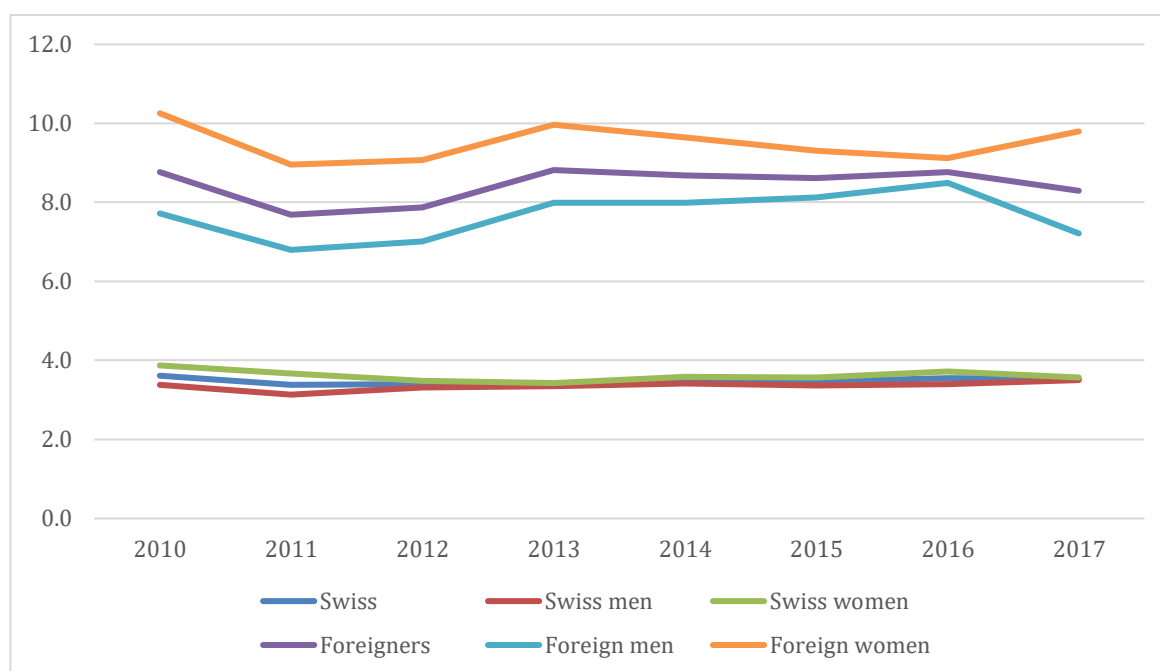


Figure 2: Unemployment rate based on ILO definition by nationality and gender (based on Bundesamt für Statistik, 2018a)

Not only do foreigners in Switzerland face higher unemployment, they also (on average) earn less than natives. Swiss men have the highest gross monthly wages, followed by foreign men with a C-permit (the long-term residence permit), and all foreign men. Women's gross monthly wages are lower in all categories, yet the differences are less pronounced and foreign women with a B permit (initial residence permit) earn slightly more than other foreign women (based on Bundesamt für Statistik, 2015).

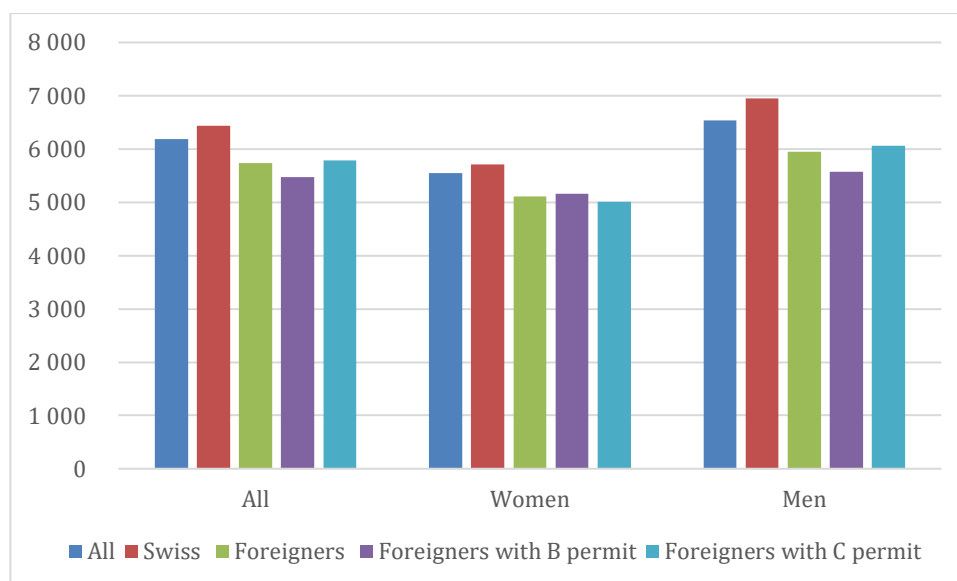


Figure 3: Gross monthly wage in 2014 in CHF in the private sector (based on Bundesamt für Statistik, 2015)

As Liebig et al. (2012) have pointed out, by the time their report was published in 2012 discrimination against immigrants had not been given much attention in the media or in politics in Switzerland. Indeed, most of the studies which will be discussed below are quite recent. One reason for this lack of research and information could be the lack of awareness when it comes to the topic of discrimination of immigrants in the Swiss labour market. This lack of awareness is coupled with missing institutional structures to combat discrimination and the fact that Switzerland does not have specific legislation that addresses discrimination against immigrants. The lack of comprehensive anti-discrimination legislation has also been admonished in the MIPEX report on Switzerland. While Switzerland ranks in the middle of the 38 countries studied when it comes to the overall score on integration policies, it is judged as “slightly unfavourable” in the field of anti-discrimination policy ranking at the bottom of the list, just above Turkey and Japan (Huddleston et al., 2015a). As the key finding points out, Switzerland is

“one of the very few countries without a comprehensive anti-discrimination law and equality body with legal standing; [and] a sizeable number of potential victims are poorly protected against racial, ethnic, religious and nationality discrimination.” (Huddleston et al., 2015b, p. 40).

Coupled with the strong principle of contractual freedom, this lack of legislation makes it easy for employers to make hiring decisions based on candidates’ origins (OECD, 2012). Several international organisations including the Committee on the Elimination of Racism or the UN Special Rapporteur on Racism have repeatedly urged Switzerland to adopt a comprehensive antidiscrimination and anti-racism law⁹.

As immigrants are quite well integrated in the Swiss labour market compared to other OECD countries, it is not surprising that the topic of ethnic discrimination has never caused a lot of public debate or media coverage. As the OECD report pointed out, “Even today, the existence of systematic discrimination on the labour market is doubted by many actors of Swiss society” (OECD, 2012, p. 252). Discrimination is rather seen as acceptable to protect Swiss workers, as the narrow adoption of the mass immigration initiative on 9 February 2014 has shown. It proposed ending the free movement of EU citizens and reintroducing a quota system and to

⁹ An overview of the international feedback can be found on <https://www.humanrights.ch/en/switzerland/recommendations/discrimination/legislation-discrimination/> (last accessed 16.03.2018)

oblige employers to offer positions first to Swiss natives or those already residents in Switzerland. As Freitag et al. (2015) pointed out, this initiative was in particular targeted towards highly qualified Western Europeans immigrants.

6.4 Evidence of labour market discrimination from previous studies

Even if discrimination has not been a part of the media or political debate and many actors in Switzerland doubt its existence, research on ethnic discrimination in the Swiss labour market has increased in the recent years. Following the structure of the previous section on methodologies to measure discrimination, this section discusses the state of knowledge on labour market discrimination based on origins in Switzerland.

Well known examples of statistical analysis of the position of immigrants and their children in the labour market are the reports by the OECD, that were already cited above. For Switzerland these include the reports “Jobs for Immigrants” (OECD, 2012), or “The labour market integration of immigrants and their children in Switzerland” (Liebig et al., 2012), and labour market integration usually also features in their International Migration Outlook (e.g. OECD, 2013). Reports on the labour market published by the Swiss Statistical Office also provides information about the situation of immigrants and their children (e.g. Bundesamt für Statistik, 2017a), as shown above.

Next to these official reports, researchers have used statistical datasets to analyse specific situations in the labour market (e.g. transition from education to employment, unemployment duration, or occupational integration) for the differences between natives, immigrants and children of immigrants. Using the Transition from Education to Employment survey (TREE), Laganà, Chevillard, and Gauthier (2014) or Seibert, Hupka-Brunner, and Imdorf (2009) find that children of immigrants that came to Switzerland as mostly low-skilled workers often follow lower career trajectories and leave the educational system earlier than children of immigrants from higher skilled migrants or native children and that the dual path vocational training system, which is widespread in German speaking countries, did not create equal chances for youths with a migration background. In particular foreign men face disadvantages in accessing vocational training. Focusing on the unemployment duration of immigrants in the Swiss canton of Vaud, Auer, Bonoli, and Fossati (2017) use a dataset of all newly unemployed individuals in the canton of Vaud combined with administrative data. They find a strong impact of

nationality on unemployment duration: while EU nationals (excluding Portuguese) have quite similar unemployment durations to the Swiss, former Yugoslavians and non-EU immigrants experience longer spells of unemployment. They conclude that “[t]he fact that none of the factors [they] measured explains much of the disadvantage that these groups of immigrants experience, suggests that their longer unemployment spells are probably due, at least in part, to discrimination by employers” (p. 170). In a comparison of the occupational incorporation of early and recent Italian and Spanish immigrants in the Swiss labour market, Vidal - Coso and Ortega - Rivera (2017) rely on data from the 1980 Population Census as well as the Structural Survey for 2010 and 2011 to show that the Swiss labour market is still “segmented by national origin, and the foreign labour force continues to be polarized” (p.18).

Research on attitudes towards foreigners has been well-established in Switzerland (e.g. K. Ackermann & Ackermann, 2015; M. Ackermann & Freitag, 2015; Longchamp et al., 2014; Rapp, 2015; Raymann, 2003; Ruedin, D’Amato, Wichmann, & Pecoraro, 2013) and some studies focus specifically on the situation of foreigners in the labour market or include information on attitudes towards immigrants in the labour market (Helbling, 2011; Longchamp et al., 2014; Pecoraro & Ruedin, 2016, 2017; Raymann, 2003).

Two surveys that focused on attitudes towards foreigners more broadly were conducted by Raymann (2003) and Longchamp et al. (2014)¹⁰. Both showed that ethnic hierarchies exist among immigrants in Switzerland, with foreigners who have an Albanian or Former Yugoslavian immigration background encountering the most negative attitudes, in contrast Italians, Spaniards, French, Austrians, Germans, and Portuguese are most often regarded favourably (Longchamp et al., 2014, pp. 59, 89; Raymann, 2003, p. 21). Looking also at questions focusing particularly on the workplace, they found that even though xenophobia is still a marginal phenomenon in the workplace, it is one that is growing in its magnitude. While 75-85% of respondents can imagine to work with a colleague from a neighbouring EU country or Portugal, this number drops to 45% for potential colleagues with Russian backgrounds, and to less than 31% of acceptance for co-workers with Turkish, African, Arab, or Albanian origins (Longchamp et al., 2014, p. 89). In the study by Raymann (2003), more than 60% of respondents agreed (agreed strongly 33%, agreed mostly 27%) with the statement that in

¹⁰ Boulila (2018) points towards the problematic theoretical basis of the concepts that were used in this survey to study racism and the political decisions that were taken in the study design.

situations where equally qualified Swiss and non-Swiss candidates applied for a job, the Swiss person should be favoured” (2003, p. 6). Longchamp et al. (2014) also showed that the number of people that report feeling discriminated against in their search for employment or an apprenticeship has grown considerably from 2010 to 2014, with only 14% of those who had already experienced discrimination mentioning their search for employment in 2010 rising to 24% in 2014 (p. 125).

Focusing on attitudes towards German immigrants in the city of Zurich, Helbling (2011) found that also higher-educated individuals feel threatened by immigration. He observes that “[a]mong West Europeans [...] Germans have by far the most negative image” (p.7) and argues that despite their cultural closeness, “the cultural difference between Germans and Swiss-Germans is considered to be very large in Switzerland” (p.8). Furthermore, he emphasises that the debate focuses on “well-educated people from a neighboring country that speak (basically) the same language” (p.10). Helbling points out, that in the case of highly skilled German migration, it is in particular highly educated people that speak the same language who are opposed to their immigration, as they consider them competitors in the job market (p.20). Similar findings of anti-German attitudes have also been observed in Austria (Greth & Köllen, 2016).

While the previous examples had a broader focus, with some information on attitudes towards foreigners in the labour market, Pecoraro and Ruedin (2016, 2017) focus specifically on the questions how values, beliefs and unemployment risks are associated with the opposition to foreigners, and on how the attitude towards foreigners is related to the share of foreigners in one’s occupation. They use data from the Swiss Household panel, combined with the European Social Survey and the World Values Survey or the Swiss Labour Force survey respectively. Results of their 2016 study show that lower-educated people are not more likely to have anti-foreigner attitudes than medium skilled people, but that values and beliefs play an important role. Workers with a higher education are however more prone to show negative attitudes towards foreigners if the unemployment risk is higher and those foreigners are perceived as competition (2016, p.659). The second study finds that 1) people working in jobs with a high share of foreigners have more negative attitudes towards foreigners (2017, p.10), that 2) attitudes towards foreigners are more positive in occupations where the share of recent immigrants is high (2017, p.12), and that 3) “a higher rate of unemployment at the occupational level lowers the propensity to exhibit positive attitudes towards foreigners” (2017, p.13). They

conclude that workers' attitudes to foreigners are nuanced depending on the share of foreigners in the occupation, the share of recently arrived immigrants among the foreigners in the occupation, and the competition in the occupation.

These studies using surveys on attitudes towards foreigners show that attitudes vary for each immigrant group considered, with migrants from the Balkans or with a Muslim background being least accepted (Longchamp et al., 2014; Rapp, 2015; Raymann, 2003), that it is not only lower educated people that hold negative attitudes towards foreigners (Helbling, 2011; Pecoraro & Ruedin, 2016, 2017), and that individuals who feel culturally or economically threatened are more likely to express such negative attitudes (Helbling, 2011; Pecoraro & Ruedin, 2016, 2017; Rapp, 2015).

Recent years have also shown an increase in vignette experiments that focused on questions of ethnic discrimination in the Swiss labour market and hiring decisions in particular. Krings and Olivares (2007) conducted a vignette experiment with Swiss university students asking them to indicate whether they would invite a fictitious candidate with a Kosovo Albanian, Spanish or Swiss background for a job interview for the positions of electricians or bank assistants. Their research showed that intentions to invite a candidate for an interview changed according to the job and the ethnicity, especially “when foreign applicants who belong to a disliked ethnic group apply for jobs that require high interpersonal skills” (p. 406). In particular, for a bank assistant position, which was perceived as requiring such high interpersonal skills, students were less likely to invite a Kosovo-Albanian candidate compared to a Spanish or Swiss candidate (p.414). Also conducting a vignette experiment Helbling and Kriesi (2014) used an online survey to test with pairs of vignettes which applicants should be given a work permit – a high-skilled immigrant or a low-skilled immigrant. Looking at labour market competition, the welfare state and deservingness as explanatory models, they found that it depended on the group membership of respondents whether they preferred high- or low-skilled applicants.

Two vignette experiments that did not use convenient student or online participant samples, but were conducted with more representative participants, i.e. hotel managers and human resource professionals, were conducted by Auer et al. (2018) and Fossati, Liechti, Auer, and Bonoli (2017). The former study combined occupational hierarchies and ethnic hierarchies in the vignette experiment, by using fictional candidates with Portuguese, former Yugoslavian and Senegalese background for the two positions of cleaners and receptionists. Results show that

immigrants face a disadvantage in the higher positioned jobs, but that “individuals with a non-Swiss background have an easier route than Swiss nationals in terms of accessing the least desirable positions in the occupational hierarchy” (Auer et al., 2018, p. 20). They emphasise that a foreign nationality can become a “source of double disadvantage” (ibid, p.20), as it increases the danger of not being hired for higher status jobs, but of being trapped in lower status jobs. The latter study focuses on the effect of listing a foreign language spoken by the parents of second generation immigrants and being active in a cultural association linked to either the host country or the country of origin. Their results show that having a foreign name (Spanish, Polish, or Turkish) instead of a Swiss name already disadvantages second generation individuals and that it follows a pattern of perceived distance from the host society. Furthermore,

“individuals who speak a foreign language and/or are engaged in a cultural association associated with a different national background face even more disadvantages – even though from a purely economic perspective both of these activities increase a candidate’s employability” (Fossati et al., 2017, p. 13).

They conclude that the high level of discrimination that is observed for individuals with foreign names is worrisome and that listing cultural activities that are linked to the country of origin may be a disadvantage for non-native candidates.

Finally, the fact that Switzerland is a federal country which practices direct democracy provides a lot of information on the attitudes towards foreigners. Between 1848 and 2017, 43 referendums and popular initiatives on immigration related issues were held only on the federal level in Switzerland, most of them being restrictive on migrants’ rights (Arrighi, 2018). Overall, these federal referendums and popular initiatives, which became more frequent in the last years, showed the negative attitude towards foreigners and, even if votes were lost, empowered anti-immigrant parties to dominate the political agenda. The detailed results of such votes might be even better to measure attitudes towards foreigners than attitude surveys where only anticipated behaviour and opinions towards groups are measured.

Qualitative interviews with employers on their selection processes can also yield surprisingly frank results. Imdorf (2007, 2008) conducted interviews in small and medium enterprises in Switzerland on their selection practices for apprentices. In many cases employers openly voiced

their preferences for native Swiss candidates, claiming doubts about language skills of candidates with immigrant origins, concerns about new undesired groups of customers foreign apprentices might attract, or the desire to avoid potential conflict amongst the employees based on the fit of a new apprentice. Applicants with an immigrant background are often perceived as potentially disruptive, especially if their parents immigrated from non-EU countries, with countries of the former Yugoslavia and Turkey being considered the least favourably.

To date there are only few studies that focus on the experiences of victims of discrimination in Switzerland. Two studies have focused on specific minority groups' experiences, while one report addressed cases brought to the attention of outreach centres where counselling on discrimination cases is offered. The two studies addressing discrimination experiences focused on blacks (Efionayi-Mäder, Ruedin, Pétrémont, Michel, & Jain, 2017) and Muslims (Golder, Mousson, Tschöpe, Herzog, & Bohn, 2017). In both studies, the labour market is mentioned as one of the key domains in which respondents experienced recurring discrimination (Efionayi-Mäder et al., 2017, p. 26; Golder et al., 2017, p. 5). Respondents in the study by Efionayi-Mäder et al. (2017) specifically mentioned access to the labour market, the relationships to colleagues and superiors, but also the frequent controls carried out by the authorities in enterprises run by black people (p. 34). A community leader cited in their study emphasises the obstacle that in particular African blacks encounter when trying to enter the labour market irrespective of their qualifications. He stresses that black people in Switzerland usually find a job through their networks, not by submitting applications to advertised vacancies, or only on the merit of their qualifications (p.35). In the study by Golder et al. (2017) 46 percent of the respondents mentioned that they felt discriminated against on the ground of their religion at least once when they were asked specifically about their job search experiences. More than two thirds of those who reported perceived discrimination in the job hunting process were convinced that their religion was a reason for not being hired, while 55 percent mentioned that the name, too, was a decisive factor (p.6). The respondents identified the hiring stage as the moment when their religion played the most important role (p.32).

Information about discrimination cases in the Swiss legal system is very scarce. To my knowledge, the only study addressing the access to the judicial system in discrimination cases has been conducted under the lead of legal scholar Walter Kälin, which addresses discrimination on several grounds (gender, sexual orientation, disability, race/ethnicity) (Kälin & Locher, 2015). Two sub-projects of this study specifically addressed the issue of racial

discrimination from a legal (Matthey & Steffanini, 2015) and a sociological perspective (Probst, 2015). These studies identified only four legal cases on racial discrimination under civil law procedures at Swiss courts, compared to approximately 100 cases that were brought forward under criminal law provisions. The sociological study in particular showed that people working in centres offering advice in instances of discrimination are very sceptical about chances of winning a discrimination case and that a lot of insecurity regarding the chances of success in legal proceedings makes victims of discrimination very reluctant to start a legal case.

Since there is almost no information about discrimination cases in the Swiss legal system, similar information can only be obtained from human rights organizations that offer advice on discrimination cases and publish statistics about the cases that were brought to their attention. The association *Verein humanrights CH* has published reports based on cases that were reported in their outreach centres (Mühlemann, 2017). They only count cases where contact between the victim and one of the centres was established, where a concrete situation has been described and categorised as a case of racial discrimination, and where counselling services had been offered (p.6). Thus, they report 199 cases of racial discrimination in 2016, slightly less, than in 2014 or 2015, yet still higher than in the years before (p.6). Here, too, the labour market and public spaces are most frequently mentioned as areas in which discrimination had been encountered (33 cases each) (p.7). In twelve cases the hiring process was mentioned specifically (p.11). Most cases were brought forward by people holding the Swiss nationality or a B or C permit, indicating that the assumed origin is more important than nationality or residence status (p.17).

Field experiments studying ethnic discrimination on the Swiss labour market are rare, to my knowledge there are only two examples. Although Switzerland was initially on the list of countries proposed by Bovenkerk for the ILO project on labour market discrimination (1992, p. 41), it did not participate in the project. It was only with the study of Fibbi et al. (2003) that a correspondence test using the ILO methodology was conducted in Switzerland. They studied discrimination of second generation youths transitioning from apprenticeships to their first position, who were born in their home country but had completed all their education in Switzerland. The minority candidates were Portuguese, former Yugoslavs (Albanian speaking and born in Kosovo), and Turks. Applications consisting of a cover letter and a CV were submitted to vacancies in both the German and the French speaking parts of Switzerland. The results showed significant discrimination against the minority candidates from Turkey and

former Yugoslavia, while results were not significant for the Portuguese (Fibbi, Lerch, & Wanner, 2006, p. 357). Similar to the ILO methodology they also provided information on cases of “equal but different treatment” in which the minority candidate is only invited once the majority candidate is no longer available. This equal but different treatment rate was highest in the French speaking region, where the minimal discrimination rate had been lower (p. 357).

More recently Diekmann et al. (2014) conducted five field experiments on discrimination in Switzerland with their students as a part of a university seminar. One of these experiments focused on the labour market. This experiment adapted the correspondence testing methodology to send unsolicited applications to Swiss German companies in the industrial or financial sectors with more than 50 employees. In total 300 applications were sent, 150 with a Swiss German name and 150 with a name indicating a former Yugoslavian background. The number of positive replies was quite low: only 11 of these 300 applications were not rejected immediately or did not receive a reply. In 9 of these 11 cases the candidate with the Swiss German name was treated more favourably.

In the Swiss context, research from social psychologists looking at attitudes against neighbouring countries is also of interest. This work does not focus specifically on the labour market, but also on attitudes towards linguistically similar but much bigger nations, i.e. Germany and France. Research by van Oudenhoven, Selenko, and Otten (2010) and Matser et al. (2010) has shown that the larger countries are perceived as a threat by the smaller countries, especially when they share the same language, and that inhabitants of the larger countries are perceived as more arrogant. While respondents acknowledged the similarities between both countries, they “keep defending their social identity by expressing a dislike for this perceived similarity (p.143). These findings are in line with results by Helbling (2011) on “Why Swiss-Germans dislike Germans” in Zurich.

Binggeli et al. (2014a); Binggeli, Krings, and Sczesny (2014b) address the perception of different immigrant groups according to the dimensions of warmth and competence following the stereotype content model. Their analysis shows that “immigrant groups perceived as lacking both competence and warmth (e.g., immigrants from the Balkans in Switzerland) should be more likely targets of blatant discrimination than immigrant groups with more mixed stereotypes (e.g., high warmth/low competence, such as Italian immigrants in Switzerland). Immigrant groups perceived as highly competent but lacking warmth may be targeted by more

subtle, interpersonal discrimination (2014b, p. 131). They find that groups that are more similar to Swiss natives (e.g. German and French immigrants) and should thus be easier to integrate, face more dislike because they are perceived as competition.

Using a similar approach, Krings et al. (2014) look at the perceptions of warmth and competence ascribed to each immigrant group in Switzerland and classify the groups accordingly:

Group 1: low in competence, low in warmth	Balkans, Turkey, Eastern Europe
Group 2: moderately warm but incompetent	Africa
Group 3: highly warm, but moderately competent	Southern Europe
Group 4: highly competent, lacking warmth	Germans and French

Table 1: Warmth and competence of immigrant groups in Switzerland (Krings et al., 2014, p. 493)

They found that immigrants from neighbouring countries that were perceived as highly competitive were more likely to report perceived discrimination at work, as they faced a higher likelihood of workplace incivility (p.491, 495). This is worrisome, as these groups are expected to integrate easily and are usually not taken into consideration in the discourse of immigrant discrimination in the workplace (p.497).

6.5 Conclusion

The issue of labour market discrimination in Switzerland has garnered more attention by researchers over the last years. The growing body of literature shows that although immigrants are well integrated in the labour market in international comparison, they still face disadvantages vis-à-vis the Swiss natives, such as (on average) higher unemployment rates and lower wages. Work on attitudes towards foreigners has consistently shown high levels of negative attitudes towards immigrants from the Balkans and with Muslim backgrounds. These negative attitudes against foreigners have also come to light in the high number of referenda and popular initiatives on immigration related topics. However, they are not only targeted towards low skilled migrants, but, as the mass immigration initiative which was adopted in February 2014 has shown, also against immigration from EU countries under the free movement provisions. Swiss citizens voted in favour of giving preference to citizens or

residents of Switzerland in application decisions, which mirrors findings from attitude research where the majority of respondents agreed that Swiss candidates should be preferred in hiring decisions. It is therefore not surprising, that the first field experiment conducted in Switzerland (Fibbi et al., 2003) found that discrimination existed in the Swiss labour market and that ethnic hierarchies that correspond to hierarchies in attitude surveys play a role. However, this field experiment was conducted fifteen years ago and only focused at a very specific segment of the labour market. Given the political anti-immigrant sentiment that is often portrayed in Swiss media and politics, it is therefore interesting to conduct a new correspondence test that focuses on a broader part of the labour market to see if discrimination can also be documented in hiring decision and how it changed (or remained consistent) in relation to the previous correspondence test.

7. A correspondence test in the Swiss labour market – the research design

As discussed previously, discrimination is hard to disentangle from other possible factors leading to labour market disadvantages and many research methodologies have addressed the issue, but – so far – correspondence testing is the only method that allows the researchers to a) quantify the extent of discrimination taking place in the labour market and to b) observe real life hiring decisions. While the first correspondence test in Switzerland in 2003 (Fibbi et al., 2003) has shown discrimination against minority candidates transitioning from apprenticeships to their first position, this new field experiment addresses a broader spectrum of the labour market. The research design and the choices made are discussed in detail in this section.

7.1 The Ethical Dimension

Given the choice of a correspondence test as the main methodology used in this research, the ethical dimension plays an important role. Ethical concerns focus in particular on the lack of informed consent, the deception of employers as well as the legal responsibility of researchers. While the research ethics of correspondence testing as a methodology are the topic of Paper 3 of this thesis “Research Ethics in Correspondence Testing: an Update”, the problems and decisions taken for this correspondence test on the Swiss labour are the focus here.

Researchers such as Bovenkerk (1992), Banton (1997), or Riach and Rich (2004a) have addressed the ethical concerns inherent in field experiment and have concluded that the use of field experiments is justified as they are the most appropriate method to measure discrimination in the labour market. Bursell (2007) and Pager (2007) refer to the respective legislation in Sweden and the US, where it is recognized that certain research goals can only be obtained without informing the participants of their involvement in an experiment. Thus, “a human subjects institutional review board (IRB) ‘may ... waive ... informed consent provided (1) the research involves no more than minimal risk to human subjects; (2) the waiver or alteration will not adversely affect the rights and welfare of the subjects; (3) the research could not practicably be carried out without the waiver or alteration; and (4) whenever appropriate, the subjects will be provided with additional information after participation.’ Each of these conditions can

arguably be satisfied in the context of audit studies of discrimination [including correspondence tests].”(Pager, 2007, p. 126).

Taking these ethical concerns seriously a thorough dossier on the ethical concerns regarding a correspondence test on the Swiss labour market was submitted to the Ethics Commission of the University of Neuchâtel. In this dossier, the arguments pro and contra correspondence testing were carefully evaluated and justifications for the use of deception and for forgoing informed consent were given.

Specifically, the following precautions were taken:

- Tests were conducted only in writing and thus at a very early stage of the hiring process, to limit the inconvenience for employers.
- The number of applications sent per employer was restricted to two to limit the burden on employers.
- The information on enterprises and decision makers was rendered anonymous and original vacancies are not accessible to anyone not involved in the core research team. Research assistants signed confidentiality agreements.
- The data was aggregated for the analysis, so no links to individual employers are possible.
- We cooperate closely with the NCCR Data Manager to ensure that the data is handled the best way possible (including archiving the data).

The research proposal was evaluated by an internal ethics commission of the NCCR on the move, the framework in which this research takes place, and an expert on research ethics from the UK. Taking into account these two positive reviews the Ethical Commission of the University of Neuchâtel approved the research project. It was acknowledged that the proposed research would not be possible without the use of deception, that the research was in the public interest, and that steps were taken to minimize potential risks to participants. The Ethical Commission emphasised two points specifically: 1) the anonymization, so that participants could not be identified and that they would not easily be able to discern upon publication that they had been participants in the research, and 2) the issue of consent, inviting the research team to consider the possibility of obtaining the informed consent of participants after the experiment.

After careful consideration of the second point, we decided to neither to inform participants of their study participation or to collect their informed consent post-hoc. We do acknowledge that this would be possible and know that Midtbøen (2014) contacted participants of his Norwegian correspondence test. However, the goal of this contact was not to inform all participants of their role in the field experiment but to recruit participants that had responded to at least one of the applicants to conduct further qualitative interviews on recruitment and hiring practices. We decided to follow Pager's argument "that for human resource personnel or managers who are thought to be discriminating, the consequences may be more serious than if no attention were brought to the audit whatsoever" (2007, p. 127). The research team of the Expert Council of German Foundations used arguments similar to Pager's to explain why they decided against informing employers that were part of their correspondence test post-hoc (Schneider et al., 2014). First, it would not improve the chances of minority applicants in the future, second, it might be problematic for future researchers if the methodology of correspondence testing became too well-known, and third, it could have negative repercussions on the employees that were responsible for the specific hiring decision. Thus, we decided not to obtain informed consent post-hoc, to minimize the potential risks for individual participants. We are aware of only one recent study where all employers were sent a debriefing letter informing them about the research project (Liebkind et al., 2016).

7.2 Identifying groups for testing

Correspondence tests vary with the number of immigrant groups that are included in the testing, ranging from only one racial or ethnic minority (e.g. Turks in Germany: Goldberg et al. (1995); Kaas and Manger (2012); Schneider et al. (2014); Weichselbaumer (2016a)), to five or six minority groups (e.g. in Canada: Oreopoulos (2011), or in the UK: Firth (1981); Wood, Hales, Purdon, Sejersen, and Hayllar (2009)).

Switzerland with its high share of immigrants in the resident population is a great case to study discrimination in the labour market. In 2015 2.05 of the 8.33 million residents held a foreign nationality (Bundesamt für Statistik, 2017c). Furthermore, anti-immigration referenda, previous research on attitudes against foreigners, and studies on discrimination against foreigners discussed above have shown a high degree of distrust against foreigners in Switzerland. Yet, this previous research has also shown that the diversity of the Swiss immigrant population also leads to different attitudes and stereotypes towards individual

groups. Immigrants from the former Yugoslavia and the Balkan are consistently perceived as the least welcome group in Switzerland (e.g. Krings & Olivares, 2007; Longchamp et al., 2014; Rapp, 2015; Raymann, 2003). Longchamp et al. (2014) show that xenophobia at the workplace is still a marginal phenomenon, but one that is increasing in its magnitude (p.89). They argue that in particular non-European nationals are being rejected by their co-workers, with an obvious connection to islamophobia, but also some European nationals are met with great reluctance. According to their findings, EU nationals (in particular Germans, Italians and French) are mostly accepted, while only a minority of the survey participants can imagine to work with Albanians, Arabs, Africans, Russians or Turks (p.89). When asked, more generally, which population groups are perceived as objectionable, those who mentioned specific groups most frequently listed foreigners in general, Albanians, Africans/blacks, and former Yugoslavs/people from the Balkan (p.59). However, negative attitudes towards foreigners are not only found amongst lower educated people. Helbling (2011) documented negative attitudes towards highly skilled German immigrants in the city of Zurich, in particular amongst highly educated people. Overall, results from attitude research give an indication that ethnic hierarchies exist among immigrant groups in Switzerland.

While other correspondence tests have usually only focused on the biggest minority groups, in the European context often Turkish or people of Muslim decent, there is only one study that also looks at a Western European minority group, Germans in Ireland (McGinnity & Lunn, 2011). Given the big presence of EU nationals from neighbouring countries in Switzerland, we also test for discrimination against Western Europeans. Building on statistical information on the biggest immigrant groups in Switzerland and the results from previous research on attitudes towards foreigners and on discrimination in Switzerland, we chose the following groups for our correspondence test: Germans in the German speaking part of the Swiss labour market, Kosovars, and Turks¹¹. This way our groups are also similar to those previously chosen by Fibbi et al. (2003), with people with a migration background from former Yugoslavia and Turkey present in both study, and another Western European immigrant group, whereas the Portuguese are now replaced by Germans.

¹¹ In parallel to the correspondence test in the Swiss German labour market that is the subject of this dissertation, another testing is still being conducted on the French speaking labour market.

7.3 Signalling ethnicity: construction of names and portrayal of citizenship

We are aware of the problems that signalling ethnicity based on names can entail (Crabtree & Chykina, 2018; Fryer & Levitt, 2004; Gaddis, 2017a, 2017b), which were discussed in detail in the section on correspondence test¹². For this Swiss correspondence test names were constructed using a variety of sources. First of all, official government statistics (Switzerland) and quasi-official statistics (Germany) were used as well as information on common names found on miscellaneous websites as well as on Wikipedia. Swiss first names were chosen based on information provided by the Swiss Federal Office for Statistics, which until recently, offered the possibility to access information on the first names most frequently given to new-born children by gender, year, and linguistic region¹³. Unfortunately, similar statistics are not available for last names. However, information on the most frequent Swiss German last names (Leybold-Johnson, 2014) and on the most common last names by Swiss cantons (Skinner, 2015) has been provided by news media. As many of the most frequent Swiss German last names are also prominent in Germany, the map of the cantonal distribution provides more distinct Swiss last names, which are not common in Germany. German names were constructed similarly. In Germany, it is not the statistical office, but the *Gesellschaft für Deutsche Sprache* (GfdS, Society for German language), that publishes the list of the most frequent German first names each year, as well as information about the fifty most common family names in Germany (Gesellschaft für deutsche Sprache e.V., 2017). While there is a certain overlap with first names in Switzerland, there were still some names, which were only used frequently in Germany. Finding official statistics on Kosovar names proved rather difficult. Here, Wikipedia provided a starting point on common surnames in Kosovo (Wikipedia, 2017). We also relied on data from a previous study conducted by the Swiss Forum for Migration and Population Studies (SFM), where register data for the cities of Zurich and Basel had been analysed for the group of 18-35 year olds that were clustered by their migration background. Based on these list, we chose first and last names for the fictitious Kosovar candidates. Turkish names were again constructed using Wikipedia for common Turkish names. Furthermore, in 2015, the *Gesellschaft für deutsche Sprache* published, for the first time, a list of Turkish and Arab first names, comparing the position of first names in Turkey with the frequency in Germany

¹² While the section on names in the section on correspondence test focused in particular on the problems that socio-economic connotations of names can pose, potential religious connotations were not discussed.

¹³ <http://www.bfs.admin.ch/bfs/portal/de/index/themen/01/02/blank/dos/prenoms/02.html>; accessed in Summer 2016, not available anymore

(Gesellschaft für deutsche Sprache e.V., 2017). In addition, the SFM study used for Kosovar names also provided information on the names of people with a Turkish migration background.

Finally, the origins and meanings of the chosen names were checked using popular baby naming websites to avoid a strong religious connotation of the names that names such as Maria or Muhammed might signal. While employers might assume that Turkish or Kosovar applicants are Muslim, candidates do not give any indication about a religious affiliation in their CVs.

After identifying frequent names for each ethnic group, male and female names were randomly matched with frequent last names. Using Facebook, it was then checked that these combinations are credible and that numerous profiles with the same names exists to make it impossible for potential employers trying to look up the fictitious candidates on social media to link them to a specific profile.

Based on these sources, the names chosen for the correspondence test are:

Origin	Female	Male
Swiss German	Fabienne Kälin	Pascal Kälin
German	Kathrin Hoffmann	Alexander Hoffmann
Kosovo	Shpresa Krasniqi	Bekim Krasniqi
Turkey	Filiz Yilmaz	Cem Yilmaz

Table 2: Names used in the correspondence test

After identifying frequent names for each ethnic group, male and female names were randomly matched with frequent last names.

While the careful selection of names is important to signal the race or ethnicity of the candidate, it is not the only cue provided in this correspondence test. Since it is common in Switzerland to list the citizenship in the CV, all the fictitious applicants list being Swiss citizens, with the ethnic minority candidates listing also the citizenship of their (assumed) country of origin. Given the very strict requirements to acquire Swiss citizenship by naturalization, and the CV indicating that all schooling and work experience has been obtained in Switzerland, they signal that they have spent a long time in the country. Furthermore, all of them list German as their first native language, in the case of the candidates with Turkish and Kosovar names, Turkish and Albanian are also listed as native languages. Being dual citizens, educated in Switzerland, and native speakers of German fits the profile of many second generation youths in Switzerland.

Our meta-analysis has shown that many studies do not explicitly claim that candidates are members of the second generation, but that this is implied by CVs that list that schooling was completed in the country where the testing was conducted (Zschirnt & Ruedin, 2016).

This research design of using multiple cues about the applicants' citizenship and ethnicity was chosen to make sure that employers will not use work or residence permits as reason for rejecting a foreign-named candidate. It also leaves no doubt about the ethnic background of the applicants. The only drawback of listing dual citizenship for the fictitious minority candidates is that they are portrayed as perfectly integrated members of the second generation. Yet, recalling the statistics discussed in section 6.1 above, 35% of the second generation have not been naturalized and are still legally considered foreigners (Bundesamt für Statistik, 2017b). We are thus measuring discrimination against very well integrated and educated candidates, and it can be expected that discrimination against less "ideal" candidates will be higher.

7.4 Identifying locations for testing

The majority of previously conducted correspondence tests have had a regional focus or targeted specific cities within the country, yet there are some examples where vacancies all over the country were included in the research design. These are usually smaller countries (Sweden: Agerström et al. (2012); Netherlands: Andriessen et al. (2012), Blommaert et al. (2014); Germany: Kaas and Manger (2012); Schneider et al. (2014); Austria: Weichselbaumer (2016b)) with the exception of one study conducted nationwide in the US (Widner & Chicoine, 2011). Although the focus within these studies is national, there are still some geographic areas where the majority of the testing has taken place, usually around the capitals. This Swiss correspondence test also uses a national scope. The results of the correspondence test in the German-speaking part of Switzerland that are discussed in this dissertation will be complemented by a still ongoing correspondence test that follows the same research design in the French-speaking area. Only the Italian speaking part of the country is excluded. German and French are by far the dominant language groups in Switzerland, so the number of vacancies published in Italian is very low. Data of the Stellenmarktmonitor Schweiz¹⁴ shows that between

¹⁴ ¹⁴ The "Stellenmarkt Monitor Schweiz (SMM)" is a project by Buchmann et al. at the University of Zürich which has been documenting the development of the publication of vacancies in Switzerland going back to 1950. The SMM draws a representative sample of vacancies from a number of publication channels during one week each year, since 2001 it is available for all of Switzerland and since 2006 the database also includes vacancies published in online job portals (Buchmann et al., 2015).

2006-2014 78.8% of the posted vacancies were located in the German language area, followed by 14.4% in the French language area. Given the comparably small size of Switzerland and the high number of people who commute to work, we apply for positions all over the country, indicating that candidates are willing to relocate (as was done by Blommaert et al. (2014) in the Netherlands). This nationwide approach allows us to analyse if there are differences between linguistic regions or rural and urban areas. Looking at the greater regions, 25.5% of the positions analysed by the *Stellenmarktmonitor Schweiz* were located in Zurich, 21.6% in the “Espace midland”, followed by North-western Switzerland with 13.1%, the Lake Geneva Region with 12.5% and Eastern Switzerland with 11.9% (Buchmann et al., 2015, own calculations).

7.5 Selecting occupations

The selection of occupations was based on several criteria: the application process takes place in writing and online (thus excluding the rare positions where written applications were required to be submitted by post), the applications can be standardised to fit a great number of jobs, and vacancies had to be published on a steady rate, which given the small size of the Swiss labour market was of particular importance. Finally, the occupations chosen should not require too many qualifications or too much work experience. The focus has therefore been laid on lower and medium skilled occupations, where the construction of application materials is still feasible.

Suitable occupations were selected in several steps. First, the above mentioned data set of the *Stellenmarktmonitor Schweiz* (Buchmann et al., 2015) was analysed to identify the 25 most frequent occupation in this data set. The *Stellenmarktmonitor Schweiz* draws a representative sample of vacancies from a number of publication channels during one week each year, since 2001 it is available for all of Switzerland, and since 2006 the database also includes vacancies published in online job portals. Thus only data published after 2006 was used to identify the most frequent occupations. These occupations were then compared to search results on six different Swiss job search websites, where the top ten of these 25 positions were identified for each website. Finally, we compared which occupations were in the top 10 of (almost) all of these job boards. Using this shortlist of occupations, we considered each occupation to decide if it would be possible to construct the necessary credible and convincing fictitious application documents. Several positions, such as architects could be immediately rejected, since

applications usually require submitting a portfolio of previous work. We finally narrowed down the list of occupations to two positions where only an apprenticeship diploma was needed, i.e. sales assistants (*Detailhandelskauffrau/-mann*), electricians (*Elektroinstallateur*), and two occupations that required intermediate qualifications, i.e. nurses (*Pflegefachfrau*), and HR clerks (*HR Fachmann/-frau*). In the case of sales assistant and HR clerks both male and female candidate pairs were prepared¹⁵, while applicants for the position of nurses (electricians) were only female (male).

	Sales	Electricians	Nurses	HR clerks
Unemployment rate total	4.91%	3.01%	1.65%	3.66%
Unemployment rate men	7.27%	2.99%	2.64%	0.52%
Unemployment rate females	4.11%	3.85%	1.52%	4.32%
Unemployment rate Suisse	4.29%	2.74%	1.48%	3.25%
Unemployment rate foreigners	7.96%	4.26%	2.25%	5.76%
Share of foreigners in occupation	15.80%	16.09%	20.67%	15.08%
Share of women in occupation	75.09%	2.69%	89.3%	81.98%

Table 3: Characteristics of occupations chosen, based on the Swiss Structural Survey 2014 and weighted. Economically active persons aged 15-64 (15-65 for women). Due to the low number of electricians in the sample, the results are for all the electronic and electro technical positions (Codes between 2501 and 25108).

7.6 Sources for vacancies

The corresponding testing methodology relies on vacancies that are publicly advertised, either in news-papers, or in more recent correspondence tests online. That means that other channels, such as finding a new job via networks, cannot be considered in a correspondence test.

Again, the data from the *Stellenmarktmonitor Schweiz* provides valuable information on the recruitment channels used in Switzerland. A survey with Swiss employers shows that 75% of the companies try to recruit through their networks, i.e. using the networks of employees, business partners or customers, with employee's networks being the most important channel (Buchs & von Ow, 2017, p. 3). For publicly advertised positions the internet has become the most important channel – with publications on company websites and job portals being used by more than 50% of employers, usually combined with other advertising channels (p.4). 35% of companies also mentioned unsolicited applications as a channel to find new personnel (p.5).

¹⁵ At the time of choosing the occupations we did not realize that HR clerks and sales positions were also strongly female dominated occupations.

Furthermore, the database provides information on the job portals used (from 2006 onwards), with www.jobs.ch being the most frequently used job portal in the cases where this information was available (Buchmann et al., 2015, own calculations).

Vacancies were collected online using two websites in particular for the German speaking part of Switzerland: www.jobs.ch as well as www.job-room.ch, the latter being hosted by the Swiss public employment services.

7.7 Constructing the applications

The application material is the core element of the correspondence test. Candidate profiles have to be realistic, representative and equivalent for the fictitious candidates, but must not be associated to actual job seekers. CVs and cover letters that are used in a correspondence test need to be comparable while at the same time avoiding being too similar to minimize the risk of detection by the employers. Finally, the application material should be as standardised as possible, to allow the researchers to send it to a large number of job offers.

Most correspondence test and particular those conducted in English speaking countries only sent resumes and cover letters for their fictitious applicants. The preparation of application material is more complicated in the German speaking context, where substantive application packages are the norm. In German speaking countries application dossiers usually contain (at least) a photograph, diplomas, and work certificates next to the standard cover letter and CV. In Switzerland sending such comprehensive application material is also the norm and conversations with HR specialists have shown that incomplete applications are almost certainly thrown out without being given further attention. This amount of detail provides its own challenges in the creation of application material for a correspondence test.

Cover letters and resumes and references

As Bertrand and Mullainathan (2004) pointed out the first step in a correspondence test is the creation of resumes to be sent, the challenge being “to produce a set of realistic and representative resumes without using resumes that belong to actual job seekers” (p. 994). The resumes have to provide all relevant information on one or two pages and portray equivalent human capital for the fictitious applicants. To be considered as plausible and qualified,

applicants have to meet the job requirements that a potential employer is looking for, which might require tailoring the application to specific requirements (e.g. driver's license or specific computer skills).

For the Swiss correspondence test we use rather young applicants, born between 1990 and 1996. All candidates have finished an apprenticeship and have gained some work experience in the same company before now starting to apply for a new job. The exception to this rule is the HR clerk profile, since the training for a "*HR-Fachperson*" involves more steps. Similar to the other candidates the HR clerk also still works in the same company where he/she obtained his/her last degree and is now looking for a career change. This set-up was chosen to avoid creating too many references from employers. Resumes were constructed using publicly available profiles on LinkedIn for the respective occupations, resumes of actual job seekers which were uploaded to job platforms, and information from career advice websites that explained job profiles and the apprenticeship requirements. With this information career trajectories were created that served as the basis for our resumes (e.g. the years of schooling required for an apprenticeship, the length of the apprenticeship itself, etc.). Furthermore, vacancies for the chosen occupations were examined closely to see which skills employers demanded in each occupation. To make sure that the CVs and cover letters created based on this information were credible we met with HR specialists who helped us fine-tune our application material. The final documents use two different layouts to avoid detection. We also made sure to include slight variations in the CVs (e.g. a year of work experience more for one candidate) to be able to conduct the Neumark test afterwards (Neumark, 2012). As mentioned in section 5.4 the Neumark test allows to examine the robustness of the results of a field experiment by using the variation in resume quality.

References from previous employers were constructed in a similar fashion. Reference letters available online as well as text books on how to write reference letters were consulted to create first drafts of reference letters for our fictitious candidates, which were then fine-tuned with the help of our HR contacts.

Photographs

The standard Swiss application procedure, in particular in the German-speaking part of the country, also includes a photograph in the application material, usually on the CV. So far correspondence tests on ethnic discrimination that made use of photographs have been conducted in Austria (Weichselbaumer, 2016b), Germany (Kaas & Manger, 2012; Schneider et al., 2014; Weichselbaumer, 2016a) and Mexico (Arceo-Gómez & Campos-Vázquez, 2013). While Goldberg et al. (1995) decided against the use of photographs, both Kaas and Manger (2012) as well as Schneider et al. (2014) used two photographs that could each be used for either a German or a Turkish applicant and randomly assigned the picture and resume type to the names. In order to enable this random use of pictures no religious symbols can be seen on the picture (e.g. necklaces with a cross, headscarves...). A detailed discussion on the use of photographs in application procedures and the problems they can introduce in terms of unobservable characteristics or their impact on correspondence testing results can be found in Rich (2018).

Since Weichselbaumer (2016b) tested several ethnic minority groups in Austria, including African (Nigerian), Asian (Chinese), Serbian, Turkish and Austrian applicants, this approach of randomly assigning one of two pictures was not possible here. As Weichselbaumer explains “the fictitious applicants on the photos needed to be comparable in terms of e.g. attractiveness, charisma or age” and four photographs per gender were needed for her experiment in Austria (Weichselbaumer, 2016b, p. 9). This involved a time-consuming process including numerous models, pre-testing with students and digital manipulation to ensure that the photos were rated similarly in their perception of “looks, likability, intelligence, reliability as well as in their overall score” (Weichselbaumer, 2016b, p. 10). To obtain the photos students from different ethnic backgrounds were recruited as models, and after a first selection round invited to a meeting and a subsequent photo session, where they received specific instructions on their outfit and styling. These photos were then pretested asking confederates to identify photos they considered equal in terms of attractiveness and likeability. A second larger pre-test was then conducted with university students. Each student evaluated one applicant “with respect to looks, intelligence, reliability and their likeability on a scale from 1 (High) to 5 (low)” (Weichselbaumer, 2016b, p. 10). Furthermore, digitally manipulations were made, until the photos scored comparably across all dimensions tested as well as their overall score.

For this Swiss correspondence test, we were able to avoid the big challenge of creating credible and equivalent photographs for the fictitious applicants that are included in the CVs by obtaining the permission of Doris Weichselbaumer to reuse the pictures that she had prepared and used in her Austrian correspondence test (Weichselbaumer, 2016b).

Diplomas

Next to cover letters, resumes, reference letters, and photographs, it is also customary to include copies or scans of diplomas obtained by the applicants in a Swiss application package. While it was possible to collect several examples of diplomas online, using google image searches and freely accessible uploaded diplomas on platforms such as LinkedIn, the lacking diplomas were obtained with the help of our personal networks. These documents were then digitally manipulated to match our candidate profiles.

It could be argued that creating these diplomas constitutes forging of official documents. For a correspondence test in Germany, the Sachverständigenrat Deutscher Stiftungen für Integration und Migration (SVR) ordered a detailed legal analysis of the methodology. Kühn, Liebscher, and Klose (2013) specifically addressed issues of criminal law, concerning the forging of documents, as well as questions on civil law regarding the liability of researchers. In the end, they came to the conclusion that the freedom of the researcher is protected by the law and that a very carefully designed correspondence test would not infringe on personal rights of the enterprises. The issue is also discussed in the paper in research ethics.

Contact details

The final elements needed for a credible and complete application are the applicant's contact details, i.e. their address, email address and a phone number. Even if applications are only send online, fictitious applicants need to provide their postal address. Researchers argue that it is reasonable to expect potential employers to respond to applications submitted by email also electronically and not by posted letters (e.g. Bursell, 2007), but street addresses are still an important part of the contact details provided in order not to raise suspicion about the application and are usually a required field in online applications. While Eid (2012) used addresses of his research team and colleagues for his Canadian experiment, Wood et al. (2009) argued against such an approach in their UK study out of ethical considerations. Like most

other studies, they constructed fictitious addresses, by making them appear as real as possible, e.g. by using real street names, but house or flat numbers that were higher than the highest existing number and using the appropriate postcode. The areas chosen for the study were ethnically diverse and the postcodes were chosen based on the diversity shown in census data (Wood et al., 2009, p. 23). Another approach used by Bursell (2007) was to use real addresses and residential blocks, but making sure that nobody with a similar name lived there. In both cases responses per post were lost to the experiment, but since they were believed to be in very low numbers, this risk was taken into account.

In the Swiss context, setting up actual postal addresses would provide a challenge. As Fibbi et al. (2003, p. 54) have pointed out, numerous social controls can be carried out in Switzerland when a new name appears on the mailbox, ranging from the building manager, the postman, the rental agency, the owner of the building or the residents registration offices. In this experiment we used real addresses in several Swiss cities, making sure to choose apartment buildings with numerous tenants. While this means that responses sent by letter will be lost to the experiment, it can be assumed that this number will be minimal. In fact, during the experiment we received email responses from only three employers (of 560) who indicated that postal replies had been returned to the sender, and were subsequently scanned and emailed to the respective candidates. In these cases, applicants were always rejected.

The second part of the contact details is the phone number. While we would have liked to use virtual phone numbers, which enable the forwarding of mailbox messages to email addresses, this option did not seem feasible in Switzerland, since such virtual phone numbers are linked to landline area codes. HR specialists informed us that the majority of candidates only lists mobile phone numbers on their CVs, which would make two applications with landline area codes stand out. Thus, we decided to use two prepaid mobile phone numbers, one for the majority candidate and one for the minority candidate. Calls were not answered, but led to a mailbox with the provider's default message. Using these default messages allowed us to use the same phone numbers for male and female job candidates and to avoid potential issues of recording voice mail messages in Swiss German dialects, which can be very specific to certain regions or cities. Responses were carefully recorded and linked to the specific vacancies using a reverse phone number search¹⁶. In total, there were 14 phone numbers that could not be linked

¹⁶ <https://tel.local.ch/en/phone-number-search> allows to reverse search for landline phone numbers in Switzerland.

to a vacancy. In three cases these numbers showed up on both the majority and the minority phone, seven numbers only called the majority candidate and four numbers only appeared on the minority phone.

Finally, email addresses were created for all fictitious applicants. Looking at the studies included in our meta-analysis (Zschirnt & Ruedin, 2016) the most frequently used email providers were gmail.com, Hotmail.com, and yahoo.com. The only study not relying on these popular providers is to my knowledge Neumark's field experiment on age discrimination in California (Neumark, Burn, & Button, 2015), where he and his team were able to set up their own email provider. While Hotmail and Yahoo services have very detailed terms of services, which require the user to "provide true, accurate, current and complete information" (Yahoo, 2012), Gmail's terms of services are quite vague. We therefore used Gmail to create the email addresses needed for our correspondence test. Since we chose common combinations of first and last names, email addresses using only the first and the last name were usually not available on Gmail, a problem that real users encounter as well. The addresses always contained the last name of the applicant, in most cases the first name or initials or abbreviation of the first name, numbers, or a "Herr" (Mr) or "Frau" (Ms) before the name. Thus, the email addresses look like serious personal email addresses, avoid nicknames, are not too uniform to make employers suspicious and are realistic as only first and last name user names are often not available anymore.

7.8 Overview of the sample

Between October 2017 and April 2018, 1120 applications were sent in pairs to 560 vacancies all over the German speaking part of Switzerland. The numbers presented in this overview of the sample always refer to one applicant pair per vacancy.

	Sales Assistant	Electrician	HR clerk	Nurse	Total
Germany	45	45	48	48	186
Kosovo	46	46	48	48	188
Turkey	45	45	48	48	186
Total	136	136	144	144	560

Table 4: Number of paired applications by nationality tested in the pair and occupation

		Sales Assistant	Electrician	HR clerk	Nurse	Total
Female	Germany	24	0	24	48	96
	Kosovo	26	0	24	48	98
	Turkey	24	0	24	48	96
	Total	74	0	72	144	290
Male	Germany	21	45	24	0	90
	Kosovo	20	46	24	0	90
	Turkey	21	45	24	0	90
	Total	62	136	72	0	270
Total		136	136	144	144	560

Table 5: Number of paired applications by gender, nationality tested in the pair and occupation

	No information	Rural	Urban
Sales Assistant	0	53	83
Electrician	4	81	51
HR Clerk	0	60	84
Nurse	0	68	76
Total	4	262	294

Table 6: Number of paired applications in rural/urban areas by occupation

Canton	Sales Assistant	Electrician	HR clerk	Nurse	Total	% of the sample
Aargau	10	13	15	17	55	9.8
Appenzell Ausserrhoden	1	2	1	2	6	1.1
Basel Land	4	6	2	8	20	3.6
Basel Stadt	7	2	8	10	27	4.8
Bern	20	16	16	28	80	14.3
Fribourg	0	0	2	2	4	0.7
Glarus	0	1	0	1	2	0.4
Graubünden	10	5	4	4	23	4.1
Liechtenstein	0	1	1	0	2	0.4
Luzern	8	14	14	6	42	7.5
Nidwalden	1	1	1	0	3	0.5
Obwalden	1	0	0	2	3	0.5
Schaffhausen	2	0	1	3	6	1.1
Schwyz	3	3	4	3	13	2.3
Solothurn	4	10	3	3	20	3.6
St. Gallen	12	14	11	8	45	8.0
Thurgau	7	5	3	5	20	3.6
Uri	0	3	0	0	3	0.5
Vaud	0	0	1	0	1	0.2
Wallis	2	3	0	0	5	0.9
Zug	4	2	2	3	11	2.0
Zürich	40	30	55	39	164	29.3
No information	0	5	0	0	5	0.9
Total	136	136	144	144	560	100%

Table 7: Number of paired applications by Canton and occupation

	No information	Private	Public
Sales Assistant	1	135	0
Electrician	0	133	3
HR Clerk	2	126	16
Nurse	40	80	24
Total	43	474	43

Table 8: Number of paired applications in the private or public sector by occupation

	Not specified	Customer contact
Sales Assistant	4	132
Electrician	68	68
HR Clerk	136	8
Nurse	2	142
Total	210	350

Table 9: Number of paired applications where customer contact was mentioned in the vacancy

	Full time	Part time	No information
Sales Assistant	78	55	3
Electrician	133	0	3
HR Clerk	99	45	0
Nurse	66	78	0
Total	376	178	6

Table 10: Number of paired applications for full time or part time positions

	Temporary	Unlimited
Sales Assistant	13	123
Electrician	1	135
HR Clerk	11	133
Nurse	7	137
Total	32	528

Table 11: Number of paired applications by contract duration

	Email	Online
Sales Assistant	112	24
Electrician	108	28
HR Clerk	90	54
Nurse	107	37
Total	417	143

Table 12: Number of paired applications by application method

8. Summary of the Papers

The extensive body of literature already available on the subject of labour market discrimination, which was discussed in the introductory section of this dissertation, and in particular the theories on labour market discrimination and the methodologies used to study labour market discrimination from different perspectives, formed the basis for the development of the research design for a correspondence test on ethnic discrimination in hiring decisions in Switzerland.

So far, five papers emerged from this research project which are all part of this dissertation. These five papers build on each other, starting with more methodological work in the first paper that provides a literature review of measuring ethnic and racial discrimination using field experiments from a historical perspective. While most researchers stop after a comprehensive literature review, we used the collected information to conduct a meta-analysis of the correspondence tests that had been carried out in OECD countries between 1990 and 2015. This paper also contributed to the debate whether discrimination is due to distaste or statistics, by placing the findings into the context of this debate. A third paper that evolved from the labour intensive preparatory phase of the correspondence test focuses on the question of research ethics in correspondence testing and how this important issue is usually, with notable exceptions, ignored in published correspondence tests and a recent book on the methodology (Gaddis, 2018b).

The last two papers focus on the results of the correspondence test in the Swiss German labour market, presenting both the rather traditional presentation of the correspondence test results as well as more qualitative findings that were obtained from analysing email correspondence with employers. These are the main papers that address the research questions which were presented in the introduction. Following a more comprehensive summary of each of the five papers, the answers to the research questions will be discussed in the concluding chapter of this dissertation.

Paper I “Measuring Hiring Discrimination: A History of Field Experiments on Discrimination” provides a literature review of measuring discrimination using field experiments, i.e. both in-person audit studies as well as written correspondence tests. It focuses on the historical development of the methodology in response to emerging anti-discrimination legislation since

it was first used in the UK labour market in the 1960s. It can thus be seen that field experiments evolved considerably in the last 50 years. While early studies from the late 1960s onwards established the methodology of testing for discrimination in the labour market and continuously adapted and improved the methodology, it was in particular the establishment of a more systematic approach to testing in the 1990s that created a more consistent approach of testing for discrimination. In Europe, this development was led by the International Labour Organisation in an international comparative project for which Bovenkerk (1992) developed a detailed manual on how the testing was to be conducted. Simultaneously the Urban Institute in Washington, DC pushed forward the US research agenda on testing for labour market discrimination, although US experiments were mostly conducted as in-person audit studies. After a short period in the beginning of the 2000s, the number of field experiments on ethnic or racial discrimination in the labour market grew quickly. This recent wave of in-person audit studies and correspondence tests covered almost all Western European countries – except for Portugal and Denmark¹⁷, but including non-EU countries such as Norway and Switzerland –, the US, Canada and Australia, as well as Chile, China, Georgia, India, Israel, Malaysia or Mexico. Next to the geographical extension, field experiment began to focus on specific labour market sectors, e.g. for recent university graduates, and started to enhance the traditional matched-pair design. Furthermore, numerous new variables have been added to the research design and increasingly field experiments have been combined with interdisciplinary approaches to better understand underlying mechanisms of why employers discriminate. The literature review shows that over the last 50 years in-person audit studies and written correspondence tests have provided a great amount of “Clear and Convincing Evidence” (Fix & Struyk, 1993) of ethnic and racial hiring discrimination across space and time.

While most reviews of the literature on hiring discrimination stop at a similar point as the literature review conducted in Paper I, Paper II “Ethnic Discrimination in Hiring: A Meta-Analysis of Correspondence Tests 1990-2015” (co-authored with Didier Ruedin) takes the review a step further by conducting a meta-analysis of 43 written correspondence tests identified in literature review which were conducted in 18 OECD countries between 1990 and 2015, thus excluding in-person audit studies. One of the main challenges encountered in the preparation of the data was the non-standardised way in which results were reported in the individual studies varying between relative call-back rates or net discrimination rates, or if

¹⁷ A correspondence test in Denmark is currently ongoing and was the subject of a presentation at the IMISCOE Annual Conference 2018 in Barcelona (Dahl, 2018)

results were presented as absolute numbers or percentages. Wherever possible we recalculated the absolute numbers and calculated both the resulting net discrimination rate as well as the relative call-back rates and presented the results both for the overall study level as well as on sub-group levels for each individual study. The mean call-back ratio among all studies where it was possible to calculate this measure was 1.6. The robustness of the results was tested across specific subgroups, i.e. regions, gender, and economic context. While some minor differences exist, these findings should be treated with caution. In the following analysis we exploit the variations in the individual research designs and compared for example the situation of first and second generation applicants finding no clear evidence that discrimination might be lower for second generation candidates, call back rates for specific ethnic groups, finding clear ethnic hierarchies, or the amount of application material that was prepared, finding lower relative call-back rates in German-speaking countries, where applications are very comprehensive. If possible, we interpreted these findings in light of both statistical and taste-based discrimination theory. Overall, discrimination is still prevalent in all countries where testing has been conducted and results are relatively robust across countries and contexts.

In preparation for the correspondence test on the Swiss labour market, an extensive research ethics dossier was prepared for the Ethics Committee of the University of Neuchâtel. This endeavour led to Paper III “Research Ethics in Correspondence Testing: an Update”. The preparation of the research ethics dossier for the University of Neuchâtel had clearly shown that the issue of research ethics, which is crucial in correspondence testing, has been neglected in many publications of correspondence test, which usually only refer to the paper on research ethics by Riach and Rich (2004a). Given the fundamental changes that arose following the change from written applications sent by post to using the internet to find positions and sent applications by email or online, the paper discusses the new ethical questions that these technological advances have raised. It addressed objections that critics of correspondence testing often bring forward: 1) that correspondence testing infringes the principles of voluntary participation and informed consent, 2) that researchers deceive their research participants, and 3) that correspondence testing can have negative consequences for employers who unwillingly participated in the experiment. Despite these objections several national and international research ethics guidelines have recognized the merit of correspondence testing and have argued that it is an example of covert research, where it would not be possible to obtain data of the same quality in another way. Thus, the societal value of the research results is judged as more important than the potential harm caused to a potential employer, especially if precautions have

been taken to limit the burden on employers. The paper offers examples from countries where more information about the research ethics approval process has been available, to show how scientists dealt with these ethical issues in the research.

Following these first three papers, Paper IV “Evidence of Hiring Discrimination against the Second Generation: Results from a Correspondence Test in the Swiss Labour Market” presents the results of the correspondence test conducted in Switzerland. It briefly explains the methodology of correspondence testing and introduces the Swiss context and the current state of knowledge on ethnic discrimination in the Swiss labour market, before turning to the research design of this correspondence test. It outlines the decisions that were taken in the preparation of the experiment (e.g. the occupations and ethnic groups chosen) and the challenges encountered especially relating to the creation of comprehensive application materials (i.e. cover letters and CVs, but also references, diplomas and photos), before presenting the results of the experiment. Focusing on the German-speaking part of the country, the results show that ethnic discrimination also occurs on the Swiss labour market. Furthermore, it is shown that discrimination varies strongly depending on the occupation and the ethnicity tested. Even though the relative call back rates are rather low in the international comparison (as far as such a comparison is possible), they are in line with the results of correspondence tests conducted in other German speaking contexts. The discussion places the findings in the Swiss context.

Finally, Paper V “Equal outcomes, but different treatment – Subtle discrimination in Email responses from a correspondence test in Switzerland” adds a further dimension to the correspondence test results. It shows that while the results presented in the correspondence test are clear and concise evidence of discrimination, there are also more subtle forms of discrimination that can be observed in the hiring process. The paper argues that instances of subtle discrimination have so far largely been ignored in correspondence tests, despite the wealth of information that responses received by email provide. The focus is therefore laid on differences in the timing of responses, and how the content and tone of emails responses differed. While the majority of candidate pairs received the same or very similar invitation or rejection messages, there is still a lot of variation in the other messages. It is therefore argued that correspondence testing results show only the tip of the iceberg of the extent of discrimination in hiring decisions, and that more subtle forms of discrimination are also at play, even if the eventual outcome whether an applicant is invited or rejected is the same.

9. Conclusion

This conclusion provides a combined discussion of the results presented in the individual articles. It picks up the main research question – whether candidates with an immigrant background face discrimination in the Swiss labour market. It puts the findings into perspective, both with the correspondence test conducted in Switzerland fifteen years ago, as well as with correspondence tests conducted in other countries by drawing in particular on the results of our meta-analysis. The conclusion points out that correspondence tests can offer more information about discrimination in employment processes and that this information can be a valuable addition to the customary presentation of correspondence test results. Furthermore, it discusses the implications of the results, both in terms of the international academic debate about hiring discrimination, as well as regarding the implications for the Swiss policy context. It acknowledges the limitations of this study before concluding with suggestions for further research.

Key findings

Several key findings have emerged from this work. First of all, the meta-analysis, which was the first meta-analysis that was conducted on this topic, showed that discrimination occurs across countries, time, and national contexts. Neither applicant gender, belonging to the first or second generation or the economic context had a strong impact on the discrimination rates measured. However, the meta-analysis confirmed findings of national studies of ethnic hierarchies in the labour market, finding rather similar rankings across Western societies. One of the most interesting findings of the meta-analysis concerned the lower discrimination rates in German speaking countries.

This relates closely to one of the key findings from the correspondence test. Similarly, to the other German speaking countries that were included in the meta-analysis, the discrimination rates reported in the Swiss German labour market are also quite low in international comparison, but are rather similar to findings from Austria or Germany. So far, we have been able to provide two possible explanations for these findings, that take into account peculiarities of the German speaking labour markets. First of all, German speaking countries use much more extensive application packages, which include a lot of information about an applicant. As

argued in the papers, this could make statistical discrimination less relevant. A second explanation could be the apprenticeship system that is prevalent in German speaking countries. Most studies on German speaking labour markets have presented employers with candidates that already completed an apprenticeship. A completed apprenticeship could signal that the applicant has already at least three years of experience in the field and has passed a final examination that attests his qualification. Apprenticeships are still very well regarded in German speaking countries and providing comprehensive information on a completed apprenticeship could also reduce the likelihood that employers have to resort to statistical discrimination.

A third key finding emerging from the research is that discrimination rates measured vary strongly depending on the ethnic group and occupation tested. Somewhat surprisingly the highest discrimination rates in the Swiss German labour market were reported for German applicants for sales positions. German candidates encountered the greatest range of discrimination rates, ranging from being preferred as nurses or HR clerks to being the least invited for sales positions. While discrimination rates for Turkish or Kosovar applicants also varied, the range was not as big as for German applicants. Thus, this research project shows that including immigrant groups in a correspondence test that are normally perceived as socially and culturally close, can also yield interesting and surprising results.

Finally, the analysis of the email correspondence obtained from potential employers, which in this way has not been done in previous labour market correspondence test, shows that discrimination not only takes the form of inviting a candidate for a job interview or not, but can also be much subtler. Even in cases that are recorded as equal outcome in the correspondence test, the timing and wording of the contact can vary a lot.

The rest of this concluding section will now answer each of the research questions presented in the introduction before discussing the implications of this work, its limitations and possibilities for future research.

Do we find ethnic discrimination in the Swiss labour market?

The correspondence test for ethnic discrimination in hiring decisions that is the basis of this dissertation was conducted in the Swiss German labour market between October 2017 and April

2018. Following a paired design, fictitious applications were sent to real vacancies that Swiss employers had advertised online, using two prominent job search websites in Switzerland. The occupations chosen for the testing were electricians and sales assistants for apprenticeship level occupations, and nurses and HR clerks for intermediately skilled positions that required more than an apprenticeship. Next to the native Swiss candidate, applicants had a German, Turkish or Kosovar migration background, which was signalled by their name, by listing dual citizenship in the CV and by including Turkish or Albanian as a second native language next to German for the respective candidates. All applicants indicated that they had obtained all their education and work experience in Switzerland and were naturalised citizens, thus showing that they are well integrated in Switzerland.

The aggregated results from applications to 560 vacancies (i.e. 1120 individual applications) show that ethnic discrimination does occur in the Swiss labour market. While the native Swiss candidates were invited for a job interview in 40% of the application procedures, the candidates with an immigrant background were only successful in 35% of their applications. This results in a relative call back rate of 1.13, that was significant at the 5% level. At a first glance these discrimination rates appear to be quite low. However, looking at specific occupations discrimination rates vary quite a lot. The highest discrimination rates were measured for sales positions (1.48, significant at the 5% level), while discrimination rates for nurses and HR clerks were close to 1.0, although these results were not significant. Furthermore, there seems to be (on the aggregate level) no difference between male and female candidates.

The strong variation of discrimination rates in this correspondence test both by ethnic group and by occupation tested show that it is not possible to generalise the results of one correspondence test for the whole Swiss labour market or other immigrant groups. To get a more encompassing picture, it would be necessary to include more ethnic groups, occupations, and skill levels in the correspondence test, which is hardly feasible. In particular, in the Swiss context, where applications are accompanied by extensive application materials, it is much more complicated to create fictitious candidates than in countries where applications only contain cover letters and CVs.

The fact that ethnic discrimination in the Swiss labour market is also encountered by fictitious minority candidates that were presented as naturalised, well integrated, and well qualified for the positions they applied for, is worrying. This shows that the meritocratic principle that is

often reiterated in Switzerland by people denying the problem of ethnic discrimination has been proven wrong. The fact that discrimination is usually not a singular experience for ethnic minority candidates can contribute to their feeling of not-belonging and eventually threaten social cohesion, as Abrams already cautioned in 1968.

Do ethnic hierarchies exist in the Swiss labour market?

Looking at the result only by ethnicity, we find that ethnic hierarchies also exist in the Swiss labour market. Candidates with a Kosovar name face the highest discrimination with a relative call back rate of 1.26 (significant at the 1% level), while Turkish named candidates have a relative call back rate of 1.14 (not significant). We do not find discrimination against German named candidates when we only look at the level of ethnicity (not significant). However, the case of candidates with a German name bears closer attention, since it is for these candidates that we find the strongest variation in the relative call back rates. The highest and the lowest relative call back rates in this correspondence test were reported for candidates with a German name. While candidates with a German migration background applying for a position of sales assistant experience discrimination at a relative call back rate of 1.8, they are preferred over the Swiss natives when it comes to positions as HR clerks (relative call back rate 0.44). Both results are significant at the 5% level. Adding applicants' gender to these ethnic hierarchies does not change this situation. Men and women with a Kosovar name encounter the highest discrimination rates (8.9%, significant at the 5% level and 8.2% significant at the 10% level). The only other significant finding by gender and ethnicity is reported for Turkish named females (7.3% significant at the 10% level).

The findings of the correspondence test mirror results of research on attitudes towards foreigners in Switzerland. Candidates with an immigration background from the former Yugoslavia or the Balkans usually encounter the most negative attitudes. While the negative results for Germans applying for sales positions might seem surprising at first glance and to readers not familiar with the Swiss context, research on attitudes towards groups that are perceived as culturally close (in this case Germans), has shown negative attitude by the host society. In these cases, comparatively small differences become elevated until they seem like insurmountable differences, in particular in cases where the other is perceived as direct competition.

Did discrimination in the Swiss labour market change over time compared to Fibbi et al. (2003)?

The only other comprehensive study of ethnic discrimination on the Swiss labour market using a correspondence test design was conducted by Fibbi et al. (2003). This naturally leads to the question if the extent of discrimination changed over the last fifteen years. However, even though both correspondence tests were conducted in Switzerland, there are several differences between the two field experiments that make a comparison very difficult. First, the candidates in Fibbi et al. (2003) were all foreign born youths who came to the country at a young age and had completed their education in Switzerland. In contrast, the present study clearly signalled that candidates were well integrated, naturalised members of the second generation. They had completed all their education and training in Switzerland and listed German as (one of) their native language(s). A second substantial difference was the moment in the career of the applicants when they are looking for a new position. In the earlier study candidates had just finished their vocational training and were now looking for their first position in the labour market. In contrast, in the current testing applicants had already worked several years in the firm where they completed their apprenticeship. In the case of HR clerks, they had also already changed their employer once and thus had work experience from two different companies. Third, Fibbi et al. (2003) also chose different occupations, the applicants in their study were looking for work as accountants, receptionists, sales assistants, bakers, or hotel personnel. The only occupation that was tested in both correspondence tests was that of sales assistants. Apart from this occupation the current study focused on electricians, nurses, and HR clerks. Fourth, the regional focus of the two studies was different. While Fibbi et al. (2003) chose certain economic areas (the cantons of Zurich and Aargau in the German speaking part), the recent study did not use any regional restrictions. Fifth, Fibbi et al. (2003) sent three applications per vacancy, while the current testing only sent pairs of two fictitious applications per vacancy. The final difference between the two studies and probably the most substantial one, is the difference in the amount of application material that was prepared for the fictitious applicants. Fibbi et al. (2003) only submitted a cover letter and a CV in their applications, while the current study prepared all the documents that are habitually included in an application in Switzerland next to the cover letter and CV, i.e. the photograph, diplomas and references from former employers. All of these differences make it very hard to compare the two correspondence tests. The only conclusions which can be drawn is to say that both studies found that ethnic discrimination occurs in hiring decisions in the Swiss labour market, and that similar ethnic

hierarchies were found, with candidates from Kosovo being the most discriminated against, followed by Turkish candidates, and little or no discrimination against Western European candidates.

How do the results compare to other countries in which correspondence tests have been conducted?

Next to the comparison of the two Swiss studies over time, the question also arises how the Swiss results compare internationally. Again, comparisons of correspondence tests are difficult given the different contexts and decisions taken by the researchers. Drawing on the results from our meta-analysis (Zschirnt & Ruedin, 2016), we can place the Swiss results in the international context, although doing so has to be taken with caution. For this study on the German-speaking labour market in Switzerland, the relative call back rate on the study level was 1.13. Findings of the meta-analysis showed a relative call back rate across studies of 1.6 (Zschirnt & Ruedin, 2016), with the highest levels of discrimination on the study level being measured in France with a call back rate of 3.6 (Cediey & Foroni, 2008). Thus, in the international comparison these new Swiss results rank at the lower end of measured discrimination.

However, when we look only at studies conducted in German speaking countries, the Swiss results are not that different, since the relative call back rates reported in these studies are often lower than those in other countries. In our meta-analysis the relative call-back rate across German speaking countries was 1.4, compared to 1.8 across the non-German speaking studies (Zschirnt & Ruedin, 2016). Apart from Fibbi et al (2003) all of these studies included comprehensive application packages, thus reducing the incentive for employers to make decision according to statistical discrimination. Furthermore, as already mentioned in the key findings, all of these countries are characterised by a strong apprenticeship system in the labour market. The results from this recent Swiss testing are at a similar level as findings from Austria and Germany. Like Weichselbaumer (2016b) we have argued in the meta-analysis, that the lower rates of discrimination in countries where extensive application materials are the norm, could indicate that employers are less likely to resort to statistical discrimination in their decision making due to a lack of information about the applicants.

Does discrimination only occur in hiring decision?

As Paper 5 on subtle discrimination in the hiring process has shown, discrimination does not only occur at the moment when an employer invites or rejects a candidate. Differential treatment is also observed in the way employers contact candidates, e.g. regarding the timing of responses or the content of the messages. In the ILO studies of the 1990s, these cases were categorized as “equal but different treatment”. While the correspondence test results show discrimination in hiring decisions, the qualitative differences could indicate that there are more subtle forms of discrimination occurring in the Swiss labour market. These instances of subtle discrimination are even harder to observe and study than instances of hiring discrimination that are traditionally reported in correspondence test. In order to observe such instances of subtle discrimination, experiences made by fictitious applicants with one employer have to be compared side by side, to identify small differences in treatment. Given the increasingly big sample sizes in correspondence tests and the possibility to archive all communication with employers (voice mail messages and emails), exploiting this research material could provide valuable information on subtle discrimination¹⁸.

Looking also at instances of subtle discrimination shows that discrimination is not an isolated incidence in the hiring process. There are several instances and ways in which applicants might encounter discriminatory treatment. As Blank et al. (2004) explicitly cautioned in their discussion of cumulative discrimination,

“Studies might measure small effects of discrimination at each stage in a domain (e.g. hiring [...]), thus leading one to conclude that discrimination is relatively unimportant because the effects at any point in time are small. Over time, however, small effects could cumulate into substantial differences.” (p. 223).

Research in other domains of the labour market in Switzerland, such as accessing vocational training (Imdorf, 2007, 2008) or on work place incivilities experienced by immigrants (Krings et al., 2014), has clearly shown that discrimination happens at multiple stages within the domain of the labour market and appears to be in line with the concept of cumulative discrimination.

¹⁸ Crabtree (2018) proposes automated analysis of reply emails. The R code is made available at <https://github.com/cdcrabtree/emailaudits>

Theoretical contributions

The results of this dissertation contribute to the theoretical debate whether discrimination is due to distaste or statistics in two ways. First, the research design of the correspondence test is very similar to Weichselbaumer (2016b) since it also tests for discrimination against several immigrant groups and uses applications with a lot of accompanying documentation for each candidate. Like Austrian and German correspondence tests, rather low discrimination rates are measured in the Swiss case and the existence of ethnic hierarchies is confirmed. The second theoretical contribution is delivered in the meta-analysis presented in Paper II, which classifies its findings according to the two theories of taste-based and statistical discrimination. While the existence of similar discrimination rates against first and second generation applicants as well as the existence of ethnic hierarchies point towards the theory of taste-based discrimination, in particular the lower discrimination rates in German speaking countries point rather towards mechanisms of statistical discrimination.

However, clearly pointing to one, or the other theory seems to be too limited. I would rather argue, that e.g. in the case of German speaking countries the fact that employers have much more information compared to employers in other countries, does reduce discrimination according to the logic of statistical discrimination theory, but, at the same time the remaining discrimination measured as well as ethnic hierarchies found point towards a taste for discrimination. Furthermore, the correspondence test cannot explain where such a taste for discrimination comes from – here work on attitudes towards foreigners and work by social psychologists on stereotypes and prejudices provides valuable insights.

Implications

Next to its academic contribution, the research is also relevant on a political and societal level. The results presented in this dissertation have shown, that discrimination against foreign named candidate occurs in the Swiss labour market. Even if candidates are naturalized and list only Swiss educational and work history, they still get disadvantaged in their search for employment. Combined with previous research on discrimination, it becomes evident that discrimination is not only suffered at one point in time, but repeatedly. Victims of discrimination also seem to slowly start to give voice to their experience of discrimination, as the increasing number of cases that were brought to the attention of NGOs advising on discrimination cases has shown.

One issue that was mentioned in studies on perceived discrimination was the increasingly negative portrayal of immigrants in the Swiss political debate and in the media. The last decades have shown a strengthening of anti-immigrant political campaigning and a rising number of popular initiatives and referenda on restricting the rights to immigrate and the rights of migrants already residing in Switzerland. In particular campaign posters of the politically right SVP/UDC have continuously portrayed immigrants as a threat. In the public discourse this focus on the negative perception of immigrants has largely taken hold, a trend that can also be observed in other European countries with the rise of right-wing parties. What is somewhat new, are the strong negative attitudes to Western European immigrants and in particular those from neighbouring countries where immigrants usually already speak one of official languages of Switzerland and should be expected to integrate easily. However, this group of highly skilled migrants was one of the key targets of the initiative against mass immigration that was adopted (narrowly) in February 2014. This initiative sought to limit the entry of foreign workers into the Swiss labour market by forcing employers to give preference to Swiss candidate or people who were already residents in Switzerland, the *Inländervorrang* as it is called in the German debate. However, the results of this study have shown, that there already is a *de facto* preference of native Swiss candidates. Even though the candidates in our study were clearly well integrated into Swiss society, as their Swiss work and education history and the fact that they had been naturalised shows, they still face disadvantages in the labour market, although differences according to the ethnic background of candidates persist.

Given the high share of people with a migration background in Switzerland, these developments are worrisome. There have been tentative attempts to introduce the topic of ethnic discrimination in the political debate, yet so far, they have not been successful. The results of this research can raise awareness of the issue of ethnic discrimination in Switzerland and contribute to a debate in the media and in politics. The findings could strengthen proponents of a Swiss anti-discrimination law, by providing evidence that discrimination occurs in the Swiss labour market, a fact that has been doubted by many actors in Swiss politics.

Limitations of this study

There are, of course, limitations to this study and most of them are inherent to the chosen methodology of a correspondence test and were already addressed in Heckman's and

Siegelman's critics discussed above. With this methodology it is only possible to observe discrimination at one very specific point in time, i.e. whether an applicant is invited for a job interview or not, and it can only be used for publicly advertised positions for which applications can be submitted online and in writing. Another limitation is the lack of generalisability. Due to the amount of preparation required to create credible application packages given the number of documents required, it is impossible to test for discrimination in numerous occupations. We therefore only included four occupations in our research design, which were either at apprenticeship level or intermediate skilled occupations. The variation of relative call back rates by occupations and ethnicity clearly show the importance of carefully choosing occupations and ethnic groups and the problem of generalizing from our results to the Swiss labour market in general. Finally, the sample size is a clear limitation of this study. Given the small size of the Swiss labour market and the limitation that companies can only be included in the testing once, it became very challenging to find enough vacancies, which could be used in the experiment. To reach a bigger sample size we would (a) need more time and rely strongly on vacancies posted by small and more locally oriented firms, or (b) add more occupations, which is challenging due to the amount of preparation required for the application materials.

Suggestions for future research

None of the more than 50 correspondence tests on ethnic and hiring discrimination conducted until this date have found no evidence of discrimination against minority candidates. As long as this is the case, there is still a need to conduct research on labour market discrimination and correspondence tests in particular.

There are several interesting avenues that future research in this area could take. It could include more occupations and/or more immigrant groups to increase the external validity of results. The results presented in this study have also shown the value of including non-stigmatised immigrant groups that do not originate from lower income countries. It would be great to see more research including such groups, to study the extent of ethnic hierarchies and the influence of social and cultural distance. Furthermore, it would be interesting to see more studies that use a standardized research design in different national contexts to study the impact of environmental factors, such as differences in legal regimes, differences in labour market structures, or differences in the economic situation of a country. In the labour market context, the first internationally standardized research project was the ILO Project of the 1990s which

followed the research design of (Bovenkerk, 1992). More recently the ongoing GEMM Project endeavoured to research discrimination following a quite standardised research design in five European countries. Yet, this research design is based on sending out single applications to each employer and using a very large number of possible ethnic minorities.

Discrimination research could also benefit greatly from more studies using mixed methods to assess discrimination in the labour market and by becoming more interdisciplinary. As was pointed out during the discussion on the theories explaining discrimination, there is a whole body of literature on discrimination and attitudes in psychology that has developed in parallel to the work of economists and sociologists. These insights could provide valuable inputs, particular when it comes to the theoretical debate on why discrimination occurs. Overall, the field could benefit from adding more qualitative information to correspondence test experiments, e.g. interviews with employers or systematic analysis of email or voicemail messages. The quantitative approaches employed in correspondence tests on ethnic discrimination have shown that discrimination on the labour market exists, yet they only document the extent of discrimination, not *why* discrimination occurs. Including more quantitative and interdisciplinary methods could be very helpful to study the mechanisms underlying labour market discrimination. Adding to the call for more mixed methods and more qualitative research on labour market discrimination it would also be interesting to not only focus on the mechanisms explaining hiring discrimination on the side of the employer, but to also study how applicants, i.e. the victims of discrimination, anticipate discriminatory treatment and adapt their job search strategies to this fact.

Finally, it would be great to see research on discrimination that does not only look at one isolated instance, but manages to combine experiences of discrimination. This goes into the direction of the concept of cumulative discrimination, which is hard to measure, but finding interdisciplinary ways to measure discrimination over life-courses, across domains and within domains could provide great information how seemingly small instances of discrimination have long-term effects.

Abstracts of the Papers

Paper 1: Measuring Hiring Discrimination – A History of Field Experiments in Discrimination Research

Zschirnt, E. (2016) “Measuring Hiring Discrimination – A History of Field Experiments in Discrimination Research”. *NCCR Working Papers*, 7, 1-32.

Abstract

Ethnic and racial discrimination in the labour market is a common and documented problem. Scientists from different backgrounds and numerous countries have tried to measure the extent of this form of discrimination, often by using field experiments at the hiring stage. This paper provides an overview of the literature on measuring discrimination with field experiments. It focuses on methodological issues such as the difference between in-person and written tests, before addressing the historical political context in which field experiments have emerged and how the technique was developed further over time. It shows that today’s field experiments not only cover a wider group of countries, professions or disadvantaged groups, but also increasingly add more variables to the testing. Despite this variety in the research designs, certain trends can be observed in all experiments and hiring discrimination can be found in all countries where field experiments were conducted.

Paper 2: Ethnic Discrimination in Hiring: A Meta-Analysis of Correspondence Tests 1990-2015 (with Didier Ruedin)

Zschirnt, E. and Ruedin, D. (2016) “Ethnic Discrimination in Hiring: A Meta-Analysis of Correspondence Tests 1990-2015”, *Journal of Ethnic and Migration Studies*, (42) 7, 1115-1134.

Abstract

For almost 50 years field experiments have been used to study ethnic and racial discrimination in hiring decisions, consistently reporting high rates of discrimination against minority applicants – including immigrants – irrespective of time, location, or minority groups tested.

While Peter A. Riach and Judith Rich [2002. “Field Experiments of Discrimination in the Market Place.” *The Economic Journal* 112 (483): F480–F518] and Judith Rich [2014. “What Do Field Experiments of Discrimination in Markets Tell Us? A Meta Analysis of Studies Conducted since 2000.” In *Discussion Paper Series*. Bonn: IZA] provide systematic reviews of existing field experiments, no study has undertaken a meta-analysis to examine the findings in the studies reported. In this article, we present a meta-analysis of 738 correspondence tests in 43 separate studies conducted in OECD countries between 1990 and 2015. In addition to summarising research findings, we focus on groups of specific tests to ascertain the robustness of findings, emphasising differences across countries, gender, and economic contexts. Moreover we examine patterns of discrimination, by drawing on the fact that the groups considered in correspondence tests and the contexts of testing vary to some extent. We focus on first and second-generation immigrants, differences between specific minority groups, the implementation of EU directives, and the length of job application packs.

Paper 3: Research Ethics in Correspondence Testing: an Update

Zschirnt, E. (2016) “Revisiting Ethics in Correspondence Testing ». *NCCR Working Papers*, 8. 1-26.

Abstract

Correspondence testing researching discrimination in the market place has become more widespread and the use of internet applications has allowed researchers to send greater numbers of applications. While questions of research ethics always arise when planning a correspondence test, the issue receives relatively little attention in published correspondence tests. This paper addresses the question of ethics in correspondence testing in the age of ready internet access. It focuses on the ethical issues that arise in correspondence testing, looking at potential problems (regarding voluntary participation, informed consent, deception, entrapment of employers, employers’ rights) and possible solutions, and technical challenges. European country examples show that the ethical questions raised in correspondence testing have to be renegotiated depending on the national context. The paper argues that correspondence testing, if planned carefully and executed responsibly, can meet most of the ethical requirements of social sciences ethics guidelines.

Paper 4: Evidence of Hiring Discrimination against the Second Generation: Results from a Correspondence Test in the Swiss Labour Market

Abstract

While there is ample evidence of discrimination against ethnic minority candidates in hiring, most existing studies have focused on stigmatized immigrant groups. We use a correspondence test to enumerate ethnic discrimination in the Swiss labour market, varying the *a priori* stigma of the immigrant groups. The field experiment compares candidates with Swiss names against candidates with German, Kosovar and Turkish names in a paired correspondence test spanning four occupations. Between October 2017 and April 2018 applications were sent in response to 560 real vacancies in the German-speaking area of Switzerland. Across the minority groups, the relative call back rate was 1.13, meaning that minority candidates have to write 1.13 times as many applications as majority candidates to be invited for a job interview. The relative call back rates differ by the ethnic origin: Germans experience almost no discrimination across all occupations; Turks face a relative call back rate of 1.14; and Kosovars encounter the highest relative call back rate across occupations (1.26). We conclude that existing studies may give the false impression that all immigrants suffer from substantive discrimination in the labour market because they focus on stigmatized groups.

Paper 5: Equal outcomes, but different treatment – subtle discrimination in email responses from a correspondence test in Switzerland

Abstract

Correspondence tests on discrimination usually only report whether an applicant was invited for a job interview or not. While in-person audit studies can observe differences in treatment during a job interview, e.g. the length or the tone of an interview, this has hardly been done in correspondence studies. Data from a field experiment in Switzerland demonstrate that the commonly reported results of correspondence tests only show one side of the picture. Candidates with the same outcome (invited, not invited) are not necessarily treated equally. The paper complements correspondence test results with information on the time elapsed until candidates were contacted, as well as qualitative differences in invitation or rejection emails.

References

- Abrams, M. (1968). Introduction. In W. W. Daniel (Ed.), *Racial Discrimination in England - Based on the PEP Report* (pp. 9-15). Middlesex: Penguin Books
- Ackermann, K., & Ackermann, M. (2015). The Big Five in Context: Personality, Diversity and Attitudes toward Equal Opportunities for Immigrants in Switzerland. *Swiss Political Science Review*, 21(3), 396-418. doi:10.1111/spsr.12170
- Ackermann, M., & Freitag, M. (2015). What Actually Matters? Understanding Attitudes toward Immigration in Switzerland. *Swiss Political Science Review*, 21(1), 36-47. doi:10.1111/spsr.12146
- Agerström, J., Björklund, F., Carlsson, R., & 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*, 34(4), 359-366. doi:10.1080/01973533.2012.693438
- Ahmed, A. M., Andersson, L., & Hammarstedt, M. (2010). Can discrimination in the housing market be reduced by increasing the information about the applicants? *Land Economics*, 86(1), 79-90.
- Aigner, D. J., & Cain, G. G. (1977). Statistical theories of discrimination in labor markets. *Industrial and Labor Relations Review*, 175-187.
- Akintola, A. R. (2011). *Hiring Discrimination in Racially Diverse Labour Markets: A Cross Country Study*. (Master), Karstad Business School.
- Andriessen, I., Dagevos, J., & Iedema, J. (2008). Onbemind of ongemeten? Over de betekenis van discriminatie als factor in de werkloosheid van niet-westerse allochtonen. *Mens & Maatschappij*, 83(3), 279-301.
- Andriessen, I., Fernee, H., & Wittebrood, K. (2014). *Perceived Discrimination in the Netherlands*. Retrieved from Den Haag: http://www.scp.nl/Publicaties/Alle_publicaties/Publicaties_2014/Perceived_discrimination_in_the_Netherlands
- Andriessen, I., Nievers, E., Dagevos, J., & Faulk, L. (2012). Ethnic Discrimination in the Dutch Labor Market: Its Relationship With Job Characteristics and Multiple Group Membership. *Work and Occupations*, 39(3), 237-269.
- Andriessen, I., Nievers, E., Faulk, L., & Dagevos, J. (2010). *Liever Mark dan Mohammed?* Retrieved from The Hague:
- Arceo-Gómez, E. O., & Campos-Vázquez, R. (2013). *Race and Marriage in the Labor Market: A Discrimination Correspondence Study in a Developing Country*. Retrieved from
- Arrighi, J.-T. (2018). *Dataset on Migration Referendums and Initiatives in Switzerland*. Retrieved from: <https://indicators.nccr-onthemove.ch/did-federal-referendums-and-initiatives-affect-immigrants-rights/>
- Arriijn, P., Feld, S., & Nayer, A. (1998). *Discrimination in access to employment on grounds of foreign origin: the case of Belgium*. Geneva: International Labour Office, Conditions of Work Branch.
- Arrow, K. J. (1973). The Theory of Discrimination. In O. Ashenfelter & A. Rees (Eds.), *Discrimination in Labor Markets* (pp. 15-42). Princeton: Princeton University Press.
- Arrow, K. J. (1998). What has economics to say about racial discrimination? *The Journal of Economic Perspectives*, 91-100.
- Asali, M., Pignatti, N., & Skhirtladze, S. (2017). Employment Discrimination in Georgia: Evidence from a Field Experiment. *MPRA Paper No. 78230*.

- Attström, K. (2007). *Discrimination against native Swedes of immigrant origin in access to employment*. Geneva: International Labor Organization.
- Auer, D., Bonoli, G., & Fossati, F. (2017). Why do immigrants have longer periods of unemployment? Swiss evidence. *International Migration*, 55(1), 157-174. doi:10.1111/imig.12309
- Auer, D., Bonoli, G., Fossati, F., & Liechti, F. (2018). The Matching Hierarchies Model: Evidence from a Survey Experiment on Employers' Hiring Intent Regarding Immigrant Applicants. *International migration review*, 0197918318764872.
- Auspurg, K., & Hinz, T. (2015). *Factorial Survey Experiments*: Sage.
- Baert, S. (2018). Hiring Discrimination: An Overview of (Almost) All Correspondence Experiments Since 2005. In M. Gaddis (Ed.), *Audit Studies: Behind the Scenes with Theory, Method and Nuance* (pp. 63-77): Springer.
- Baert, S., Albanese, A., du Gardein, S., Ovaere, J., & Stappers, J. (2017). Does work experience mitigate discrimination? *Economics Letters*, 155, 35-38. doi:<http://dx.doi.org/10.1016/j.econlet.2017.03.011>
- Baert, S., Cockx, B., Gheyle, N., & Vandamme, C. (2015). Is there less discrimination in occupations where recruitment is difficult? *International Labour Review*. doi:0019793915570873
- Baert, S., & De Pauw, A.-S. (2014). Is ethnic discrimination due to distaste or statistics? *Economics Letters*, 125(2), 270-273.
- Baert, S., & Vujić, S. (2016). Immigrant volunteering: A way out of labour market discrimination? *Economics Letters*, 146, 95-98.
- Bagley, C., & Abubaker, M. (2017). Muslim Woman Seeking Work: An English Case Study with a Dutch Comparison, of Discrimination and Achievement. *Social Sciences*, 6(1), 17.
- Banerjee, A., & Duflo, E. (2014). The experimental approach to development economics. In D. L. Teele (Ed.), *Field experiments and their critics: Essays on the uses and abuses of experimentation in the social sciences* (pp. 78-114). New Haven: Yale University Press.
- Banton, M. (1997). The ethics of practice - testing. *Journal of Ethnic and Migration Studies*, 23(3), 413-420.
- Barrett, C. B., & Carter, M. R. (2014). A retreat from radical skepticism: rebalancing theory, observational data, and randomization in development economics. In D. L. Teele (Ed.), *Field experiments and their critics: Essays on the uses and abuses of experimentation in the social sciences* (pp. 58-77). New Haven: Yale University Press.
- Becker, G. (1957). *The Economics of Discrimination* (2nd ed.). Chicago: University of Chicago Press.
- Becker, G. (1964). *Human Capital: a Theoretical and Empirical Analysis, with Special Reference to Education*. Chicago: University of Chicago Press.
- Beigang, S., Fetz, K., Foroutan, N., Kalkum, D., & Otto, M. (2016). *Diskriminierungserfahrungen in Deutschland - Erste Ergebnisse einer repräsentativen Erhebung und einer Betroffenenbefragung* Retrieved from Berlin:
- Bendick, M. (1996). *Discrimination Against Racial-ethnic Minorities in Access to Employment in the United States: Empirical Findings from Situation Testing*: Employment Department, International Labour Office.
- Bendick, M. (2007). Situation testing for employment discrimination in the United States of America. *Horizons stratégiques*(3), 17-39.

- Bendick, M., Jackson, C. W., Reinoso, V. A., & Hodges, L. E. (1991). Discrimination against Latino job applicants: A controlled experiment. *Human Resource Management*, 30(4), 469-484.
- Bendick, M., & Nunes, A. P. (2012). Developing the Research Basis for Controlling Bias in Hiring. *Journal of Social Issues*, 68(2), 238-262. doi:10.1111/j.1540-4560.2012.01747.x
- Bendick, M., Rodriguez, R. E., & Jayaraman, S. (2010). Employment discrimination in upscale restaurants: Evidence from matched pair testing. *The Social Science Journal*, 47(4), 802-818.
doi:<http://dx.doi.org/10.1016/j.soscij.2010.04.001>
- Bergmann, B. R. (1989). Does the market for women's labor need fixing? *Journal of Economic Perspectives*, 3(1), 43-60.
- Berthoud, R. (2000). Ethnic employment penalties in Britain. *Journal of Ethnic and Migration Studies*, 26(3), 389-416.
- Bertrand, M., Chugh, D., & Mullainathan, S. (2005). Implicit Discrimination. *The American Economic Review*, 95(2), 94-98. doi:10.2307/4132797
- Bertrand, M., & Duflo, E. (2016). *Field Experiments on Discrimination*. Retrieved from London:
http://cepr.org/active/publications/discussion_papers/dp.php?dpno=11123
- 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. doi:10.2307/3592802
- Binggeli, S., Krings, F., & Sczesny, S. (2014a). Perceived competition explains regional differences in the stereotype content of immigrant groups. *Social Psychology*, 45(1), 62.
- Binggeli, S., Krings, F., & Sczesny, S. (2014b). Stereotype content associated with immigrant groups in Switzerland. *Swiss journal of psychology*, 73(3), 123-133.
- Birkelund, G. E., Heggebø, K., & Rogstad, J. (2017). Additive or Multiplicative Disadvantage? The Scarring Effects of Unemployment for Ethnic Minorities. *European Sociological Review*, 33(1), 17-29. doi:10.1093/esr/jcw030
- Birkelund, G. E., Ugreninov, E., Chan, W., Midtbøen, A., & Rogstad, J. (2017). *Do Terrorist Attacks Affect Ethnic Discrimination in the Labour Market? Evidence from Two Randomised Field*. Retrieved from
<http://econpapers.repec.org/RePEc:qss:dqsswp:1702>
- Blank, R. M., Dabady, M., & Citro, C. F. (2004). *Measuring racial discrimination*. Washington, DC: National Academies Press.
- Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, 436-455.
- Blommaert, L. (2013). *Are Joris and Renske more employable than Rshid and Samira? A study on the prevalence of sources of ethnic discrimination in recruitment in the Netherlands using experimental survey data*. Utrecht University Retrieved from <http://dspace.library.uu.nl/handle/1874/276002>
- Blommaert, L., Coenders, M., & van Tubergen, F. (2014). Discrimination of Arabic-Named Applicants in the Netherlands: An Internet-Based Field Experiment Examining Different Phases in Online Recruitment Procedures. *Social Forces*, 92(3), 957-982. doi:10.1093/sf/sot124
- Bonoli, G., & Fossati, F. (2018). More than noise? Explaining instances of minority preference in correspondence studies of recruitment. *Journal of Ethnic and Migration Studies*, 1-17.

- Booth, A. L., Leigh, A., & Varganova, E. (2012). Does ethnic discrimination vary across minority groups? Evidence from a field experiment. *Oxford Bulletin of Economics and Statistics*, 74(4), 547-573.
- Bosch, M., Carnero, M. A., & Farre, L. (2010). Information and discrimination in the rental housing market: Evidence from a field experiment. *Regional Science and Urban Economics*, 40(1), 11-19.
- Boulila, S. C. (2018). Race and racial denial in Switzerland. *Ethnic and Racial Studies*. doi:DOI: 10.1080/01419870.2018.1493211
- Bovenkerk, F. (1992). *Testing discrimination in natural experiments: a manual for international comparative research on discrimination on the grounds of" race" and ethnic origin*. Geneva: International Labour Office.
- Bovenkerk, F., Gras, M., Ramsoedh, D., Dankoor, M., & Havelaar, A. (1995). *Discrimination against migrant workers and ethnic minorities in access to employment in the Netherlands*. Retrieved from Geneva:
- Buchmann, M., Buchs, H., Gnehm, A.-S., Hevenstone, D., Klarer, U., Müller, M., . . . Salvisberg, A. (2015). *Erhebung von Stelleninseraten 1950-2014 [Dataset]*.
- Buchs, H., & von Ow, A. (2017). *Personalsuche und Stellenbesetzung - Kurzreport zur Unternehmensbefragung 2017*. Retrieved from Zürich:
http://www.stellenmarktmonitor.uzh.ch/dam/jcr:5abe78f7-9813-4d8b-b25f-08762530d24c/Report_Unternehmensbefragung_2017_dt.pdf
- Bundesamt für Statistik. (2015). Gross monthly wage, Swiss and foreigners - Private sector - Switzerland. Retrieved from
<https://www.bfs.admin.ch/bfs/de/home/statistiken/arbeit-erwerb/loehne-erwerbseinkommen-arbeitskosten/lohniveau-schweiz/personenbezogene-merkmale.assetdetail.304061.html>
- Bundesamt für Statistik. (2017a). *Arbeitsmarktindikatoren 2017*. Retrieved from Neuchâtel:
- Bundesamt für Statistik. (2017b, 24.10.2017). Bevölkerung nach Migrationsstatus 2016. Retrieved from
<https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung/migration-integration.assetdetail.1521883.html>
- Bundesamt für Statistik. (2017c). Demographic balance of the permanent resident population by citizenship 2010-2016. Retrieved from
<https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung.assetdetail.300195.html>
- Bundesamt für Statistik. (2018a). Erwerbslosenquote gemäss ILO nach Geschlecht, Nationalität und anderen Merkmalen. Retrieved from
<https://www.bfs.admin.ch/bfs/en/home/statistics/work-income/unemployment-underemployment-vacancies/ilo-unemployed.assetdetail.4463045.html>
- Bundesamt für Statistik. (2018b). Statistischer Atlas der Schweiz - Ständige ausländische Wohnbevölkerung, 2016. *Statistik Schweiz*. Retrieved from
https://www.atlas.bfs.admin.ch/maps/13/de/12732_10479_89_70/20951.html
- Bursell, M. (2007). » *What's in a name?-A field experiment test for the existence of ethnic discrimination in the hiring process* «, . Retrieved from
- Bursell, M. (2012). *Ethnic Discrimination, Name Change and Labor Market Inequality - Mixes approaches to ethnic exclusion in Sweden*. Stockholm University, Stockholm.
- Calvès, G. (2007). Au service de la connaissance et du droit: le testing. *Horizons stratégiques*(3), 8-16.

- Carlsson, M. (2010). Experimental Evidence of Discrimination in the Hiring of First- and Second-generation Immigrants. *Labour*, 24(3), 263-278.
- Carlsson, M., & Eriksson, S. (2014). Discrimination in the rental market for apartments. *Journal of Housing Economics*, 23, 41-54.
- Carlsson, M., & Eriksson, S. (2017). Do attitudes expressed in surveys predict ethnic discrimination? *Ethnic and Racial Studies*, 40(10), 1739-1757.
- Carlsson, M., Fumarco, L., & Rooth, D.-O. (2014). Does the design of correspondence studies influence the measurement of discrimination? *IZA Journal of Migration*, 3(1), 1-17. doi:10.1186/2193-9039-3-11
- Carlsson, M., & Rooth, D.-O. (2007). Evidence of ethnic discrimination in the Swedish labor market using experimental data. *Labour Economics*, 14(4), 716-729. doi:<http://dx.doi.org/10.1016/j.labeco.2007.05.001>
- Carlsson, M., & Rooth, D.-O. (2012). Revealing taste-based discrimination in hiring: a correspondence testing experiment with geographic variation. *Applied Economics Letters*, 19(18), 1861-1864. doi:10.1080/13504851.2012.667537
- Carlsson, R., & Agerström, J. (2016). A closer look at the discrimination outcomes in the IAT literature. *Scandinavian journal of psychology*, 57(4), 278-287.
- Cediey, E., & Foroni, F. (2008). *Discrimination in access to employment on grounds of foreign origin in France: A national survey of discrimination based on the testing methodology of the International Labour Office* (Vol. 85E). Geneva: International Labor Organisation (ILO).
- CERD. (2014). *Concluding Observations on the Combined Seventh to Ninth Periodic Reports of Switzerland*, CERD/C/CHE/CO/7-9 (13 March 2014). Retrieved from Geneva:
- Cherry, F., & Bendick, M. (2018). Making It Count: Discrimination Auditing and the Activist Scholar Tradition *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 45-62): Springer.
- Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin, 180 C.F.R. (2000a).
- Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation (2000b).
- Crabtree, C. (2018). An introduction to conducting email audit studies *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 103-117): Springer.
- Crabtree, C., & Chykina, V. (2018). Last Name Selection in Audit Studies. *Sociological Science*, 5, 2-21.
- Dahl, M. R. (2018). *Cultural determinants of ethnic discrimination: Evidence from a field experiment in the labour market*. Paper presented at the IMISCOE Annual Conference, Barcelona.
- Damelang, A., & Abraham, M. (2016). You can take some of it with you! *Zeitschrift für Soziologie*, 45(2), 91-106.
- Daniel, W. W. (1968). *Racial discrimination in England: based on the PEP report* (Vol. 1084). Middlesex: Penguin.
- Darity, W. A., & Mason, P. L. (1998). Evidence on discrimination in employment: Codes of color, codes of gender. *Journal of Economic Perspectives*, 12(2), 63-90.
- Darolia, R., Koedel, C., Martorell, P., Wilson, K., & Perez-Arce, F. (2016). Race and gender effects on employer interest in job applicants: new evidence from a resume field experiment. *Applied Economics Letters*, 23(12), 853-856.

- de Prada, M. A., Actis, W., Pereda, C., & Molina, R. P. (1995). *Labour market discrimination against migrant workers in Spain* (Vol. 9). Geneva: International Labour Office, Employment Department.
- Dechief, D., & Oreopoulos, P. (2012). Why do some employers prefer to interview Matthew, but not Samir? New Evidence from Toronto, Montreal, and Vancouver. *Canadian Labour Market and Skills Researcher Network, Working Paper No. 95*.
- Di Stasio, V. (2014). Education as a signal of trainability: Results from a vignette study with Italian employers. *European Sociological Review*, 30(6), 796-809.
- Diekmann, A., Jann, B., & Näf, M. (2014). Wie fremdenfeindlich ist die Schweiz? Fünf Feldexperimente über prosoziales Verhalten und die Diskriminierung von Ausländern in der Stadt Zürich und der Deutschschweiz. *Soziale Welt*, 65, 185-199.
- Drydakis, N. (2014). Sexual orientation discrimination in the Cypriot labour market. Distastes or uncertainty? *International Journal of Manpower*, 35(5), 720-744. doi:doi:10.1108/IJM-02-2012-0026
- Drydakis, N. (2017). Measuring labour differences between natives, non-natives, and natives with an ethnic-minority background. *Economics Letters*, 161, 27-30.
- Efionayi-Mäder, D., Ruedin, D., Pétrémont, M.-E., Michel, N., & Jain, R. (2017). *Anti-Schwarzen-Rassismus in der Schweiz—Eine Bestandsaufnahme*. Retrieved from
- Eid, P. (2012). *Mesurer la discrimination à l'embauche subie par les minorités racisées: résultats d'un "testing" mené dans le grand Montréal*. Quebec: Commission des droits de la personne et des droits de la jeunesse.
- European Union Agency for Fundamental Rights. (2017). *Second European Union Minorities and Discrimination Survey - Main results* Retrieved from Luxembourg: <http://fra.europa.eu/en/publication/2017/eumidis-ii-main-results>
- Ewens, M., Tomlin, B., & Wang, L. C. (2014). Statistical discrimination or prejudice? A large sample field experiment. *Review of Economics and Statistics*, 96(1), 119-134.
- Fibbi, R. (2006). Discrimination dans l'accès à l'emploi des jeunes d'origine immigrée en Suisse. *Formation-emploi*(94), 45-58.
- Fibbi, R., Kaya, B., & Piguet, E. (2003). *Le passeport ou le diplôme ? Etude des discriminations à l'embauche des jeunes issus de la migration*. Neuchâtel: Forum suisse pour l'étude des migrations et de la population.
- Fibbi, R., Lerch, M., & Wanner, P. (2006). Unemployment and Discrimination against Youth of Immigrant Origin in Switzerland: When the Name Makes the Difference. *Journal of International Migration and Integration*, 7(3), 351–366.
- Field, A., & Hole, G. (2002). *How to design and report experiments*: Sage.
- Firth, M. (1981). Racial Discrimination in the British Labor Market. *Industrial and Labor Relations Review*, 34(2), 265-272. doi:10.2307/2522541
- Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. *The handbook of social psychology*, 2, 357-411.
- Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878-902.
- Fix, M., & Struyk, R. J. (Eds.). (1993). *Clear and Convincing Evidence: Measurement of Discrimination in America*. Washington, D.C.: The Urban Institute Press.

- Flinn, C. J. (2015). GARY BECKER'S CONTRIBUTIONS TO THE ANALYSIS OF DISCRIMINATION. *Journal of Demographic Economics*, 81(01), 45-50. doi:doi:10.1017/dem.2014.5
- Fossati, F., Liechti, F., Auer, D., & Bonoli, G. (2017). *Discrimination Multipliers - How Immigrants' integration affects labour market disadvantage*. MIM Working Papers Series. Malmö Institute for Studies of Migration, Diversity and Welfare (MIM). Malmö. Retrieved from https://www.mah.se/upload/Forskningscentrum/MIM/WPS_17.2_Fossatti.pdf
- Freitag, M., Vatter, A., & Mueller, S. (2015). Switzerland's Immigration Challenge - Viewpoints and insights in the aftermath of the mass immigration initiative. *Swiss Political Science Review*, 21(1), 1-4. doi:10.1111/spsr.12149
- Fryer, R. G., & Levitt, S. D. (2004). The causes and consequences of distinctively black names. *The Quarterly Journal of Economics*, 119(3), 767-805.
- Gaddis, S. M. (2014). Discrimination in the Credential Society: An Audit Study of Race and College Selectivity in the Labor Market. *Social Forces*, 1-29.
- Gaddis, S. M. (2017a). How Black are Lakisha and Jamal? Racial Perceptions from Names Used in Correspondence Audit Studies. *Sociological Science*, 4, 469-489.
- Gaddis, S. M. (2017b). Racial/Ethnic Perceptions from Hispanic Names: Selecting Names to Test for Discrimination. *Socius*, 3, 2378023117737193. doi:10.1177/2378023117737193
- Gaddis, S. M. (2018a). An introduction to audit studies in the social sciences *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 3-44): Springer.
- Gaddis, S. M. (Ed.) (2018b). *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*: Springer.
- Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. New York: WW Norton & Company.
- Gesellschaft für deutsche Sprache e.V. (2017). Die beliebtesten Vornamen. Retrieved from <http://gfds.de/vornamen/beliebteste-vornamen/>
- Ghumman, S., & Ryan, A. M. (2013). Not welcome here: Discrimination towards women who wear the Muslim headscarf. *Human Relations*, 66(5), 671-698.
- Givens, T. E., & Case, R. E. (2014). *Legislating Equality: The Politics of Antidiscrimination Policy in Europe*. Oxford: Oxford University Press.
- Goldberg, A., Mourinho, D., & Kulke, U. (1995). *Labour market discrimination against foreign workers in Germany*: International Labour Office, Employment Department.
- Golder, L., Mousson, M., Tschöpe, S., Herzog, N., & Bohn, D. (2017). *Hinweise auf erlebte, aber kaum registrierte Hasskriminalität - Pilotstudie Diskriminierungserfahrungen Muslime in der Schweiz*. Retrieved from Bern: <http://www.gfsbern.ch/de-ch/Detail/hinweise-auf-erlebte-aber-kaum-registrierte-hasskriminalitaet>
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464.
- Greth, J., & Köllen, T. (2016). Perceived Anti-Germanism in Austria. *Studies in Ethnicity and Nationalism*, 16(1), 40-62. doi:10.1111/sena.12165
- Guryan, J., & Charles, K. K. (2013). Taste - based or Statistical Discrimination: The Economics of Discrimination Returns to its Roots. *The Economic Journal*, 123(572), F417-F432.

- Hagendoorn, L. (1993). Ethnic categorization and outgroup exclusion: cultural values and social stereotypes in the construction of ethnic hierarchies. *Ethnic and Racial Studies*, 16(1), 26-51.
- Hainmueller, J., Hangartner, D., & Yamamoto, T. (2015). Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences*, 112(8), 2395-2400.
- Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 1009-1055.
- Heath, A. F., & Cheung, S. Y. (2006). *Ethnic penalties in the labour market: Employers and discrimination* (1841239968). Retrieved from Norwich:
- Heath, A. F., & Cheung, S. Y. (Eds.). (2007). *Unequal chances: Ethnic minorities in Western labour markets*. Oxford: Oxford University Press.
- Heath, A. F., Rothon, C., & Kilpi, E. (2008). The second generation in Western Europe: Education, unemployment, and occupational attainment. *Annu. Rev. Sociol.*, 34, 211-235.
- Heckman, J. J. (1998). Detecting discrimination. *The Journal of Economic Perspectives*, 101-116.
- Heckman, J. J., & Siegelman, P. (1993). The Urban Institute Audit Studies: Their Methods and Findings. In M. Fix & R. Struyk (Eds.), *Clear and Convincing Evidence: Measurements of Discrimination in America* (pp. 187-258). Washington, DC: The Urban institute Press.
- Helbling, M. (2011). WHY SWISS-GERMANS DISLIKE GERMANS. *European Societies*, 13(1), 5-27. doi:10.1080/14616696.2010.533784
- Helbling, M., & Kriesi, H. (2014). Why Citizens Prefer High- Over Low-Skilled Immigrants. Labor Market Competition, Welfare State, and Deservingness. *European Sociological Review*, 30(5), 595-614. doi:10.1093/esr/jcu061
- Helland, H., & Støren, L. A. (2006). Vocational Education and the Allocation of Apprenticeships: Equal Chances for Applicants Regardless of Immigrant Background? *European Sociological Review*, 22(3), 339-351. doi:10.1093/esr/jci061
- Huddleston, T., Bilgili, Ö., Joki, A.-L., & Vankova, Z. (2015a). *Migrant Integration Policy Index 2015: Anti-discrimination*. Barcelona/Brussels: CIDOB and MPG.
- Huddleston, T., Bilgili, Ö., Joki, A.-L., & Vankova, Z. (2015b). *Migrant Integration Policy Index 2015: Switzerland*. Barcelona/Brussels: CIDOB and MPG.
- Humburg, M., & Van der Velden, R. (2015). Skills and the graduate recruitment process: Evidence from two discrete choice experiments. *Economics of Education Review*, 49, 24-41.
- Imdorf, C. (2007). *Lehrlingsselektion in KMU*. Retrieved from Freiburg: http://lehrlingsselektion.de/documents/selektion_d.pdf
- Imdorf, C. (2008). Migrantenjugendliche in der betrieblichen Ausbildungsplatzvergabe – auch ein Problem für Kommunen. In M. Bommes & M. Krüger-Potratz (Eds.), *Migrationsreport 2008: Fakten - Analysen - Perspektiven* (pp. 113-158). Frankfurt am Main: Campus Verlag.
- Jackson, M., & Cox, D. R. (2013). The Principles of Experimental Design and Their Application in Sociology. *Annual Review of Sociology*, 39(1), 27-49. doi:10.1146/annurev-soc-071811-145443
- Jacquemet, N., & Yannelis, C. (2012). Indiscriminate discrimination: A correspondence test for ethnic homophily in the Chicago labor market. *Labour Economics*, 19(6), 824-832. doi:<http://dx.doi.org/10.1016/j.labeco.2012.08.004>

- Kaas, L., & Manger, C. (2012). Ethnic Discrimination in Germany's Labour Market: A Field Experiment. *German Economic Review*, 13(1), 1-20. doi:10.1111/j.1468-0475.2011.00538.x
- Kälin, W., & Locher, R. (2015). *Der Zugang zur Justiz in Diskriminierungsfällen (Synthesebericht)*. Retrieved from Bern:
- Keuschnigg, M., & Wolbring, T. (2016). The Use of Field Experiments to Study Mechanisms of Discrimination *Analyse & Kritik* (Vol. 38, pp. 179).
- Koopmans, R., Veit, S., & Yemane, R. (2018). Ethnische Hierarchien in der Bewerberauswahl: Ein Feldexperiment zu den Ursachen von Arbeitsmarktdiskriminierung. *WZB Discussion Paper, SP VI 2018-104*.
- Krings, F., Johnston, C., Binggeli, S., & Maggiori, C. (2014). Selective incivility: Immigrant groups experience subtle workplace discrimination at different rates. *Cultural Diversity and Ethnic Minority Psychology*, 20(4), 491-498. doi:10.1037/a0035436
- Krings, F., & Olivares, J. (2007). At the doorstep to employment: Discrimination against immigrants as a function of applicant ethnicity, job type, and raters' prejudice. *International Journal of Psychology*, 42(6), 406-417. doi:10.1080/00207590701251721
- Kühn, K., Liebscher, D., & Klose, A. (2013). *Expertise zu den rechtlichen Aspekten sozialwissenschaftlicher Testing-Verfahren im Rahmen der vom SVR geplanten Studie "Diskriminierung am Ausbildungsmarkt?"*. Retrieved from http://svr-migration.dymagics.de/wp-content/uploads/2014/11/Expertise_fuer_Studie_Diskriminierung_am_Ausbildungsmarkt.pdf
- Laganà, F., Chevillard, J., & Gauthier, J.-A. (2014). Socio-economic Background and Early Post-compulsory Education Pathways: A Comparison between Natives and Second-generation Immigrants in Switzerland. *European Sociological Review*, 30(1), 18-34. doi:10.1093/esr/jct019
- Lahey, J., & Beasley, R. (2009). Computerizing audit studies. *Journal of Economic Behavior & Organization*, 70(3), 508-514. doi:<http://dx.doi.org/10.1016/j.jebo.2008.02.009>
- Lahey, J., & Beasley, R. (2018). Technical aspects of correspondence studies. In M. Gaddis (Ed.), *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 81-101): Springer.
- Lalive, R., & Lehmann, T. (2017). The labor market in Switzerland, 2000–2016. *IZA World of Labor*, 402-402.
- Lane, T. (2016). Discrimination in the laboratory: A meta-analysis of economics experiments. *European Economic Review*, 90, 375-402.
- LaPiere, R. T. (1934). Attitudes vs. Actions. *Social Forces*, 13(2), 230-237. doi:10.2307/2570339
- Larja, L., Warius, J., Sundbäck, L., Liebkind, K., Kandolin, I., & Jasinskaja-Lahti, I. (2012). *Discrimination in the Finnish Labour Market: An Overview and a Field Experiment on Recruitment*. Retrieved from <https://yhdenvertaisuus-fi-bin.directo.fi/@Bin/070c0305f220b492e9b19da85cb314b5/1508422136/application/pdf/469926/discrimination-in-the-finnish-labour-market.pdf>
- Lee, H.-A., & Khalid, M. A. (2016). Discrimination of high degrees: race and graduate hiring in Malaysia. *Journal of the Asia Pacific Economy*, 21(1), 53-76.
- Lee, T. L., & Fiske, S. T. (2006). Not an outgroup, not yet an ingroup: Immigrants in the stereotype content model. *International Journal of Intercultural Relations*, 30(6), 751-768.

- Levitt, S. D., & List, J. A. (2009). Field experiments in economics: the past, the present, and the future. *European Economic Review*, 53(1), 1-18.
- Leybold-Johnson, I. (2014, 28.01.2014). What's in a Swiss name? *Swissinfo.ch*. Retrieved from https://www.swissinfo.ch/eng/culture/schmid--favre--ferrari_what-s-in-a-swiss-name-/37706646
- Liebig, T., Kohls, S., & Krause, K. (2012). *The labour market integration of immigrants and their children in Switzerland*. Retrieved from Paris
- Liebkind, K., Larja, L., & Brylka, A. (2016). Ethnic and Gender Discrimination in Recruitment: Experimental Evidence From Finland. *Journal of Social and Political Psychology*, 4(1), 403-426. doi:10.5964/jspp.v4i1.433
- List, J. A. (2009). An introduction to field experiments in economics. *Journal of Economic Behavior & Organization*, 70(3), 439-442. doi:<http://dx.doi.org/10.1016/j.jebo.2008.10.013>
- List, J. A. (2011). Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off. *The Journal of Economic Perspectives*, 25(3), 3-15.
- Lodder, L., McFarland, S., & White, D. (2003). Racial Preference and Suburban Employment Opportunities. *Chicago: Legal Assistance Foundation of Chicago*.
- Longchamp, C., Imfeld, M., Tschöpe, S., Müller, M., Rochat, P., & Schwab, J. (2014). *Verbreitung und Entwicklung von Rassismus, Fremdenfeindlichkeit, Muslimfeindlichkeit und Judenfeindlichkeit - Schlussbericht zur Studie "Zusammenleben in der Schweiz 2010-2014"*. Retrieved from Bern: <http://www.gfsbern.ch/de-ch/Detail/zusammenleben-in-der-schweiz-2010-2014>
- Matser, C., van Oudenhoven, J. P., Askevis - Leherpeux, F., Florack, A., Hannover, B., & Rossier, J. (2010). Impact of relative size and language on the attitudes between nations and linguistic groups: The case of Switzerland. *Applied Psychology*, 59(1), 143-158.
- Matthey, F., & Steffanini, F. (2015). *Partie 6: Racisme - Analyse juridique*. Retrieved from Neuchâtel:
- McGinnity, F., & Lunn, P. D. (2011). Measuring discrimination facing ethnic minority job applicants: an Irish experiment. *Work, Employment & Society*, 25(4), 693-708.
- Mercat-Bruns, M. (2016). *Discrimination at Work: Comparing European, French, and American Law*. Oakland, California: University of California Press.
- Midtbøen, A. H. (2014). The Invisible Second Generation? Statistical Discrimination and Immigrant Stereotypes in Employment Processes in Norway. *Journal of Ethnic and Migration Studies*, 40(10), 1657-1675. doi:10.1080/1369183X.2013.847784
- Midtbøen, A. H. (2015a). Ethnic Penalties In Western Labour Markets. *Nordic Journal of Migration Research*, 5(4), 185.
- Midtbøen, A. H. (2015b). Ethnic penalties in western labour markets. *Nordic Journal of Migration Research*, 5(4), 185-193.
- Midtbøen, A. H., & Rogstad, J. (2012). Discrimination. Methodological controversies and sociological perspectives on future research. *Nordic Journal of Migration Research*, 2(3), 203-212.
- Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of political economy*, 66(4), 281-302.
- Modood, T., & Khattab, N. (2016). Explaining Ethnic Differences: Can Ethnic Minority Strategies Reduce the Effects of Ethnic Penalties? *Sociology*, 50(2), 231-246. doi:10.1177/0038038515575858

- Mühlemann, D. (2017). *Rassismusvorfälle aus der Beratungspraxis - Januar bis Dezember 2016*. Retrieved from Bern:
http://www.ekr.admin.ch/pdf/Rassismusbericht_2016_D_web.pdf
- Neumark, D. (2012). Detecting Discrimination in Audit and Correspondence Studies. *Journal of Human Resources*, 47(4), 1128-1157.
- Neumark, D. (2016). Experimental Research on Labor Market Discrimination. *NBER Working Paper No. 22022*.
- Neumark, D., Burn, I., & Button, P. (2015). Is it harder for older workers to find jobs? New and improved evidence from a field experiment. *National Bureau of Economic Research Working Paper*, 21669.
- Neumark, D., & Rich, J. (2018). Do field experiments on labor and housing markets overstate discrimination? A re-examination of the evidence. *ILR Review*.
- Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, 693-709.
- OECD. (2010). *Equal Opportunities? The Labour market Integration of Children of Immigrants*. Retrieved from Paris:
- OECD. (2012). *Jobs for Immigrants (Vol. 3): Labour Market Integration in Austria, Norway and Switzerland*. Retrieved from Paris:
<http://dx.doi.org/10.1787/9789264167537-en>
- OECD. (2013). *International Migration Outlook 2013*. Retrieved from
http://dx.doi.org/10.1787/migr_outlook-2013-en
- OECD. (2018a). Foreign-born unemployment (indicator). Retrieved from
<https://data.oecd.org/migration/foreign-born-unemployment.htm>
- OECD. (2018b). Native-born unemployment (indicator). Retrieved from
<https://data.oecd.org/migration/native-born-unemployment.htm> - indicator-chart
- OECD. (2018c). Unemployment rate (indicator). Retrieved from
<https://data.oecd.org/unemp/unemployment-rate.htm>
- 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), 148-171. doi:10.2307/41330445
- Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, 105(2), 171.
- Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance.[Comment/Reply]. *Journal of Personality and Social Psychology*, 108(4), 562-571. doi:doi: 10.1037/pspa0000023
- Pager, D. (2003). The Mark of a Criminal Record. *American Journal of Sociology*, 108(5), 937-975.
- Pager, D. (2007). The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future. *Annals of the American Academy of Political and Social Science*, 609, 104-133. doi:10.2307/25097877
- Pager, D. (2016). Are Firms That Discriminate More Likely to Go Out of Business? *Sociological Science*, 3, 849-859.
- Pager, D., Bonikowski, B., & Western, B. (2009). Discrimination in a Low-Wage Labor Market: A Field Experiment. *American Sociological Review*, 74(5), 777-799. doi:10.2307/27736094

- Pager, D., & Quillian, L. (2005). Walking the talk? What employers say versus what they do. *American Sociological Review*, 70(3), 355-380.
- Pager, D., & Western, B. (2012). Identifying discrimination at work: The use of field experiments. *Journal of Social Issues*, 68(2), 221-237.
- Pecoraro, M., & Ruedin, D. (2016). A foreigner who does not steal my job: The role of unemployment risk and values in attitudes toward equal opportunities. *International migration review*, 50(3), 628-666.
- Pecoraro, M., & Ruedin, D. (2017). *The Share of Foreigners in One's Occupation and Attitudes Towards Foreigners*. Retrieved from Neuchâtel: [https://www.unine.ch/files/live/sites/irene/files/shared/documents/Publications/Working papers/2017/WP17-06.pdf](https://www.unine.ch/files/live/sites/irene/files/shared/documents/Publications/Working%20papers/2017/WP17-06.pdf)
- Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review*, 62(4), 659-661.
- Pierné, G. (2018). Hiring discrimination, ethnic origin and employment status. *International Journal of Manpower*, 39(1), 152-165.
- Probst, J. (2015). *Teilstudie 7: Rassismus - Sozialwissenschaftliche Erhebungen*. Retrieved from Neuchâtel:
- Protsch, P., & Solga, H. (2017). Going across Europe for an apprenticeship? A factorial survey experiment on employers' hiring preferences in Germany. *Journal of European Social Policy*, 27(4), 387-399. doi:10.1177/0958928717719200
- Quillian, L. (2006). New Approaches to Understanding Racial Prejudice and Discrimination. *Annual Review of Sociology*, 32, 299-328. doi:10.2307/29737741
- Quillian, L., Pager, D., Hexel, O., & Midtbøen, A. H. (2017). Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. *Proceedings of the National Academy of Sciences*, 114(41), 10870-10875. doi:10.1073/pnas.1706255114
- Rapp, C. (2015). More diversity, less tolerance? The effect of type of cultural diversity on the erosion of tolerance in Swiss municipalities. *Ethnic and Racial Studies*, 38(10), 1779-1797. doi:10.1080/01419870.2015.1015582
- Raymann, U. (2003). *Meinungen und Einstellungen gegenüber AusländerInnen in der Schweiz*. Retrieved from Zürich:
- Reskin, B. (2012). The race discrimination system. *Annual Review of Sociology*, 38, 17-35.
- Riach, P. A., & Rich, J. (2002). Field Experiments of Discrimination in the Market Place. *The Economic Journal*, 112(483), F480-F518. doi:10.2307/798458
- Riach, P. A., & Rich, J. (2004a). Deceptive field experiments of discrimination: are they ethical? *Kyklos*, 57(3), 457-470.
- Riach, P. A., & Rich, J. (2004b). Fishing for Discrimination. *Review of Social Economy*, 62(4), 465-486. doi:10.1080/0034676042000296227
- Rich, J. (2014a). What Do Field Experiments of Discrimination in Markets Tell Us? A Meta Analysis of Studies Conducted since 2000. *IZA Discussion Paper*(8584).
- Rich, J. (2014b). *What Do Field Experiments of Discrimination in Markets Tell Us? A Meta Analysis of Studies Conducted since 2000*. Retrieved from Bonn:
- Rich, J. (2018). Do photos help or hinder field experiments of discrimination? *International Journal of Manpower*, 39(4), 502-518.
- Rooth, D.-O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labor economics*, 17, 523-534.

- Rorive, I. (2009). *Proving Discrimination Cases.: The Role of Situation Testing*. Bruxelles and Stockholm: Centre for Equal Rights.
- Ross, S. L. (2017). Measuring trends in discrimination with field experiment data. *Proceedings of the National Academy of Sciences*, 114(41), 10815-10817.
- Ruedin, D., D'Amato, G., Wichmann, N., & Pecoraro, M. (2013). *Migrationsängste der Schweizer Bevölkerung*. Retrieved from Neuchâtel:
- Schneider, J., Yemane, R., & Weinmann, M. (2014). *Diskriminierung am Ausbildungsmarkt: Ausmass, Ursachen und Handlungsperspektiven* Retrieved from Berlin
- Seibert, H., Hupka-Brunner, S., & Imdorf, C. (2009). Wie Ausbildungssysteme Chancen verteilen - Berufsbildungschancen und ethnische Herkunft in Deutschland und der Schweiz unter Berücksichtigung des regionalen Verhältnisses von betrieblichen und schulischen Ausbildungen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 61(4), 595-620.
- Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge, MA: Cambridge University Press.
- Sidanius, J., Pratto, F., Van Laar, C., & Levin, S. (2004). Social dominance theory: Its agenda and method. *Political Psychology*, 25(6), 845-880.
- Skinner, B. (2015, 05.01.2015). Alles Müller. *Blog*. Retrieved from <https://blog.tagesanzeiger.ch/datenblog/index.php/6859/alles-mueller-oder-was>
- Teele, D. L. (2014a). Reflections on the ethics of field experiments. In D. L. Teele (Ed.), *Field experiments and their critics: Essays on the uses and abuses of experimentation in the social sciences*. New Haven: Yale University Press.
- Teele, D. L. (Ed.) (2014b). *Field experiments and their critics: essays on the uses and abuses of experimentation in the social sciences*. New Haven: Yale University Press.
- Thijssen, L. (2016). *Taste-based versus Statistical Discrimination: Placing the Debate into Context*. Retrieved from <http://gemm2020.eu/wp-content/uploads/2016/12/GEMM-Report-Review-taste-vs-statistical-discrimination.pdf>
- International Convention on the Elimination of All Forms of Racial Discrimination, (1965).
- van der Plancke, V. (2007). Les tribulations du testing en Belgique: quels enseignements? *Horizons stratégiques*(3), 40-59.
- van Oudenhoven, J. P., Selenko, E., & Otten, S. (2010). Effects of country size and language similarity on international attitudes: A six - nation study. *International Journal of Psychology*, 45(1), 48-55.
- Veenman, J. (2010). Measuring labor market discrimination: An overview of methods and their characteristics. *American behavioral scientist*, 53(12), 1806-1823.
- Vidal - Coso, E., & Ortega - Rivera, E. (2017). Skill Composition and Occupational Incorporation of Early and Recent Immigrants in Switzerland: the case of Italians and Spaniards. *International Migration*, 55(S1), 86-111.
- Vuolo, M., Uggen, C., & Lageson, S. (2016). Statistical power in experimental audit studies: Cautions and calculations for matched tests with nominal outcomes. *Sociological Methods & Research*, 45(2), 260-303.
- Vuolo, M., Uggen, C., & Lageson, S. (2018). To match or not to match? Statistical and substantive considerations in audit design and analysis *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 119-140): Springer.

- 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.
- Weichselbaumer, D. (2016a). Discrimination against Female Migrants Wearing Headscarves. *IZA Discussion Paper Series*(No. 10217).
- Weichselbaumer, D. (2016b). Discrimination Against Migrant Job Applicants in Austria: An Experimental Study. *German Economic Review*, 18(2), 237-265. doi:10.1111/geer.12104
- Wessendorf, S. (2008). Culturalist discourses on inclusion and exclusion: the Swiss citizenship debate 1. *Social Anthropology*, 16(2), 187-202.
- Widner, D., & Chicoine, S. (2011). It's All in the Name: Employment Discrimination Against Arab Americans. *Sociological Forum*, 26(4), 806-823. doi:10.1111/j.1573-7861.2011.01285.x
- Wikipedia. (2017). List of most common surnames in Europe. Retrieved from https://en.wikipedia.org/wiki/List_of_most_common_surnames_in_Europe
- Wood, M., Hales, J., Purdon, S., Sejersen, T., & Hayllar, O. (2009). *A Test for Racial Discrimination in Recruitment Practice in British Cities*. Retrieved from Norwich: <http://webarchive.nationalarchives.gov.uk/20130314010347/http://research.dwp.gov.uk/asd/asd5/rrs-index.asp>
- Wrench, J. (2011). Data on discrimination in EU countries: statistics, research and the drive for comparability. *Ethnic and Racial Studies*, 34(10), 1715-1730.
- Wrench, J. (2016). Discrimination against immigrants in the labour market: An overview and a typology. In A. Triandafyllidou (Ed.), *Routeledge Handbook of Immigration and Refugee Studies* (pp. 118-124). London: Routledge.
- Yahoo. (2012). Yahoo Terms of Service (Last updated March 16, 2012). Retrieved from <https://policies.yahoo.com/us/en/yahoo/terms/utos/>
- Zschirnt, E. (2016). Measuring Hiring Discrimination – A History of Field Experiments in Discrimination Research. *NCCR Working Papers*, 7, 1-32.
- Zschirnt, E., & Ruedin, D. (2016). Ethnic Discrimination in Hiring Decisions: A Meta-Analysis of Correspondence Tests 1990–2015. *Journal of Ethnic and Migration Studies*, 42(7), 1115-1134. doi:10.1080/1369183X.2015.1133279

Paper I:

Measuring Hiring Discrimination – A History of Field Experiments on Discrimination

First Published as NCCR Working Paper:

Zschirnt, E. (2016) "Measuring Hiring Discrimination – A History of Field Experiments in Discrimination Research". NCCR Working Papers, 7, 1-32.

Revised version

Measuring Hiring Discrimination – A History of Field Experiments on Discrimination

Eva Zschirnt
University of Neuchâtel and NCCR “on the move”
eva.zschirnt@unine.ch
Tel. +41 32 718 39 49

University of Neuchâtel
Swiss Forum for Migration and Population Studies SFM
Rue A.L. Breguet 2
2000 Neuchâtel, Switzerland

ORCID iD: 0000-0002-8918-4146

Abstract

Ethnic and racial discrimination in the labour market is a common and documented problem. Scientists from different backgrounds and numerous countries have tried to measure the extent of this form of discrimination, often by using field experiments at the hiring stage. This paper provides an overview of the literature on measuring discrimination with field experiments. It focuses on methodological issues such as the difference between in-person and written tests, before addressing the historical political context in which field experiments have emerged and how the technique was developed further over time. It shows that today’s field experiments not only cover a wider group of countries, professions or disadvantaged groups, but also increasingly add more variables to the testing. Despite this variety in the research designs, certain trends can be observed in all experiments and hiring discrimination can be found in all countries where field experiments were conducted.

Keywords

Ethnic discrimination, hiring, measuring discrimination, field experiments, methodology, labour market

Acknowledgments

This work was supported by the nccr – on the move, which is financed by the Swiss National Science Foundation. I would like to thank Rosita Fibbi, Pieter Bevelander, Flavia Fossati and Didier Ruedin for comments on earlier versions of this text

1. Introduction

Due to increased mobility, modern Western societies have diversified considerably. As the labour market holds a key position in the integration of migrants and their children, discrimination in this social sphere has far reaching consequences. While wage disparities between natives, immigrants and the second generation can be quantified using available economic data, measuring discrimination in hiring is more difficult. Since the late 1960s researchers have used field experiments in which two candidates with interchangeable qualifications that differ only in the characteristic to be analysed apply for the same job. While a growing amount of literature exists which reviews the discrimination of disadvantaged groups in their access to credit, products, or accommodation (e.g. Baert, 2018; Bertrand & Duflo, 2017; Gaddis, 2018a; Riach & Rich, 2002; Rich, 2014), this article focuses on field experiments that address the discrimination of racial or ethnic minorities in the hiring process¹.

Concentrating on racial or ethnic discrimination in hiring and the use of field experiments, three grand strands of literature can be identified. The first is methodological as it addresses the use of field experiments and their development in measuring discrimination in hiring. It focuses on the differences between in-person audit tests and written CV-based correspondence tests, discusses their merits and limitations, and addresses ethical issues regarding testing situations. The second strand of literature is result-focused and consists of the in-person audit and correspondence studies. A wide range of studies has been published with great variations in survey design. While early studies focused on showing that ethnic or racial discrimination in hiring exists, more recent studies use field experiments as a basis but then add new variables or combine these findings with further research on the underlying reasons of discrimination. Thus, a third strand of literature has emerged more recently which tries to address the limits of current testing methods by offering a combination of mixed and interdisciplinary methods to study not only *whether* but also *why* discrimination exists in today's hiring decisions.

This paper starts by defining discrimination and the historical development of field experiments, before addressing the methodological literature on testing and focusing on the studies already published. It turns towards recent developments and trends observed in field experiments on hiring discrimination.

¹ The working paper version of this article was published in 2016, thus before Gaddis' book Gaddis (2018b) including his introductory chapter (Gaddis, 2018a) that also has a historical focus was published.

2. Ethnic and Racial Discrimination: Definition and Delimitation of the Field

Studies of discrimination have been conducted on numerous discrimination grounds, the two most frequent being discrimination on the ground of gender or on the ground of race or ethnicity. However, the distinction between race and ethnicity is not clear cut (Blank, Dabady, & Citro, 2004). It can be observed that the US debate focuses on racial minorities (e.g. African-Americans or Hispanics), while Canadians speak of visible minorities², and European countries (except for the UK) have focused on ethnic minorities. No matter which term is used, both race and ethnicity are prohibited grounds of discrimination. Looking at ethnic or racial discrimination, several definitions of discrimination exist, depending on the scholarly background.

Based on Arrow (1998), Bendick summarises that “Economists define employment discrimination as valuation in the labor market of workers’ characteristics which are not related to the workers’ productivity” (2007, p. 18). Furthermore, economists distinguish between discrimination at the individual and at the group level. Looking at the context of ethnic discrimination in hiring, discrimination is defined as “a causal effect defined by a hypothetical *ceteris paribus* conceptual experiment – varying race but keeping all else constant” (Heckman, 1998, p. 102).

Within the social sciences definitions of discrimination range from a very broad understanding of discrimination, which “take[s] all inequality among racial groups as discrimination, assuming all inequality that exists among groups must be the result of current or past discriminatory practices” (Quillian, 2006, p. 300), to a more narrow definition, where discrimination is understood “only [as] acts that are intended to harm the target group” (p.300). According to Quillian, most definitions of discrimination used by social scientists fall within the scope of the US National Research Council’s definition that discrimination is understood as “(1) differential treatment on the basis of race that disadvantages a racial group and (2) treatment on the basis of inadequately justified factors other than race that disadvantages a racial group” (Blank et al., 2004, p. 39).

Legal definitions of discrimination can be found in national and international anti-discrimination legislation, such as the EU’s Directive 2000/43/EC, better known as the “Race Directive” implementing the principle of equal treatment between persons irrespective of racial or ethnic origin. In the Race Directive the concept of discrimination is defined in Article 3 (2):

² According to Statistics Canada, visible minorities are non-Caucasian in race or non-white in colour and are not Aboriginal.

- (a) Direct discrimination shall be taken to occur where one person is treated less favourably than another is, has been or would be treated in a comparable situation on grounds of racial or ethnic origin;
- (b) Indirect discrimination shall be taken to occur where an apparently neutral provision, criterion or practice would put persons of a racial or ethnic origin at a particular disadvantage compared with other persons, unless that provision, criterion or practice is objectively justified by a legitimate aim and the means of achieving that aim are appropriate and necessary.

Despite the differences between the economic, sociological and legal definitions, what they all have in common is the notion that one person is treated less favourably than another because of a certain characteristic that cannot be changed and that this differential treatment is not based on any other differences that influence productivity.

While discrimination can occur in almost every interaction between individuals, it has been mostly studied in relation to the housing, labour and product markets³. This paper addresses ethnic discrimination in the labour market and more specifically ethnic discrimination in the hiring process. Hiring is only a small part of the labour market, yet the focus on this specific point is important as it can prevent candidates from entering the labour market, and most discrimination has been found to occur at the initial hiring stage (Riach & Rich, 1991). Furthermore, discrimination is not a phenomenon that only happens at one point in time, but may become cumulative, starting with education and continuing in hiring, wages or promotion or termination decisions. Thus small incidences of discrimination, for example in the hiring process, can have “substantial effects on aggregate outcomes” (Pager, Bonikowski, & Western, 2009, p. 778).

3. Measuring Discrimination with Field Experiments

While discrimination used to be a blatantly obvious phenomenon, in particularly in the US, it has since become a more subtle phenomenon which is not as easily observed as during the middle of the 20th century. Researchers are confronted with questions such as: “How can we measure discrimination when it is an often illegal and hidden practice?” (Quillian, 2006, p. 299), “What is the actual *extent* of discrimination in different spheres of social life? And what are the *causes* of discriminatory treatment?” (emphasis in the original: Midtbøen & Rogstad, 2012, p. 203).

³ Product markets include for example access to mortgages, markets for (used) cars or insurances. For a comprehensive overview see e.g. Riach and Rich (2002), Rich (2014), or Bertrand and Duflo (2017).

Academics from different disciplines have addressed the issue of discrimination using a variety of methods, such as statistical analyses usually focused on employment or wages in economics, analyses of court proceedings or complaints in law, and surveys (e.g. victim surveys, surveys on perceived or observed discrimination, or attitude surveys) and experimental methods in the social sciences⁴.

Faced with the limitations of existing research designs, field experiments, both in-person audit studies and written correspondence tests, have become increasingly popular. Using these field experiments allows the researcher to measure the effect of ethnicity or race in the application process and to draw statistically significant results on the extent of discrimination in the labour market (Midtbøen & Rogstad, 2012; Pager, 2007; Quillian, 2006). In almost 50 years, field experiments have become an important means to quantify the extent of discrimination in several countries, at different points in time and for numerous minority groups. While several comprehensive reviews (Bertrand & Duflo, 2017; Pager, 2007; Pager & Shepherd, 2008; Riach & Rich, 2002; Rich, 2014) have been published of studies already carried out in the labour, product and housing market on discrimination based on ethnicity or race, sex, disability, age or sexual orientation, these reviews stop short of further analysing the data they compiled from the individual studies. Looking at correspondence tests on ethnic and racial discrimination in hiring decisions, this gap has been addressed in a meta-analysis by Zschirnt and Ruedin (2016) which systematically analysed 43 correspondence tests that were conducted in OECD countries between 1990 and 2015 as well as in a meta-analysis on field experiments conducted in the US by Quillian, Pager, Hexel, and Midtbøen (2017).

Types of Field Experiments: In-Person Audit Studies and Correspondence Tests

Field experiments are based on the idea that two applicants who are as closely matched as possible regarding their qualification and presentation and only differ in the characteristic to be studied apply to the same vacancies. The results of the application process are carefully recorded thus enabling the researcher to observe actual hiring decisions (Midtbøen & Rogstad, 2012, p. 205). They vary however in how employers are contacted, either in person or in writing and each method has its advantages and disadvantages.

⁴ Veenman (2010) provides a good overview of these methods and their limitations.

In-person Audit Studies

In-person audit studies use carefully matched and trained testers who apply in-person either directly at the business or by phone. There are several advantages of conducting in-person audits. First, race or ethnicity is easily signalled by the physical characteristics of the applicants (Pager, 2007, p. 111). Second, in-person audits allow testing for ethnic discrimination in low qualified or entry-level jobs where written applications are not common. Third, in-person audits may cover the whole application process since candidates are able to attend job interviews. And finally, by attending interviews, in-person audits enable researchers to also collect qualitative data on how both applicants were treated during the interview, thus documenting cases of “equal but different treatment” (Bovenkerk, Gras, Ramsoedh, Dankoor, & Havelaar, 1995, p. 20; Midtbøen & Rogstad, 2012). Pager shows how durations of interviews differ or how minority applicants are channelled into lower paying jobs than those initially applied for, while the opposite might occur for majority candidates – yet both would be counted as job offers (cf. Pager et al., 2009, p. 787).

However, critics of in-person audits are quite vocal in listing the problems inherent with this approach: It is time and resource consuming and requires extensive supervision of the testers. Focusing on in-person audits conducted in the US, Heckman and Siegelman (1993) identify three main limitations with this test design. First, the small number of tests carried out and the limited sample of occupations tested do not allow for generalisation of the results as the studies are not representative for the whole labour market. Second, leaving out the cases in which both applicants were rejected when calculating the net discrimination rate distorts the results significantly, an argument which not only applies to in-person audit studies, but also to correspondence tests discussed below⁵. Third, unobservable variables in the selection of candidates may impact the selection procedure. Furthermore, there is a danger of experimenter effects, since the testers might influence the outcome (Heckman, 1998). Finally, by presenting two almost identical candidates, employers “may be forced to privilege relatively minor characteristics simply out of necessity of breaking the tie” (Pager, 2007, p. 116). Thus, summarising Heckman’s and Siegelman’s arguments, different degrees of success in the hiring process might be attributed to the “failure by the researchers to match the testers on some subtle productivity-related characteristics” (Bendick & Nunes, 2012, p. 248). A detailed discussion of Heckman’s criticism and an approach to analyse the robustness of results found in field experiments for discrimination and the impact that

⁵ For a detailed discussion on the calculation of discrimination rates see Riach and Rich, 2002 and in particular pp. F486.

unobservables might have on these results can be found in Neumark (2012). The test that Neumark proposes in this paper has subsequently been applied by using data from previously published correspondence tests (M. Carlsson, Fumarco, & Rooth, 2014; Neumark & Rich, 2018). Neumark and Rich show that results in field experiments on the labour market are much less robust than those found on the housing market. They therefore caution researchers to take the variation in resume quality needed to apply the Neumark test to correspondence test results into consideration when designing field experiments.

Written correspondence studies

Correspondence studies address several of these key concerns by giving the researcher complete control over the content of the fictitious written applications. By forgoing actual testers, the process is much less time and resource-intensive and allows to apply to a greater number of vacancies. Furthermore, the opportunity to apply for a wider range of jobs and the possibility to assign ethnically distinct names randomly to the applications increases the representativeness of the studies and meets most of Heckman's and Siegelman's concerns (Midtbøen & Rogstad, 2012, p. 207).

Still, there are also limitations to this approach. Most importantly discrimination is only measured in the first phase of the hiring process, that is, the response to written applications. Yet, "it does highlight one quite decisive form of discrimination – that of denying the applicant the chance to even compete for a job" (Riach & Rich, 1991, p. 241) and the findings of the ILO studies confirm that the majority of discrimination (more than 80%) happens in the first stage of hiring (Riach & Rich, 2002; Rich, 2014). Second, all information about race or ethnicity has to be conveyed by the name, in some cases memberships in specific organisations, or, where this is common, by attaching a picture to the application. Regarding the names, a problem arises, as many names not only signal race or ethnicity but have socio-economic connotations, "thus confounding the effects of race and class" (Pager, 2007, p. 111), an issue that has recently garnered more attention particularly in the US (Crabtree & Chykina, 2018; Gaddis, 2017a, 2017b). Finally, correspondence tests are typically reserved for occupations in which applications are accepted in writing, thus excluding those entry-level or unskilled jobs where applications are usually made in-person.

As the planning of a field experiments, either in-person or written, is challenging, the design of field experiments and the implications of decisions taken by research teams have increasingly received

the attention of researchers. These papers focus amongst others on the ethical implications of conducting audit studies (Banton, 1997; Riach & Rich, 2004a; Zschirnt, 2016), the signalling value of names (Crabtree & Chykina, 2018; Gaddis, 2017a, 2017b), the statistical power of experimental audit studies (Vuolo, Uggen, & Lageson, 2016), or on computerising audit studies and technical aspects (Lahey & Beasley, 2009, 2018).

While both in-person audits as well as correspondence tests have some limitations, they are, so far, the best way to measure the existence of discrimination in hiring (e.g. Schneider, Yemane, & Weinmann, 2014, p. 14). One of the greatest advantages of field experiments is that this “innovative research technique of matched-pair testing offers laboratory-like controlled conditions in quasi-experiments in real-world hiring situations” (Bendick & Nunes, 2012, p. 238). Interestingly, in-person audits are hardly employed in Europe. While in the US the use of in-person audit tests is still frequent, especially in the low-qualified sector, correspondence tests are also growing in importance, the first example being the often-cited study by Bertrand and Mullainathan (2004).

4. Testing for Discrimination: 50 Years of Audit and Correspondence Studies

4.1 Early Developments of Discrimination Research

Before the rise of social rights movements, discrimination was overtly practiced in many countries. In the US, discrimination “existed in perfectly open form, with no need for subtle economic analysis” (Arrow, 1998, p. 92). Furthermore, segregation and discrimination were also promoted by the law. Thus, the “fact of discrimination would not have needed testing” since “the presence of racial discrimination throughout American society was [...] a fact ‘too evident for detection and too gross for aggravation’” (p. 92). In the UK racial discourses gained prominence in the political debate of the 1940s and 1950s (Small & Solomos, 2006, p. 239). It took the development of a notion of equality irrespective of a racial or ethnic background to prompt legislation in this field. Therefore, it is not surprising that racial discrimination was first studied in the UK and the US, where in the beginning these studies mostly monitored compliance with new anti-discrimination legislation. It is, however, remarkable that despite a lack of contact between US and UK researchers the methodology of field experiments was developed independently on both sides of the Atlantic (Bovenkerk, 1992, p. 11). Yet differences persist in the methodological developments on both continents.

Field experiments on hiring discrimination first emerged in the UK in the 1960s. Rising Commonwealth immigration, border control, and racial issues became closely connected in a very politicized and racialized public debate (Karapin, 1999; Small & Solomos, 2006). Following the adoption of stricter immigration controls by the Conservatives (Bleich, 2003), it was the Labour Party that adopted the first Race Relations Act in 1965, prohibiting discrimination in public places. Due to inherent weaknesses, the Act was extended in 1968 to specifically address the issue of employment. Given the close link between immigration and race relation policies “Labour’s policy in the 1960s was that of the ‘balancing act’, stressing the mutual relationship between immigration controls on the one hand and measures to tackle racial discrimination on the other” (Sooben, 1990, p. 8). In the 1970s the focus changed and the Race Relation Act of 1976, which again addressed shortcomings of the previous versions, was adopted with almost no opposition from the Conservatives. This “indicat[ed] a new approach to race relations, equal opportunities and tackling discrimination” (Sooben, 1990, p. 6). The whole Part II focused on Discrimination in the Employment Field and established the Commission for Racial Equality that was also given the power to investigate and prosecute.

With the emergence of anti-discrimination legislation in the 1960s UK researchers began to study discrimination. Starting with the Political and Economic Planning (PEP) research office (Daniel, 1968), Jowell and Prescott-Clarke (1970), McIntosh and Smith (1974), Hubbock and Carter (1980), Firth (1981) and Brown and Gay (1985) documented discrimination against ethnic minority job seekers and advanced the methodology from purely in-person audits to written correspondence tests, or, if applicable to field experiments combining in-person audits and correspondence tests. All of these early studies in the UK showed discrimination against minority candidates, despite the entry into force and strengthening of anti-discrimination legislation. Brown and Gay explain the continued presence of discrimination in the labour market by the low risk of detection for discriminatory employers claiming that “the heart of the problem is that employers know that cases rarely get as far as legal action because the victim is very unlikely to be aware that he or she has been discriminated against” (1985, p. 32).

Following the British research, similar studies using this methodology were also conducted in other countries. In the Netherlands (Bovenkerk, 1977; Bovenkerk & Breuning-van Leeuwen, 1978), France (Raveau and Kilbourne in Bovenkerk, Kilborne, Raveau, and Smith (1979)), Canada (Ginsberg & Henry, 1985; Henry, 1989) and Australia (Riach & Rich, 1991), researchers carried out field experiments showing significant levels of discrimination against minority candidates. As

Riach and Rich concluded for their Australian study, “what is quite remarkable and disturbing is the similarity between our results and those of British researchers testing for discrimination against ‘non-white’ job applicants”(1991, p. 247). Even if national legislation prohibited discrimination, discrimination still existed in the labour markets. Similar to Brown and Gay, Riach and Rich argue that “even where an applicant suspects he/she has been the victim of discrimination, current labour market practices make it extremely difficult to present a *prima facie* case to the courts” (Riach & Rich, 1991, p. 256).

In the US before the 1960s common, open, and daily discrimination was the norm and it was even enforced by state legislation, the so-called Jim Crow laws. It was only with the Civil Rights movement of the 1960s that “many blatant discriminatory practices were prohibited, and whites increasingly repudiated discrimination and overt forms of prejudice” (Quillian, 2006, p. 299). Several landmark acts of US Civil Rights legislation were adopted, such as the Civil Rights Act of 1964, the Voting Rights Act of 1965, or the Civil Rights Act of 1968. Title VIII of the 1968 Civil Rights Act soon became known as the Fair Housing Act as it extended previous provisions and addressed housing discrimination and included provisions on the enforcement of these rights. While overt discrimination declined sharply after the introduction of these acts, great racial disparities could still be observed, especially in housing, employment, or in the criminal justice system.

Given the prominent position of the Fair Housing Act and the problems with racial segregation encountered in everyday life, researchers in the US started to use field experiments to study housing discrimination and conducted in-person audit studies (for overviews see Blank et al., 2004; Cherry & Bendick, 2018; Fix & Struyk, 1993; Gaddis, 2018a). Evidence from these in-person audit tests on the application of fair housing laws was even admitted in courts. At the end of the 1970s Newman (1978) first applied situation tests to the US labour market. However, contrary to the situation in the UK, studies on employment discrimination were not very common in the US of the 1980s. As Bendick points out, with the election of President Reagan in 1981 “political support for vigorous action against employment discrimination began to falter [...] and] under his administration, enforcement of anti-discrimination laws by the federal Equal Employment Opportunity Commission dramatically weakened” (2007, p. 19). Field experiments were then used by advocates of further anti-discrimination policies to “generat[e] additional knowledge through research, influenc[e] public opinion through dramatic findings, and provid[e] evidence for enforcement litigation” (Bendick, 1996, p. 10).

Summarising these early experiences with field experiments, it is interesting to note that in the UK and the countries influenced by the British approach the focus quickly turned to testing discrimination in employment predominantly by using correspondence testing (sometimes combined with smaller in-person audit studies), while US researchers rather looked at the issue of discrimination in housing and favoured the in-person test design.

4.2. The Establishment of a Systematic Comparative Approach to Testing: the ILO and Urban Institute in the 1990s

Following the early studies on discrimination, the 1990s saw an increase in field experiments with systematic approaches. Numerous studies with a comparative focus were conducted under the auspices of the International Labour Office (ILO) in the European context and the Urban Institute (UI) based in Washington, DC, for the US. Thus, by the end of the 1990s a considerable body of literature existed on labour discrimination on the ground of ethnicity and race in Europe and the US⁶.

The ILO studies were based on the methodology developed in Bovenkerk's manual on "Testing Discrimination in Natural Experiments" (1992). He clearly outlines the testing procedure to be followed to ensure comparability of the results. Following these guidelines, tests were conducted in Germany (Goldberg, Mourinho, & Kulke, 1995), the Netherlands (Bovenkerk et al., 1995), Spain (de Prada, Actis, Pereda, & Molina, 1995), and Belgium (Arriijn, Feld, & Nayer, 1998). A Swedish contribution was rejected by the Swedish Social Research Council on ethical grounds⁷. The US contribution to the project (Bendick, 1996) consisted of a discussion of tests conducted by the Urban Institute which followed a similar research design. Finally, the Italian (Allasino, Reyneri, Venturini, & Zincone, 2006), French (Cediey & Foroni, 2008) and – eventually – the Swedish (Attström, 2007) studies were conducted more than a decade after the start of the project.

The Urban Institute research, that was subsequently used as the basis for the US contribution to the ILO project, was already conducted around 1990, using multiple pairs of applicants to test discrimination against Hispanics (in San Diego and Chicago: Cross, Kenney, Mell, & Zimmermann, 1990) and blacks (in Washington DC and Chicago: Turner, Fix, & Struyk, 1991). In addition to these Urban Institute studies, Bendick (1996) also includes two of his studies for the Fair

⁶ For an overview on field experiments until 2000 see Riach and Rich (2002)

⁷ See Banton (1997) for details.

Employment Council, which also measured discrimination against blacks and Hispanics in Washington, DC (Bendick, Jackson, & Reinoso, 1994; Bendick, Jackson, Reinoso, & Hodges, 1991). These tests focused on the in-person audit approach by using matched testers who continued as far in the application process as possible, even if their matched partner had already been rejected, thus differing from the ILO approach. As Bovenkerk points out, it is interesting to note that despite the lack of contact between European and US researchers, Cross et al. (1990) took almost the same methodological decisions as European scholars (1992, p. 11).

Taking into account all studies mentioned since the first testing was conducted by the PEP in the UK until the end of the 1990s, a similar trend can be observed in all countries. As Riach and Rich put it:

“The results of the racial discrimination tests have extended over a period of thirty years and nine countries, in Europe, North America and the Pacific; all are members of the OECD. The minority groups include black, Asian, Arab, Turkish and other white non-nationals. The extent of discrimination varies temporally, spatially and between the various minority groups.” (Riach & Rich, 2002, p. F499)

Yet, despite these variations, all studies showed that discrimination against the tested minority occurred at statistically significant rates (Riach & Rich, 1991, p. F499).

4.3 Recent Developments in Field Experiments

After the wave of studies completed in the 1990s, the topic of discrimination received renewed attention from academics. After 2000 the number of studies in the US and Europe increased quickly. Particularly in Europe, where the European Union had adopted anti-discrimination directives which EU member states had to transpose into national law, correspondence tests were carried out in many EU countries, covering almost all countries of Western Europe. In the 2000s the last three ILO studies on discrimination in the labour market were published for Italy (Allasino et al., 2006), France (Cediey & Foroni, 2008) and Sweden (Attström, 2007). Once the Swedish ethical approval to join the ILO project was granted, numerous correspondence tests were conducted in Sweden, Norway, and Finland, where researchers had previously assumed that their research would not be approved either (Larja et al., 2012; Liebkind, Larja, & Brylka, 2016; Midtbøen, 2013, p. 52). Next to the Scandinavian countries, studies were, amongst others, implemented in Austria (Weichselbaumer, 2016b), Belgium (e.g. Baert, Cockx, Gheyle, & Vandamme, 2015), France (Duguet, Leandri, L'Horty, & Petit, 2010), Germany (Kaas & Manger, 2012; Schneider et al., 2014), Greece (Drydakis & Vlassis, 2010), Ireland (McGinnity & Lunn, 2011), the Netherlands (Andriessen, Nievers, Dagevos, & Faulk, 2012), Switzerland (Fibbi, Lerch, & Wanner, 2006), or

the UK (Wood, Hales, Purdon, Sejersen, & Hayllar, 2009). Despite the strong focus on North America, Europe and Australia, the methodology of testing for ethnic or racial discrimination was also employed in countries such as Chile (Bravo, Sanhueza, & Urzúa, 2008), China (Maurer-Fazio, 2012), Georgia (Asali, Pignatti, & Skhirtladze, 2017), India (Banerjee, Bertrand, Datta, & Mullainathan, 2009), Israel (Ariel et al., 2015), Malaysia (Lee & Khalid, 2016), or Mexico (Arceo-Gomez & Campos-Vazquez, 2014).

Field experiments have also addressed specific segments of the labour market. While Pager et al. (2009) focused on the low wage labour market, Bendick, Rodriguez, and Jayaraman (2010) tested for racial discrimination of waiters in up-scale restaurants. Two of the correspondence tests conducted in Germany also focus on very specific labour markets, i.e. the market for apprenticeships (Schneider et al., 2014) and the market for student internships (Kaas & Manger, 2012). University graduates have increasingly become the focus of correspondence studies, as can be observed in the US with studies conducted by Nunley, Pugh, Romero, and Seals (2015), Gaddis (2014), or Deming, Yuchtman, Abulafi, Goldin, and Katz (2016).

Moreover, field experiments have evolved from the original paired within subject design which varies only the dimension of race or ethnicity. Oreopoulos (2011) conducted one of the most comprehensive and complex field experiment on ethnic discrimination with almost 13,000 resumes on the Toronto labour market. By differentiating the fictitious CVs by name, place of education and place of work experience, he was able to show that Canadian employers discriminate by name and that Canadian work experience was more important than Canadian education. This experiment was repeated by Dechief and Oreopoulos (2012) with a greater regional scope, but overall similar results. Furthermore, field experiments no longer focused on only one minority group, but started to compare different ethnic or racial groups (e.g. Wood et al., 2009), and even added indigenous populations (Booth, Leigh, & Varganova, 2012).

Furthermore, research designs have become increasingly complex trying to go beyond showing the mere existence of discrimination in the hiring process, but attempting to narrow down underlying reasons. New variables that have been added next to race or ethnicity include further information included the CV, studying the intersection of gender and/or religion, or focusing on the context of the application. Specifically, researchers tested the impact of:

- varying the resume quality or residential area of an applicant (Bertrand & Mullainathan, 2004; Jacquemet & Yannelis, 2012),

- varying the resume quality, language skills and quality of the university (Lee & Khalid, 2016),
- indicating a criminal record on the CV (Pager, 2003),
- whether university degrees were obtained at an elite or less selective university (Gaddis, 2014), or from a public or for profit institution (Deming et al., 2016),
- including a reference letter from a previous employer (Kaas & Manger, 2012),
- including productivity variables for college graduates e.g. information on unemployment periods or internships (Nunley, Pugh, Romero, & Seals, 2014),
- signalling candidates status as first generation immigrant or second generation youth (M. Carlsson, 2010),
- adding foreign names⁸ whose origins were not easily recognised by Americans (Jacquemet & Yannelis, 2012),
- the intersection of race and gender (Darolia, Koedel, Martorell, Wilson, & Perez-Arce, 2016),
- whether applicants wore a head-scarf (in the US Ghumman and Ryan (2013), in Germany Weichselbaumer (2016a)),
- CV whitening (Kang, DeCelles, Tilcsik, & Jun, 2016),
- Facebook profile pictures (Baert, 2015), indicating work experience (Baert, Albanese, du Gardein, Ovaere, & Stappers, 2017), or volunteering (Baert & Vujčić, 2016)
- whether employers were in the public or private sector (e.g. Wood et al., 2009),
- submitting applications as CV or if application forms were used e.g. Wood et al. (2009),
- the gender of the contact person in the vacancy (M. Carlsson, 2010), or
- whether labour markets were tight or not (Baert et al., 2015).

Drydakis and Vlassis (2010) also complemented their Greek correspondence test with information on wages and insurance coverage offered. Blommaert, Coenders, and van Tubergen (2014) conducted a field experiment on hiring discrimination in the Netherlands, in which they posted resumes on online job search websites and measured response rates from employers, by counting the number of times the profile was viewed as well as the times candidates were contacted by potential employers. While this is not a classical correspondence test, the reported results of discrimination were rather similar.

⁸ Names were constructed by using Albanian, Armenian and Georgian first names, groups which are relatively unknown in the area of Chicago. These first names were combined with male last names from the same groups. Reasoning that these names cannot be attributed to a specific ethnic group, Jacquemet and Yannelis expected them to be just categorised as foreign and unfamiliar by employers (Jacquemet & Yannelis, 2012, p. 826).

As well as the added variables in the research design, mixed-methods and interdisciplinary approaches have gained importance in the last years. Midtbøen (2014) complements his Norwegian correspondence test with employer interviews, thus combining different research designs. In a more interdisciplinary approach that leans strongly on the discipline of psychology, other researchers focus on unconsciously working stereotypes, implicit attitudes and automatically activated associations focusing mainly on recruiters to study the reasons of discriminatory behaviour in depths. Examples of these studies include Bertrand, Chugh, and Mullainathan (2005) who study implicit prejudices and behaviour, Agerström and Rooth (2009) who use Implicit Association Tests (IAT) to test whether employers in Sweden discriminate automatically against Arab Muslim applicants, or Rooth (2010) who also focuses on Swedish recruiters and automatically activated associations that influence their behaviour and cause discriminatory treatment. However, more recent work on IATs has been more sceptical regarding the IAT's ability to predict discrimination (R. Carlsson & Agerström, 2016; Frederick L Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013; Frederick L.; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2015)

Recently research on the consequences of discriminatory behaviour on the employers has also emerged. Pager (2016) was able to show that employers who had been found to discriminate in her field experiment (Pager et al., 2009) were more likely to be out of business six years later. French research (Bon-Maury et al., 2016) also looked at the potential costs of discrimination to employers.

5. Discussion

Looking at these more recent studies that have evolved considerably since the 1960s, it is possible to observe certain trends in the recent field experiment literature. First, most recent studies chose written correspondence tests over the in-person audit tests. This is especially true in Europe. Apart from the ILO studies in the 1990s and the later studies in the 2000s that used the in-person audit approach in combination with correspondence tests, there is only one European study by Andriessen et al. (2012) in the Netherlands that uses an in-person audit approach with applications by telephone as a small addition to the larger correspondence test. While correspondence tests are also increasingly used in the US, several well-known in-person audit tests have still been conducted, especially for low-skilled or restaurant positions (Bendick et al., 2010; Ghumman & Ryan, 2013; Pager, 2003; Pager et al., 2009).

Second, almost all studies have a regional focus and are conducted in a specific region, the biggest cities, the biggest labour markets or in regions with a high percentage of immigrants. Exceptions are the studies conducted in Austria (Weichselbaumer, 2016b), Germany (Kaas & Manger, 2012; Schneider et al., 2014), the Netherlands (Andriessen et al., 2012) or Sweden (Agerström, Björklund, Carlsson, & Rooth, 2012) and some studies in the US (e.g. Widner & Chicoine, 2011). Still, despite their national focus, these studies are often heavily weighted towards certain regions.

Third, some trends in the choice of minority groups can be observed. Trends comparing different groups of nationalities can be witnessed especially in countries with a colonial history, such as the Netherlands (Bovenkerk et al., 1995) and the UK (Firth, 1981; Wood et al., 2009) or those with a tradition of guest-workers (e.g. Switzerland (Fibbi, Kaya, & Piguet, 2003) or Germany (Kaas & Manger, 2012; Schneider et al., 2014; Weichselbaumer, 2016a)). In recent European studies the groups most frequently chosen are Moroccans and Turks, or in more general terms immigrants with a Middle Eastern background. They often constitute the biggest ethnic minorities in the country and also, according to surveys, encounter the biggest amount of prejudice and negative experiences when looking for jobs (Bursell, 2007; M. Carlsson, 2010; M. Carlsson & Rooth, 2007; Kaas & Manger, 2012; Midtbøen, 2012; Schneider et al., 2014).

Fourth, testing has become more complex as more variables and ethnic or racial groups have been included in the experiments. While early studies – with the exception of Firth (1981) – usually focused on one group in one segment of the labour market, recent studies use several groups⁹, numerous qualifications and jobs¹⁰, address differences between private and public employers¹¹, or focus on gender differences for ethnic minority applicants¹².

Fifth, in European studies a focus on children of immigrants, the so-called second generation, has emerged. Fibbi et al. (2003) or Midtbøen (2012) studied these members of ethnic minority groups as they have been educated in their host country and are fluent in the language. However, differences in treatment still occur, which, all other things being equal, can be attributed to their ethnic names. M. Carlsson (2010) emphasises this point by comparing the differences in hiring for majority candidates, first-generation immigrants and second-generation youths, and finds almost no

⁹ e.g. Wood et al. (2009); McGinnity and Lunn (2011); Booth et al. (2012); Weichselbaumer (2016b)

¹⁰ e.g. Bursell (2007); M. Carlsson and Rooth (2007); Midtbøen (2012)

¹¹ e.g. Wood et al. (2009); Midtbøen (2012)

¹² e.g. Bursell (2007); Arai, Bursell, and Nekby (2015); Andriessen et al. (2012); Liebkind et al. (2016)

differences between the first and second generation. Thus, he concludes, that “the factor driving discrimination seems to be ethnicity *per se*” (M. Carlsson, 2010, p. 272).

Sixth, there is an emerging trend to combine field experiments with multi-method research that also studies the behaviour of employers. While early studies have already included some employer interviews (Firth, 1981), especially Scandinavian researchers have placed an emphasis on trying to explain discriminatory behaviour of employers. Researchers have included interviews or collected comprehensive data on the businesses and recruiters involved in the testing (M. Carlsson & Rooth, 2007; Midtbøen, 2013, 2014; Pager & Quillian, 2005). Furthermore, Agerström and Rooth (2009) combined the results from correspondence tests with Implicit Association Tests to examine the attitudes of HR decision makers. Thus, field experiments are not only becoming more complex in their research design on the actual testing conducted, but also encompass multi-method and interdisciplinary approaches to tackle the complex phenomenon of discrimination. However, by conducting multi-method and multisite research, these projects quickly become large international endeavors that require a lot of funding (e.g. the GEMM project), which might become a limitation to conducting such studies.

Sixth, it can be observed that all studies conducted on discrimination in the labour market, despite differences in survey designs, minority groups selected for the study, segment of the labour market, country or point in time, reported significant discrimination of the minority candidates. These results have been analysed more closely in a meta-analysis by Zschirnt and Ruedin (2016) or by Quillian et al. (2017) who focus on audit studies in the US. Both meta-analyses have shown that discrimination rates are persistent and change very little over time and place. They thus confirm the trends identified across the individual studies.

Seventh, field experiments on discrimination that have traditionally been designed using matched pairs or sets of resumes are more and more altered to designs using single applications, arguably to avoid detection of the experiments (e.g. Koopmans, Veit, & Yemane, 2018; Weichselbaumer, 2015, 2016a). However, critics of this approach argue that this change to the methodology means that it is no longer discrimination that is measured, but “only” preferential treatment in the labour market (Cherry & Bendick, 2018; Riach & Rich, 2004b).

Finally, researchers increasingly build variation in their resume design to allow for the use of the Neumark test by using the difference in resume quality to test the robustness of the results obtained.

These tests have also been conducted with existing data to test the reliability of their results, but these analyses have found mixed results. While approximately half of the labour market studies have robust results, in the other half findings of discrimination become insignificant, disappear or even change direction (Neumark & Rich, 2018). Thus, Neumark and Rich caution researchers to include variation in the resumes in the research design state to allow this control for the impact of unobservable characteristics.

6. Conclusion

Given the substantial body of literature that focuses on field experiments on racial or ethnic discrimination in the hiring process that have been conducted for fifty years and in numerous OECD countries, several trends can be observed. Although there is a great variation by ethnic or racial group chosen, the geographic location of the studies, the time, and the occupations tested, it is striking that the results found in all studies prove that discrimination is a problem that minority applicants have to face, despite the development of anti-discrimination legislation in these last 50 years. The persistently high rates of discrimination and the low number of official complaints or court cases dealing with the issue support the argument that discrimination in the hiring process has become much subtler, but still has significant effects. While there are differences between the countries where testing was conducted, minority applicants are 50 percent less likely to be invited for a job interview than applicants belonging to the majority, although variations between occupations, degree of qualification or sometimes between minorities exist (Zschirnt & Ruedin, 2016).

Going back to the first field experiment on discrimination in the UK labour market by Daniel (1968) shows that the problem remains very similar. In the introduction to Daniel's research, Abrams (1968) points out that discrimination in the labour market represents a threat to social cohesion. He emphasises that it is the duty of a democratic society to provide equal opportunities to all its members. Not doing so not only means a waste of manpower but could lead to unrest among the disadvantaged groups. Sadly, these arguments sound very current. In western, supposedly meritocratic societies, discrimination of minority applicants still presents a threat to social cohesion. Despite the adoption of anti-discrimination laws in most western countries, labour market discrimination still persists and there are still significant barriers to the labour market access of immigrants and their offspring.

Moving forward, interdisciplinary studies on hiring discrimination could provide valuable information, not only whether discrimination exists, but focusing rather on the question why discrimination occurs. Here adding qualitative approaches could be particularly useful, e.g. to interview employers about their hiring strategies, to interview job applicants on their experiences with discrimination and their coping strategies, or to analyse communication between employers and job candidates, e.g. in the published job vacancies or the emails received in reply to applications in correspondence tests. Such interdisciplinary research teams would most likely require more time and funding, but could potentially offer great insights into the mechanisms underlying discrimination. However, experiments that become increasingly complex also need to keep the ethical dimension of conducting field experiments in mind. Sending multiple CVs per vacancy for example might be a convenient way to reach a sufficient sample size, however, it increases the burden on employers.

References

- Abrams, M. (1968). Introduction. In W. W. Daniel (Ed.), *Racial Discrimination in England - Based on the PEP Report* (pp. 9-15). Middlesex: Penguin Books
- Agerström, J., Björklund, F., Carlsson, R., & 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*, 34(4), 359-366. doi:10.1080/01973533.2012.693438
- Agerström, J., & Rooth, D. O. (2009). Implicit prejudice and ethnic minorities. *International Journal of Manpower*, 30(1/2), 43-55. doi:doi:10.1108/01437720910948384
- Allasino, E., Reyneri, E., Venturini, A., & Zincone, G. (2006). *Labour market discrimination against migrant workers in Italy* (Vol. No 67). Geneva: ILO.
- Andriessen, I., Nievers, E., Dagevos, J., & Faulk, L. (2012). Ethnic Discrimination in the Dutch Labor Market: Its Relationship With Job Characteristics and Multiple Group Membership. *Work and Occupations*, 39(3), 237-269.
- Arai, M., Bursell, M., & Nekby, L. (2015). The Reverse Gender Gap in Ethnic Discrimination: Employer Stereotypes of Men and Women with Arabic Names. *International migration review*, n/a-n/a. doi:10.1111/imre.12170
- Arceo-Gomez, E. O., & Campos-Vazquez, R. M. (2014). Race and marriage in the labor market: A discrimination correspondence study in a developing country. *American Economic Review*, 104(5), 376-380.
- Ariel, B., Tobby-Alimi, I., Cohen, I., Ben Ezra, M., Cohen, Y., & Sosinski, G. (2015). Ethnic and Racial Employment Discrimination in Low-Wage and High-Wage Markets: Randomized Controlled Trials Using Correspondence Tests in Israel. *The Law & Ethics of Human Rights*, 9(1), 113-139.
- Arriijn, P., Feld, S., & Nayer, A. (1998). *Discrimination in access to employment on grounds of foreign origin: the case of Belgium*. Geneva: International Labour Office, Conditions of Work Branch.
- Arrow, K. J. (1998). What has economics to say about racial discrimination? *The Journal of Economic Perspectives*, 91-100.
- Asali, M., Pignatti, N., & Skhirtladze, S. (2017). Employment Discrimination in Georgia: Evidence from a Field Experiment. *MPRA Paper No. 78230*.
- Attström, K. (2007). *Discrimination against native Swedes of immigrant origin in access to employment*. Geneva: International Labor Organization.
- Baert, S. (2015). Do They Find You on Facebook? Facebook Profile Picture and Hiring Chances. *IZA Discussion Papers*, No. 9584.
- Baert, S. (2018). Hiring Discrimination: An Overview of (Almost) All Correspondence Experiments Since 2005. In M. Gaddis (Ed.), *Audit Studies: Behind the Scenes with Theory, Method and Nuance* (pp. 63-77): Springer.
- Baert, S., Albanese, A., du Gardein, S., Ovaere, J., & Stappers, J. (2017). Does work experience mitigate discrimination? *Economics Letters*, 155, 35-38. doi:<http://dx.doi.org/10.1016/j.econlet.2017.03.011>
- Baert, S., Cockx, B., Gheyle, N., & Vandamme, C. (2015). Is there less discrimination in occupations where recruitment is difficult? *International Labour Review*. doi:0019793915570873
- Baert, S., & Vujić, S. (2016). Immigrant volunteering: A way out of labour market discrimination? *Economics Letters*, 146, 95-98.
- Banerjee, A., Bertrand, M., Datta, S., & Mullainathan, S. (2009). Labor market discrimination in Delhi: Evidence from a field experiment. *Journal of Comparative Economics*, 37(1), 14-27.
- Banton, M. (1997). The ethics of practice - testing. *Journal of Ethnic and Migration Studies*, 23(3), 413-420.

- Bendick, M. (1996). *Discrimination Against Racial-ethnic Minorities in Access to Employment in the United States: Empirical Findings from Situation Testing*. Employment Department, International Labour Office.
- Bendick, M. (2007). Situation testing for employment discrimination in the United States of America. *Horizons stratégiques*(3), 17-39.
- Bendick, M., Jackson, C. W., & Reinoso, V. A. (1994). Measuring employment discrimination through controlled experiments. *The Review of Black Political Economy*, 23(1), 25-48.
- Bendick, M., Jackson, C. W., Reinoso, V. A., & Hodges, L. E. (1991). Discrimination against Latino job applicants: A controlled experiment. *Human Resource Management*, 30(4), 469-484.
- Bendick, M., & Nunes, A. P. (2012). Developing the Research Basis for Controlling Bias in Hiring. *Journal of Social Issues*, 68(2), 238-262. doi:10.1111/j.1540-4560.2012.01747.x
- Bendick, M., Rodriguez, R. E., & Jayaraman, S. (2010). Employment discrimination in upscale restaurants: Evidence from matched pair testing. *The Social Science Journal*, 47(4), 802-818. doi:<http://dx.doi.org/10.1016/j.soscij.2010.04.001>
- Bertrand, M., Chugh, D., & Mullainathan, S. (2005). Implicit Discrimination. *The American Economic Review*, 95(2), 94-98. doi:10.2307/4132797
- Bertrand, M., & Duflo, E. (2017). Field Experiments on Discrimination *Handbook of Economic Field Experiments* (Vol. 1, pp. 309-393): Elsevier.
- 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. doi:10.2307/3592802
- Blank, R. M., Dabady, M., & Citro, C. F. (2004). *Measuring racial discrimination*. Washington, DC: National Academies Press.
- Bleich, E. (2003). *Race politics in Britain and France: Ideas and policymaking since the 1960s*. Cambridge: Cambridge University Press.
- Blommaert, L., Coenders, M., & van Tubergen, F. (2014). Discrimination of Arabic-Named Applicants in the Netherlands: An Internet-Based Field Experiment Examining Different Phases in Online Recruitment Procedures. *Social Forces*, 92(3), 957-982. doi:10.1093/sf/sot124
- Bon-Maury, G., Bruneau, C., Dherbécourt, C., Diallo, A., Flamand, J., Gilles, C., & Trannoy, A. (2016). *Le coût économique des discriminations*. Retrieved from Paris:
- Booth, A. L., Leigh, A., & Varganova, E. (2012). Does ethnic discrimination vary across minority groups? Evidence from a field experiment. *Oxford Bulletin of Economics and Statistics*, 74(4), 547-573.
- Bovenkerk, F. (1977). Rasdiscriminatie op de Amsterdamse arbeidsmarkt. *Sociologische Gids*, 24(1/2).
- Bovenkerk, F. (1992). *Testing discrimination in natural experiments: a manual for international comparative research on discrimination on the grounds of "race" and ethnic origin*. Geneva: International Labour Office.
- Bovenkerk, F., & Breuning-van Leeuwen, E. (1978). Rasdiscriminatie en rasvooordeel op Amsterdamse arbeidsmarkt. In F. Bovenkerk (Ed.), *Omdat zij anders zijn. Patronen van rasdiscriminatie Nederland*. (pp. 17-35). Meppel Boom.
- Bovenkerk, F., Gras, M., Ramsoedh, D., Dankoor, M., & Havelaar, A. (1995). *Discrimination against migrant workers and ethnic minorities in access to employment in the Netherlands*. Retrieved from Geneva:
- Bovenkerk, F., Kilborne, B., Raveau, F., & Smith, D. (1979). Comparative aspects of research on discrimination against non-white citizens in Great Britain, France and the Netherlands. *Problems in international comparative research in the social sciences*. Pergamon, Oxford.

- Bravo, D., Sanhueza, C., & Urzúa, S. (2008). An Experimental Study of Labor Market Discrimination: Gender, Social Class and Neighborhood in Chile. *IDB Working Paper*, No. 226. doi:<http://dx.doi.org/10.2139/ssrn.1815907>
- Brown, C., & Gay, P. (1985). *Racial discrimination: 17 years after the Act*. London: Policy Studies Institute.
- Bursell, M. (2007). What's in a name? A field experiment test for the existence of ethnic discrimination in the hiring process. *SULCIS Working Paper*(2007:7).
- Carlsson, M. (2010). Experimental Evidence of Discrimination in the Hiring of First-and Second-generation Immigrants. *Labour*, 24(3), 263-278.
- Carlsson, M., Fumarco, L., & Rooth, D.-O. (2014). Does the design of correspondence studies influence the measurement of discrimination? *IZA Journal of Migration*, 3(1), 1-17. doi:10.1186/2193-9039-3-11
- Carlsson, M., & Rooth, D.-O. (2007). Evidence of ethnic discrimination in the Swedish labor market using experimental data. *Labour Economics*, 14(4), 716-729. doi:<http://dx.doi.org/10.1016/j.labeco.2007.05.001>
- Carlsson, R., & Agerström, J. (2016). A closer look at the discrimination outcomes in the IAT literature. *Scandinavian journal of psychology*, 57(4), 278-287.
- Cediey, E., & Foroni, F. (2008). *Discrimination in access to employment on grounds of foreign origin in France: A national survey of discrimination based on the testing methodology of the International Labour Office* (Vol. 85E). Geneva: International Labor Organisation (ILO).
- Cherry, F., & Bendick, M. (2018). Making It Count: Discrimination Auditing and the Activist Scholar Tradition *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 45-62): Springer.
- Crabtree, C., & Chykina, V. (2018). Last Name Selection in Audit Studies. *Sociological Science*, 5, 2-21.
- Cross, H., Kenney, J., Mell, J., & Zimmermann, W. (1990). *Employer hiring practices: Differential treatment of Hispanic and Anglo job seekers*. Washington, DC: The Urban Institute Press.
- Daniel, W. W. (1968). *Racial discrimination in England: based on the PEP report* (Vol. 1084). Middlesex: Penguin.
- Darolia, R., Koedel, C., Martorell, P., Wilson, K., & Perez-Arce, F. (2016). Race and gender effects on employer interest in job applicants: new evidence from a resume field experiment. *Applied Economics Letters*, 23(12), 853-856.
- de Prada, M. A., Actis, W., Pereda, C., & Molina, R. P. (1995). *Labour market discrimination against migrant workers in Spain* (Vol. 9). Geneva: International Labour Office, Employment Department.
- Dechief, D., & Oreopoulos, P. (2012). Why do some employers prefer to interview Matthew, but not Samir? New Evidence from Toronto, Montreal, and Vancouver. *Canadian Labour Market and Skills Researcher Network, Working Paper No. 95*.
- Deming, D. J., Yuchtman, N., Abulafi, A., Goldin, C., & Katz, L. F. (2016). The value of postsecondary credentials in the labor market: An experimental study. *The American Economic Review*, 106(3), 778-806.
- Drydakis, N., & Vlassis, M. (2010). Ethnic Discrimination in the Greek Labour market: Occupational Access, Insurance Coverage and Wage Offers. *The Manchester School*, 78(3), 201-218. doi:10.1111/j.1467-9957.2009.02132.x
- Duguet, E., Leandri, N., L'Horty, Y., & Petit, P. (2010). Are young French jobseekers of ethnic immigrant origin discriminated against? A controlled experiment in the Paris area. *Annals of Economics and Statistics/Annales d'Économie et de Statistique*, 187-215.

- Fibbi, R., Kaya, B., & Piguet, E. (2003). *Le passeport ou le diplôme ? Etude des discriminations à l'embauche des jeunes issus de la migration*. Neuchâtel: Forum suisse pour l'étude des migrations et de la population.
- Fibbi, R., Lerch, M., & Wanner, P. (2006). Unemployment and Discrimination against Youth of Immigrant Origin in Switzerland: When the Name Makes the Difference. *Journal of International Migration and Integration*, 7(3), 351–366.
- Firth, M. (1981). Racial Discrimination in the British Labor Market. *Industrial and Labor Relations Review*, 34(2), 265-272. doi:10.2307/2522541
- Fix, M., & Struyk, R. J. (Eds.). (1993). *Clear and Convincing Evidence: Measurement of Discrimination in America*. Washington, D.C.: The Urban Institute Press.
- Gaddis, S. M. (2014). Discrimination in the Credential Society: An Audit Study of Race and College Selectivity in the Labor Market. *Social Forces*, 1-29.
- Gaddis, S. M. (2017a). How Black are Lakisha and Jamal? Racial Perceptions from Names Used in Correspondence Audit Studies. *Sociological Science*, 4, 469-489.
- Gaddis, S. M. (2017b). Racial/Ethnic Perceptions from Hispanic Names: Selecting Names to Test for Discrimination. *Socius*, 3, 2378023117737193. doi:10.1177/2378023117737193
- Gaddis, S. M. (2018a). An introduction to audit studies in the social sciences *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 3-44): Springer.
- Gaddis, S. M. (Ed.) (2018b). *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*: Springer.
- Ghumman, S., & Ryan, A. M. (2013). Not welcome here: Discrimination towards women who wear the Muslim headscarf. *Human Relations*, 66(5), 671-698.
- Ginsberg, E., & Henry, F. (1985). *Who gets the work? A test of racial discrimination in employment* Retrieved from Toronto:
- Goldberg, A., Mourinho, D., & Kulke, U. (1995). *Labour market discrimination against foreign workers in Germany*: International Labour Office, Employment Department.
- Heckman, J. J. (1998). Detecting discrimination. *The Journal of Economic Perspectives*, 101-116.
- Heckman, J. J., & Siegelman, P. (1993). The Urban Institute Audit Studies: Their Methods and Findings. In M. Fix & R. Struyk (Eds.), *Clear and Convincing Evidence: Measurements of Discrimination in America* (pp. 187-258). Washington, DC: The Urban institute Press.
- Henry, F. (1989). *Who gets the Work in 1989?* Retrieved from Ottawa:
- Hubbock, J., & Carter, S. (1980). *Half a Chance? A Report on Job Discrimination against Young Blacks in Nottingham*. Retrieved from London:
- Jacquemet, N., & Yannelis, C. (2012). Indiscriminate discrimination: A correspondence test for ethnic homophily in the Chicago labor market. *Labour Economics*, 19(6), 824-832. doi:<http://dx.doi.org/10.1016/j.labeco.2012.08.004>
- Jowell, R., & Prescott-Clarke, P. (1970). Racial discrimination and white-collar workers in Britain. *Race & Class*, 11(4), 397-417.
- Kaas, L., & Manger, C. (2012). Ethnic Discrimination in Germany's Labour Market: A Field Experiment. *German Economic Review*, 13(1), 1-20. doi:10.1111/j.1468-0475.2011.00538.x
- Kang, S. K., DeCelles, K. A., Tilcsik, A., & Jun, S. (2016). Whitened resumes: Race and self-presentation in the labor market. *Administrative Science Quarterly*, 61(3), 469-502.
- Karapin, R. (1999). The politics of immigration control in Britain and Germany: subnational politicians and social movements. *Comparative Politics*, 423-444.
- Koopmans, R., Veit, S., & Yemane, R. (2018). Ethnische Hierarchien in der Bewerberauswahl: Ein Feldexperiment zu den Ursachen von Arbeitsmarktdiskriminierung. *WZB Discussion Paper, SP VI 2018-104*.
- Lahey, J., & Beasley, R. (2009). Computerizing audit studies. *Journal of Economic Behavior & Organization*, 70(3), 508-514. doi:<http://dx.doi.org/10.1016/j.jebo.2008.02.009>

- Lahey, J., & Beasley, R. (2018). Technical aspects of correspondence studies. In M. Gaddis (Ed.), *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 81-101): Springer.
- Larja, L., Warius, J., Sundbäck, L., Liebkind, K., Kandolin, I., & Jasinskaja-Lahti, I. (2012). *Discrimination in the Finnish Labour Market: An Overview and a Field Experiment on Recruitment*. Retrieved from <https://yhdenvertaisuus-fi-bin.directo.fi/@Bin/070c0305f220b492e9b19da85cb314b5/1508422136/application/pdf/469926/discrimination-in-the-finnish-labour-market.pdf>
- Lee, H.-A., & Khalid, M. A. (2016). Discrimination of high degrees: race and graduate hiring in Malaysia. *Journal of the Asia Pacific Economy*, 21(1), 53-76.
- Liebkind, K., Larja, L., & Brylka, A. (2016). Ethnic and Gender Discrimination in Recruitment: Experimental Evidence From Finland. *Journal of Social and Political Psychology*, 4(1), 403-426. doi:10.5964/jspp.v4i1.433
- Maurer-Fazio, M. (2012). Ethnic discrimination in China's internet job board labor market. *IZA Journal of Migration*, 1(1), 1-24.
- McGinnity, F., & Lunn, P. D. (2011). Measuring discrimination facing ethnic minority job applicants: an Irish experiment. *Work, Employment & Society*, 25(4), 693-708.
- McIntosh, N., & Smith, D. J. (1974). *The extent of racial discrimination* (Vol. 40): PEP London.
- Midtbøen, A. H. (2012). Discrimination of the Second Generation: Evidence from a Field Experiment in Norway. *Journal of International Migration and Integration*, 1-20.
- Midtbøen, A. H. (2013). *Determining Discrimination* University of Oslo.
- Midtbøen, A. H. (2014). The context of employment discrimination: interpreting the findings of a field experiment. *The British journal of sociology*.
- Midtbøen, A. H., & Rogstad, J. (2012). Discrimination. Methodological controversies and sociological perspectives on future research. *Nordic Journal of Migration Research*, 2(3), 203-212.
- Neumark, D. (2012). Detecting Discrimination in Audit and Correspondence Studies. *Journal of Human Resources*, 47(4), 1128-1157.
- Neumark, D., & Rich, J. (2018). Do field experiments on labor and housing markets overstate discrimination? A re-examination of the evidence. *ILR Review*.
- Newman, J. M. (1978). Discrimination in Recruitment: An Empirical Analysis. *Indus. & Lab. Rel. Rev.*, 32, 15.
- Nunley, J. M., Pugh, A., Romero, N., & Seals, R. A. (2014). An Examination of Racial Discrimination in the Labor Market for Recent College Graduates: Estimates from the Field. *Working Paper Series, Department of Economics, Auburn University*(AUWP 2014-06).
- Nunley, J. M., Pugh, A., Romero, N., & Seals, R. A. (2015). Racial discrimination in the labor market for recent college graduates: Evidence from a field experiment. *The BE Journal of Economic Analysis & Policy*, 15(3), 1093-1125.
- 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), 148-171. doi:10.2307/41330445
- Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, 105(2), 171.
- Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance.[Comment/Reply]. *Journal of Personality and Social Psychology*, 108(4), 562-571. doi:doi: 10.1037/pspa0000023
- Pager, D. (2003). The Mark of a Criminal Record. *American Journal of Sociology*, 108(5), 937-975.

- Pager, D. (2007). The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future. *Annals of the American Academy of Political and Social Science*, 609, 104-133. doi:10.2307/25097877
- Pager, D. (2016). Are Firms That Discriminate More Likely to Go Out of Business? *Sociological Science*, 3, 849-859.
- Pager, D., Bonikowski, B., & Western, B. (2009). Discrimination in a Low-Wage Labor Market: A Field Experiment. *American Sociological Review*, 74(5), 777-799. doi:10.2307/27736094
- Pager, D., & Quillian, L. (2005). Walking the talk? What employers say versus what they do. *American Sociological Review*, 70(3), 355-380.
- Pager, D., & Shepherd, H. (2008). The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets. *Annual Review of Sociology*, 34(1), 181-209.
- Quillian, L. (2006). New Approaches to Understanding Racial Prejudice and Discrimination. *Annual Review of Sociology*, 32, 299-328. doi:10.2307/29737741
- Quillian, L., Pager, D., Hexel, O., & Midtbøen, A. H. (2017). Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. *Proceedings of the National Academy of Sciences*, 114(41), 10870-10875. doi:10.1073/pnas.1706255114
- Riach, P. A., & Rich, J. (1991). Testing for racial discrimination in the labour market. *Cambridge Journal of Economics*, 239-256.
- Riach, P. A., & Rich, J. (2002). Field Experiments of Discrimination in the Market Place. *The Economic Journal*, 112(483), F480-F518. doi:10.2307/798458
- Riach, P. A., & Rich, J. (2004a). Deceptive field experiments of discrimination: are they ethical? *Kyklos*, 57(3), 457-470.
- Riach, P. A., & Rich, J. (2004b). Fishing for Discrimination. *Review of Social Economy*, 62(4), 465-486. doi:10.1080/0034676042000296227
- 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.-O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labor economics*, 17, 523-534.
- Schneider, J., Yemane, R., & Weinmann, M. (2014). *Diskriminierung am Ausbildungsmarkt: Ausmass, Ursachen und Handlungsperspektiven* Retrieved from Berlin
- Small, S., & Solomos, J. (2006). Race, Immigration and Politics in Britain Changing Policy Agendas and Conceptual Paradigms 1940s–2000s. *International Journal of Comparative Sociology*, 47(3-4), 235-257.
- Sooben, P. N. (1990). *The Origins of the Race Relations Act*: Centre for Research in Ethnic Relations, University of Warwick Coventry.
- Turner, M. A., Fix, M., & Struyk, R. J. (1991). *Opportunities denied, opportunities diminished: Racial discrimination in hiring*. Washington, DC: The Urban Institute.
- Veenman, J. (2010). Measuring labor market discrimination: An overview of methods and their characteristics. *American behavioral scientist*, 53(12), 1806-1823.
- Vuolo, M., Uggen, C., & Lageson, S. (2016). Statistical power in experimental audit studies: Cautions and calculations for matched tests with nominal outcomes. *Sociological Methods & Research*, 45(2), 260-303.
- 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.
- Weichselbaumer, D. (2016a). Discrimination against Female Migrants Wearing Headscarves. *IZA Discussion Paper Series*(No. 10217).
- Weichselbaumer, D. (2016b). Discrimination Against Migrant Job Applicants in Austria: An Experimental Study. *German Economic Review*, 18(2), 237-265. doi:10.1111/geer.12104

- Widner, D., & Chicoine, S. (2011). It's All in the Name: Employment Discrimination Against Arab Americans. *Sociological Forum*, 26(4), 806-823. doi:10.1111/j.1573-7861.2011.01285.x
- Wood, M., Hales, J., Purdon, S., Sejersen, T., & Hayllar, O. (2009). *A Test for Racial Discrimination in Recruitment Practice in British Cities*. Retrieved from Norwich: <http://webarchive.nationalarchives.gov.uk/20130314010347/http://research.dwp.gov.uk/asd/asd5/rrs-index.asp>
- Zschirnt, E. (2016). Revisiting Ethics in Correspondence Testing. *NCCR Working Papers*, 8.
- Zschirnt, E., & Ruedin, D. (2016). Ethnic Discrimination in Hiring Decisions: A Meta-Analysis of Correspondence Tests 1990–2015. *Journal of Ethnic and Migration Studies*, 42(7), 1115-1134. doi:10.1080/1369183X.2015.1133279

Paper II:

Ethnic Discrimination in Hiring: A Meta-Analysis of Correspondence Tests 1990-2015 (Co-authored with Didier Ruedin)

“This is an Accepted Manuscript of an article published by Taylor & Francis in the Journal of Ethnic and Migration Studies on 22 January 2016, available online: <https://www.tandfonline.com/doi/full/10.1080/1369183X.2015.1133279>

Zschirnt, E. and Ruedin, D. (2016) “Ethnic Discrimination in Hiring: A Meta-Analysis of Correspondence Tests 1990-2015”, *Journal of Ethnic and Migration Studies*, (42) 7, 1115-1134.

Ethnic Discrimination in Hiring Decisions: A Meta-Analysis of Correspondence Tests 1990–2015

Eva Zschirnt

University of Neuchâtel and NCCR “on the move”

eva.zschirnt@unine.ch

Tel. +41 32 718 39 49

Didier Ruedin

University of Neuchâtel and University of Witwatersrand

didier.ruedin@wolfson.oxon.org

Tel. +41 32 718 39 32

University of Neuchâtel

Swiss Forum for Migration and Population Studies SFM

Faubourg de l’Hôpital 1006

2000 Neuchâtel, Switzerland

Acknowledgements and Author Contributions

We would like to thank Arnfinn Midtbøen, Marco Pecoraro, and Rosita Fibbi for comments. An earlier version was presented at the 2015 IMISCOE Annual Conference in Geneva. Author contributions: DR designed the study, EZ collected the data, DR analysed the data, DR and EZ wrote the paper.

Funding

This research was supported by the *nccr - on the move* funded by the Swiss National Science Foundation

Ethnic Discrimination in Hiring Decisions: A Meta-Analysis of Correspondence Tests 1990–2015

Abstract

For almost 50 years field experiments have been used to study ethnic and racial discrimination in hiring decisions, consistently reporting high rates of discrimination against minority applicants – including immigrants –, irrespective of time, location, or minority groups tested. While Riach and Rich (2002) and Rich (2014) provide systematic reviews of existing field experiments, no study has undertaken a meta-analysis to examine the findings in the studies reported. In this article we present a meta-analysis of 738 correspondence tests in 43 separate studies conducted in OECD countries between 1990 and 2015. In addition to summarizing research findings, we focus on groups of specific tests to ascertain the robustness of findings, emphasizing differences across countries, gender, and economic contexts. Moreover we examine patterns of discrimination, by drawing on the fact that the groups considered in correspondence tests and the contexts of testing vary to some extent. We focus on first- and second-generation immigrants, differences between specific minority groups, the implementation of EU directives, and the length of job application packs.

Keywords: ethnic discrimination, hiring, correspondence test, meta-analysis, immigration

Introduction

Whenever members of one minority group are less likely to obtain paid work, or do so under unfavourable conditions, some people are quick to shout ‘discrimination’. Social scientists tend to be more cautious and highlight that there are many reasons why one group is more likely to obtain paid work than others apart from discrimination (Pager 2007). To rule out these alternative explanations, field experiments were devised in the United Kingdom in the 1960s, allowing researchers to draw inferences about racial discrimination in hiring decisions (Daniel 1968, Jowell and Prescott-Clarke 1970). Fifty years after the first British Race Relations Act of 1965, prohibiting racial discrimination in public places, interest in discrimination and hiring decisions remains high. Indeed, in recent years numerous studies using field experiments have been carried out to test whether discrimination in terms of race, ethnicity, immigration background, or other minority statuses remains a problem (Bendick 2007, Pager 2007, Pager and Shepherd 2008, Riach and Rich 2002, Rich 2010, 2014).

Field experiments offer strong evidence of discriminatory behaviour in the labour market, using either in-person audit tests or written correspondence tests. Since discrimination in hiring decisions usually cannot be observed directly, researchers resort to fictitious candidates with equivalent and thus exchangeable qualifications. One employer is presented with two substantially identical job applications. The only difference is the characteristic of interest: the ethnic or racial group of the applicant. This results in controlled experiments on discrimination in hiring decisions in a real world setting. It can plausibly be argued that differences in call-back rates of equally qualified minority and majority candidates can be attributed to discrimination (Jackson and Cox 2013, Midtbøen and Rogstad 2012), especially in the case of correspondence tests where the experimental manipulation can be controlled better.

Studies employing correspondence tests find systematic evidence for discrimination in hiring decisions. At first sight, there are no apparent differences across time, location, and minority groups tested. These findings suggest that while overt racial and ethnic discrimination are no longer practised as much as it was in the past – consider racial segregation in the US –, ethnic and racial discrimination remains a common phenomenon, albeit a more subtle and covert one: “Today, it is harder to assess the degree to which everyday experiences and opportunities may be shaped by ongoing forms of discrimination” (Pager and Shepherd 2008, 6, see also Arrow 1998). In most countries under study there is anti-discrimination legislation in place that prohibits ethnic and racial discrimination in hiring, but field experiments highlight that current legislation seems to be inefficient and that discrimination remains commonplace. Indeed, using field experiments it is possible to enumerate the degree of discrimination members of ethnic and racial minorities face when applying for jobs.

As is common with experiments, however, a single audit or correspondence test is unable to explain *why* discrimination occurs. To overcome this limitation, studies increasingly resort to finer distinctions of carefully chosen groups, or seek other methods. In this article we draw inferences from various studies by contrasting comparable groups from different correspondence tests in a meta-analysis.

Theory and Background

Racial or ethnic discrimination can be defined in various ways, often depending on the research question and scientific tradition of the study. For this meta-analysis we use the US National Research Council's definition, which focuses on '*differential treatment on the basis of race* that disadvantages a racial group and *treatment on the basis of inadequately justified factors other than race* that disadvantages a racial group (differential effect)' (Blank, Dabady, and Citro 2004, 39, italics in original) thus covering groups such as immigrants. This definition is similar to the one used in the European Union's Directive 2000/43/EC, commonly known as the 'Race Directive', which differentiates between direct and indirect discrimination (Art. 2) and prohibits both forms.

Given that racial and ethnic discrimination are outlawed in many jurisdictions and have thus become hidden, questions of how to measure discrimination have taken centre stage in recent years (Quillian 2006). We refer readers to Veenman (2010) for a thorough review of the approaches used: statistical analysis of observational data, behavioural research, attitudes research, and victim research. The field experiments focused on in this article are a form of behavioural research.

The literature offers different explanations why discrimination occurs in hiring processes. A classic distinction is that between taste-based discrimination and statistical discrimination. Taste-based discrimination describes the situation where the employer has racial or ethnic preferences (Becker 1957). This includes xenophobia and racism, but also personal preferences of other kinds; the employer will discriminate against a group irrespective of other information he or she has about the applicants. Because of racial or ethnic preferences, the employer is willing to pay a higher price to hire a person who matches the desired racial or ethnic profile. Put differently, employers do not act in a purely profit-maximizing manner, but "an avoidance of the psychic cost of contact with the 'wrong' race [...] takes precedence" (Riach and Rich 1991, 247). Following this logic, employers without racial preferences have a competitive advantage, which, in the long-run, should lead to the elimination of racial discrimination in the market place.

By contrast, statistical discrimination describes the situation where members of a specific group are discriminated against because the employer is lacking information (Phelps 1972, Arrow 1973).

The employer who seeks to maximize expected profit will discriminate against blacks or women if he believes them to be less qualified, reliable, long-term, etc. on the average than whites and men, respectively, and if the cost of gaining information about the individual applicants is excessive (Phelps 1972, 659).

This is a characteristic of the hiring process where the employer will never be able to obtain all the information about the candidate, or obtaining such information is too costly. The employer will thus rely on signals and other cues from the application and CV (Pager 2007). Ethnic minority status may be such a signal that members of a particular group are less skilled or otherwise unsuited – or in some cases more skilled, harder working, and so on. Drawing on stereotypes, hearsay, or previous experience with a small number of group members, the employer discounts the applicant because of his or her ethnicity – ethnicity acts as a proxy for unobserved information. With more information, the employer would not discriminate against the minority candidate. As a consequence of statistical discrimination, an employer will not always succeed in hiring the most qualified applicants. If hiring decisions

are taken on a regular basis the employer may regard statistical discrimination as an acceptable trade-off between the effort to obtain more information about an applicant and the recruitment of a productive employee (Bursell 2007).

Besides these predominant economic theories of discrimination, other explanations analyse discriminatory treatment of minority groups more generally. Many researchers have become more cautious, preferring terms such as ‘ethnic penalty’ to describe differential treatment on the basis of race and ethnicity simply because the act of discrimination or the intention to discriminate are not observed. For instance, Heath and Cheung (2006) highlight that certain differentials between ethnic minority groups and majority groups in the labour market cannot be explained by age, education, or country of origin.

Human capital theory, by contrast, focuses on factors like age, education, work experience, or health. The theory highlights the often lower human capital of members of minority groups compared to their majority competitors to explain their disadvantaged position on the labour market (e.g. Andriessen, Dagevos, and Iedema 2008). It is argued that members of ethnic minority groups on average are less educated, are unfamiliar with host-country institutions, are not fluent in the language, or lack relevant networks for job searching. Differences in economic outcomes persist, however, when human capital differences are controlled for (Blommaert, Coenders, and van Tubergen 2014).

Theories of social dominance highlight that groups are not only distinguished but also ranked according to their social position and negative stereotypes connected with these groups, resulting in status hierarchies. Men tend to be ‘ranked’ higher than women, and natives are usually ranked higher than immigrants (Andriessen et al. 2010). Closely related to this theory, the notion of ethnic hierarchies is often discussed in the Dutch context, where Moroccans are consistently ‘ranked’ at the bottom and Surinamese immigrants are regarded more favourably (Andriessen et al. 2012). To some extent, ethnic hierarchies draw on cultural distance, where groups perceived as ‘more different’ tend to have less status and thus rank lower in the hierarchy. Cultural distance can reflect social distance (Parrillo and Donoghue 2005), but it frequently draws on visible markers like skin colour and dress as signals of cultural distance (Fetzer 2013). Ethnic hierarchies may play a role in taste-based discrimination and statistical discrimination, and they serve to remind us that discrimination in the hiring process is not a binary decision: the hiring decision may be context-dependent and depend on the other applicants for the same job.

The literature further highlights factors like the size and composition of the minority population, the economic situation and outlook, policies, media reporting, as well as attitudes in the population. The way minorities are presented in the media and how they are politicized in public debates is likely to play an important role (van der Brug et al. 2015, Klingeren et al. 2014). The mediatized debate provides and reinforces stereotypes that can be used as shortcuts in statistical discrimination. At the same time, employers gain additional knowledge about different minority groups when immigration is politicized – making them less likely to (have to) resort to shortcuts. Taste-based discrimination may also be affected by the public debate and attitudes in the population (Pecoraro and Ruedin 2015, Pettigrew and Tropp 2006). Because of ‘in-group loyalty’ and ‘out-group rejection’, it can similarly be expected that applicants from one’s in-group are more likely to be invited for a job interview (c.f. Ford 2015).

Expectations

Based on the existing theories outlined, we have identified four expectations. Obviously, many other expectations could be stated, given the numerous variables that are potentially related to patterns of discrimination, but in this article we will focus on those related to taste-based and statistical discrimination:

E1: According to statistical discrimination theory, employers are expected to react to signals like education completed in the country under study. Similarly, children of immigrants – second-generation immigrants – tend to have more social ties in the country under study. Employers are thus likely to perceive them more positively, with generation serving as a signal for civic integration. It can therefore be expected that *discrimination is lower for second-generation immigrants than first-generation immigrants*. More generally, the more established an immigrant group is in a country, the more information can be expected to be available, translating into lower rates of statistical discrimination.

E2: Taking taste-based discrimination seriously, because of ethnic and status hierarchies, it can be assumed that *more distant and visible minority groups are discriminated against more than other groups*. Ostensible difference is used as a reason to discriminate, including the degree to which a particular minority group is established in a country. Immigrant groups associated with guest-worker programmes or colonial ties tend to be more established and are expected to face less discrimination than newly arrived groups.

E3: Two EU directives adopted in 2000 were designed to reduce discrimination. Irrespective of the effectiveness of the ensuing policies, it can be assumed that awareness of discrimination in hiring and the labour market has increased due to the political and public debates at the time. Hence *discrimination is likely to be lower after 2000 than before*.

E4: Depending on the country, job applications require different details. If statistical discrimination prevails, it can be expected that discrimination is lower in countries where more details are the norm in job applications, like diplomas or transcripts. In these contexts employers have less need to resort to mechanisms that can result in statistical discrimination (Weichselbaumer 2015b). More detailed application packages are widespread in German-speaking countries. It can therefore be expected that *discrimination rates are lower in German-speaking countries than in other European countries*.

Methods and Data

Correspondence tests are well suited for identifying discrimination in hiring, especially because they are able to minimize other influences (Jackson and Cox 2013, Bendick and Nunes 2012, Midtbøen and Rogstad 2012). In correspondence tests researchers apply in writing for actual positions at real companies, and thus capture real hiring decisions. They are much easier to implement than in-person audits, and allow more control over the application process. Correspondence tests can be repeated in relatively great numbers – especially now that electronic applications are commonplace –, and enable researchers to apply for a wider variety of jobs with different skill levels. They allow some conclusions about discrimination in the hiring process.

However, there are limits to correspondence tests. First, they usually rely exclusively on the applicant's name to convey information about race or ethnicity: stereotypical ethnic names may lead to different responses than lesser-known names from the same group, some ethnic

names may be misattributed to other ethnic groups, and names may have connotations of class or socio-economic status the researcher is unaware of (Bertrand and Mullainathan 2004, Pager 2007). These are confounding effects beyond the control of the researcher.¹ Second, correspondence tests are only suited for occupations where written applications are the norm. This excludes many entry-level and unskilled jobs where applications are typically made in person. Third, correspondence tests can only be used for publicly announced jobs and exclude informally or internally filled vacancies. Fourth, since correspondence tests rely on deception to obtain results, correspondence tests also face ethical challenges – in some cases also legal constraints. Today researchers take ethical questions increasingly seriously and obtain serious ethical clearance.

By design, correspondence tests only cover the first step of the hiring process and it is impossible to observe the behaviour of employers as is done during in-person audit studies. The second step is not unimportant, but estimates suggest that the first step may account for as much as 90 per cent of the discrimination levels measured (Riach and Rich 2002).

In this article, we use meta-analysis to summarize existing research in a systematic manner, drawing on the fact that all correspondence tests are conducted in a similar fashion (Weichselbaumer and Winter-Ebmer 2005). Meta-analyses use statistics to combine the reported findings across studies, offering a quantitative means to synthesize research with less reliance on the subjective assessment of the reviewing authors (Wolf 1986, Petticrew and Roberts 2006). We will benefit from the fact that correspondence tests have been carried out for different kinds of groups and sub-groups to draw inferences about taste-based and statistical discrimination where possible.

We carried out systematic searches using *Web of Knowledge* and *Google Scholar*, limiting the search to ethnic and racial discrimination in hiring and correspondence tests, which includes the discrimination of immigrant groups. We chose not to include in-person audit studies as written correspondence tests have become the dominant method in recent years. We further narrowed down the focus to correspondence studies in OECD countries between 1990 and 2015 to increase comparability. The following keywords were used: ‘discrimination’, ‘correspondence test’, ‘ethnic discrimination’, ‘racial discrimination’, ‘discrimination in hiring’, ‘discrimination AND labour market’ ‘discrimination AND field experiment’ and ‘discrimination AND employment’. We also relied on the often extensive bibliographies provided in the literature, especially in the systematic reviews conducted by Riach and Rich (2002) and Rich (2014). Furthermore, we carefully checked the bibliographies of every correspondence study and broadened our search from there. We were able to include studies published in English, French, German, and Dutch.

We note that there is no standard for reporting the results of correspondence tests and a wide variety of approaches are found (see supplementary material S4). Many studies report discrimination using relative call-back rates as the sole measure, other studies focus on net discrimination rates. Often only absolute numbers or only percentages are presented; we recalculated the absolute numbers wherever possible because this allows the calculation of corresponding call-back rates and odds ratios, drawing on four categories: ‘positive treatment minority’, ‘negative treatment minority’, ‘positive treatment majority’, ‘negative treatment majority’. The vocabulary here reflects the fact that meta-analyses are more established in the

¹ For a detailed discussion on unobservable characteristics in field experiments we refer readers to Heckman and Siegelman (1993), Heckman (1998) and Neumark (2012).

medical sciences (Petticrew and Roberts 2006). Majority applicants constitute our control group, while minority applicants are considered the treatment group. In studies that combine in-person audit tests with correspondence tests we singled out the results from the written correspondence tests and included them in our database. We are unable to identify a reason as to how a subsequent in-person test could affect the preceding correspondence test. Generally speaking, we note that the level of data provided in the studies is often incomplete, and for that reason we often rely on relative call-back rates to maximize the number of cases considered (see supplementary material S6 to S10 for odds ratios).

Data and Variables

The present article includes data from 43 studies conducted in 18 countries, looking at over 20 minority groups.² In Table 1 each study presents one data-point. For most analyses, each study can be broken down into several subgroups, namely specific minority or immigrant groups, depending on the level of detail provided in the data included in the articles. We treat Akintola (2011) as two separate studies because it covers both Canada and Sweden. There are in total 738 subgroups, and to some extent each can be treated as an independent experiment, given that hiring decisions were made by different employers and are thus unlikely to influence each other. While a study may aggregate discrimination rates across say Serbian, or Turkish applicants, a subgroup is more specific, like Chinese men applying to be cooks. At the subgroup level we gain variance in otherwise relatively homogeneous setups. This variance is used as a test of robustness for the overall meta-analysis, but also to test expectations related to the nature of discrimination. The supplementary material includes considerations of publication bias (S12).

Table 1: Data for Ethnic and Racial Discrimination in Hiring Decisions

	1990-95	1996-2000	2001-05	2006-10	2011-15	Total
Number of studies	6	2	3	9	23	43
Number of subgroups	78	4	66	192	398	738
Geographic location						
North America	1	0	2	0	9	12
European Union	4	2	0	9	11	26
Other European	0	0	1	0	1	2
Other OECD	1	0	0	0	2	3
Mean call-back ratio	1.29	1.50	1.47	1.98	1.47	1.55

Notes: Unless otherwise indicated, the numbers refer to the number of cases included; the mean call-back ratio refers to the study level

The variable of interest in this article is discrimination in hiring decisions. Two measures are available: relative call-back ratios and odds ratios. Relative call-back ratios compare how often a majority applicant is called back for an interview (control) to how often a minority applicant is called back for an interview (treatment). The call-back ratio is available for most subgroups. Odds ratios compare the odds of being invited for a job interview, drawing on a different means to express probabilities. By necessity, we were forced to accept that definitions of race and ethnicity vary across studies. For the comparisons across specific

² See supplementary material S1 for a complete list of studies. We also note the studies by Duguet et al. (2015), Agerström et al. (2012) and Adida et al. (2010), but their measurements are not comparable to the other studies. In the study by Weichselbaumer (2015a) we did not include the manipulations with headscarves to maintain comparability across studies. Also not considered in the analysis were studies using unsolicited applications; see Diekmann, Jann, and Näf (2014) and Ariel et al. (2015) for recent examples.

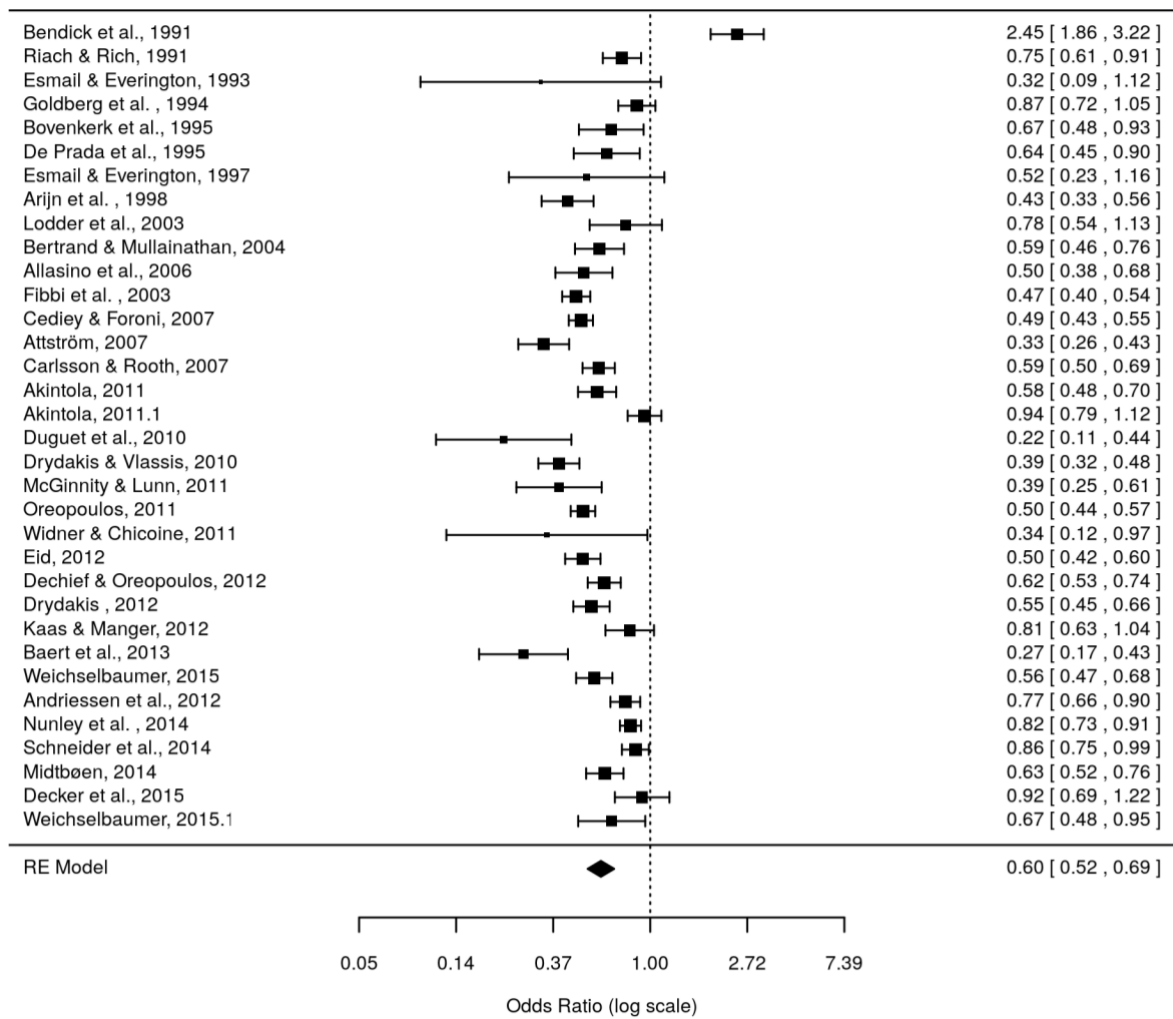
minority groups it was necessary to reclassify some of these groups, like when we included ‘Swedes of Middle Eastern origin’ in the category ‘Arabs and people of Middle Eastern origin’. These coding decisions are apparent in the supplementary material (S2, S3).

Discrimination across Studies

As a first step a meta-analysis of all studies is presented. Using a random-effects model, the forest plot in Figure 1 presents the odds ratios for the studies for which the data to calculate odds ratios was available on a log scale. With the exception of Bendick et al. (1991) in the US,³ who used CVs with enhanced credentials for Latino applicants, but not for Anglo applicants, most studies found significant evidence of discrimination against the minority applicants. Notable are also Akintola (2011) who found only little discrimination against minority applicants in Canada, and Decker et. al (2015) who reported very low rates of discrimination against black minority applicants in their US study. These are among the few studies where the two standard deviations cross the line at 1, indicating that the interpretation of ‘no discrimination’ cannot be ruled out. Across all studies for which sufficient details are available to calculate odds ratios, the odds ratio is 0.51, indicated by the rhomboid at the bottom of the figure: minority applicants have 49 per cent lower odds to be invited for an interview, compared to the equally qualified majority candidate. Given that each study covers several subgroups, the result of a model on subgroups is of equal interest: the odds ratio in this case is 0.60, around the same order of magnitude (supplementary material S6).

³ We only looked at the part of the study where applications were sent by mail. While response rates were higher for Latino applicants, the differences were not statistically significant (Bendick et al. 1991, 8).

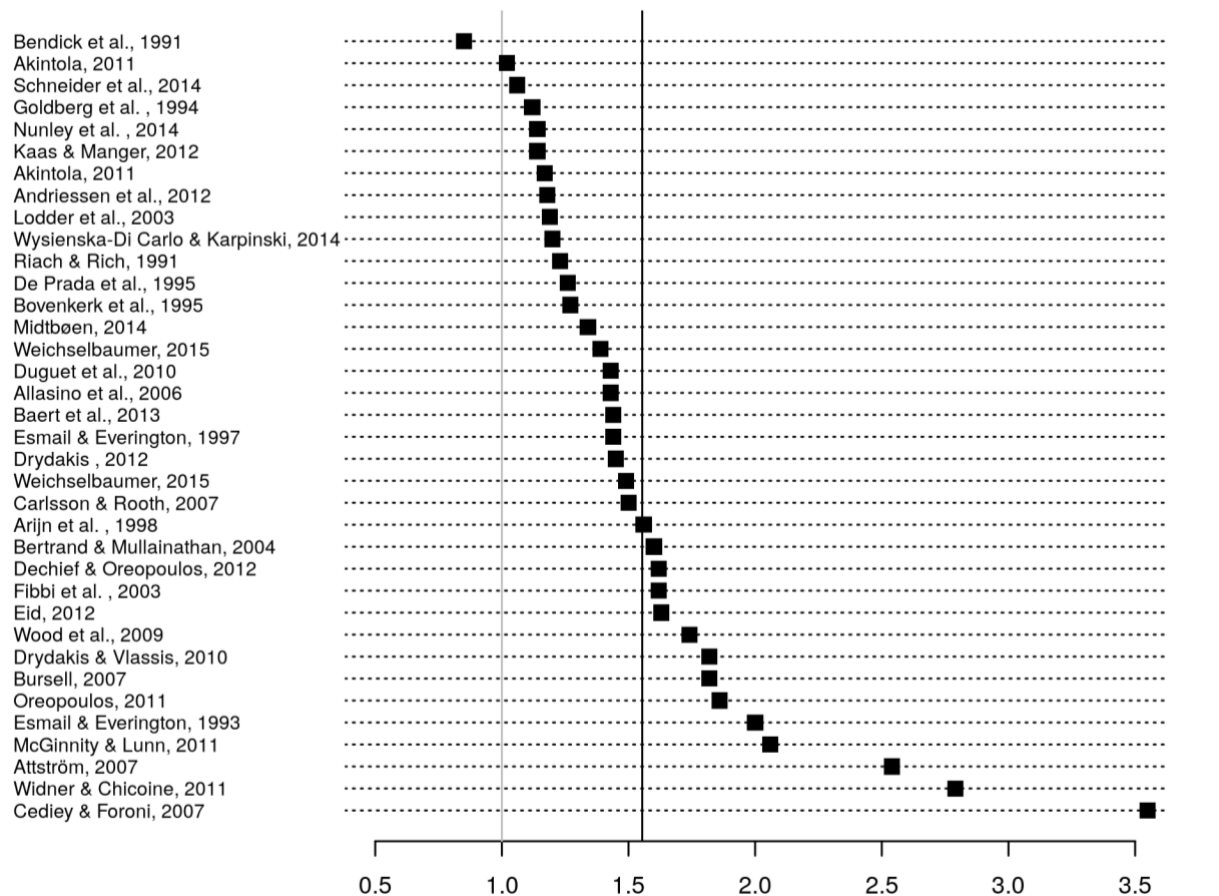
Figure 1: Ethnic and Racial Discrimination in Hiring Decisions



Notes: given are the odds ratios of each study (point estimate as squares, two standard errors as lines) on the left, along with the numerical representation on the right. At the bottom, the rhomboid summarizes the effect size across all studies. $N = 34$ studies (study level).

In some studies insufficient details are reported to calculate odds ratios, so a comparison of the relative call-back rates is necessary to cover more studies. Figure 2 shows the relative call-back rates reported in the studies. It ranges from Bendick et al. (1991) in the US to Cédiey and Foroni (2007) in France, where the highest relative call-back rates were measured. The mean relative call-back rate is 1.55 at the study level (indicated with a straight black line in the figure) and 1.75 at the level of subgroups. The median values are 1.44 for studies and 1.49 at the subgroup level. This means that minority applicants have to write around 50 per cent more applications to be invited for a job interview.

Figure 2: Relative Call-Back Rates for Studies



Notes: $N=36$ studies; the grey line at 1 indicates equal treatment, the black line gives the mean relative call-back rate across all studies

When interpreting these numbers, however, it must be borne in mind that the ethnic groups studied in correspondence tests are rarely chosen at random: Often researchers suspect discrimination for specific groups, or they examine the most salient minority groups – usually groups considered ‘different’ or with historical ties to the country, and not necessarily the largest minority groups in society. This may mean a focus on visible minority groups while ignoring immigrants from other European countries. The reported rates of discrimination may thus *overestimate* the extent of discrimination.

In a second step, the robustness of the meta-analysis is tested by examining specific subgroups. For instance, comparing European and North American correspondence studies indicates that minority applicants may be facing more discrimination in Europe than in the US and Canada, as far as it is possible to compare these groups (consider for example the tradition of strong anti-discrimination legislation). Discrimination occurs on both sides of the Atlantic, irrespective of whether we consider racial discrimination in North America or ethnic discrimination in Europe. At the subgroup level, the relative call-back rate is 1.84 in Europe and 1.69 in the US/Canada. These results, however, do not take into consideration that in-person audits are still prevalent in the US and often report high rates of discrimination (Pager 2007). The reported differences should be interpreted with caution.

The second dimension we focus on is gender. Stereotypes and media images of immigrant women tend to be less radical than those of immigrant men (Bovenkerk 1992, Andriessen et

al. 2012). This may lead to women being perceived as better integrated into and less threatening to society than immigrant men, and thus lowering discrimination for women. The opposite expectation can be drawn from status hierarchies, where men tend to be ranked ‘higher’ (Andriessen et al. 2010). Indeed, women seem to fare slightly worse than men (relative call-back rate 1.74 for women and 1.63 for men). However, these small differences are not statistically significant ($p>0.1$) and may be related to the particular occupations and positions chosen in the correspondence test, where gender stereotypes of ‘typical’ male or female jobs may influence the results. Substantively, there is no indication of systematic gender differences on a large scale.

A third dimension in which studies may be differing in a systematic way is the economic context. During times of economic boom and labour shortage, employers are likely to take more risks when hiring. It can be assumed that this affects discrimination rates: employers become more likely to ‘give a candidate a chance’, irrespective of past experience with other members of the same group or prevailing stereotypes. It can therefore be expected that discrimination is lower during times with low unemployment and high GDP growth (Baert et al. 2013). By contrast, Carlsson, Fumarco, and Rooth (2015) showed that for Sweden ethnic discrimination increases when the labour market improves. Focusing our analysis on GDP growth and unemployment rates we find no systematic association between the economic situation and ethnic discrimination in hiring. While a higher level of discrimination can be observed at times of high unemployment (mean call-back rates of 2.03 and 1.50), when considering median call-back rates, the differences disappear (supplementary material S11). Looking at the correlation between unemployment rates and call-back rates, there is no clear association ($r=-0.05$, $p>0.1$). Similarly, the correlation between annual GDP growth rates and call-back rates is not significant ($r=0.04$, $p>0.1$). Taken together, there is no evidence that rates of discrimination vary according to the national economic situation – although the relevant level of analysis may be occupation-specific and region-specific and unattainable in this analysis.

Rather than looking at the influence of individual factors, the supplementary material also includes multivariate regression meta-analysis to examine the influence of different factors (S13). In particular, the skills level may be of interest, and the regression coefficient for high skills is positive (0.28, $p<0.05$), while the regression coefficient for low skills is negative (-0.16 , $p<0.05$). The substantive patterns reported in this section remain unchanged when controlling for gender or whether first- or second-generation applicants are considered, suggesting that the reported findings are robust.

Taste-Based and Statistical Discrimination

Having established that ethnic discrimination in hiring exists across contexts in a fairly robust manner, we now make use of the variation in the studies. First we focus on the difference between first- and second-generation immigrants. While some studies explicitly mention if their candidates belong to the first or second generation, most studies just mention that applicants have been schooled in the country where the testing is conducted. We treat these minority applicants as second-generation immigrants. As summarized in Table 2, the relative call-back rate for first-generation immigrants on the subgroup level is 1.93 (mean), while it is 1.71 for second-generation immigrants. There is no clear pattern across studies, and no evidence that discrimination would generally be lower for the second generation in

substantive terms. In the multivariate models presented in the supplementary material (S13), the coefficient for the second generation is negative (-0.35 , $p < 0.05$).

Table 2: Relative Call-Back Rates for First- and Second-Generation Applicants

	<i>Mean</i>	<i>Median</i>	<i>N</i>
First generation	1.93	1.64	97
Second generation	1.71	1.46	448
Second generation (explicit only)	1.82	1.62	184

Notes: N indicates the number of cases included, $p > 0.05$

The minority groups selected for testing have become more diverse in recent years, but there are some groups which are included frequently, especially in European correspondence tests. By focusing on specific ethnic groups, we are able to minimize the influence of unobserved variables on call-back rates. We focus on the ethnic groups most commonly studied: Arabs and people of Middle Eastern origin; Chinese; Indians, Pakistani, and Bangladeshi; and Turks.⁴ The results in Table 3 make apparent a clear hierarchy of minority groups: Discrimination is highest for Arabs and people of Middle Eastern origin, followed by Chinese, Indians, Pakistani, and Bangladeshi; it is lowest for Turks. Similar patterns are reported in individual studies where more than one minority group was included. For instance, in Austria Serbs face the lowest relative call-back rate (1.31), followed by Chinese (1.37), Turks (1.46) and Nigerians (1.98) (Weichselbaumer 2015b, see also Booth, Leigh, and Varganova 2012, McGinnity and Lunn 2011). Multivariate regression analysis in the supplementary material suggests that these differences are robust to differences in skill levels (S13). Taken together, the results suggest clear ethnic hierarchies, but hierarchies that are specific to a place and probably time.

Table 3: Relative Call-Back Rates for Specific Groups

<i>Specific Group</i>	<i>Mean</i>	<i>Median</i>	<i>N</i>
Arabs and people of Middle Eastern origin	2.35	2.13	68
Chinese	1.64	1.57	39
Indians, Pakistani, and Bangladeshi	1.77	1.69	42
Turks	1.33	1.23	73

Notes: N indicates the number of cases included, $p < 0.01$

As the issue of racial and ethnic discrimination appeared on the European political agenda at the end of the last century, two EU directives were adopted in record time (Directive 2000/43/EC and 2000/78/EC). Table 4 presents the discrimination rates in the European Union (thus excluding Switzerland and Norway) before and after the adoption of these directives in order to examine the impact of these anti-discrimination policies. Interestingly, the reported level of discrimination has increased since the adoption of the EU directives, with the relative call-back rate rising from 1.36 to 1.96. The observed increase is probably a reflection of the groups included in the correspondence tests and may be due to the fact that most European studies were conducted after the adoption of the directives, but there is certainly no evidence that the EU directives would have led to a direct reduction in discrimination.

⁴ Many US studies focus on Hispanics, but most of them are audit studies rather than correspondence tests.

Table 4: Relative Call-Back Rates in Europe Before and After EU Directives

<i>Time</i>	<i>Mean</i>	<i>Median</i>	<i>N</i>
Before	1.36	1.21	47
After	1.96	1.65	364

Notes: N indicates the number of cases included, $p < 0.01$

Rather than looking at the EU directives, the level of discrimination in German-speaking countries is of particular interest because it allows direct inferences about statistical discrimination. German-speaking countries are known for their extensive application packs, requiring detailed documentation about job candidates. In order to be considered as a serious applicant in German-speaking countries, it is customary to compile an application package that contains not only a cover letter and a CV, but also at least a photograph and school reports for entry level positions such as apprenticeships, or university transcripts and diplomas as well as reference letters from former employers for people with more experience (Kaas and Manger 2012, Schneider, Yemane, and Weinmann 2014, Weichselbaumer 2015b). This amount of detailed information provides employers with more knowledge about candidates than in other contexts, and is thus likely to reduce statistical discrimination. The results in Table 5 suggest that this is the case, with levels of discrimination being lower in German-speaking countries than elsewhere. Multivariate regression analysis in the supplementary material shows that this difference is robust and not just a reflection of the skills level tested (S13). The implications are two-fold. On the one hand, the difference suggests that statistical discrimination indeed plays a role, something that could be addressed with more information or different application packs. On the other hand, the call-back rates in the German-speaking countries suggest that there is a high degree of discrimination even where application packs are more substantial, indicating that statistical discrimination is not the only factor explaining discriminatory behaviour in hiring decisions. In this case we are looking at preferences and attitudes, and remedies are less obvious.

Table 5: Relative Call-Back Rates in German-Speaking Countries

<i>Country</i>	<i>Mean</i>	<i>Median</i>	<i>N</i>
German-speaking	1.43	1.31	119
Other	1.83	1.55	421

Notes: N indicates the number of cases included, $p < 0.01$

In several studies using correspondence tests it is suggested that discrimination is higher in private companies and that the chances of minority applicants to be invited for a job interview are greater in public companies (e.g. Wood et al. 2009, Eid 2012, Midtbøen 2014). Our analysis confirms the assumption that public employers are less likely to discriminate against minority applicants, with the mean call-back rate for private employers at 1.65, and a corresponding call-back rate for the public sector at 1.19. However, the number of studies included is relatively small and further research is needed to confirm this relationship. Public employers bear a special responsibility and are often bound by specific procedures to ensure equal opportunities during employment (e.g. the use of standardized application forms; Wood et al. 2009).

Discussion

Across OECD countries, members of ethnic and racial minority groups face discrimination in the hiring process. Most studies report discrimination of minority groups, and across studies the difference amounts to minority groups having 49 per cent lower odds to be invited for a

job interview than their majority competitor. Looking at relative call-back ratios, members of minority groups need to send around 3 applications for every 2 applications a member of the majority group needs to send in order to be called back for an interview. These patterns of discrimination are relatively robust across countries and economic situations. The fact that discrimination is still prevalent in all countries where testing has been conducted, despite the adoption of anti-discrimination legislation, shows that there is still much room for future research, especially concerning the underlying reasons for discrimination and how the reported differentials come into existence.

For instance, many more correspondence tests focus on male candidates than on female candidates, something in part attributable to the ILO studies of the 1990s (Goldberg, Mourinho, and Kulke 1995, Bovenkerk et al. 1995, Arriijn, Feld, and Nayer 1998, de Prada et al. 1995). Recent Scandinavian studies (e.g. Arai, Bursell, and Nekby 2011, Bursell 2014) suggest that men with foreign names are less likely to be invited for a job interview than women with foreign names. It is unclear whether women are perceived as being lower qualified and thus are considered for lower quality work, or men are discriminated against because they are perceived as more threatening (Bovenkerk 1992). While across studies there appear to be no systematic differences between the discrimination of minority men and minority women, further research in this area is warranted to identify relevant mechanisms, especially because most existing studies were not designed to test the stipulated gender differences.

There is no systematic difference between the relative call-back rates for first- and second-generation applicants, suggesting that taste-based discrimination dominates – second-generation candidates have local qualifications so employers have no need to use ethnicity to guess. As Carlsson (2010, 272) highlighted, “the factor driving discrimination seems to be ethnicity *per se*.” In this case, as Heath and Cheung (2006) emphasize, disadvantage is unlikely to disappear between generations. There is some evidence that levels of discrimination decrease over time, but the lack of a clear substantive difference between the first- and second-generation candidates is problematic in as much as many immigrant integration policies in Western Europe are based on what is perceived as a meritocratic society, where qualifications and language skills should allow for equal chances. This is also the case for the EU directives that do not appear to have lowered discriminatory practices in hiring directly. More research is needed to understand *how* these policies fail to make a dent on discrimination in hiring, including considerations of indirect and lagged effects.

Further evidence for taste-based discrimination comes from the fact that different minority groups fare differently in hiring decisions. Research is necessary to make sense of patterns of ethnic hierarchies, because correspondence tests often contrast more established minority groups with more recent arrivals, an approach recommended by Bovenkerk (1992). As a result, studies may confound different mechanisms. Nonetheless, numerous explanations are provided in the different studies, ranging from ethnic hierarchies, to social distance between the minority groups tested and the majority (e.g. Andriessen et al. 2010). While these all point towards status hierarchies, the differences across countries and time indicate that these hierarchies are neither universal nor purely based on skin-colour. Research linking discriminatory behaviour towards certain immigrant or minority groups with attitudes towards these minority groups would be fruitful to further understand what characteristics of the minority candidates lead to discrimination. For instance, most studies on Arab applicants have been conducted in Scandinavia after 2006, at a time when attitudes towards Arabs have become negative, stereotypes threatening, and Islamophobia widespread (Helbling 2014,

Dolezal, Helbling, and Hutter 2012, Ruedin and Berkhout 2012). While our focus has been on discrimination by employers, (anticipated) discrimination by colleagues and/or customers might also play an important role (Baert and De Pauw 2014).

Evidence that statistical discrimination plays a role comes from German-speaking countries where more extensive application material is the norm and from public sector employers where non-discriminatory hiring practices are often explicitly sought. Discrimination is higher in the private sector and in countries without the extensive application packs commonplace in German-speaking countries. With more information, there is less room for statistical discrimination. Results by Weichselbaumer (2015b) highlight that simply providing more information is no cure for discrimination: the photograph required in German-speaking application packs seems to be used to systematically discriminate against applicants with headscarves (i.e. taste-based discrimination). The situation is somewhat different in the public sector where more careful selection of candidates with regard to diversity may play a role – possibly deliberate demonstrative action to forward political agenda –, aspects perhaps less valued in the private sector where efficiency and productivity may be overruling other concerns. Moreover, standardized application procedures are more widespread in the public sector (Wood et al. 2009). The introduction of standardized procedures and requirements for more detailed application packs or other means to increase the information employers receive – for example by officially vetting foreign qualifications – are readily actionable.

Conclusion

This article provided a meta-analysis of ethnic discrimination in hiring decisions, showing that such discrimination has remained widespread across OECD countries in the last 25 years. Correspondence tests clearly indicate that the discrimination of ethnic and racial minority groups in hiring decisions is still commonplace: Equivalent minority candidates need to send around 50 per cent more applications to be invited for an interview than majority candidates. In a second step we used the variation across studies to draw inferences on the presence of taste-based and statistical discrimination as far as possible. There are many indications that taste-based discrimination remains dominant, although in some instances there is evidence that statistical discrimination also plays a role. This is important since the two forms of discrimination require different interventions: more extensive and standardized procedures seem to reduce statistical discrimination, albeit at the cost of adding bureaucracy, while awareness and consciousness may help reduce taste-based discrimination.

It lies in the nature of a meta-analysis that no detailed examination of discrimination can be provided. We identified much scope for further research, particularly with regard to identifying the underlying mechanisms that lead to discriminatory practices: *how* it is that discrimination takes place. Carefully designed correspondence tests may play a role here, and differences in response rates across minority groups merit further examination, given that these differences seem to follow patterns, albeit complex patterns that seem to depend on time and place. It is likely that insights from work on attitudes towards foreigners and minority groups and other related research can help understand why there are differences in discrimination and which groups are likely to be discriminated against. With discrimination found across countries and time, there seems to be plenty of research material out there, so to speak. What is needed are studies that go beyond showing that ethnic discrimination in hiring exists, to identifying the exact mechanisms and how more equitable hiring can be achieved – unless we want to keep wasting talents.

References

- Adida, Claire L, David D Laitin, and Marie-Anne Valfort. 2010. "Identifying barriers to Muslim integration in France." *Proceedings of the National Academy of Sciences* 107 (52):22384-22390.
- Agerström, J., F. Björklund, R. Carlsson, and D.-O Rooth. 2012. "Warm and Competent Hassan = Cold and Incompetent Eric: A Harsh Equation of Real-Life Hiring Discrimination." *Basic and Applied Social Psychology* 34 (4):359-366. doi: 10.1080/01973533.2012.693438.
- Akintola, Abayomi Rasheed. 2011. "Hiring Discrimination in Racially Diverse Labour Markets: A Cross Country Study." Master, Karstad Business School.
- Allasino, E., E. Reyneri, A. Venturini, and G. Zincone. 2006. *Labour market discrimination against migrant workers in Italy*. Vol. No 67, *ILO International Migration Papers*. Geneva: ILO.
- Andriessen, Iris, Jaco Dagevos, and Jurjen Iedema. 2008. "Onbemind of ongemeten? Over de betekenis van discriminatie als factor in de wekloosheid van niet-westerse allochtonen." *Mens & Maatschappij* 83 (3):279-301.
- Andriessen, Iris, Eline Nievers, Jaco Dagevos, and Laila Faulk. 2012. "Ethnic Discrimination in the Dutch Labor Market: Its Relationship With Job Characteristics and Multiple Group Membership." *Work and Occupations* 39 (3):237-269.
- Andriessen, Iris, Eline Nievers, Laila Faulk, and Jaco Dagevos. 2010. *Liever Mark dan Mohammed?* edited by The Netherlands' Institute for Social Research (SCP). The Hague: The Netherlands' Institute for Social Research (SCP).
- Arai, Mahmood, Moa Bursell, and Lena Nekby. 2011. "The Reverse Gender Gap in Ethnic Discrimination: Employer Priors against Men and Women with Arabic Names." *Working Papers CEB* 11.
- Arceo-Gómez, Eva Olimpia, and Raymundo Campos-Vázquez. 2013. *Race and Marriage in the Labor Market: A Discrimination Correspondence Study in a Developing Country*. In *Serie documentos de trabajo del Centro de Estudios Económicos*: El Colegio de México, Centro de Estudios Económicos.
- Ariel, Barak, Ilanit Tobby-Alimi, Irit Cohen, Mazal Ben Ezra, Yafa Cohen, and Gabriela Sosinski. 2015. "Ethnic and Racial Employment Discrimination in Low-Wage and High-Wage Markets: Randomized Controlled Trials Using Correspondence Tests in Israel." *The Law & Ethics of Human Rights* 9 (1):113-139.
- Arriijn, Peter, Serge Feld, and André Nayer. 1998. *Discrimination in access to employment on grounds of foreign origin: the case of Belgium*. Geneva: International Labour Office, Conditions of Work Branch.
- Arrow, Kenneth J. 1973. "The Theory of Discrimination." In *Discrimination in Labor Markets*, edited by O. Ashonfelter and A. Rees, 15-42. Princeton: Princeton University Press.
- Arrow, Kenneth J. 1998. "What has economics to say about racial discrimination?" *The Journal of Economic Perspectives*:91-100.
- Attström, Karin. 2007. *Discrimination against native Swedes of immigrant origin in access to employment*, *ILO International Migration Papers* Geneva: International Labor Organization.
- Baert, Stijn, Bart Cockx, Niels Gheyle, and Cora Vandamme. 2013. *Do employers discriminate less if vacancies are difficult to fill? Evidence from a field experiment*. Discussion Paper Series, Forschungsinstitut zur Zukunft der Arbeit.

- Baert, Stijn, and Ann-Sophie De Pauw. 2014. "Is ethnic discrimination due to distaste or statistics?" *Economics Letters* 125 (2):270-73.
- Becker, Gary. 1957. *The Economics of Discrimination*. Chicago: University of Chicago Press.
- Bendick, Marc. 2007. "Situation testing for employment discrimination in the United States of America." *Horizons stratégiques* (3):17-39.
- Bendick, Marc, Charles W Jackson, Victor A Reinoso, and Laura E Hodges. 1991. "Discrimination against Latino job applicants: A controlled experiment." *Human Resource Management* 30 (4):469-484.
- Bendick, Marc, and Ana P. Nunes. 2012. "Developing the Research Basis for Controlling Bias in Hiring." *Journal of Social Issues* 68 (2):238-262. doi: 10.1111/j.1540-4560.2012.01747.x.
- Bertrand, Marianne, and Sendhil Mullainathan. 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. doi: 10.2307/3592802.
- Blank, Rebecca M, Marilyn Dabady, and Constance F Citro. 2004. *Measuring racial discrimination*. Washington, DC: National Academies Press.
- Blommaert, Lieselotte, Marcel Coenders, and Frank van Tubergen. 2014. "Discrimination of Arabic-Named Applicants in the Netherlands: An Internet-Based Field Experiment Examining Different Phases in Online Recruitment Procedures." *Social Forces* 92 (3):957-982. doi: 10.1093/sf/sot124.
- Booth, Alison L, Andrew Leigh, and Elena Varganova. 2012. "Does ethnic discrimination vary across minority groups? Evidence from a field experiment." *Oxford Bulletin of Economics and Statistics* 74 (4):547-573.
- Bovenkerk, Frank. 1992. *Testing discrimination in natural experiments: a manual for international comparative research on discrimination on the grounds of "race" and ethnic origin*. Geneva: International Labour Office.
- Bovenkerk, Frank, MJI Gras, Dhurender Ramsoedh, M Dankoor, and A Havelaar. 1995. *Discrimination against migrant workers and ethnic minorities in access to employment in the Netherlands*. Geneva: Employment Department, International Labour Office.
- Bursell, Moa. 2007. "What's in a name? A field experiment test for the existence of ethnic discrimination in the hiring process." *SULCIS Working Paper* (2007:7).
- Bursell, Moa. 2014. "The Multiple Burdens of Foreign-Named Men—Evidence from a Field Experiment on Gendered Ethnic Hiring Discrimination in Sweden." *European Sociological Review* 30 (3):99-409.
- Carlsson, Magnus. 2010. "Experimental Evidence of Discrimination in the Hiring of First- and Second-generation Immigrants." *Labour* 24 (3):263-278.
- Carlsson, Magnus, Luca Fumarco, and Dan-Olof Rooth. 2015. "Does Labor Market Tightness Affect Ethnic Discrimination in Hiring?" *Linnaeus University Labour Market and Discrimination Studies Centre Working Paper* 2015:1.
- Cédiey, E., and F. Foroni. 2007. *Les discriminations à raison de "l'origine" dans les embauches en France. Une enquête nationale par tests de discrimination selon la méthode du Bureau International du Travail*. Genève: Bureau International du Travail.
- Daniel, William Wentworth. 1968. *Racial discrimination in England: based on the PEP report*. Vol. 1084. Middlesex: Penguin.

- de Prada, Miguel Angel, W Actis, C Pereda, and R Perez Molina. 1995. *Labour market discrimination against migrant workers in Spain*. Vol. 9, *International Migration Papers* Geneva: International Labour Office, Employment Department.
- Dechief, Diane, and Philip Oreopoulos. 2012. "Why do some employers prefer to interview Matthew, but not Samir? New Evidence from Toronto, Montreal, and Vancouver." *Canadian Labour Market and Skills Researcher Network Working Paper* No. 95.
- Decker, Scott H., Natalie Ortiz, Cassia Spohn, and Eric Hedberg. 2015. "Criminal stigma, race, and ethnicity: The consequences of imprisonment for employment." *Journal of Criminal Justice* 43 (2):108-121. doi: <http://dx.doi.org/10.1016/j.jcrimjus.2015.02.002>.
- Diekmann, Andreas, Ben Jann, and Mathias Näf. 2014. "Wie fremdenfeindlich ist die Schweiz? Fünf Feldexperimente über prosoziales Verhalten und die Diskriminierung von Ausländern in der Stadt Zürich und der Deutschschweiz." *Soziale Welt* 65:185-199.
- Dolezal, Martin, Marc Helbling, and Swen Hutter. 2012. "Debating Islam in Austria, Germany and Switzerland: Ethnic Citizenship, Church-State Relations and Right-Wing Populism." *West European Politics* 33:171-190.
- Drydakis, Nick. 2012. "Estimating ethnic discrimination in the labour market using experimental data." *Southeast European and Black Sea Studies* 12 (2):335-355. doi: 10.1080/14683857.2012.686016.
- Drydakis, Nick, and Minas Vlassis. 2010. "Ethnic Discrimination in the Greek Labour market: Occupational Access, Insurance Coverage and Wage Offers." *The Manchester School* 78 (3):201-218. doi: 10.1111/j.1467-9957.2009.02132.x.
- Duguet, Emmanuel, Loïc du Parquet, Yannick L'Horty, and Pascale Petit. 2015. "New Evidence of Ethnic and Gender Discriminations in the French Labor Market Using Experimental Data: A Ranking Extension of Responses from Correspondence Tests." *Annals of Economics and Statistics* (117-118):21-39. doi: 10.15609/annaeconstat2009.117-118.21.
- Duguet, Emmanuel, Yannick L'Horty, L du Parquet, Pascale Petit, and Florent Sari. 2010. "Discriminations à l'embauche des jeunes franciliens et intersectionnalité du sexe et de l'origine: les résultats d'un testing." *Rapport pour l'Agence nationale pour la Cohésion Sociale et l'Egalité des chances (ACSé), Document de recherche de l'EPEE*, .
- Eid, Paul. 2012. *Mesurer la discrimination à l'embauche subie par les minorités racisées: résultats d'un testing" mené dans le grand Montréal*. Quebec: Commission des droits de la personne et des droits de la jeunesse.
- Esmail, Aneez, and Sam Everington. 1993. "Racial discrimination against doctors from ethnic minorities." *BMJ: British Medical Journal* 306 (6879):691.
- Esmail, Aneez, and Sam Everington. 1997. "Asian Doctors Are Still Being Discriminated Against." *BMJ: British Medical Journal* 314 (7094):1619. doi: 10.2307/25174744.
- Fetzer, Joel S. 2013. "The paradox of immigration attitudes in Luxembourg: A pan-European comparison." In *Immigration and Public Opinion in Liberal Democracies*, edited by Gary P. Freeman, Randall Hansen and David L. Leal. London: Routledge.
- Fibbi, Rosita, Bülent Kaya, and Etienne Piguet. 2003. *Le passeport ou le diplôme ? Etude des discriminations à l'embauche des jeunes issus de la migration, PNR 43 Rapport de recherche au FNS*. Neuchâtel: Forum suisse pour l'étude des migrations et de la population.

- Ford, Robert. 2015. "Who Should We Help? An Experimental Test of Discrimination in the British Welfare State." *Political Studies*. doi: 10.1111/1467-9248.12194.
- Gaddis, S Michael. 2014. "Discrimination in the Credential Society: An Audit Study of Race and College Selectivity in the Labor Market." *Social Forces*:1-29.
- Goldberg, Andreas, Dora Mourinho, and Ursula Kulke. 1995. *Labour market discrimination against foreign workers in Germany*: International Labour Office, Employment Department.
- Heath, Anthony, and Sin Yi Cheung. 2006. Ethnic penalties in the labour market: Employers and discrimination. Norwich: Department for Work and Pensions.
- Heckman, James J. 1998. "Detecting discrimination." *The Journal of Economic Perspectives*:101-116.
- Heckman, James J, and Peter Siegelman. 1993. "The Urban Institute Audit Studies: Their Methods and Findings." In *Clear and Convincing Evidence: Measurements of Discrimination in America*, edited by Michael Fix and Raymond Struyk, 187-258. Washington, DC: The Urban institute Press.
- Helbling, Marc. 2014. "Opposing Muslims and the Muslim Headscarf in Western Europe." *European Sociological Review* 30 (2):242-257. doi: 10.1093/esr/jct038.
- Jackson, Michelle, and D. R. Cox. 2013. "The principles of experimental design and their application in sociology." *Annual Review of Sociology* 39:27-49.
- Jacquemet, Nicolas, and Constantine Yannelis. 2012. "Indiscriminate discrimination: A correspondence test for ethnic homophily in the Chicago labor market." *Labour Economics* 19 (6):824-832. doi: <http://dx.doi.org/10.1016/j.labeco.2012.08.004>.
- Jowell, Roger, and Patricia Prescott-Clarke. 1970. "Racial discrimination and white-collar workers in Britain." *Race & Class* 11 (4):397-417.
- Kaas, Leo, and Christian Manger. 2012. "Ethnic Discrimination in Germany's Labour Market: A Field Experiment." *German Economic Review* 13 (1):1-20. doi: 10.1111/j.1468-0475.2011.00538.x.
- Klingeren, Marijn van, Hajo G. Boomgaarden, Rens Vliegthart, and Claes H. de Vreese. 2014. "Real World is Not Enough: The Media as an Additional Source of Negative Attitudes Toward Immigration, Comparing Denmark and the Netherlands." *European Sociological Review*:jcu089. doi: 10.1093/esr/jcu089.
- Lodder, LeeAnn, Scott McFarland, and Diana White. 2003. "Racial Preference and Suburban Employment Opportunities." *Chicago: Legal Assistance Foundation of Chicago*.
- McGinnity, Frances, and Peter D Lunn. 2011. "Measuring discrimination facing ethnic minority job applicants: an Irish experiment." *Work, Employment & Society* 25 (4):693-708.
- Midtbøen, Arnfinn H, and Jon Rogstad. 2012. "Discrimination. Methodological controversies and sociological perspectives on future research." *Nordic Journal of Migration Research* 2 (3):203-212.
- Midtbøen, Arnfinn H. 2014. "Discrimination of the Second Generation: Evidence from a Field Experiment in Norway." *Journal of International Migration and Integration*:1-20. doi: 10.1007/s12134-014-0406-9.
- Neumark, David. 2012. "Detecting Discrimination in Audit and Correspondence Studies." *Journal of Human Resources* 47 (4):1128-1157.
- Nunley, John M, Adam Pugh, Nicholas Romero, and R Alan Seals. 2014. "An Examination of Racial Discrimination in the Labor Market for Recent College Graduates: Estimates from the Field." *Working Paper Series, Department of Economics, Auburn University* (AUWP 2014-06).

- Oreopoulos, Philip. 2011. "Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes." *American Economic Journal: Economic Policy* 3 (4):148-171. doi: 10.2307/41330445.
- Pager, Devah. 2007. "The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future." *Annals of the American Academy of Political and Social Science* 609:104-133. doi: 10.2307/25097877.
- Pager, Devah, and Hana Shepherd. 2008. "The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets." *Annual Review of Sociology* 34 (1):181-209.
- Parrillo, Vincent N., and Christopher Donoghue. 2005. "Updating the Bogardus social distance studies: a new national survey." *The Social Science Journal* 42:257-271. doi: 10.1016/j.soscij.2005.03.011.
- Pecoraro, Marco, and Didier Ruedin. 2015. "A Foreigner Who Doesn't Steal My Job: The Role of Unemployment Risk and Values in Attitudes towards Equal Opportunities." *International Migration Review*. doi: 10.1111/imre.12162.
- Petticrew, Mark, and Helen Roberts. 2006. *Systematic Reviews in the Social Sciences: A Practical Guide*. New York: Wiley-Blackwell.
- Pettigrew, T., and L. Tropp. 2006. "A meta-analytic test of intergroup contact theory." *Journal of Personality and Social Psychology* 90:751-83.
- Phelps, Edmund S. 1972. "The statistical theory of racism and sexism." *The American Economic Review* 62 (4):659-661.
- Quillian, Lincoln. 2006. "New Approaches to Understanding Racial Prejudice and Discrimination." *Annual Review of Sociology* 32:299-328. doi: 10.2307/29737741.
- Riach, Peter A, and Judith Rich. 1991. "Testing for racial discrimination in the labour market." *Cambridge Journal of Economics*:239-256.
- Riach, Peter A, and Judith Rich. 2002. "Field Experiments of Discrimination in the Market Place." *The Economic Journal* 112 (483):F480-F518. doi: 10.2307/798458.
- Rich, Judith. 2010. "Measuring Discrimination: What Do Field Experiments of Markets Tell Us?" In *Making Equality Count: Irish and International Research Measuring Equality and Discrimination*, edited by Laurence Bond, Frances McGinnity and Helen Russell, 48-63. Dublin: The Liffey Press.
- Rich, Judith. 2014. What Do Field Experiments of Discrimination in Markets Tell Us? A Meta Analysis of Studies Conducted since 2000. In *Discussion Paper Series*. Bonn: IZA
- Ruedin, Didier, and Joost Berkhout. 2012. "Patterns of claims-making on civic integration and migration in Europe: Are Muslims different?" *SOM Working Paper* 8:1-30.
- Schneider, Jan, Ruta Yemane, and Martin Weinmann. 2014. Diskriminierung am Ausbildungsmarkt: Ausmass, Ursachen und Handlungsperspektiven Berlin Forschungsbereich beim Sachverständigenrat deutscher Stiftungen für Integration und Migration (SVR).
- van der Brug, Wouter, Gianni D'Amato, Joost Berkhout, and Didier Ruedin. 2015. *The Politicisation of Migration*. Abingdon: Routledge.
- Veenman, Justus. 2010. "Measuring labor market discrimination: An overview of methods and their characteristics." *American Behavioral Scientist* 53 (12):1806-1823.
- Weichselbaumer, Doris. 2015a. "Beyond the veil: Discrimination against female migrants wearing a headscarf in Germany." *University of Linz Working Paper*.

- Weichselbaumer, Doris. 2015b. "Discrimination against migrants in Austria: An experimental study." *Working Paper, Department of Economics, Johannes Kepler University of Linz* (No. 1501).
- Weichselbaumer, Doris, and Rudolf Winter-Ebmer. 2005. "A Meta-Analysis of the International Gender Wage Gap." *Journal of Economic Surveys* 19 (3):479-511.
- Widner, Daniel, and Stephen Chicoine. 2011. "It's All in the Name: Employment Discrimination Against Arab Americans." *Sociological Forum* 26 (4):806-823. doi: 10.1111/j.1573-7861.2011.01285.x.
- Wolf, Fredric Marc. 1986. *Meta-analysis: quantitative methods for research synthesis, Quantitative Applications in the Social Sciences (QASS)*. London: Sage Publications.
- Wood, Martin, John Hales, Susan Purdon, Tanja Sejersen, and Oliver Hayllar. 2009. *A Test for Racial Discrimination in Recruitment Practice in British Cities*. Norwich: Department for Work and Pensions.

Paper III:

Research Ethics in Correspondence Testing: an Update

First Published as NCCR Working Paper:

Zschirnt, E. (2016) “Revisiting Ethics in Correspondence Testing ». NCCR Working Papers, 8. 1-26.

Submitted to Research Ethics

Accepted with minor revisions (revisions already included in this version)

Research Ethics in Correspondence Testing: an Update

Abstract

Correspondence testing researching discrimination in the market place has become more widespread and the use of internet applications has allowed researchers to send greater numbers of applications. While questions of research ethics always arise when planning a correspondence test, the issue receives relatively little attention in published correspondence tests. This paper addresses the question of ethics in correspondence testing in the age of ready internet access. It focuses on the ethical issues that arise in correspondence testing, looking at potential problems (regarding voluntary participation, informed consent, deception, entrapment of employers, employers' rights) and possible solutions, and technical challenges. European country examples show that the ethical questions raised in correspondence testing have to be renegotiated depending on the national context. The paper argues that correspondence testing, if planned carefully and executed responsibly, can meet most of the ethical requirements of social sciences ethics guidelines.

Keywords

Research Ethics, Correspondence Testing, Discrimination, Field Experiments

Introduction

Open discrimination has decreased with the adoption of anti-discrimination legislation. Yet, discrimination continues to occur in more subtle and hidden ways. Field experiments in the market place, such as audit and correspondence studies, provide important information on the extent of systematic differential treatment in the labour market and are currently seen as the best way to measure discrimination. The methodology of correspondence testing exists since the late 1960s (Jowell & Prescott-Clarke, 1970) and correspondence tests relying on fictitious candidates have been carried out most prominently in the labour and housing market on various grounds of discrimination, e.g. ethnicity/race, gender, disability, or sexual orientation¹. Questions of research ethics arise in the planning stages of such an experiment, because correspondence testing relies on covert research where participants are not aware that they are part of an experiment. While this violates core research principles such as informed consent and voluntary participation that have been enshrined in ethical guidelines across disciplines, sociological research ethics guides argue that covert research can be justified under certain strict provisions.

Despite the growing number of correspondence tests on various grounds of discrimination, ethical questions are rarely thoroughly addressed in published correspondence tests on ethnic discrimination in hiring decisions. They were for example largely left out of the recent book on the methodology of audit studies edited by Gaddis (2018b). Most authors only refer to Riach and Rich (2004a), who discuss the deceptive nature of field experiments in detail, and there are rarely any references to Banton (1997) who focuses on the rejection of Swedish research proposals for a correspondence test on ethical grounds. Contributions on research ethics in the US context by Edley Jr. (1993) or Fix, Galster, and Struyk (1993) on the Urban Institute Studies or by Pager (2007) who discusses ethical issues concerning in-person audit studies are also hardly mentioned. Similarly, the teaching case presented by Connor (2000) that presents ethical challenges and arguments against a proposed testing that were voiced by an Internal Review Board (IRB) is also largely ignored. In more recent methodology publications, Lahey and Beasley (2018) only very briefly mention ethical issues related to the number of correspondence that researcher send to employers, while Crabtree (2018) explicitly states that getting IRB approval is not discussed in his chapter on the steps included in conducting audit studies.

Does this mean that a consensus over the legitimacy of using correspondence testing has emerged or, is it rather the case that most researchers focus more thoroughly on the questions of ethics in the preparation of their experiments, but do not include further information in the publication of their results? Almost fifteen years have passed since the publication of Riach and Rich's article and since then correspondence testing has become more widespread and diverse². Furthermore, researchers increasingly use the internet to find vacancies and send applications. This greater number of CVs being sent out at relatively low costs is a great opportunity for researchers, because it allows to test numerous new variables or combinations of variables that require larger samples and to obtain more nuanced results. However, it also means that more employers are affected by a correspondence test. It is therefore worth to revisit the question of ethics, to account for recent technological

¹ An overview of the historical development of field experiments with a focus on ethnic or racial discrimination in the labour market can be found in e.g. Cherry and Bendick (2018), Gaddis (2018a), or Zschirnt (2016).

² For recent reviews and meta-analyses of correspondence tests see e.g. Baert (2018), Gaddis (2018a), Bertrand and Duflo (2017), Quillian, Pager, Hexel, and Midtbøen (2017), Neumark (2016), Zschirnt and Ruedin (2016), or Rich (2014).

developments, and to address challenges that the computerisation of field experiments has created.

This paper investigates ethical concerns involved in correspondence testing and argues that the thorough preparatory work required in the planning stages of a correspondence test can mitigate most ethical concerns frequently voiced by opponents of the technique. Looking at (mostly) sociological research ethics guidelines, experiments on discrimination have been acknowledged as instances where conducting covert research can be justified under strict circumstances. The fact that correspondence tests have been approved by ethical commissions³ in numerous countries shows that this methodology has been recognized as, so far, the best way to measure discrimination in hiring decisions and that ethical concerns can be minimized.

Correspondence testing – an introduction to the technique

Field experiments, of which correspondence tests are only one method, allow researchers to observe behaviour in real life situations. One of the earliest field experiments was conducted by LaPiere (1934), who travelled with two Chinese friends through the US and documented if they were accepted as guests in hotels or restaurants. After half a year had passed he contacted these establishments again and asked if they would accept Chinese customers. While most establishments had previously welcomed them, almost all expressed their refusal to do so in a written questionnaire. This shows that certain behaviour cannot be detected by simply asking people “How would you react?”, but that it can be observed in real life situations and that considerable differences between the outcomes can occur.

Since the 1960s, field experiments have been used to study the phenomenon of discrimination in the labour market, making use of in-person audits and written correspondence testing (Cherry & Bendick, 2018; Gaddis, 2018a; Zschirnt, 2016). In recent years, and in European countries in particular, correspondence testing has been deemed to be one of the most suitable methods to identify and measure discrimination in the labour market and in hiring decisions in particular (Schneider, Yemane, & Weinmann, 2014, p. 14). In a correspondence test researchers apply in writing to real-life vacancies and present potential employers with (at least) two substantially equal and thus interchangeable candidates, who differ only in the characteristic to be studied, e.g. ethnicity. The companies included in the experiment are not named and the exclusive focus on anonymised aggregated data guarantees the protection of participants’ privacy.

Correspondence tests are carefully planned experiments and most published experiments contain a detailed research design section. These careful considerations before the actual testing takes place show that correspondence tests require a lot of time and detailed preparation before they can be conducted. The recently published book by Gaddis (2018b) provides valuable discussions and detail-oriented chapters on audit studies and is a helpful and comprehensive guide for both experienced researchers as well as those just getting acquainted with the technique. Once this labour intensive preparatory work has been done, Bendick and Nunes point out, that correspondence tests are an “innovative research technique [...] that offers laboratory-like controlled conditions in quasi-experiments in real-world hiring situations” (2012, p. 238). The preparatory work should, however, not be limited to the

³ Unfortunately, information about the composition of ethical commissions or IRBs is often not provided. A rare example is Connor (2000) who gives detailed information about the members of the IRB that rejected the research project and points out the problems with having an all-white review panel deciding on research proposals on racial discrimination.

research design, but also include the inevitably arising discussion of ethical issues. Correspondence tests constitute an ethical challenge, especially concerning the responsibilities of the researcher towards the research participants. It is therefore not surprising that correspondence tests are often met with scepticism regarding their compliance with research ethics standards, and in particular the criteria of informed and voluntary consent and the use of deception.

Ethical issues in correspondence testing

Researching hiring discrimination using correspondence tests lies at the intersection of sociology and economic research, and developments in the fields of research ethics in the social sciences also influence researchers planning correspondence tests. The most frequently voiced ethical objections are now addressed in detail.

Objection: Correspondence testing infringes the principles of voluntary participation and informed consent

Two of the fundamental ethical principles underlying research across disciplines are that “potential research subjects should be given the opportunity to refuse participating in research” (Dench, Iphofen, & Huws, 2004, p. 56) and that they make this decision based on comprehensive and accurate information (p. 63). The insistence on voluntary participation and informed consent can be traced back to medical experiments conducted by the Nazi regime and the subsequent development of the Nuremberg Code of 1947, which above all emphasises the importance of voluntary and informed consent of research participants (Israel, 2015, p. 27)⁴. This emphasis on voluntary and informed consent has been extended far beyond medical experiments and is also applied in social science research.

Research ethics guidelines both on the supranational level, such as the non-binding EU Code of Ethics (Dench et al., 2004), and on the national level emphasise the importance of ensuring voluntary participation and informed consent of research subjects since “the consent requirement is intended to prevent invasions of personal integrity” (National Committees for Research Ethics in Norway (NESH), 2006, p. 13). However, sociological research ethic guidelines recognize that there are exceptions, where research would not be possible if voluntary and informed consent had to be obtained first.

“In certain cases, participant’s freedom and self-determination can be respected even though consent has not been obtained beforehand. [...], exceptions [...] can be made in certain cases in situations in which the research does not imply physical contact with the research subjects, where the data being processed is not particularly sensitive, and *where the utility value of the research clearly exceeds any disadvantages that might be inflicted on the subjects.*” (National Committees for Research Ethics in Norway (NESH), 2006, p. 14, emphasis added)⁵

⁴ For a historical discussion on the development of research ethics see: Israel (2015), Dingwall (2012), Dench et al. (2004), Hunter (2010), Wassenaar (2006) or Nakray (2016).

⁵ The case of Norway is discussed so prominently, because it is one of the rare studies where detailed information about the ethical approval process is available.

It lies in the nature of correspondence testing that employers cannot make a voluntary choice to participate in a scientific experiment and it is not possible to inform them of their participation and the goal of the study. Starting with Bovenkerk in the 1990s, researchers have recognised that not informing the research subjects conflicted with their right to provide or refuse their consent (1992, p. 33). Yet, according to Bovenkerk three reasons justify breaking the principle of informed consent: first, hiring decisions are not a private matter and hiring discrimination is unlawful, second, if field experiments are carefully prepared and carried out there is almost no detrimental effect on the employers tested, and third, it is normal hiring decision that are observed and researchers “do not lure employers into a situation in which they are enticed to deviate from their normal course of action” (Bovenkerk, 1992, pp. 33-34).

Breaking the principle of informed consent has also been addressed by other researchers who argued that breaking it “is a crucial feature of this type of research, as informing participants would invalidate the experiment” (Blommaert, Coenders, & van Tubergen, 2014, p. 964). This has also been recognized in national laws. In the case of Sweden, Bursell (2007) refers to Swedish law which states that “research without the participant’s informed consent can still be carried out” if the research meets certain conditions, such as being “of high societal importance” (p.9). Looking at the US, Pager (2007) also refers to legal provisions on conducting research without obtaining informed consent:

“a human subjects institutional review board (IRB) ‘may ... waive ... informed consent provided (1) the research involves no more than minimal risk to human subjects; (2) the waiver or alteration will not adversely affect the rights and welfare of the subjects; (3) the research could not practicably be carried out without the waiver or alteration; and (4) whenever appropriate, the subjects will be provided with additional information after participation.’ Each of these conditions can arguably be satisfied in the context of audit studies of discrimination.” (Pager, 2007, p. 126)

While Pager explicitly refers to audit studies, these conditions are also met in correspondence tests of discrimination. The methodology of correspondence testing is usually judged admissible under certain (strict) provisions due to the higher societal interest to measure discrimination, even if it breaks the principles of voluntary participation and informed consent.

In order to mitigate the potential negative effects of breaking the principle of informed consent, the British Sociological Association proposes that this consent could be obtained post-hoc (2002, p. 4). While Midtbøen (2014) decided to contact only some of his unknowing research participants to recruit participants for follow-up interviews, I am only aware of one study in which all participants were systematically informed post-hoc (Liebkind, Larja, & Brylka, 2016). However, Pager (2007) argued that “for human resource personnel or managers who are thought to be discriminating, the consequences may be more serious than if no attention were brought to the audit whatsoever” (p.127). The research team of the Expert Council of German Foundations on Integration and Migration (SVR) decided against informing their participants. They reasoned that informing employers post-hoc would not improve the chances of minority applicants in the future and that it might pose a problem for further research by making the technique too well-known. It should still be possible for other researchers to conduct correspondence tests without a too big awareness of the methodology among employers. Following Pager’s argument, the SVR’s research team claimed that obtaining post-hoc consent could potentially prove problematic for the employees responsible

for the hiring decisions. Thus, in order to limit these potential damages incurred by individuals, it can be argued against seeking post-hoc consent. Furthermore, informing unknowing participants post-hoc will also take more time away from the HR personnel. Finally, informing participants post-hoc could give them the possibility to try to sue researchers or the ethical approval bodies and researchers should also avoid causing harm to themselves or their colleagues.

Another option would be informing employers via the media that correspondence testing will take place. Yet, Swedish research showed that employers did not change their hiring behaviour after extensive media coverage of a correspondence testing had occurred (Carlsson & Rooth, 2012). There is, however, first research on discrimination in the Belgium rental housing market that indicates that informing landlords before an experiment is being conducted can lower rates of discrimination (Van der Bracht & Verhaeghe, 2017). This effect was only found for commercial landlords and it is unclear if this is a durable and long-term effect.

Objection: Researchers are deceiving their research participants

The above-mentioned principle of informed consent includes that participants make their choice based on accurate information. But are there situations in which it is acceptable to deceive participants? Dench et al. argue in the EU Code of Ethics that “there are varying debates about whether deception is ever acceptable” and that “the conclusions vary depending on the methodological, philosophical and moral stance” (2004, p. 68) of the researchers.

Correspondence testing relies on the deception of research subjects, since employers are presented with fictitious applicants who pretend to be real candidates. This deceptive nature of field experiments has been the focus of Riach and Rich’s (2004a) seminal article where it featured prominently in the title “Deceptive field experiment – Are they ethical?”. They start their argument saying that field experiments “constitute an unequivocal procedure for charting, over the time, the effectiveness, or otherwise, of equal opportunity legislation” (2004, p. 458). They then turn towards the context in which the deception of employers occurs. Using Bovenkerk’s argument that the action performed by the researcher is “a non-genuine transaction performed in a manner which is not infrequent in the labor market” (1992, p. 34), Riach and Rich elaborate on the notion that testing “takes place in an arena where deception is a regular and acknowledged activity” (2004a, p. 461). They justify the deception of employers because

“a lack of veracity is endemic in these markets; [...] great harm is done to the social fabric by discriminatory practices in such markets; [...] minimal inconvenience is imposed on the entrepreneurs in the experiment, and [...] the technique provides evidence with a degree of accuracy and transparency which is not available from any other procedure” (Riach & Rich, 2004a, p. 463).

Similarly Edley argued that the use of testing was justified, because “the moral costs of deception are outweighed by the great benefit of developing a clearer understanding of the social disease” (1993, p. 378). The deceptive nature of field experiments is thus seen as necessary to obtain information about the socially harmful practice of discrimination.

One of the strongest arguments why the use of deception in research may be its “resemblance to an accepted method for gathering evidence for the enforcement of anti-discrimination law” (Banton, 1997, p. 416), which many courts, including the US Supreme

court, have endorsed (Banton, 1997; Pager, 2007). While some methodological differences exist between testing for research where employers are only sampled once and testing for law enforcement purposes, where one employer is tested multiple times (Pager, 2007), the methodology was initially developed in the 1960s/1970s in response to the adoption of anti-discrimination laws and to monitor their effectiveness (Pager & Western, 2012). Cherry and Bendick (2018) as well as Boggs (1998) provide great overviews of the development of scholarship and activism in the fight against (mostly) housing discrimination in the US. In 1982, the US Supreme Court strengthened the position of these scholars and activists in *Havens Realty Corp vs. Coleman*⁶, when it gave its approval for this methodology. Over the years, US courts have confirmed the legal standing of testers, and “broaden[ed] their endorsement of this methodology” (Pager, 2007, p. 127). Even though these cases concerned testing for legal reasons and not research, Pager argued that “implicit in these holdings [...] is the belief that the misrepresentation involved in testing is worth the unique benefit this practice can provide in uncovering discrimination and enforcing civil rights laws” (p.127). Similarly, the endorsement of the methodology shows that deception is seen as regrettable but unavoidable:

“ [...] we have long recognized that this requirement of deception was a relatively small price to pay to defeat racial discrimination. The evidence provided by testers [...] is a major resource in society’s continuing struggle to eliminate the subtle but deadly poison of racial discrimination.” (Boggs, Sellers, & Bendick Jr, 1993, p. 367)

The legal situations, e.g. if results obtained by testing are accepted in courts, vary depending on the national context. The case of the US is the best known, yet evidence obtained through testing is also recognized in discrimination cases in several European countries. While Rorive (2009) provides a good overview of testing in eleven European countries (Belgium, Czech Republic, Denmark, Finland, France, Hungary, Latvia, The Netherlands, Slovakia, the United Kingdom, and Sweden), van der Plancke (2007) and Calvès (2007) focus on the situation in Belgium and French courts respectively.

Researchers often provide the reasoning of the courts next to guidelines by professional scientific organisations to justify the use of audit or correspondence tests and the element of deception included in this methodology when applying for IRB approval.

Objection: Correspondence Testing can have negative consequences for employers who unwillingly participated in the experiment

Another principle that researchers should adhere to, is the “principle of no harm”. Research should not harm research participants, researchers themselves, or future researchers. Opponents of correspondence tests have objected that researchers try to trap employers and catch them in unlawful behaviour, that employers suffer a loss of time by being included in an experiment, that employers’ privacy is being breached and that an employer’s reputation might suffer from the unwilling participation in a correspondence test.

The first argument brought forward is that researchers try to trap employers if the experiment encourages research subjects to behave illegally. This argument and the fear that researchers might be held liable for such an entrapment has already been addressed in the 1990s by Bovenkerk (1992) and Edley Jr. (1993). According to Bovenkerk this “concern is

⁶ *Havens Realty Corp. v. Coleman*, 455 U.S. 363 (1982)

ill-conceived as discriminating employers break the legal rules probably more than the researcher does” (1992, p. 34). Most importantly, researchers only observe normal hiring practices, they do not lure employers into a trap of acting in a way that they would not have under different circumstances.

Secondly, opponents of correspondence testing argue that employers suffer from a loss of time by assessing fictitious applicants. This argument has e.g. been brought forward by the former US house-speaker Newt Gingrich who argued against funding for the Equal Employment Opportunity Commission (EEOC), because “the use of testers [...] causes innocent businesses to waste resources” (Gingrich, 1998). Researchers acknowledge that assessing additional fictitious applications may pose a burden on the employers’ time (e.g. Pager, 2007; Pager & Western, 2012) and most correspondence tests limit this burden by considering an employer only once, even if more matching vacancies are published. Finally, invitations to interviews are quickly and politely declined in order to keep the application process as normal as possible for genuine applicants. The loss of time should thus be considered minimal (e.g. Wood, Hales, Purdon, Sejersen, & Hayllar, 2009). Furthermore, it is assumed that employers do not spent much time on the initial screening of applications.

A third argument used against correspondence testing is the breach of employers’ privacy. However, Bovenkerk claims that “there is no question of breaking legitimate expectations of privacy. Hiring is not an entirely private matter” (1992, p. 33). He further argues that providing equal opportunities in the hiring process is in the public interest and that discrimination in these public fields has been declared unlawful. This argument is supported by Fix, Galster and Struyk who point out that the “behaviors that have been monitored [...] involve public, commercial, or professional acts. In most instances there has been a special invitation issued to the public – via a published add for a job, apartment, or loan” (1993, p. 16). Furthermore, researchers only gather data on employers that is publicly available.

Finally, concerns are voiced regarding the reputation of enterprises and possible negative effects of being part of a correspondence test. Pager emphasises that “efforts must be taken to protect employer identities so that even associations with a study on discrimination cannot be made” (2007, p. 127). Most studies point out that data is anonymized and only accessible to the core research team. Furthermore, the fact that data is only analysed aggregately also helps to avoid inference about individual employers. Finally, correspondence testing for research is not interested in accusing individual employers of discriminatory behaviour, but in reporting trends in discrimination patterns in a society.

Correspondence testing as an example of covert research

By the nature of the research design, in which employers are not aware that an experiment is being conducted, correspondence tests are an example of covert research. This becomes apparent in the definition provided by the non-binding *EU Code of Ethics for Socio-Economic Research*:

“By definition, covert research means that participation is not voluntary and participants are not able to give informed consent. To some researcher this is unacceptable. Others argue that, in some circumstances, covert research is the only way in which the necessary information can be collected or difficult situations researched.” (Dench et al., 2004, p. 12)

Numerous professional associations have acknowledged the need for deception when it is absolutely necessary. Information on the conditions under which the use of deception is justifiable are included in the ethical guidelines of i.e. the American Sociological Association (2018), the British Sociological Association (2002), the American Psychological Association (2017), the National Committees for Research Ethics in Norway (NESH) (2006), the German Sociological Association (Deutsche Gesellschaft für Soziologie, 2017), or the EU's Respect Project on professional and ethical codes for socio-economic research in the information society. The British Sociological Association for example argues that while "there are serious ethical and legal issues [...] the use of covert methods may be justified in certain circumstances" (2002, p. 4). It points out that covert research violates the principle of informed consent and may violate the privacy of research subjects, making it a method that should only be used as a last resort if it is impossible to obtain information using other methods. The *EU Code of Ethics for Socio-Economic Research* also addresses the question of deception in covert research:

"If it is only possible to obtain information through covert research (for example, studies of violent, criminal or subversive groups, or of fraudulent or *discriminatory practices*) how can the researcher balance the need for deception against the value to society of conducting the research?" (Dench et al., 2004, p. 64, emphasis added)

They clearly identify studies on discrimination as one area in which covert research is often the only way to avoid the bias of socially desirable behaviour. Dench et al. even refer to field experiments in the labour market explicitly saying that

"if a study exploring *discrimination in the recruitment process* involved researchers posing as applicants, informing the recruiters in advance may lead to their acting differently than normal." (Dench et al., 2004, p. 62, emphasis added)

Using covert research methods is a delicate matter, but, as seen above, it can be justified in situations in which information of a similar quality and richness cannot be obtained using other methodologies.

Matched pair testing vs. non-matched pair testing

As was briefly mentioned previously, the number of applications submitted for one vacancy can also have ethical implications that can become important in IRB submissions and discussions of the ethical questions of the research, the most obvious being the time an employer spends on assessing applications (Gaddis, 2018a). However, as Lahey and Beasley (2018) point out, the number of resumes could also affect hiring practices and recruiters' decisions, e.g. if more very well-qualified applications are received in an on-going hiring process. They emphasise that "unmatched sets send a less focused signal and may be less likely to harm a participant's overall view of the market" (p.91).

While most field experiments have traditionally been designed as matched-pair experiments, some more recent studies have deviated from the design of matching candidates and only sent out single applications (e.g. Koopmans, Veit, & Yemane, 2018; Weichselbaumer, 2015, 2016)⁷. Researchers using a single application design often argue that

⁷ For an overview on matched and unmatched audit studies see Table 6.1 in Vuolo, Uggen, and Lageson (2018)

they want to minimize the risk of detection of the experiment. However, Riach and Rich (2004b) argue that studies using only a single application per vacancy are tests of “preferential treatment” in the broader labour market or of a propensity among employers to discriminate rather than actual discrimination, since “employment discrimination can only occur when an individual employer is confronted with a need to choose” (p.471). Since it is not possible to attribute discriminatory treatment to specific employers Cherry and Bendick (2018) describe the findings of unpaired audits as “villainy without villains” (p.55). So far, the most thorough discussion of matched vs. unmatched designs has been provided by Vuolo, Uggen, and Lageson (2016) and Vuolo et al. (2018), who mainly focus on the statistical implications of this research design choice. The ethical implications of the number of applications send, is however, not discussed in the literature.

Technical and legal challenges in correspondence testing

Although the ethical issues discussed above are important to obtain ethical approval to conduct a correspondence test, there are also other stages in the planning phase of the experiment that can have ethical or legal implications: in particular setting up contact details, such as email addresses and phone numbers for the fictitious applicants. To my knowledge, these issues have so far not been addressed in articles on the methodology. Finally, correspondence testing may also require the preparation of photographs, diplomas, or work certificates if these are usually included in an application. This is usually required in German speaking countries and will therefore be addressed in the country examples below.

Creating Email Addresses

Each fictitious applicant requires an email addresses to send applications and receive replies. The email addresses most frequently used in the correspondence test included in the meta-analysis by Zschirnt and Ruedin (2016), were gmail.com, Hotmail.com, and yahoo.com. These providers differ considerably in their terms of services. Yahoo for example clearly spells out in Section 3 of its Terms and Services:

“In consideration of your use of the Yahoo Services, you represent that you are of legal age to form a binding contract [...]. You also agree to (a) provide true, accurate, current and complete information about yourself [...].” (Yahoo, 2012)

Similarly, Microsoft stipulates in its terms of services that “You agree not to use any false, inaccurate or misleading information” (Microsoft, 2015Section 4.a.i.). Furthermore, in its Code of Conduct it emphasises that the account is not to be used for anything illegal and that the account holder is not to “engage in activity that is false or misleading (e.g., [...] impersonating someone else [...])” (Section 3.a.i.). These terms can be potentially problematic for researchers conducting a correspondence test, since it is impossible to “provide true, accurate, current and complete information” for fictitious applicants.

Google’s Terms of Service only state “Don’t misuse our services” (Google, 2014) and do not specifically define who is allowed to open a google account and which conditions have to be fulfilled. Yet at the very end of the Terms and Services, it is stated that “The laws of California, U.S.A., [...] will apply to any disputes arising out of or relating to these terms or the Services” (Google, 2014). In 2010, the State of California adopted its first online impersonation law – the Senate Bill SB 1411, which regulates that

“(a) Notwithstanding any other provisions of law, any person who knowingly and without consent credibly impersonates another actual person through or on an Internet Web site or by other electronic means for purposes of harming, intimidating, threatening, or defrauding another person is guilty of a public offense punishable pursuant to subdivision (d).” (Simitian, 2010, Section 1)

It is therefore necessary to examine in how far correspondence testing might be considered an impersonation of another *actual* person. Since the fictitious applicants in correspondence tests do not exist in real life, it can be argued that this is not an impersonation of another *actual* person. Furthermore, it should be obvious that correspondence testing is not done for “the purpose of harming, intimidating, threatening or defrauding another person”.

So far, all published correspondence tests I am aware of, with the exception of Neumark, Burn, and Button (2017) who created their own email provider, have used free and frequently used email providers such as Gmail, Yahoo or Hotmail. To my knowledge, there have never been legal objections to their use.

Next to these legal issues, researchers can also encounter problems with the security settings of free email providers that might limit the possibility to send high numbers of emails using programming scripts. Readers interested in the technical aspects of setting up correspondence tests via email should refer to Crabtree (2018) who devotes an entire book chapter to the issue.

Generating street addresses

Street addresses are another elemental part of the contact details that might become problematic. Eid (2012) used addresses of his research team and colleagues for his Canadian experiment. Wood et al. (2009) decided against such an approach in their UK study out of ethical considerations. They argue that UK employers sometimes carry out background checks, including credit checks. Thus, like most other studies, they constructed credible fictitious addresses, e.g. by using real street names, but non-existing house numbers. The areas chosen for the study were chosen based on the ethnic diversity shown in census data (Wood et al., 2009, p. 23). Another approach used by Bursell (2007) was to use real addresses in residential blocks, but making sure that nobody with a similar name lived there. Similarly, researchers could use addresses of real apartment buildings but provide fake apartment numbers. While posted responses are lost, Eid (2012) reported that the grand majority of employers contacted potential candidates by phone and that letters were hardly used.

Providing phone numbers

The final element of the contact details is a phone number. Here almost all researchers use the same approach: phone numbers connected to a voicemail box were set up using mobile phones or online generated phone numbers. The number of phone numbers used varied, however. While Eid (2012) used only two numbers, one for the majority and one for the minority applicant, Wood et al. (2009) had twelve phone numbers, depending on the gender and ethnicity of the fictitious applicant. In all studies the voicemail messages were either standard voicemail messages by the phone provider or recorded without any discernible accent. One of the challenges of using voicemails is matching the response received with the vacancy it was connected to. Furthermore, local legal regulations need to be taken into

consideration when it comes to setting up mobile phone accounts, e.g. if a proof of ID is required to open an account.

Managing ethical issues in correspondence testing – European examples

Since this paper has so far predominantly focused on the theoretical discussion of research ethics in correspondence tests as well as arguments from an ethical perspective pro and contra correspondence testing, the last part of the paper discusses ways in which ethical commissions have dealt with correspondence tests. While most publications of correspondence tests acknowledge the question of research ethics without going into further details, some researchers specifically refer to the ethics bodies and procedures in their countries. Unfortunately, such in-depths information on the ethical approval process is only publicly available for four countries, namely, Sweden, Norway, Finland, and Germany.

Sweden, Norway and Finland

To my knowledge, the first country where a correspondence test was stopped by a research ethics commission was Sweden. Swedish researchers wanted to participate in the large ILO Project on labour market discrimination in the 1990s; they thus submitted two research proposals using the correspondence test design outlined by Bovenkerk (1992), yet their proposals were not approved by the Swedish Ethics Board. The Board claimed that “invit[ing] an innocent employer to act in a manner likely to have been made punishable by the time any such research started” (Banton, 1997, p. 415) posed too big a risk. While a first assessor had not expressed any doubts, a second reviewer concluded that while the research might be in the public interest, the potential consequences for people found guilty of discriminatory behaviour were too big: “The employer runs both a risk of injury to reputation and a financial risk. It is these risks of injury which so clearly make the proposed experiment ethically unacceptable” (in Banton, 1997, p. 415). Furthermore, potential issues of liability for the researcher or funding organisations were addressed and it was recommended that neither researcher should be funded.

As Carlsson and Rooth (2012) point out, the Swedish authorities subsequently reconsidered their position on correspondence testing: “An important event for this turnaround occurred in 2005 when law students initiated lawsuits against restaurants and night clubs based on situation tests of ethnic discrimination” (p.99). Following this change of position, three Swedish correspondence tests were published in 2007 (Attström, 2007; Bursell, 2007; Carlsson & Rooth, 2007). Except for a brief section on ethics in (Bursell, 2007) the issue of research ethics was not addressed.

The rejection of the Swedish contribution to the ILO Project also affected Norwegian researchers, as Midtbøen (2013) points out:

“Because the method was rejected by the Swedish Council for Social Research, it was assumed that the Research Council of Norway would reach the same conclusion. This is a main reason why Norwegian researchers during the 1990s never even applied for funding of experimental studies of discrimination” (p.52).

Once the Swedish research ethics boards approved the above-mentioned research projects using correspondence testing, researchers in Norway proposed to conduct a field experiment on the Norwegian labour market. Midtbøen (2013) reports that the National Committees for Research Ethics in Norway (NESH) approved the research design under three strict conditions. First, testing should be conducted in the early phase of the hiring process. Second, the privacy of the individuals in the hiring procedure was to be protected. Third, regarding the recruitment of participants for follow-up interviews, it was emphasised that this should respect the principles of voluntary participation and informed consent.

In 2011, ethical approval was also given for the first Finnish correspondence test, where the guidelines of the Finnish National Advisory Board on Research Ethics also “lists field experiments in studies about discrimination as an example of a research design where deviating from the principle of informed consent and misleading research subjects is acceptable” (Larja et al., 2012, p. 142).

Germany

In Germany, so far four correspondence tests on ethnic discrimination in hiring decisions have been conducted (Goldberg, Mourinho, & Kulke, 1995; Kaas & Manger, 2012; Schneider et al., 2014; Weichselbaumer, 2016). The Expert Council of German Foundations on Integration and Migration’s (SVR) was the only one who extensively addressed the ethical questions. Its report includes a short section on the ethical and legal challenges in correspondence testing (Schneider et al., 2014, p. 16). It emphasises that the research design was approved by the Ethical Committees of the German Sociological Society and the German Association of Sociologists as it was judged unproblematic both from a data protection and from an ethical point of view. It was argued that the aggregated analysis of the data would not allow inferences about individual employers, and that the use of fictitious applications did not infringe any personal rights (Schneider et al., 2014, p. 16). Yet, the research team went even further than obtaining ethical approval, and also addressed potential legal problems. While two legal expertise by Klose and Kühn (2009, 2010) on the use of correspondence testing had previously been commissioned by the Federal Anti-Discrimination Authority, the SVR hired these lawyers again to specifically analyse their proposed research design for a correspondence test on the German apprenticeship market (Kühn, Liebscher, & Klose, 2013). Since these expert opinions look at numerous legal concerns raised in regard to correspondence testing, they warrant a more detailed look.

In total, there are now three legal expertise on testing available for Germany: the first two expertise by Klose and Kühn (2009, 2010) focus on very specific legal questions regarding testing and racial or ethnic discrimination in the area of “*Gewerberecht*” (trade law) (2009) and the use of testing as an instrument in trials regarding the burden of proof in discrimination cases (2010). The third expertise by Kühn et al. (2013) addresses the legal questions concerning testing as a social science research method and focused explicitly on the SVR’s research design. It is therefore the most relevant publication to be considered here. Regarding criminal law, they focus on the use of certificates or copies thereof, concluding that the testing methodology is protected under the scientific freedom guaranteed by the German Basic Law, and that testing does not fulfil the crime of forgery of documents (e.g. school or university certificates) that are required to submit a complete German application. Furthermore, they claim that researchers do not have to fear being punished for fraud, since testing studies are not intended for unlawful gains of the researchers. Looking at civil law, Kühn et al. argue that claims for liability of the researcher due to the time employers invested in examining a

fraudulent application are not likely, since the loss of time is not considered a replaceable damage. Employers are also unlikely to succeed in suing for damages by arguing that the fictitious applications caused a delay or necessitated a repeated application procedure. Furthermore, Kühn et al. closely look at the German data-protection laws in relation to correspondence testing. According to them, data-protection laws do not apply if the data was anonymised and analysed quantitatively and if no inferences about individuals can be made. The use of publicly available data, such as addresses, is also permitted. This legal expertise thus enabled the researchers of the Expert Council of German Foundations on Integration and Migration to conduct their correspondence test on labour market discrimination.

These individual country examples show that the theoretical concerns regarding the ethical questions in correspondence testing discussed in previous sections of the paper are valid, but can be addressed in well-prepared research designs. In the case of Germany many of the aforementioned reservations, such as the possibility of committing fraud, of forging documents, of potential damages to employers, or the liability of researchers have been addressed by legal experts and found not to be an obstacle to conducting a correspondence test. While a similar wealth of information on ethical and legal preparation work was not publicly available for other countries, the examples of Sweden, Norway, and Finland show, that ethical commissions were quite thorough in their evaluation of the research projects, but eventually decided that a good research design could meet their concerns and that the societal interest to study discrimination was held above the inconveniences that could potentially be caused to an individual employer.

Conclusion

As shown in this paper breaking core research ethics principles, particularly those of informed consent and voluntary partition, can be justified in the case of correspondence studies on discrimination in the labour market. Using examples from different countries and different ethical committees, it can be seen that researchers were able to obtain ethical approval to conduct correspondence tests if certain strict criteria were met, the most important being to keep the inconvenience to employers at a minimum, to guarantee the confidentiality and privacy of the research subjects, to analyse data in an aggregated form to avoid inferences being made about individual research subjects, and to adhere again to the principles of research ethics in any follow-up research.

Given the rapid growth in the numbers of correspondence tests conducted in recent years, it could be questioned if more studies measuring discrimination are actually needed. A recent meta-analysis has shown that in the case of ethnic discrimination in hiring, minority candidates have to write on average 50% more applications than equally qualified majority candidates (Zschirnt & Ruedin, 2016). It could therefore also be argued that not studying discrimination would be unethical, as the data provided can help in “society’s continuing struggle to eliminate the subtle but deadly poison of racial discrimination” (Boggs et al., 1993, p. 367). There is still a need to provide data on the extent of discrimination in hiring decisions to make employers aware of this issues, which may also occur unintentionally, and to lobby for policy changes and stronger and more effective anti-discrimination laws. Considering the strong power imbalance in the hiring process, it is necessary to strengthen the position of applicants to balance the scales. Providing better data of the extent of discrimination of minorities is just a first step in this direction. Using data that was obtained through correspondence testing on the broader labour market to strengthen legal cases against discriminatory employers, as it is already possible in some countries, could be a next step.

References

- American Psychological Association. (2017). *Ethical Principles of Psychologists and Code of Conduct*. Retrieved from <http://www.apa.org/ethics/code/ethics-code-2017.pdf>
- American Sociological Association. (2018). *Code of Ethics*. Retrieved from http://www.asanet.org/sites/default/files/asa_code_of_ethics-june2018.pdf
- Attström, K. (2007). *Discrimination against native Swedes of immigrant origin in access to employment*. Geneva: International Labor Organization.
- Baert, S. (2018). Hiring Discrimination: An Overview of (Almost) All Correspondence Experiments Since 2005. In M. Gaddis (Ed.), *Audit Studies: Behind the Scenes with Theory, Method and Nuance* (pp. 63-77): Springer.
- Banton, M. (1997). The ethics of practice-testing. *Journal of Ethnic and Migration Studies*, 23(3), 413-420.
- Bendick, M., & Nunes, A. P. (2012). Developing the Research Basis for Controlling Bias in Hiring. *Journal of Social Issues*, 68(2), 238-262. doi:10.1111/j.1540-4560.2012.01747.x
- Bertrand, M., & Duflo, E. (2017). Field Experiments on Discrimination *Handbook of Economic Field Experiments* (Vol. 1, pp. 309-393): Elsevier.
- Blommaert, L., Coenders, M., & van Tubergen, F. (2014). Discrimination of Arabic-Named Applicants in the Netherlands: An Internet-Based Field Experiment Examining Different Phases in Online Recruitment Procedures. *Social Forces*, 92(3), 957-982. doi:10.1093/sf/sot124
- Boggs, R. V. O. (1998). The Future of Civil Rights Testing: Current Trends and New Directions. In M. Fix & M. A. Turner (Eds.), *A National Report Card on Discrimination in America: The Role of Testing*. Washington, D.C.: The Urban Institute.
- Boggs, R. V. O., Sellers, J. M., & Bendick Jr, M. (1993). Use of testing in civil rights enforcement. In M. Fix & R. Struyk (Eds.), *Clear and Convincing Evidence: Measurement of Discrimination in America*. Washington, D.C.: The Urban Institute Press.
- Bovenkerk, F. (1992). *Testing discrimination in natural experiments: a manual for international comparative research on discrimination on the grounds of "race" and ethnic origin*. Geneva: International Labour Office.
- British Sociological Association. (2002). Statement of Ethical Practice for the British Sociological Association. Retrieved from <http://www.britisoc.co.uk/media/27107/StatementofEthicalPractice.pdf>
- Bursell, M. (2007). What's in a name? A field experiment test for the existence of ethnic discrimination in the hiring process. *SULCIS Working Paper*(2007:7).
- Calvès, G. (2007). Au service de la connaissance et du droit: le testing. *Horizons stratégiques*(3), 8-16.
- Carlsson, M., & Rooth, D.-O. (2007). Evidence of ethnic discrimination in the Swedish labor market using experimental data. *Labour Economics*, 14(4), 716-729. doi:<http://dx.doi.org/10.1016/j.labeco.2007.05.001>
- Carlsson, M., & Rooth, D.-O. (2012). The Power of Media and Changes in Discriminatory Behavior Among Employers. *Journal of Media Economics*, 25(2), 98-108. doi:10.1080/08997764.2012.676583
- Cherry, F., & Bendick, M. (2018). Making It Count: Discrimination Auditing and the Activist Scholar Tradition *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 45-62): Springer.
- Connor, J. M. (2000). Studying racial bias: Too hot to handle? *Journal of College Science Teaching*, 30(1), 26.
- Crabtree, C. (2018). An introduction to conducting email audit studies *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 103-117): Springer.
- Dench, S., Iphofen, R., & Huws, U. (2004). *An EU Code of Ethics for Socio-Economic Research*. Retrieved from <http://www.respectproject.org/ethics/412ethics.pdf>

- Deutsche Gesellschaft für Soziologie. (2017). Ethik-Kodex der deutschen Gesellschaft für Soziologie (DSG) und des Berufsverbandes deutscher Soziologinnen und Soziologen (BDS). Retrieved from <https://www.soziologie.de/de/die-dgs/ethik/ethik-kodex/>
- Dingwall, R. (2012). How did we ever get into this Mess? The Rise of Ethical Regulation in the Social Sciences. In K. Love (Ed.), *Ethics in Social Research* (Vol. 12, pp. 3-26). Bingley: Emerald Group Publishing Limited.
- Edley Jr., C. (1993). Implications of Empirical Studies on Race and Discrimination. In M. Fix & R. Struyk (Eds.), *Clear and Convincing Evidence: Measurement of Discrimination in America*. Washington, D.C.: The Urban Institute Press.
- Eid, P. (2012). *Mesurer la discrimination à l'embauche subie par les minorités racisées: résultats d'un 'testing' mené dans le grand Montréal*. Quebec: Commission des droits de la personne et des droits de la jeunesse.
- Fix, M., Galster, G., & Struyk, R. (1993). An Overview of Auditing for Discrimination. In M. Fix & R. Struyk (Eds.), *Clear and Convincing Evidence: Measurement of Discrimination in America* (pp. 1-49). Washington DC: The Urban Institute Press.
- Gaddis, S. M. (2018a). An introduction to audit studies in the social sciences *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 3-44): Springer.
- Gaddis, S. M. (Ed.) (2018b). *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*: Springer.
- Gingrich, N. (1998). *Testimony of House Speaker Newt Gingrich before the House Subcommittee on Employer-Employee Relations on "The Future Direction of the Equal Employment Opportunity Commission"*. Retrieved from <http://archives.republicans.edlabor.house.gov/archive/hearings/105th/eeer/eeoc3398/ginrich.htm>.
- Goldberg, A., Mourinho, D., & Kulke, U. (1995). *Labour market discrimination against foreign workers in Germany*: International Labour Office, Employment Department.
- Google. (2014). Google Terms of Service (Last updated April 14, 2014). Retrieved from <http://www.google.com/intl/en/policies/terms/>
- Hunter, D. (2010). Locating ethics in research. In J. Hughes (Ed.), *European Textbook on Ethics in Research* (pp. 11-32). Luxembourg: Publication Office of the European Union.
- Israel, M. (2015). *Research Ethics for Social Scientists* (2nd ed.). London: Sage.
- Jowell, R., & Prescott-Clarke, P. (1970). Racial discrimination and white-collar workers in Britain. *Race & Class*, 11(4), 397-417.
- Kaas, L., & Manger, C. (2012). Ethnic Discrimination in Germany's Labour Market: A Field Experiment. *German Economic Review*, 13(1), 1-20. doi:10.1111/j.1468-0475.2011.00538.x
- Klose, A., & Kühn, K. (2009). *Gewerberecht, rassistische/ethnische Diskriminierung und Testingverfahren*. Berlin: Antidiskriminierungsnetzwerks Berlin des Türkischen Bundes in Berlin-Brandenburg.
- Klose, A., & Kühn, K. (2010). Die Anwendbarkeit von Testing-Verfahren im Rahmen der Beweislast, § 22 Allgemeines Gleichbehandlungsgesetz. *Expertise im Auftrag der Antidiskriminierungsstelle des Bundes, Berlin*.
- Koopmans, R., Veit, S., & Yemane, R. (2018). Ethnische Hierarchien in der Bewerberauswahl: Ein Feldexperiment zu den Ursachen von Arbeitsmarktdiskriminierung. *WZB Discussion Paper, SP VI 2018-104*.
- Kühn, K., Liebscher, D., & Klose, A. (2013). *Expertise zu den rechtlichen Aspekten sozialwissenschaftlicher Testing-Verfahren im Rahmen der vom SVR geplanten Studie "Diskriminierung am Ausbildungsmarkt?"*. Retrieved from http://svr-migration.dymagics.de/wp-content/uploads/2014/11/Expertise_fuer_Studie_Diskriminierung_am_Ausbildungsmarkt.pdf

- Lahey, J., & Beasley, R. (2018). Technical aspects of correspondence studies. In M. Gaddis (Ed.), *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 81-101): Springer.
- LaPiere, R. T. (1934). Attitudes vs. Actions. *Social Forces*, 13(2), 230-237. doi:10.2307/2570339
- Larja, L., Warius, J., Sundbäck, L., Liebkind, K., Kandolin, I., & Jasinskaja-Lahti, I. (2012). *Discrimination in the Finnish Labour Market: An Overview and a Field Experiment on Recruitment*. Retrieved from <https://yhdenvertaisuus-fi-bin.directo.fi/@Bin/070c0305f220b492e9b19da85cb314b5/1508422136/application/pdf/469926/discrimination-in-the-finnish-labour-market.pdf>
- Liebkind, K., Larja, L., & Brylka, A. (2016). Ethnic and Gender Discrimination in Recruitment: Experimental Evidence From Finland. *Journal of Social and Political Psychology*, 4(1), 403-426. doi:10.5964/jspp.v4i1.433
- Microsoft. (2015). Microsoft Services Agreement (Published June 4, 2015). Retrieved from <https://www.microsoft.com/en-us/servicesagreement/>
- Midtbøen, A. H. (2013). *Determining Discrimination* University of Oslo.
- Midtbøen, A. H. (2014). The Invisible Second Generation? Statistical Discrimination and Immigrant Stereotypes in Employment Processes in Norway. *Journal of Ethnic and Migration Studies*, 40(10), 1657-1675. doi:10.1080/1369183X.2013.847784
- Nakray, K. (2016). Social Science Research Ethics for a Globalizing World - A Critical Overview of Interdisciplinary and Cross-Cultural Perspectives. In K. Nakray, M. Alston, & K. Whittenbury (Eds.), *Social Science Research Ethics for a Globalizing World - Interdisciplinary and Cross-Cultural Perspectives*. New York: Routledge.
- National Committees for Research Ethics in Norway (NESH). (2006). Guidelines for Research Ethics in the Social Sciences, Law and the Humanities. Retrieved from <https://www.etikkom.no/globalassets/documents/publikasjoner-som-pdf/guidelines-for-research-ethics-in-the-social-sciences-law-and-the-humanities-2006.pdf>
- Neumark, D. (2016). Experimental Research on Labor Market Discrimination. *NBER Working Paper No. 22022*.
- Neumark, D., Burn, I., & Button, P. (2017). *Age Discrimination and Hiring of Older Workers*. Retrieved from <http://www.frbsf.org/economic-research/files/el2017-06.pdf>
- Pager, D. (2007). The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future. *Annals of the American Academy of Political and Social Science*, 609, 104-133. doi:10.2307/25097877
- Pager, D., & Western, B. (2012). Identifying discrimination at work: The use of field experiments. *Journal of Social Issues*, 68(2), 221-237.
- Quillian, L., Pager, D., Hexel, O., & Midtbøen, A. H. (2017). Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. *Proceedings of the National Academy of Sciences*, 114(41), 10870-10875. doi:10.1073/pnas.1706255114
- Riach, P. A., & Rich, J. (2004a). Deceptive field experiments of discrimination: are they ethical? *Kyklos*, 57(3), 457-470.
- Riach, P. A., & Rich, J. (2004b). Fishing for Discrimination. *Review of Social Economy*, 62(4), 465-486. doi:10.1080/0034676042000296227
- 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).
- Rorive, I. (2009). *Proving Discrimination Cases.: The Role of Situation Testing*. Bruxelles and Stockholm: Centre for Equal Rights.
- Schneider, J., Yemane, R., & Weinmann, M. (2014). *Diskriminierung am Ausbildungsmarkt: Ausmass, Ursachen und Handlungsperspektiven* Retrieved from Berlin
- Simitian, J. (2010). *An act to add Section 528.5 to the Penal Code, relating to impersonation*. Retrieved from http://leginfo.ca.gov/pub/09-10/bill/sen/sb_1401-1450/sb_1411_bill_20100927_chaptered.html.

- Van der Bracht, K., & Verhaeghe, P.-P. (2017). *Can Discrimination be reduced by Inducing Compliance? A Quasi-Experiment in Rental Housing Discrimination*
- van der Plancke, V. (2007). Les tribulations du testing en Belgique: quels enseignements? *Horizons stratégiques*(3), 40-59.
- Vuolo, M., Uggen, C., & Lageson, S. (2016). Statistical power in experimental audit studies: Cautions and calculations for matched tests with nominal outcomes. *Sociological Methods & Research*, 45(2), 260-303.
- Vuolo, M., Uggen, C., & Lageson, S. (2018). To match or not to match? Statistical and substantive considerations in audit design and analysis *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 119-140): Springer.
- Wassenaar, D. (2006). Ethical issues in social science research. In M. Terre Blanche, K. Durrheim, & D. Painter (Eds.), *Research in Practice: Applied methods for the Social Sciences* (2nd ed.). Capetown: University of Capetown Press.
- 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.
- Weichselbaumer, D. (2016). Discrimination against Female Migrants Wearing Headscarves. *IZA Discussion Paper Series*(No. 10217).
- Wood, M., Hales, J., Purdon, S., Sejersen, T., & Hayllar, O. (2009). *A Test for Racial Discrimination in Recruitment Practice in British Cities*. Retrieved from Norwich: <http://webarchive.nationalarchives.gov.uk/20130314010347/http://research.dwp.gov.uk/asd/asd5/rrs-index.asp>
- Yahoo. (2012). Yahoo Terms of Service (Last updated March 16, 2012). Retrieved from <https://policies.yahoo.com/us/en/yahoo/terms/utos/>
- Zschirnt, E. (2016). Measuring Hiring Discrimination – A History of Field Experiments in Discrimination Research. *NCCR Working Papers*, 7, 1-32.
- Zschirnt, E., & Ruedin, D. (2016). Ethnic Discrimination in Hiring Decisions: A Meta-Analysis of Correspondence Tests 1990–2015. *Journal of Ethnic and Migration Studies*, 42(7), 1115-1134. doi:10.1080/1369183X.2015.1133279

Paper IV:

Evidence of Hiring Discrimination against the Second Generation: Results from a Correspondence Test in the Swiss Labour Market

Submitted to the Journal of International Migration and Integration

(under review)

Evidence of Hiring Discrimination against the Second Generation: Results from a Correspondence Test in the Swiss Labour Market

Eva Zschirnt
University of Neuchâtel
eva.zschirnt@unine.ch
Tel. +41 32 718 39 49

University of Neuchâtel
Swiss Forum for Migration and Population Studies SFM
Rue A.L. Breguet 2
2000 Neuchâtel, Switzerland

ORCID iD: 0000-0002-8918-4146

ABSTRACT

While there is ample evidence of discrimination against ethnic minority candidates in hiring, most existing studies have focused on stigmatized immigrant groups. We use a correspondence test to enumerate ethnic discrimination in the Swiss labour market, varying the *a priori* stigma of the immigrant groups. The field experiment compares candidates with Swiss names against candidates with German, Kosovar and Turkish names in a paired correspondence test spanning four occupations. Between October 2017 and April 2018 applications were sent in response to 560 real vacancies in the German-speaking area of Switzerland. Across the minority groups, the relative call back rate was 1.13, meaning that minority candidates have to write 1.13 times as many applications as majority candidates to be invited for a job interview. The relative call back rates differ by the ethnic origin: Germans experience almost no discrimination across all occupations; Turks face a relative call back rate of 1.14; and Kosovars encounter the highest relative call back rate across occupations (1.26). We conclude that existing studies may give the false impression that all immigrants suffer from substantive discrimination in the labour market because they focus on stigmatized groups.

Keywords: Ethnic Discrimination, Labour Market, Hiring, Correspondence Testing, Switzerland

Acknowledgements: This research was supported by the nccr – on the move funded by the Swiss National Science Foundation. I would like to thank Daniel Auer, Rosita Fibbi,, Flavia Fossati and Didier Ruedin for their comments.

Introduction

Switzerland has one of the biggest shares of foreign citizens in Europe, with about a quarter of the population not holding a Swiss passport. In Europe, similarly high numbers of foreigners are only recorded in Luxembourg. Although the labour market is one of the essential places for the integration of these foreigners into Swiss society, relatively little is known about discrimination that foreigners or citizens with immigrant origins experience in the Swiss labour market.

While ethnic discrimination in hiring decisions in the labour market has been well documented in most OECD countries, with minority candidates usually having to write 50% more applications to be called back for a job interview (Zschirnt and Ruedin 2016), this information is largely missing for the case of Switzerland. So far only one field experiment on ethnic discrimination in hiring decisions in the Swiss labour market has been conducted by Fibbi et al. (2003), yet this experiment only looked at the transition from an apprenticeship to the first job. They found significant levels of discrimination which varied between the ethnic minority groups and suggested that ethnic hierarchies exist in the Swiss labour market.

In contrast to previous studies, this paper does not only focus on stigmatised groups of people with a migration background that come from low-income countries, but also includes more recent and higher qualified immigrants from neighbouring EU countries, i.e. naturalised Germans. Between October 2017 and April 2018 paired applications were sent in response to 560 real vacancies across four occupations in the German-speaking area of Switzerland.

Theory

Discrimination against ethnic or racial minority candidates in the labour market is a well-documented phenomenon. Since the late 1960s, researchers have documented and tried to measure its occurrence (Zschirnt 2016a; Gaddis 2018). Discrimination has been studied in many countries, with ethnic or racial minority candidates having to write about 1.6 times as many applications to be invited for a job interview (Zschirnt and Ruedin 2016). While the occurrence and extent of ethnic and racial discrimination in the labour market has been documented, the reasons for it are harder to establish.

Economic theories mostly focus on whether discrimination is taste-based (Becker 1957) or statistical (Phelps 1972; Arrow 1973). The theory of taste-based discrimination (Becker 1957) is based on the notion that employers prefer working with members of the majority group and are willing to face financial disadvantages in order to hire candidates that correspond to their taste. They sometimes justify discriminatory hiring decisions by pointing out that their employees or customers would not accept minority candidates. In contrast, statistical discrimination theory (Arrow 1973; Phelps 1972) argues that discrimination can also be due to rational decision making. To compensate for a lack of knowledge,

employers resort to signals such as race and ethnicity and make assumptions about the productivity and skills based on the group membership of a candidate. The “employer who seeks to maximize expected profit will discriminate against blacks or women if he believes them to be less qualified, reliable, long-term, etc. [...] and if the cost of gaining information about the individual applicant is excessive. Skin color or sex is taken as a proxy for relevant data not sampled.” (Phelps 1972). Thus, according to statistical discrimination theory, discrimination should decline if employers have more information about candidates and do not have to make inferences based on a candidate’s group membership. The debate if discrimination is due to taste or statistics is ongoing (e.g. Flinn 2015; Thijssen 2016; Keuschnigg and Wolbring 2016; Guryan and Charles 2013) and researchers have tried to incorporate it in their research designs (e.g. Baert and De Pauw 2014).

Methodology of Correspondence Testing

Field experiments, such as correspondence tests, have become a popular way to study discrimination in hiring decision (Jackson and Cox 2013; Gaddis 2018). As List (2009) pointed out, they “are a useful marriage between laboratory and naturally occurring data in that they represent a mixture of control and realism not usually achieved in the lab or with naturally occurring data” (p.439). Field experiments on discrimination are conducted in the research subjects’ natural environment and observe natural behaviour, without informing the research subjects that they are part of an experiment and that their actions are being recorded and analysed. They give researchers the possibility to study discrimination, which is usually illegal and hidden, and to attempt to measure its extent towards specific disadvantaged groups.

To date, field experiments have been carried out to study hiring discrimination on the grounds of race or ethnicity, gender, age, disability, sexual orientation, caste, religion, or obesity (for detailed overviews see Riach and Rich 2002; Rich 2014; Bertrand and Duflo 2016; Baert 2018). Only looking at correspondence tests on ethnic and racial discrimination in hiring decisions, Zschirnt and Ruedin (2016) have identified 43 studies that were conducted in OECD countries between 1990 and 2015. In all of these studies fictitious written applications were sent in response to real vacancies posted by employers. In the simplest design, each employer receives two fictitious applications, one from a majority candidate and one from a minority candidate, in our case a candidate belonging to a racial or ethnic minority. The candidates are equally qualified, matched as closely as possible and only differ in the characteristic to be tested – their ethnicity or race. By carefully recording replies (invitation or rejection), researchers are able to analyse racial or ethnic differences in invitation rates which can then be attributed to discrimination.

While correspondence tests are often considered one of the best ways to measure discrimination, the methodology also faces limitations and critique. On the one hand, criticism of in particular in-person

audit studies focused on the difficulties of matching testers and the problem of unobserved variables, the limited representativeness of the results if only low skilled or entry level positions are considered, and the way that results are presented (Heckman and Siegelman 1993; Heckman 1998; Neumark 2012). Furthermore, these studies usually focus on stigmatised groups for which researchers expect to find discrimination, certain regions or occupations (Lahey and Beasley 2018). The findings are therefore not representative for the whole labour market and they are clearly not representative for the whole immigrant population, as ethnic hierarchies found in studies testing several groups have shown (e.g. Booth et al. 2012; for an overview see Zschirnt and Ruedin 2016). Most of this criticism has been mitigated in correspondence tests, which give researchers complete control over the application material, thus making it possible to match the fictitious candidates very closely. Furthermore, it is possible to send greater numbers of applications and to test more occupations on different skill levels. However, it is only possible to measure if a candidate gets an invitation for a job interview, not if he actually is offered the job (Midtbøen and Rogstad 2012).

On the other hand, correspondence tests are often criticised on research ethics concerns. By receiving applications from fictitious candidates, employers are deceived and unable to give their informed and voluntary consent to their participation in the research and could potentially suffer negative consequences. These ethical concerns are rarely addressed in published correspondence tests. Readers are usually referred to an article by Riach and Rich (2004) that discusses whether field experiments are ethical. Recently Zschirnt (2016b) reconsidered the ethical questions, in particular regarding the technological changes that being able to send great numbers of application per email has brought. The value of the methodology has been acknowledged in several international research ethics guidelines, where it is listed as an example where covered research is justified.

For this correspondence test on the Swiss labour market the ethical approval was obtained from the Ethics Commission at the University [removed for review]. It was judged that the ethical problems of breaking the principles of informed consent and voluntary participation and possible losses for employers were addressed in detail and concrete steps on addressing these issues in the research project were proposed.

Switzerland: A country with a sizeable foreign population

Switzerland has a high share of immigrants in its resident population. Of the approximately 8.33 million residents, 2.05 million hold foreign nationalities. The biggest share of immigrants is made up by people from European countries (84.6%). Immigrants from EU-28 countries account for 66.3% of the immigrant population with Italians, Germans and Portuguese constituting the biggest immigrant groups (each between 15.2% and 13.1% of the immigrant population). French, Kosovars, Spanish, Serbians and Turks also each account for 6% to 3.4% of the immigrants in Switzerland (Bundesamt für Statistik

2017). Among the share of foreigners, the majority are first generation immigrants. Due to the restrictive Swiss naturalization policy, almost 40% of the second generation are still legally foreigners.

In contrast to its neighbouring countries Switzerland does not have a comprehensive anti-discrimination law. Since it is not a member of the European Union, it did not have to implement the two EU anti-discrimination directives adopted in 2000. In 2014 Switzerland was ranked 35th of 38 countries analysed for the MIPEX Index in the field of anti-discrimination policy (Huddleston et al. 2015). They reported that since Switzerland was “One of the very few countries without a comprehensive anti-discrimination law and equality body with legal standing; a sizeable number of potential victims are poorly protected against racial, ethnic, religious and nationality discrimination” (p.40). It is therefore not surprising that numerous international organisations have urged Switzerland to adopt a comprehensive anti-discrimination law¹.

Despite the lack of legal protection and the strong presence of immigrants in Switzerland, the topic has received little attention in the media, in politics or in research. There is relatively little research on labour market discrimination and most of it is appeared in the last years. Recent research on the labour market position of ethnic minorities has focused on the statistical analysis of observational data to look at unemployment duration or the occupational incorporation of immigrants (Auer et al. 2017; Vidal-Coso and Ortega-Rivera 2017), or on attitudes towards foreigners shown in surveys (Helbling 2011; Longchamp et al. 2014; Raymann 2003; Rapp 2015; Ruedin et al. 2013) or vignette experiments (Helbling and Kriesi 2014; Auer et al. 2016; Fossati et al. 2017). Studies on attitudes towards foreigners show that attitudes vary for each immigrant group, with migrants from the Balkans or with a Muslim background being least accepted (Longchamp et al. 2014; Raymann 2003; Rapp 2015), but also German immigrants being met with negative stereotypes (Helbling 2011; Binggeli et al. 2014a), that it is not only lower educated people that hold negative attitudes towards foreigners (Pecoraro and Ruedin 2016, 2017; Helbling 2011), and that individuals who feel culturally or economically threatened are more likely to express such negative attitudes (Rapp 2015; Helbling 2011; Pecoraro and Ruedin 2016, 2017).

During the last year studies with victims of (perceived) discrimination have appeared, focusing specifically on the experiences of black people (Efionayi-Mäder et al. 2017), Muslims (Golder et al. 2017) or case reports from human rights outreach centres that document discrimination cases that were brought to their attention (Mühlemann 2017). Social psychologists have also addressed the perception of and the stereotypes ascribed to different immigrant groups in Switzerland (Krings and Olivares 2007; Binggeli et al. 2014b; Krings et al. 2014; Matser et al. 2010). All of these studies find that immigrants face a disadvantage in the Swiss labour market and that this disadvantage differs by the ethnic group an immigrant belongs to. Negative attitudes are strongest against immigrants from the Balkans and those

¹ An overview of the international feedback can be found on <https://www.humanrights.ch/en/switzerland/recommendations/discrimination/legislation-discrimination/> (last accessed 16.03.2018)

with a Muslim background, but also highly skilled German immigrants in the city of Zurich are considered as problematic, especially by well-educated Swiss people (Helbling 2011).

The only correspondence test so far in Switzerland by Fibbi et al. (2003) found strong evidence of discrimination against youths from non-EU countries, with Albanian speaking youths from former Yugoslavia faring the worst in the German speaking part of Switzerland (discrimination rate: 59%), followed by Turks in the German speaking parts (30%) and former Yugoslavs in the French speaking regions (24%). Compared to previous ILO studies (de Prada et al. 1995; Bovenkerk et al. 1995; Goldberg et al. 1995; Arriijn et al. 1998) net discrimination rates documented in Switzerland for second generation youths were higher than those found in other countries, with the exception of the Portuguese.

Based on the statistical information on the biggest immigrant groups in Switzerland and the results from research on attitudes towards foreigners in Switzerland, we chose the following groups for this correspondence test: Germans, Kosovars, and Turks. This way the immigrant groups are similar to those of the previous study by Fibbi et al. (2003), changing only the group of Western Europeans from Portuguese to German.

Research Design

Planning a correspondence test on the Swiss labour market is challenging. Similar to other German-speaking countries, a complete application consists not only of a cover letter and CV, but of copies of high school or university diplomas, references from previous employers, and a photograph. This means that researchers have to prepare a great number of accompanying documents to submit a credible application package.

Correspondence tests usually rely on conveying the ethnic or racial identity via the name of the applicant. This raises the problem that names can not only signal ethnicity, but can have unintended and unobserved socio-economic connotations (Crabtree and Chykina 2018; Gaddis 2017a, 2017b; Fryer Jr and Levitt 2004). For this study names were constructed using government statistics, quasi-official statistics, miscellaneous websites on common names and Wikipedia. Furthermore, it was checked that name combinations had not been used in previous correspondence tests and, using Facebook, it was ensured that several persons were registered with this name, thus making it impossible for a potential employer to try to identify a fictitious applicant using social media. While the deliberate choice of names is important to convey ethnicity, applicants also list being dual nationals as citizenship is normally mentioned in a Swiss CV. This way residence or work permits cannot be used as an excuse for rejection. Furthermore, the restrictive Swiss naturalization laws provide assurance that candidates are well integrated, while leaving no doubts about their ethnic background.

Vacancies were collected online in the German speaking part of the Swiss labour market without further geographical limitations. Data of the *Stellenmarktmonitor Schweiz*² show that between 2006-2014 78.8% of the vacancies posted in Switzerland were located in the German language area, with 25.5% of the positions located in Zurich, 21.6% in the “Espace midland”, followed by North-western Switzerland with 13.1%, the Lake Geneva Region with 12.5% and Eastern Switzerland with 11.9% (Buchmann et al. 2015, own calculations). Furthermore, only vacancies where applicants were asked to apply in writing and electronically were considered. Finally, positions that require more than the habitually included documents were excluded, e.g. positions asking for portfolios of previous work or police background checks.

We calculated the required sample size according to the methodology proposed by Vuolo et al. (2016) and used the discrimination rates observed by Fibbi et al. (2003) as a reference for the calculation. According to these calculations we would need between 28 and 60 cases in which only one candidate was invited for a job interview. We lowered these numbers based on preliminary findings after three weeks. However, having obtained our results, we double checked the outcome on the aggregate study level with the sample sizes suggested by Vuolo et al. (2016). Based on our findings from the Swiss German labour market the suggested sample size is 357, thus on the aggregate level, we surpassed the required sample size. We found that discrimination occurs in the Swiss labour market on the study level, and despite the lower number for the combinations of ethnicity and occupation, we were able to obtain significant results on some of these combinations.

In contrast to the first Swiss correspondence test on ethnic discrimination in hiring by Fibbi et al. (2003) this correspondence focuses not only on the entry into the labour market after an apprenticeship but combines different education levels, ranging from lower qualified sales assistants and electricians to intermediate positions of nurses and HR clerks. The share of foreigners for these occupations ranged from 15% (sales, HR) to 21% (nurses), and unemployment rates for Swiss in these occupations were all below 5%, while the occupational unemployment rates for foreigners were always higher, ranging from 2.3% for foreigners working as nurses to 8% for foreigners in sales positions. This range of occupations is meant to establish whether discrimination varies by qualification levels. The occupations selected are in the private sector and include occupations with varying degrees of customer contact. Furthermore, these positions are frequently advertised ensuring a big enough sample of vacancies. Since part-time work is quite common in Switzerland (35.6% hold part-time contracts) and frequent among women, of

² The “Stellenmarkt Monitor Schweiz (SMM)” is a project by Buchmann et al. at the University of Zürich which has been documenting the development of the publication of vacancies in Switzerland going back to 1950. The SMM draws a representative sample of vacancies from a number of publication channels during one week each year, since 2001 it is available for all of Switzerland and since 2006 the database also includes vacancies published in online job portals (Buchmann et al. 2015)

which only 42.8% work fulltime (Bundesamt für Statistik 2016), we included all job offers matching the selected occupations and carefully record the percentage of the job offered.

Developing Applications

The fictitious applications are the core element of the correspondence test. Profiles have to be plausible and provide the qualifications necessary for the position, while keeping comparability in mind. To be able to send out the large numbers, applications are standardized as much as possible. In this correspondence test we use relatively young fictitious applicants. All of them have started to work in their first position after their apprenticeship, with the exception of the HR specialist, where the career path involves a few more steps. Based on publicly available LinkedIn profiles, resumes of job seekers and career advice websites, we prepared pairs of CVs and cover letters for each occupation, which were fine-tuned with HR specialists. Finally, the cover letter and CV in each pair use a different layout to avoid detection. For methodological reasons, the CVs included slight variations such as an additional year of work experience, having completed the Federal Vocational Degree, extra skills listed in the CV (e.g. computer courses), or planned additional qualifications, depending on the occupation tested.

The cover letter and CV were only the first step in preparing a convincing application package. In Switzerland it is customary to include not only a photograph of the applicant, but also copies of the degrees obtained as well as work certificates from previous employers. While some degrees and diplomas could be downloaded from the internet using thorough google images searches, the lacking diplomas were collected with the help from our personal networks. These were then professionally digitally edited to match the trajectories of our fictitious applicants. In a next step we created reference letters for previous positions. Again we worked with examples downloaded from the internet, real reference letters obtained via our networks, and detailed instructions on phrases often used in such letters. As with the CVs and cover letters, these references were discussed and adapted with HR specialists to ensure their plausibility. We also included photographs, as required in standard job applications in the German speaking part of Switzerland. We thank Doris Weichselbaumer for letting us use the pictures that she had carefully prepared, pretested and used in a correspondence test conducted in Austria. We are aware that photographs might introduce unobserved differences, e.g. due to the attractiveness of the candidates. However, the creation of the photographs and the pretesting to minimize these unobservable differences is described in detail in her section on the experimental design (Weichselbaumer 2016b)³.

³ Weichselbaumer (2016b) provides detailed information on the construction of the photographs using student models, that were pre-tested, digitally altered and rated on several dimensions such as “looks, likability, intelligence, reliability, as well as their overall score” (p. 10).

Finally, all candidates include contact details in their application consisting of an individual email address (Gmail), a mobile phone number leading to a mailbox and an existing address in Switzerland. Using Google street view addresses in several Swiss cities were chosen which show apartment houses with multiple resident parties. Our HR contacts have assured us that the likelihood of receiving answers to online applications by regular mail is extremely low⁴.

Results

Between October 2017 to April 2018, 1120 applications were sent to 560 vacancies in the Swiss German labour market. In 86% of the cases one or both of the fictitious applicants received a reply by the employer either by phone or email, while in the other 14% neither applicant was contacted. The results presented in Table 1 show the final outcome of the application procedures – meaning if a candidate was contacted for a job interview or not. They do not indicate if a minority candidate was only invited after the majority candidate had already withdrawn his or her application. They also do not consider qualitative differences in responses, or the number of times employers tried to make contact with the candidates. These qualitative differences are discussed in [reference removed for review].

Swiss candidates received a positive feedback on their application in 39.8% of the application procedures, while overall 35.2% of the minority candidates' applications elicited a positive response. Column 9 of Table 1 translates these outcomes into a ratio, the relative call back rate, which shows the factor of applications that minority candidates have to send compared to their fictitious majority counterpart. On the study level we found a relative call back rate of 1.13, meaning that minority candidates have to write 1.13 times as many applications as majority candidates to be invited for a job interview.

Among the minority candidates, Germans named applicants are the most successful with the share of positive responses almost as high as those of native Swiss candidates (38.7%), followed by Turkish named candidates (34.4%) and, finally those with Kosovar names have the lowest success rate (32.5%). The relative call back rates in the German speaking part of Switzerland differ by the ethnic origin, with Germans experiencing almost no discrimination across all occupations, Turks facing a relative call back rate of 1.14 and finally Kosovars who encounter the highest relative call back rate across occupations (1.26).

⁴ We received emails from employers in three cases (out of 560) where rejection letters sent by post were returned to sender.

	Number of jobs [1]	None invited [2]	Both invited [3]	Only majority invited [4]	Only minority invited [5]	Net discrimination rate % [6]	% call back majority [7]	% call back minority [8]	Ratio [9]	Percentage difference
Aggregate results	560	306	166	57	31	10.2%	39.8%	35.2%	1.13	7.6**
Gender										
Male	270	160	76	23	11	10.9%	36.7%	32.2%	1.14	4.5*
Female	290	146	90	34	20	9.7%	42.8%	37.9%	1.13	4.8*
Ethnicity										
Germany	186	97	56	17	16	1.1%	39.2%	38.7%	1.01	0.5
Kosovo	188	105	54	23	7	19.1%	41.0%	32.5%	1.26	8.5***
Turkey	186	105	56	17	8	11.1%	39.3%	34.4%	1.14	4.8
Occupations										
Sales	136	83	19	24	10	26.4%	31.6%	21.3%	1.48	10.3**
Germany	45	24	7	11	3	38.1%	40.0%	22.2%	1.80	17.8**
Kosovo	46	28	8	7	3	22.2%	32.6%	23.9%	1.36	8.7
Turkey	45	31	4	6	4	14.3%	22.2%	17.8%	1.25	4.4
Electrician	136	56	64	12	4	10.0%	55.9%	50.0%	1.12	5.9**
Germany	45	16	23	3	3	0.0%	57.8%	57.8%	1.00	0
Kosovo	46	16	23	6	1	16.7%	63.0%	52.2%	1.21	10.9*
Turkey	45	24	18	3	0	14.3%	46.7%	40.0%	1.17	6.7*
Nurse	144	62	63	12	7	6.1%	52.1%	48.6%	1.07	3.5
Germany	48	18	22	3	5	-6.7%	52.1%	56.3%	0.93	-4.2
Kosovo	48	27	16	5	0	23.8%	43.8%	33.3%	1.31	10.4**
Turkey	48	17	25	4	2	6.5%	60.4%	56.3%	1.07	4.2
HR clerk	144	106	20	9	10	-2.6%	20.1%	20.8%	0.97	-0.7
Germany	48	39	4	0	5	-55.6%	8.3%	18.8%	0.44	-10.4**
Kosovo	48	34	7	5	3	13.3%	25.0%	20.8%	1.20	4.2
Turkey	48	33	9	4	2	13.3%	27.1%	22.9%	1.18	4.2

Table 1: Correspondence test results, responses by ethnic background and occupation Column [1] shows the total number of vacancies. Column [2] includes both rejections and applications that received no reply. Of 560 jobs applied to, responses were sent to both candidates in 428 cases, in 54 cases only one candidate received a response, and in 78 cases no candidate was contacted. Column [3] shows the number of cases where both candidates received a positive reply. Column [4] lists the cases in which only the Swiss candidate received a positive reply. Column [5] are the cases in which only the minority candidate received a positive reply. Column [6] shows the net discrimination rate: $([4]-[5])/([3]+[4]+[5])$. Columns [7] and [8] report the success rates of the majority $([3]+[4])/[1]$ and minority candidates $([3]+[5])/[1]$, Column [9] presents the ratio between the success rates of the majority and minority candidates, also known as the relative call back rate. Column [10] shows the difference between [7] and [8], as well as the statistical significance of these differences using a chi-square test (significant at the * 10%, ** 5% and *** 1% level).

Looking at the results by the qualification level for all minority candidates aggregated, discrimination is higher in the two low-skilled occupations (1.23) than in the two medium skilled occupations (1.04). Considering the individual occupations, applications were most successful for electricians (55.9% of the majority candidates and 50.0% of the minority candidates receiving a positive response) and least successful for HR positions (20.1% and 20.9% respectively).

	Number of jobs [1]	None invited [2]	Both invited [3]	Only majority invited [4]	Only minority invited [5]	Net discrimination rate % [6]	% call back majority [7]	% call back minority [8]	Ratio [9]	Percentage difference
Low skilled	272	139	83	36	14	16.5%	43.8%	35.7%	1.23	8.1***
German	90	40	30	14	6	16.00%	48.9%	40.0%	1.22	8.9*
Kosovo	92	44	31	13	4	18.8%	47.8%	38.0%	1.26	9.8**
Turkey	90	55	22	9	4	14.3%	34.4%	28.9%	1.19	5.6
Medium skilled	288	168	83	21	17	3.3%	36.1%	34.7%	1.04	1.4
German	96	57	26	3	10	-18.0%	30.2%	37.5%	0.81	-7.3*
Kosovo	96	61	23	10	3	19.4%	34.4%	27.1%	1.27	-2.7*
Turkey	96	50	34	8	4	8.7%	43.8%	39.6%	1.11	4.2

Table 2: Correspondence test results, responses by skill level and ethnic background (chi square test, significant at the * 10%, ** 5% and ***1% level)

For the two low-skilled positions, the relative call back rates also vary between the occupational profiles. While discrimination rates for electricians are quite low (aggregate relative call back rate 1.12), those for the sales sector are the highest (1.48). It is in particular the high relative call back rate of 1.80 against German applicants for sales positions that is striking. In the German speaking part of Switzerland, German named applicants fare worse than those with a Kosovar or a Turkish name (1.36 and 1.25 respectively) when applying for a sales job and have to send almost twice as many applications to be invited for a job interview than their Swiss competitors. Looking at results for electricians, however, Swiss and German named candidates are treated equally when applying for positions as electricians – at least regarding the outcome of job invitation or rejection –, while candidates with Turkish and Kosovar names need to write more applications to get an interview invitation. Yet, the difference between the latter is very small (1.17 vs. 1.21).

For the medium skilled positions, we found a relative call back rate of 1.04, thus there is almost equal treatment of the majority and minority candidates at this skill level. It is interesting to note that at this level, German named applicants are preferred over the native Swiss candidates, while the results are the opposite for applicants with a Kosovar name (0.81 vs 1.27). For the position of HR clerks, the lowest relative call back rates were measured, indicating that there is no discrimination for this occupation. However, for HR positions the invitation rates were very low, so these results have to be treated with caution. For the second medium skilled position, nurses, the results vary again by ethnicity, with German named candidates being preferred over the Swiss, Turkish named candidates facing no significant discriminations and candidates with a Kosovar name being the most disadvantaged with a relative call back rate of 1.31.

Results by gender

If we consider the results by gender of the applicant, we find no difference between male and female applicants on the aggregate level presented in Table 1. While we have one male (electricians) and one female (nurses) dominated position, to which only fictitious candidates of the respective gender applied, the other two positions (sales, HR) received male or female pairs of applications. On the study level, we find discrimination rates that are very similar for males and females (1.14 and 1.13 respectively). If we look at the results more closely, we find that the relative call back rate is highest for men with Kosovar names (1.25), while females with a Turkish and Kosovar migration background also encounter discrimination (relative call back rates of 1.18 and 1.14 respectively).

	Number of jobs [1]	None invited [2]	Both invited [3]	Only majority invited [4]	Only minority invited [5]	Net discrimination rate % [6]	% call back majority [7]	% call back minority [8]	Ratio [9]	Percentage difference
German female	96	43	32	10	11	-1.89%	43.8%	44.8%	0.98	-1.0
German male	90	54	24	7	5	5.56%	34.4%	32.2%	1.07	2.2
Kosovo female	98	56	54	13	5	11.11%	68.4%	60.2%	1.14	8.2*
Kosovo male	90	48	30	10	2	19.05%	44.4%	35.6%	1.25	8.9**
Turkish female	96	47	34	11	4	14.29%	46.9%	39.6%	1.18	7.3*
Turkish male	90	58	22	6	4	6.25%	31.1%	28.9%	1.08	2.2

Table 3: Correspondence test results, responses by gender and ethnic background (chi square test, significant at the * 10%, ** 5% and ***1% level)

Finally, we also tested whether the variations of the photographs, the layout or the cover letter used for an applicant had an influence on the results. Yet, the results of the chi-square tests turned out non-significant.

Discussion

Looking at the results by ethnicity across all occupations, we find signs of ethnic hierarchies in the Swiss labour market. Candidates with Kosovar names experience the most discrimination followed by Turkish named candidates, while we find no discrimination against those with German names at this level. These results mirror findings of attitude research in Switzerland, which consistently shows that candidates from the Balkans or former Yugoslavia face the most negative attitudes in Switzerland (Longchamp et al. 2014; Raymann 2003; Rapp 2015; Ruedin et al. 2013). Turkish immigrants are usually regarded a little better, while Western Europeans and in particular those from neighbouring countries are usually not perceived as a threat. However, the highest relative call back rate of this correspondence test was measured for German candidates applying for sales positions. While readers who are not familiar with the Swiss context might find this result puzzling, work conducted by Helbling (2011) in the city of

Zurich as well as research by Matser et al. (2010) have shown negative attitudes towards German immigrants. Both argue that in the face of many similarities the minor differences between Germans and Swiss Germans are strongly emphasised. The fact that we find higher relative call back rates for Germans applying for sales positions, could also be due to perceived language skills (dialect) and/or expected customer discrimination. Since all sales positions involve customer contact, employers might expect German candidates to only speak high-German and not the Swiss German dialect, which could be regarded as negative by Swiss customers. However, since all candidates had completed their education in Switzerland, where the local dialect is also spoken in schools, they should be expected to be proficient in the local dialect. In all other positions German named candidates face no discrimination or are even preferred to the Swiss candidates. As Krings et al. (2014) have shown, the Swiss perceive Germans as highly competent, but lacking in warmth. This might be beneficial for the positions of electricians, nurses, or HR clerks, while the lack in warmth might contribute to the high discrimination these candidates experience in the sales positions. Interestingly, a preference for German candidates even before national candidates also seems to emerge from research on labour market discrimination in the Netherlands suggesting that German applicants might benefit from positive stereotypes in the labour market (Phlippen and van Eldert 2017).

The results show higher relative call back rates on the occupational level for the lower skilled positions. For all low skilled positions, the relative call back rate is 1.23 (at 1% significance level), with discrimination being strongest for low skilled Kosovars (1.26) and Germans (1.22). At the medium skill level, there is no significant discrimination against all minority applicants. While medium qualified Germans actually encounter positive discrimination (relative call back rate 0.81), this is the opposite for Kosovar candidates (relative call back rate 1.27). This confirms the assumption that discrimination decreases for higher skilled positions (e.g Bovenkerk 1992), but contradicts findings from Auer et al. (2016).

Separated by gender, there are almost no differences between male and female candidates on the aggregated level. Looking at the data more closely, it can be seen that male Kosovar applicants face the highest discrimination rate. Again, these results are in line with Swiss attitude research where immigrants from the Balkans are often perceived as a threat (Rapp 2015).

A relative call back rate of 1.13 in the German-speaking area of Switzerland sounds low in international comparison, as far as it is possible to compare the results of such field experiments. Zschirnt and Ruedin (2016) reported a mean call back rate of 1.6 across all studies included in their meta-analysis, with studies reporting relative call back rates of up to 3.6 (Cédiey and Foroni 2007). However, as depicted in Figure 1, relative call back rates in German speaking countries are often lower in international comparison. The high relative call back rate of 4.48 for female applicants with a Turkish background in Germany shown here can be attributed to the fact that these candidates were wearing a headscarf in the

picture attached to their CV (Weichselbaumer 2016a). As argued by Zschirnt and Ruedin (2016) and Weichselbaumer (2016b) the comprehensive information that potential employers receive about a candidate (i.e. not only cover letter and CV, but also degrees, work certificates, and photographs) make it less likely that employers have to resort to statistical discrimination and make assumptions about an applicant based on his or her ethnic background. Furthermore, at the time the testing was conducted, Switzerland had a low rate of unemployment (below 5%), which has been very stable since the turn of the millennium (Lalive and Lehmann 2017).

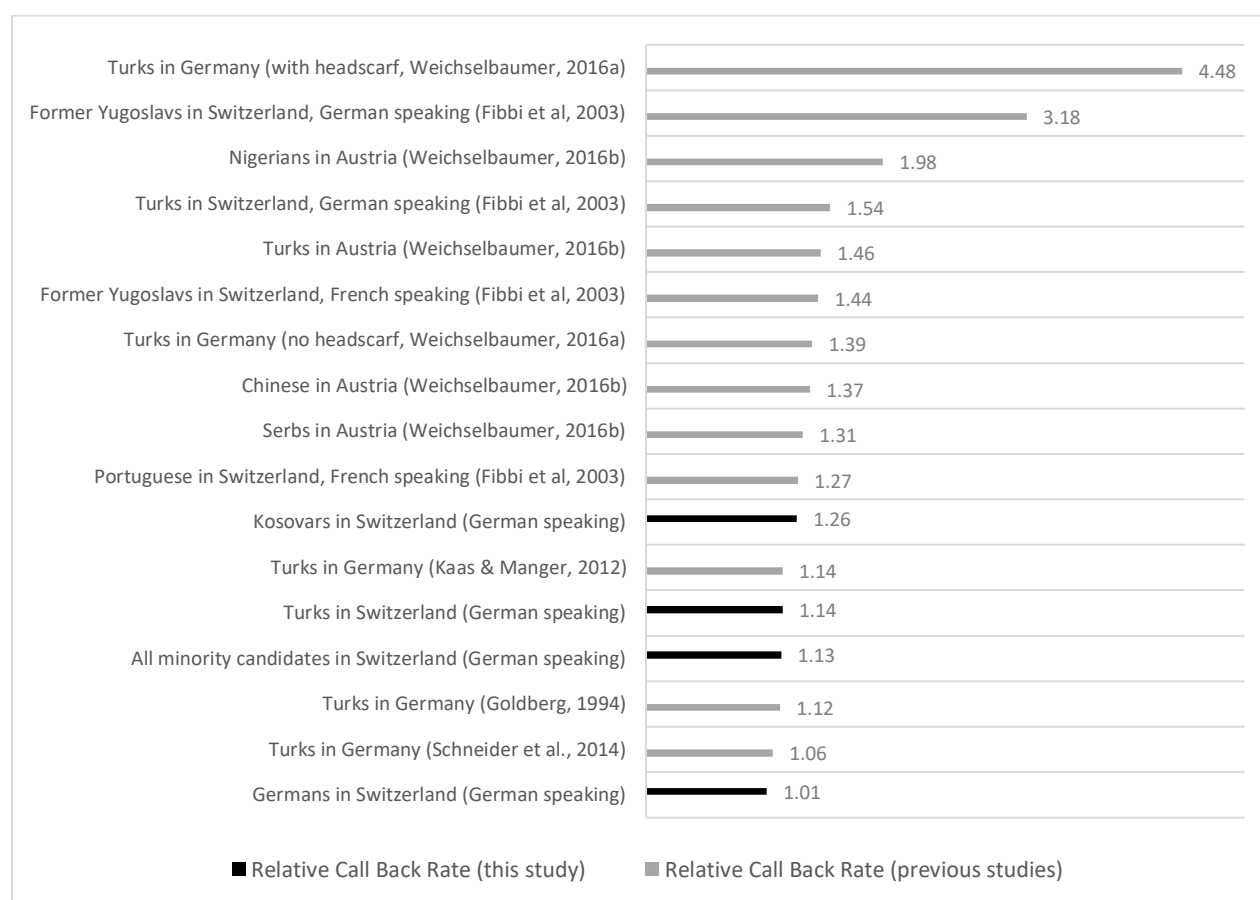


Figure 1: Relative Call Back Rates in German speaking countries, previous and current studies

Since this has been the second correspondence test conducted in the Swiss labour market it raises the question in how far the results are comparable over time. Figure 1 also provides the results of the previous study, which at the first glance are higher than the results presented in this correspondence test. However, there are several methodological differences that make it difficult to compare the two studies. First, Fibbi et al. (2003) tested discrimination against foreign-born youth, yet fully schooled in Switzerland, while the present study tested members of the second generation that were already naturalised in Switzerland and thus showed that they were well integrated in the country. Second, the applications in Fibbi et al. (2003) presented young applicants that had just finished their vocational

training and were looking for their first position. While the applicants presented in this study were also all less than 30 years old, they had already gained several years of work experience following their apprenticeships and trainings. Third, the amount of application material provided differed greatly between the two studies: in the former only cover letters and CVs were sent to employers, while the latter compiled complete application packages, including photographs, diplomas and work certificates. Fourth, the two studies did not test the same occupations, the only overlap occurred for sales assistants. Fifth, each employer tested by Fibbi et al. (2003) received three applications, while the current study only used pairs of applicants. Finally, the earlier study focused the cantons of Zurich and Aargau, while the current testing used vacancies from all over the Swiss German labour market. The only conclusions that can be drawn from this comparison is that both correspondence tests show the existence of ethnic discrimination in the Swiss labour market and that in both cases ethnic hierarchies became apparent, with candidates from today's Kosovo facing the highest discrimination, followed by those from Turkey and Western Europeans (former study Portuguese, latter study Germans) being the least discriminated.

The study faces limitations, some of which are due to the research design of a correspondence test. It is only possible to apply for publicly advertised positions for which applications can be made in writing. However, current research indicates that only 30% of the positions in Switzerland are being advertised (Berther and Casutt 2017). Furthermore, the fact that discrimination was only measured in four occupations and for three ethnic groups prevents us from making more generalised conclusions about the "real" extent of discrimination in the Swiss labour market. As the results for this study have shown, discrimination varies a lot depending on the occupation and ethnicity tested. Thus, while the internal validity of correspondence tests is high, the external validity and the generalisability for the whole labour market are limited.

Apart from McGinnity and Lunn (2011) I am not aware of other correspondence test on ethnic discrimination that included an immigrant group that did not come from a lower income country. In particular in light of easy mobility of workers within the EU and the European Free Trade Area, it would be interesting to see more studies on groups that do not face the same stigmas as traditional immigrant groups, to establish whether discrimination is caused by their status as a foreigner or rather due to cultural and social distance.

Conclusion

Ethnic discrimination exists in the Swiss labour market. Our results suggest that ethnic hierarchies that have been shown in attitude research in Switzerland, are mirrored in hiring decisions, where the group encountering the most negative attitudes, candidates with a Kosovar background, also faces the highest discrimination. However, there is also a strong variation between groups and occupations, which clearly shows that it is not possible to use the results of this testing to infer the extent of discrimination in the entire Swiss labour market. By testing for discrimination against candidates from both low as well as

high income countries, we show a more encompassing picture of discrimination against candidates with a migration background in Switzerland. We conclude that existing studies of discrimination are prone to overestimate the degree of discrimination encountered by immigrants, since they usually focus on stigmatized groups.

In a country in which over 35% of the population are either first or second generation immigrants, such findings of ethnic discrimination are worrying. While Western societies – including the Swiss – always reiterate that positions are awarded based on meritocratic principles, this has been proven wrong in discrimination studies in numerous countries and areas of life. Research that looked at in the search of apprenticeships, transitions from apprenticeships to the labour market, or at workplace incivilities faced by immigrants at the workplace, shows that discrimination is not only encountered once, but often frequently during the lifetime of immigrants in the same or different domains. These instances of cumulative discrimination (Blank et al. 2004) can become substantial over time.

The fact that discrimination is rather low in international comparison – as far as it is possible to compare results across countries and time – points in the same direction as findings by Weichselbaumer (2016b) and Zschirnt and Ruedin (2016). They both argued that the amount of information provided in complete applications in German speaking countries makes it less likely that employers have to resort to statistical discrimination, since the amount of information reduces uncertainty about the applicants. It could be possible, that the discrimination found in these experiment is rather due to taste based discrimination – a hypothesis that might be examined looking at the responses that candidates received from employers to their applications, to see if these responses show signs of biases.

References

- Arriijn, P., Feld, S., & Nayer, A. (1998). *Discrimination in access to employment on grounds of foreign origin: the case of Belgium*. Geneva: International Labour Office, Conditions of Work Branch.
- Arrow, K. J. (1973). The Theory of Discrimination. In O. Ashonfelter, & A. Rees (Eds.), *Discrimination in Labor Markets* (pp. 15-42). Princeton: Princeton University Press.
- Auer, D., Bonoli, G., & Fossati, F. (2017). Why do immigrants have longer periods of unemployment? Swiss evidence. *International Migration*, 55(1), 157-174, doi:10.1111/imig.12309.
- Auer, D., Bonoli, G., Fossati, F., & Liechti, F. (2016). The Matching Hierarchies Model: Evidence from a Survey Experiment on Employers' Hiring Intent of Immigrant Applicants. *SSRN*.
- Baert, S. (2018). Hiring Discrimination: An Overview of (Almost) All Correspondence Experiments Since 2005. In M. Gaddis (Ed.), *Audit Studies: Behind the Scenes with Theory, Method and Nuance* (pp. 63-77, Methods Series, Vol. 14): Springer.
- Baert, S., & De Pauw, A.-S. (2014). Is ethnic discrimination due to distaste or statistics? *Economics Letters*, 125(2), 270-273.
- Becker, G. (1957). *The Economics of Discrimination* (2nd ed.). Chicago: University of Chicago Press.
- Berther, L., & Casutt, J. (2017). Sieben von zehn Stellen werden nicht ausgeschrieben. https://awa.zh.ch/content/dam/volkswirtschaftsdirektion/awa/amt/wm/monitorings/201704_WiMo/AWA_WiMo_04-2017_4-6.pdf. Accessed 01.04.2018 2018.
- Bertrand, M., & Duflo, E. (2016). *Field Experiments on Discrimination*. London: Centre for Economic Policy Research.
- Binggeli, S., Krings, F., & Sczesny, S. (2014a). Perceived competition explains regional differences in the stereotype content of immigrant groups. *Social Psychology*, 45(1), 62.
- Binggeli, S., Krings, F., & Sczesny, S. (2014b). Stereotype content associated with immigrant groups in Switzerland. *Swiss journal of psychology*, 73(3), 123-133.
- Blank, R. M., Dabady, M., & Citro, C. F. (2004). *Measuring racial discrimination*. Washington, DC: National Academies Press.
- Booth, A. L., Leigh, A., & Varganova, E. (2012). Does ethnic discrimination vary across minority groups? Evidence from a field experiment. *Oxford Bulletin of Economics and Statistics*, 74(4), 547-573.
- Bovenkerk, F. (1992). *Testing discrimination in natural experiments: a manual for international comparative research on discrimination on the grounds of "race" and ethnic origin*. Geneva: International Labour Office.
- Bovenkerk, F., Gras, M., Ramsodh, D., Dankoor, M., & Havelaar, A. (1995). Discrimination against migrant workers and ethnic minorities in access to employment in the Netherlands. Geneva: Employment Department, International Labour Office.
- Buchmann, M., Buchs, H., Gnehm, A.-S., Hevenstone, D., Klarer, U., Müller, M., et al. (2015). Erhebung von Stelleninseraten 1950-2014 [Dataset]. In U. Zürich (Ed.). Lausanne: FORS.
- Bundesamt für Statistik (2016). Arbeitsmarktindikatoren 2016. *Statistik der Schweiz*. Neuchâtel: Bundesamt für Statistik BfS.
- Bundesamt für Statistik (2017). Demographic balance of the permanent resident population by citizenship 2010-2016. <https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung.assetdetail.300195.html>. Accessed 21.04.2018.
- Cédiey, E., & Foroni, F. (2007). *Les discriminations à raison de "l'origine" dans les embauches en France. Une enquête nationale par tests de discrimination selon la méthode du Bureau International du Travail*. Genève: Bureau International du Travail.
- Crabtree, C., & Chykina, V. (2018). Last Name Selection in Audit Studies. *Sociological Science*, 5, 2-21.
- de Prada, M. A., Actis, W., Pereda, C., & Molina, R. P. (1995). *Labour market discrimination against migrant workers in Spain* (Vol. 9, International Migration Papers). Geneva: International Labour Office, Employment Department.

- Efionayi-Mäder, D., Ruedin, D., Pétrémont, M.-E., Michel, N., & Jain, R. (2017). Anti-Schwarzen-Rassismus in der Schweiz—Eine Bestandsaufnahme. Universität de Neuchâtel.
- Fibbi, R., Kaya, B., & Piguet, E. (2003). *Le passeport ou le diplôme ? Etude des discriminations à l'embauche des jeunes issus de la migration* (PNR 43 Rapport de recherche au FNS). Neuchâtel: Forum suisse pour l'étude des migrations et de la population.
- Flinn, C. J. (2015). GARY BECKER'S CONTRIBUTIONS TO THE ANALYSIS OF DISCRIMINATION. *Journal of Demographic Economics*, 81(01), 45-50, doi:doi:10.1017/dem.2014.5.
- Fossati, F., Liechti, F., Auer, D., & Bonoli, G. (2017). Discrimination Multipliers - How Immigrants' integration affects labour market disadvantage. *MIM Working Papers Series* (Vol. 17, pp. 1-22). Malmö: Malmö Institute for Studies of Migration, Diversity and Welfare (MIM).
- Fryer Jr, R. G., & Levitt, S. D. (2004). The causes and consequences of distinctively black names. *The Quarterly Journal of Economics*, 767-805.
- Gaddis, S. M. (2017a). How Black are Lakisha and Jamal? Racial Perceptions from Names Used in Correspondence Audit Studies. *Sociological Science*, 4, 469-489.
- Gaddis, S. M. (2017b). Racial/Ethnic Perceptions from Hispanic Names: Selecting Names to Test for Discrimination. *Socius*, 3, 2378023117737193, doi:10.1177/2378023117737193.
- Gaddis, S. M. (2018). An introduction to audit studies in the social sciences. In *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 3-44): Springer.
- Goldberg, A., Mourinho, D., & Kulke, U. (1995). *Labour market discrimination against foreign workers in Germany*: International Labour Office, Employment Department.
- Golder, L., Mousson, M., Tschöpe, S., Herzog, N., & Bohn, D. (2017). Hinweise auf erlebte, aber kaum registrierte Hasskriminalität - Pilotstudie Diskriminierungserfahrungen Muslime in der Schweiz. In G. Bern (Ed.). Bern.
- Guryan, J., & Charles, K. K. (2013). Taste-based or Statistical Discrimination: The Economics of Discrimination Returns to its Roots. *The Economic Journal*, 123(572), F417-F432.
- Heckman, J. J. (1998). Detecting discrimination. *The Journal of Economic Perspectives*, 101-116.
- Heckman, J. J., & Siegelman, P. (1993). The Urban Institute Audit Studies: Their Methods and Findings. In M. Fix, & R. Struyk (Eds.), *Clear and Convincing Evidence: Measurements of Discrimination in America* (pp. 187-258). Washington, DC: The Urban institute Press.
- Helbling, M. (2011). WHY SWISS-GERMANS DISLIKE GERMANS. *European Societies*, 13(1), 5-27, doi:10.1080/14616696.2010.533784.
- Helbling, M., & Kriesi, H. (2014). Why Citizens Prefer High- Over Low-Skilled Immigrants. Labor Market Competition, Welfare State, and Deservingness. *European Sociological Review*, 30(5), 595-614, doi:10.1093/esr/jcu061.
- Huddleston, T., Bilgili, Ö., Joki, A.-L., & Vankova, Z. (2015). *Migrant Integration Policy Index 2015: Switzerland* (Vol. 09.07.2015). Barcelona/Brussels: CIDOB and MPG.
- Jackson, M., & Cox, D. R. (2013). The Principles of Experimental Design and Their Application in Sociology. *Annual Review of Sociology*, 39(1), 27-49, doi:10.1146/annurev-soc-071811-145443.
- Keuschnigg, M., & Wolbring, T. (2016). The Use of Field Experiments to Study Mechanisms of Discrimination. *Analyse & Kritik* (Vol. 38, pp. 179).
- Krings, F., Johnston, C., Binggeli, S., & Maggiori, C. (2014). Selective incivility: Immigrant groups experience subtle workplace discrimination at different rates. *Cultural Diversity and Ethnic Minority Psychology*, 20(4), 491-498, doi:10.1037/a0035436.
- Krings, F., & Olivares, J. (2007). At the doorstep to employment: Discrimination against immigrants as a function of applicant ethnicity, job type, and raters' prejudice. *International Journal of Psychology*, 42(6), 406-417, doi:10.1080/00207590701251721.
- Lahey, J., & Beasley, R. (2018). Technical aspects of correspondence studies. In M. Gaddis (Ed.), *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 81-101): Springer.
- Lalive, R., & Lehmann, T. (2017). The labor market in Switzerland, 2000–2016. *IZA World of Labor*, 402-402.

- List, J. A. (2009). An introduction to field experiments in economics. *Journal of Economic Behavior & Organization*, 70(3), 439-442, doi:<http://dx.doi.org/10.1016/j.jebo.2008.10.013>.
- Longchamp, C., Imfeld, M., Tschöpe, S., Müller, M., Rochat, P., & Schwab, J. (2014). Verbreitung und Entwicklung von Rassismus, Fremdenfeindlichkeit, Muslimfeindlichkeit und Judenfeindlichkeit - Schlussbericht zur Studie "Zusammenleben in der Schweiz 2010-2014". Bern: Gfs Bern.
- Matser, C., van Oudenhoven, J. P., Askevis-Leherpeux, F., Florack, A., Hannover, B., & Rossier, J. (2010). Impact of relative size and language on the attitudes between nations and linguistic groups: The case of Switzerland. *Applied Psychology*, 59(1), 143-158.
- McGinnity, F., & Lunn, P. D. (2011). Measuring discrimination facing ethnic minority job applicants: an Irish experiment. *Work, Employment & Society*, 25(4), 693-708.
- Midtbøen, A. H., & Rogstad, J. (2012). Discrimination. Methodological controversies and sociological perspectives on future research. *Nordic Journal of Migration Research*, 2(3), 203-212.
- Mühlemann, D. (2017). Rassismussvorfälle aus der Beratungspraxis - Januar bis Dezember 2016. Bern: Verein humanrights.ch, Eidgenössische Kommission gegen Rassismus (EKR).
- Neumark, D. (2012). Detecting Discrimination in Audit and Correspondence Studies. [Article]. *Journal of Human Resources*, 47(4), 1128-1157.
- Pecoraro, M., & Ruedin, D. (2016). A foreigner who does not steal my job: The role of unemployment risk and values in attitudes toward equal opportunities. *International migration review*, 50(3), 628-666.
- Pecoraro, M., & Ruedin, D. (2017). The Share of Foreigners in One's Occupation and Attitudes Towards Foreigners. In I. o. E. Research (Ed.), *IRENE Working Paper* (Vol. 17-06). Neuchâtel.
- Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review*, 62(4), 659-661.
- Niederlander heeft liefst een Duitser op de werkvloer. (2017, *Algemeen Dagblad*.
- Rapp, C. (2015). More diversity, less tolerance? The effect of type of cultural diversity on the erosion of tolerance in Swiss municipalities. *Ethnic and Racial Studies*, 38(10), 1779-1797, doi:10.1080/01419870.2015.1015582.
- Raymann, U. (2003). Meinungen und Einstellungen gegenüber AusländerInnen in der Schweiz. Zürich: GfS-Forschungsinstitut.
- Riach, P. A., & Rich, J. (2002). Field Experiments of Discrimination in the Market Place. *The Economic Journal*, 112(483), F480-F518, doi:10.2307/798458.
- Riach, P. A., & Rich, J. (2004). Deceptive field experiments of discrimination: are they ethical? *Kyklos*, 57(3), 457-470.
- Rich, J. (2014). What Do Field Experiments of Discrimination in Markets Tell Us? A Meta Analysis of Studies Conducted since 2000. *Discussion Paper Series*. Bonn: IZA
- Ruedin, D., D'Amato, G., Wichmann, N., & Pecoraro, M. (2013). Migrationsängste der Schweizer Bevölkerung. Neuchâtel: Swiss Forum for Migration and Population Studies (SFM).
- Thijssen, L. (2016). Taste-based versus Statistical Discrimination: Placing the Debate into Context. *GEMM Project*.
- Vidal-Coso, E., & Ortega-Rivera, E. (2017). Skill Composition and Occupational Incorporation of Early and Recent Immigrants in Switzerland: the case of Italians and Spaniards. *International Migration*, 55(S1), 86-111.
- Vuolo, M., Uggen, C., & Lageson, S. (2016). Statistical power in experimental audit studies: Cautions and calculations for matched tests with nominal outcomes. *Sociological Methods & Research*, 45(2), 260-303.
- Weichselbaumer, D. (2016a). Discrimination against Female Migrant Wearing Headscarves. *IZA Discussion Paper Series*. Bonn: Institute for the Study of Labor (IZA).
- Weichselbaumer, D. (2016b). Discrimination Against Migrant Job Applicants in Austria: An Experimental Study. *German Economic Review*, 18(2), 237-265, doi:10.1111/geer.12104.
- Zschirnt, E. (2016a). Measuring Hiring Discrimination – A History of Field Experiments in Discrimination Research. *NCCR Working Papers*, 7, 1-32.

Zschirnt, E. (2016b). Revisiting Ethics in Correspondence Testing. *NCCR Working Papers*, 8.

Zschirnt, E., & Ruedin, D. (2016). Ethnic Discrimination in Hiring Decisions: A Meta-Analysis of Correspondence Tests 1990–2015. *Journal of Ethnic and Migration Studies*, 42(7), 1115-1134, doi:10.1080/1369183X.2015.1133279.

Paper V:

Equal outcomes, but different treatment – subtle discrimination in email responses from a correspondence test in Switzerland

Submitted to the Swiss Journal of Sociology
Accepted Manuscript, forthcoming 2019

**Equal outcomes, but different treatment –
subtle discrimination in email responses from a
correspondence test in Switzerland**

Eva Zschirnt

ORCID iD: 0000-0002-8918-4146

University of Neuchâtel
Swiss Forum for Migration and
Population Studies SFM
Rue A.L. Breguet 2
2000 Neuchâtel, Switzerland
Email: eva.zschirnt@unine.ch

Max Weber Fellow
Department of Political and Social
Sciences
European University Institute
Via dei Roccettini, 9
50014 San Domenico di Fiesole
Tel: (+39)-055-4685-517
Email: eva.zschirnt@EUI.eu

Acknowledgements

This research was supported by the nccr – on the move funded by the Swiss National Science Foundation. I would like to thank Daniel Auer, Rosita Fibbi, Carolin Fischer, Fabienne Liechti and Didier Ruedin as well as the two anonymous reviewers for their comments.

**Equal outcomes, but different treatment –
subtle discrimination in email responses from a correspondence test in
Switzerland**

Abstract: Correspondence tests on discrimination usually report only whether an applicant was invited for a job interview or not. While in-person audit studies can observe differences in treatment during a job interview, such as in the length or the tone of an interview, this has barely been done in correspondence studies. Data from a field experiment in Switzerland demonstrate that the commonly reported results of correspondence tests show only one side of the picture. Candidates with the same outcome (invited, not invited) are not necessarily treated equally. The paper complements correspondence test results with information on the time elapsed until candidates were contacted, as well as qualitative differences in invitation or rejection emails.

Key words: hiring discrimination, correspondence test, Switzerland, labour market, subtle discrimination

Des résultats égaux, mais un traitement différent - discrimination subtile dans les réponses par courriel dans un test par correspondance en Suisse

Abstract: Les tests par correspondance se limitent généralement à indiquer si un candidat a été invité à un entretien d'embauche ou non. Les tests par des acteurs permettent d'observer des différences de traitement au cours des entretiens, mais ils n'ont presque jamais été réalisés dans les études par correspondance. Les données d'une expérience suisse montrent que les candidats avec le même résultat final ne sont pas nécessairement traités

de manière égale. L'article complète les résultats du test par correspondance avec des informations sur le temps écoulé jusqu'au moment où les candidats sont contactés, ainsi que sur les différences qualitatives dans les courriels d'invitation ou de rejet.

Mots clés : discrimination à l'embauche, test par correspondance, Suisse, marché du travail, discrimination subtile

Gleiche Ergebnisse, aber unterschiedliche Behandlung - subtile Diskriminierung bei E-Mail-Antworten aus einem Korrespondenztest in der Schweiz

Abstrakt: Korrespondenztests zeigen in der Regel nur, ob ein Bewerber zu einem Vorstellungsgespräch eingeladen wurde oder nicht. Während bei Studien mit persönlichem Kontakt Unterschiede im Laufe des Vorstellungsgesprächs beobachtet werden können, wurde dies bei Korrespondenztests kaum gemacht. Daten aus einem Schweizer Feldexperiment zeigen, dass Kandidaten mit dem gleichen Ergebnis nicht unbedingt gleich behandelt werden. Dieser Artikel ergänzt die Ergebnisse des Korrespondenztests mit Informationen über die Zeit bis zur Kontaktaufnahme mit den Kandidaten sowie mit qualitativen Unterschieden in den Einladungs- oder Ablehnungs-E-Mails.

Stichworte: Einstellungsdiskriminierung, Korrespondenztest, Schweiz, Arbeitsmarkt, subtile Diskriminierung

1. Introduction

In recent years there has been an increase in field experiments on ethnic or racial discrimination in the labour market (Bertrand & Duflo, 2017; Neumark, 2018; Riach & Rich, 2002; Rich, 2014). In these field experiments researchers present real employers with fictitious candidates applying for advertised positions either in person, on the phone, or in writing. Candidates differ only in the characteristics that are the focus of the study (e.g. ethnicity) and differences in invitation rates can then be attributed to discriminatory treatment in the hiring process. A meta-analysis by Zschirnt and Ruedin (2016) has shown that ethnic minority candidates have to write about 1.6 times as many applications to be invited for a job interview. However, field experiments focus on the final outcome alone, i.e. whether applicants were invited for a job interview or not. It is assumed that in cases where both candidates were invited or rejected, they were both treated equally. While the outcome might – eventually – be the same, it is nevertheless possible to observe differences in the treatment of applicants. These differences, for example the tone of the correspondence or the length of time until a reply was received, are, however, not usually discussed.

There are only a few studies that address these subtle forms of discrimination. These have been mostly in-person audit studies where the testers were able to observe treatment during the job interview. There is yet just one example of a correspondence test analysing email responses for subtle discrimination; this is on the US housing market (Hanson et al., 2011). Correspondence tests on hiring discrimination published in recent years have not usually made use of the information that email responses in particular provide (Crabtree, 2018). The only aspect addressed in addition to the correspondence test

results is the time difference relating to when applicants were contacted (Kaas & Manger, 2012; Weichselbaumer, 2016). Looking at employers' responses to paired application side by side therefore enables us to consider not only the outcome of the application procedure, but also to look at the subtle discrimination revealed in emails.

This paper uses results from a recent Swiss correspondence test. It compares the results of the field experiment with the way applicants were contacted by employers. Focusing on the timing of responses and the content of the emails that candidates received, I show that the simplification of correspondence test results happens at the expense of a more nuanced picture.

2. Theory

Research on labour market discrimination, in particular on the grounds of race and ethnicity, has a long tradition, in particular in economics and sociology. This literature often focuses on two classical economic theories to explain why discrimination occurs at the interpersonal level¹: the theory of taste-based discrimination (Becker, 1957) and the theory of statistical discrimination (Aigner & Cain, 1977; Arrow, 1973; Phelps, 1972). Becker's theory departs from the assumption that people (e.g. employers) have a certain distaste for working with a particular group (e.g. migrants or women) and are willing to pay a price (e.g. a higher salary to another candidate) to avoid hiring members of this disliked group. Statistical discrimination theory, however, assumes that employers act in a way to maximise profits and resort to discriminatory hiring behaviour to make up for a lack of information about an applicant. Thus, they use group membership as a proxy to make

¹ For a detailed discussion on the causes of discrimination see Pager and Shepherd (2008) or Reskin (2003).

up for this missing information. While both theories are very prominent in the academic debate, they fail to explain why discrimination still occurs at persistent levels over time and place. Both theories predict that discrimination should disappear or at least decline over time, either because discriminating employers go out of business (taste-based discrimination) or because employers receive more information about different groups the better they become known and thus no longer feel the need to infer information based on group signals (statistical discrimination) (e.g. Darity & Mason, 1998). Yet, meta-analyses of field experiments have shown that discrimination rates are quite stable over time (with a focus on the US Quillian et al., 2017; Zschirnt & Ruedin, 2016).

Next to these two well-known economic theories researchers in other fields, such as sociology and social psychology, have developed alternative explanations for the occurrence of discriminatory treatment. One approach focuses on subtle discrimination. Research on subtle discrimination in the work place or on work place incivility has been advanced mostly in the field of social psychology (for the case of Switzerland see e.g. Krings et al., 2014), for example Van Laer and Janssens (2011) define subtle discrimination as

“forms of discrimination that pervade society, are less visible, often very ambiguous for those experiencing it, not easily recognized as discrimination and often not punishable under anti-discrimination legislation. It entails interpersonal discrimination that is often enacted unconsciously or unintentionally and that is entrenched in common, everyday interactions, taking the shape of harassment, jokes, incivility, avoidance and other types of disrespectful treatment.” (p. 1205).

While the acts might seem unimportant at first glance, Rowe (1990) argues that “these mechanisms of prejudice against persons of difference are usually small in nature, but not trivial in effect. They are especially powerful taken together” (p. 153), thus alluding to the concept of cumulative discrimination (Blank et al., 2004). Similarly, Krings et al. (2014) have shown that immigrants experience instances of subtle discrimination or workplace incivility. In their findings, groups that are believed to integrate easily because they are competitive and from neighbouring countries experience “seemingly harmless discourteous behaviours” (p. 497). The possibility of cumulative effects of discriminatory treatment (apart from being invited to a job interview or not) are discussed, for example, in Bendick (1996), who points out that “taken together, these effects make the labour market experience of identically-qualified minority and majority job applicants profoundly different” (p.29). Thus, even behaviour that is often unconscious and unintentional can constitute subtle discrimination and can have harmful and cumulative long-term effects.

3. Background

Since the late 1960s, researchers have attempted to measure ethnic and racial discrimination in hiring decisions using field experiments (Daniel, 1968). Since then, the methodology has evolved considerably (Bertrand & Duflo, 2017; Neumark, 2018; Riach & Rich, 2002; Rich, 2014; Zschirnt, 2016). Furthermore, meta-analyses have shown that ethnic or racial minority job candidates encounter discrimination in all the countries studied (Zschirnt & Ruedin, 2016) and that discrimination rates remain stable over time (ibid; for the US in particular Quillian et al., 2017).

While most research on hiring discrimination focuses on the observable differences in invitation rates for minority and majority candidates, it is also possible to observe more subtle forms of discrimination. It has been shown that the data collected in a field experiment can be used for more than just the analysis of the outcome; that is, whether applicants received a call-back for a position or not. Riach and Rich (2002) already mention that some of the experimental studies discussed in their paper recorded differences in treatment despite equal outcomes in the application process, quoting the Fair Employment Commission's (FEC) and Urban Institute (UI) studies in the US or the International Labour Organisation (ILO) studies in Europe. Summarising these US studies, Bendick (1996) includes examples of discriminatory behaviour that had been observed in the tests, e.g. being invited for an interview, receiving a job offer or a referral, the compensation offered, being steered into certain positions, being offered alternative opportunities, and the cumulative effect of these forms of differential treatment. However, so far, such instances of differential treatment have been observed mostly in in-person audit studies (e.g. Bendick, 1996; Bendick et al., 1991; Bendick et al., 2010; Ghumman & Ryan, 2013; Lodder et al., 2003; Pager et al., 2009). Information that was recorded included the length of phone calls or interviews, where and by whom the interview was conducted, the number of topics discussed, the differences in compensation, hours or shifts offered, if information about the job duties was offered without being asked, how politely applicants were treated, if additional vacancies were mentioned, or if applicants were steered towards other positions.

Looking, for example, at the in-person audit study by Pager et al. (2009), the results from the field experiment are complemented with testers' observations on the differences

in treatment. While the experimental results show the importance of race in hiring, analysing these field notes provides a more nuanced picture. They group their observations of employers' responses into three categories: (1) a categorical exclusion of the minority candidate, (2) shifting standards where "employers' evaluations of applicants appear actively shaped or constructed through a racial lens" (Pager et al., 2009, p. 787), and (3) race-coded job channelling, where candidates are channelled into certain job types. They argue that their descriptive results "[reveal] mechanisms at work that observational research can rarely identify" (p. 787). Even in cases that are recorded as equal treatment in the results presented in the field experiment, discriminatory mechanisms can be at play.

Equally detailed information on the more nuanced aspects of the application process is usually missing in written field experiments. I am aware of only one correspondence test on the US housing market, where email correspondence with landlords was analysed for subtle discrimination (Hanson et al., 2011). While the ILO project on labour market discrimination included information on equal but different treatment, this applied mostly to the telephone contact stage (Arriijn et al., 1998; Attström, 2007; Bovenkerk et al., 1995; Cediey & Foroni, 2008; de Prada et al., 1995; Goldberg et al., 1995)². In their study on the Netherlands Bovenkerk et al. (1995) emphasised that

"Although the evidence of discrimination in these instances is less "hard" than the difference between acceptance and rejection (unequal treatment), it is important in practice. Equal but different treatment does not deny one the opportunity to compete for the job, but may be clearly discouraging for the applicant." (p. 12).

² Although not part of the ILO project Fibbi et al. (2003) followed the same approach.

Thus, like the other ILO studies, Bovenkerk et al. provide information about the cases where equal but different treatment occurred.

Looking at studies conducted since 2000, Rich (2014) addresses instances of “Other dimensions of differential treatment” or “dishonest concealment of rejection” and identifies them both on the labour and the housing market. For the former, she refers particularly to Pager et al. (2009), discussed above, and Drydakis and Vlassis (2010) who showed differences in wages and insurance coverage offered to applicants. Not discussed in Rich (2014) is Gaddis (2014) who included information from forwarded emails that had accidentally also been sent to applicants. Furthermore, both Kaas and Manger (2012) and Weichselbaumer (2016) provide information about the time interval in which candidates were contacted. As Weichselbaumer (2016) points out “Migrants are not only discriminated with respect to the frequency with which they are invited for an interview, they also receive these invitations more hesitantly” (p. 26). Not only are they less likely to be invited for a job interview, they also have to wait longer for an invitation.

These small differences in treatment are difficult to observe, unless cases are regarded side by side, and these instances of “equal but different treatment” as they were called in the ILO studies or “subtle discrimination” by Hanson et al. (2011) are observed not only in the hiring process but throughout a person’s working life.

4. Methods

To address this gap in research, this paper uses data from a recent correspondence test in the German speaking area of Switzerland. It studied hiring discrimination against equally qualified German, Kosovar, and Turkish candidates using a matched pair design³.

Ethnicity was signalled by the name, information on citizenship, and listing additional mother languages on the CV where applicable. As is customary in Switzerland, applicants also include their citizenship on their CV; minority candidates listed dual citizenship. Furthermore, Turkish and Kosovar candidates mentioned their respective languages as a second mother tongue. Finally, all education and work experience listed in the CV had been obtained in Switzerland.

Vacancies for two positions, requiring a completed apprenticeship, as a sales assistant (*Detailhandelskauffrau/-mann*) and electrician (*Elektroinstallateur*), and two intermediately skilled positions as a nurse (*Pflegefachfrau*), and HR clerk (*HR Fachmann/-frau*) were obtained from internet job boards. Application material, which in Switzerland includes not only a cover letter and CV, but also work certificates and diplomas, was carefully constructed and discussed with HR specialists. For the final required element, a photograph, we received the permission from Doris Weichselbaumer to use photographs she had carefully prepared and pretested for her Austrian correspondence test (Weichselbaumer, 2016). The individual parts of the application materials (CV, cover letter, photo, etc.) were randomly assigned to the candidates. The paired applications were then sent online. Responses were received by email and mobile phone and carefully recorded; invitations for job interviews were quickly and politely declined by email.

³ Details of the research design are provided in (removed for review), which discusses the quantitative results of the field experiment.

While correspondence testing allows researchers to observe real life hiring decisions, it also faces limitations, one of them being that only advertised positions can be included in the experiment (Heckman & Siegelman, 1993). If positions are filled using word of mouth or personal networks this is likely to be a disadvantage for job-seekers with a migrant background.

For this paper, the time span between sending an application and receiving the first reply by phone or email was measured in workdays and then analysed by duration until a rejection or invitation was received. In a second step, only responses received by email were considered when analysing content for signs of subtle discrimination.

5. Results

This section briefly presents and discusses the correspondence test results, i.e. the descriptive statistics as they are customarily displayed. It combines these findings with information on the time when candidates were contacted and examples of the content of email responses.

5.1 Results from the correspondence test

Focusing on the German speaking part of Switzerland and the positions of electrician and sales assistant, from October 2017 to March 2018 paired applications were sent to 136 positions for electrician and 136 positions for sales assistant, i.e. 544 individual applications. For both of these occupations, vacancies usually mentioned that customer contact was part of the position. In 204 of these 272 application procedures both candidates

received a response from the employer, in 31 cases only one applicant received a reply and in 37 cases employers did not react to either candidate⁴.

Considering the 31 cases in which only one applicant was contacted by the employer, while the second applicant received no reaction, in 15 cases the majority candidate was invited for a job interview compared to three invitations for the minority candidate. In 13 cases only one candidate received a rejection email, while the other candidate was not contacted, although there is no clear pattern here, in 6 of these cases the Swiss candidates received no reply and in 7 cases the minority candidate was not contacted.

Table 1 about here

The correspondence test results presented in Table 1 show that minority candidates in positions requiring a completed apprenticeship have to write 1.23 times as many applications to be invited for a job interview compared to majority candidates (significant at the 5% level). The relative call back rate is lower for the electricians than for the sales positions (1.12 and 1.48 respectively, both significant at the 5% level) and also varies by ethnic background.

While German candidates for sales positions show a high relative call back rate of 1.8 (significant at 5%) for the sales positions, they are treated equally when applying as electricians (relative call back rate of 1.0, not significant). The high discrimination rate against Germans in sales positions is somewhat surprising. Although German named candidates (like all non-Swiss named candidates) indicate that they are dual nationals and

⁴ A more detailed discussion of the quantitative results of this correspondence test is provided in (removed for review)

that all their schooling had been completed in Switzerland, one possible explanation could be that Swiss German employers expect that these candidates will not be fluent in the local dialect and anticipate that this would be unacceptable to their customers. This high relative call back rate is in line with attitude research conducted by Helbling (2011) in Zurich where he found strong anti-German attitudes. Turkish candidates fare better than candidates from Kosovo in both occupations (relative call back rates: Turkish electricians 1.17, Turkish sales assistants 1.25, Kosovar electricians 1.21, Kosovar sales assistants 1.36, not significant for sales assistants, significant at 10% for electricians). Because of the lower overall success rate for sales assistants, these results are statistically significant only for the position of electrician.

Overall, the differences by ethnicity mirror findings from attitude research conducted in Switzerland, where immigrants from the Balkans and former Yugoslavia are usually the least liked group in Switzerland, those from Turkey being regarded slightly more favourably and immigrants from EU countries and neighbouring countries are mostly accepted (e.g. Helbling, 2011; Longchamp et al., 2014; Rapp, 2015; Raymann, 2003; Ruedin et al., 2013). There are also differences in the discrimination rate by ethnic groups and occupations studied. The two biggest extremes were observed for German candidates, who are treated equally to their native Swiss peers when applying for positions as electricians, but face the highest rates of discrimination measured in this correspondence test when applying for sales positions. Turkish candidates face discrimination in both occupations, while Kosovar candidates are the most discriminated against for positions as electricians and fare worse than Turkish candidates for sales positions.

5.2 Timing of responses

As discussed above, correspondence tests usually stop at reporting these results, i.e. whether an applicant was invited for a job interview or not. However, both Kaas and Manger (2012) and Weichselbaumer (2016) reported that the time when candidates received a reply from the employer varied according to the candidate's background. Looking at the Swiss case the results do not show a clear pattern, except that Kosovar candidates have to wait the longest before receiving a reply (either invitation or rejection).

Figure 1 about here

5.3 Content of Email responses

In 112 of 145 cases in which both candidates were contacted by email, messages were the same or very similar. Yet, in the rest of the responses messages differ regarding the names and salutations used, the level of enthusiasm about candidates, keeping applicants in the candidate pool, or blatantly preferring the Swiss candidate. These differences in treatment are shown here with examples from the replies⁵ for both the majority and the minority candidates juxtaposed with each other.

Table 2 about here

Starting at the top of the body of the email, the salutation line is the first place where unequal treatment occurs. Here we observed misspelled names, once only for the

⁵ Own translations from German to English. Translations were kept as close to the German original as possible. Some replies were unusual, even in the original German version.

majority candidate (Kählin instead of Kälín), but more frequently for the minority candidate (e.g. Hofmann instead of Hoffmann; Krasnigi instead of Krasniqi); using the wrong gender (“*sehr geehrte Herr Hoffmann*”, although these instances might be cases of typos); confusing first and last names (Mr. Cem, instead of Mr. Yilmaz); or greeting minority candidates with the first and last name, while only the last name is used for Swiss candidates (Dear Ms. Kaelin vs. Dear Ms. Shpresa Krasniqi). Using both the first and last name in the salutation is less formal than using the last name alone and this occurs in the case of Kosovar candidates in particular. This could be a reflection of uncertainty about the given and the family name. While the mistakes about spelling and/or confusing first and last names might be due to employers being less familiar with non-Swiss names, the names chosen were among the most frequent names in these migrant populations and in particular the Turkish and Kosovar names should be easily identifiable and familiar to Swiss employers. There were also cases where no greeting was used for the minority candidate (“Krasniqi” vs. “Dear Mr. Kaelin), which is unusual and rude in German business correspondence. Finally, we observed two instances where the Swiss candidate was greeted with “*Grüezi*”, a rather informal greeting in the Swiss German dialect, while the minority candidate was greeted with the polite high-German “*Guten Tag*”.

Table 3 about here

Cases with equal outcomes

In several cases both applicants were treated equally regarding the final outcome of their application (i.e. both invited/both rejected), but employers replied more favourably to the majority candidate.

This first example shows differences in rejection emails for the same position:

Dear Mr. Kālin, We thank you for your interest in our company and for sending in your application and introducing yourself. Unfortunately, we have to reject your application today, but we have taken the liberty of making a copy of your application, because we might be interested in you at a later date. (0053, Swiss male)

Dear Mr. Yilmaz, We thank you for your interest in our company and for your application. After thorough examination of your application we are sorry to inform you that we cannot consider your application for the vacant position. Please do not feel discouraged, your application has made a very good impression on us, and we therefore believe that you will surely succeed in finding your desired job. [...] (0053, Turkish male)

Both candidates were rejected, but only the Swiss candidate was told that his application would be kept for a possible later opportunity, while the Turkish candidate received a polite, standard rejection message.

A second example also shows the same outcome (rejection) for both applicants, but there were considerable differences in the timing of the replies:

Dear Mr Yilmaz, You have entered your application into the race for the position of [...]. We thank you for your interest in our company and the work you put into your application. As promised we are contacting you

after having carefully reviewed all the applications we received and having made a first selection. We would have preferred to send you a positive message today, but unfortunately the competition was too strong this time. Other applicants were more convincing to us. [...] (0071, Turkish male)

Dear Mr Kälın, We are referring to your application of [...] and would like to give you a short update. [...] Thanks to you and the other applicants, we are in the fortunate position of holding many excellent applications in our hands after a first selection round. Your application has left a good impression and is still on our reserve list. However, in the first round we preferred other candidates whose applications convinced us even more. [...] and we would like to wait for the results of the first round of interviews and keep your application pending and to thoroughly review it again following the first interviews. This will be early January.

Do you agree? And will you give us a second chance? (0071, Swiss male)

Both candidates were contacted on the same day, the Turkish candidate was immediately told that his application did not make it into the next selection round, while the Swiss candidate was kept in the running. Although the Swiss candidate was rejected a month later, he was encouraged to keep up with their social media channels and to apply again as they are “convinced that you always meet twice in your life – at least”.

The third example of equal outcomes (invitation) shows different levels of enthusiasm about the respective candidates. The day after the applications were sent, the candidates received the following reply:

Grüezi Mr Kälin, Your application made me very happy and curious. I would be glad if we could meet for an interview. Can you offer me some dates? (0232, Swiss male)

Grüezi Mr Hoffmann, Thank you for your application, we will examine it as soon as possible and inform you about our decision. (0232, German male)

The Swiss candidate was immediately invited for a job interview and was met with very positive feedback on his application, while the German candidate received a standard message that his application had been received. Although he was invited for an interview two days later, the invitation lacked the same level of interest.

In the final example of an application that is shown as “equal treatment” in the correspondence test results, the employer tried very hard to invite the Swiss candidate for an interview. Both candidates sent their application on the same day, but only the Swiss candidate’s application was acknowledged. After eleven days, the Swiss candidate received an interview invitation by email and phone message. When the invitation was quickly and politely declined by email saying that the applicant had already found a new position, he immediately received a reply wishing him all the best for the new position, and offering this: “if [this new position] does not suit you and you would again look for another opportunity, you can contact us again. We have interesting opportunities...” (0136, Swiss). Three days after this exchange and 18 days after the application, the German candidate was informed that the preselection of candidates was still ongoing and asked for patience. After 23 days the German candidate received an invitation by phone.

Cases with different outcomes

We also observe an obvious preference for the Swiss candidate in cases with different outcomes. While the German applicant in this case received a standardised rejection email, the employer made an effort to meet the Swiss candidate for a job interview:

Dear Ms Kaelin, We are contacting you a little later than promised ...
Apologies! Your application has raised our interest and we would like to meet you. Bern and [...] are unfortunately not near each other. I will be in Bern on [...] to visit the Zwiebel Märkt with my wife. If you are still interested in this job [...], we could meet in the early afternoon somewhere close to the train station. I am looking forward to hearing from you. (0127, Swiss female)

The most striking response observed during the correspondence test also occurred in a pair of applications with different outcomes. While the German candidate was never contacted, the Swiss candidate immediately received a long email, which included the employer's mobile phone number and the explicit instruction to call, even in the evening:

Dear Mr. Kälin

Thank you very much for your application [...]. I have just looked at it and have come across a few interesting points.

I like your educational background (secondary school and vocational degree), and your experience in service work. Of course, and you are probably well aware of this yourself, *I also like your age and your origin.*

Your move to our area is also very exciting and I can highly recommend this from personal experience ;-)...

I would like to interview you in person and take the opportunity to introduce myself in detail. Already in advance, I can already tell you about some very interesting upcoming projects and, moreover, I am convinced that we are an above-average competent and motivated young team with a correspondingly demanding client base.

I am interested in a long-term cooperation and therefore an unlimited contract.

I would appreciate it if you could get back to me with a suggestion for a time for an interview (it can also be in the evening). You can reach me under [...] (0055, Swiss male, emphasis added)

Apart from the last two examples, the messages quoted above have shown that cases considered as equal treatment in a correspondence test, can be cases of subtle discrimination. These may be comparably minor differences, such as differences in names or greetings. Yet, except for one case where the Swiss name was misspelled, these mistakes happened only in replies to the minority candidates. Furthermore, it becomes evident that employers often make more of an effort and are more enthusiastic in their contact with majority candidates, even if the final outcomes recorded in the testing are the same. While these differences are likely to be unconscious, the final case was the most blatant in openly stating his preference for a young Swiss candidate (*"I also like your age and your origin"*⁶).

⁶ Original German quote „Natürlich [gefällt mir] auch, und das wissen Sie wahrscheinlich selbst zu gut, Ihr Alter und die Herkunft.“

6. Discussion

Results from correspondence tests show a compelling way of portraying “clear and concise evidence” (Fix & Struyk, 1993) of discrimination. They are intuitive to understand, even for a lay audience. Yet, while they aim to quantify discrimination in the labour market, they address only one specific point in the hiring process. By reporting only the final outcome of the application procedure, invitation or rejection, correspondence test results fail to provide a more nuanced picture.

While Heckman and Siegelman (1993) have briefly acknowledged that supplemental data such as wait times for an interview could be “enlightening and useful in illuminating potential causes of discriminatory treatment” (p. 193), they do not discuss the issue further. However, it could be argued that the unobserved differences in employers’ responses in correspondence tests, are similar to their critique about unobserved differences for candidates. The findings presented in this paper show that the binary presentation of results (job interview: yes/no) can hide bias and subtle discrimination. The results further suggest that even if candidates experienced equal outcomes in the first stage of the application process, they might still face disadvantage in the second stage. Applicants that experienced subtle discrimination, might (unknowingly) go into a job interview with a disadvantage compared to the other applicant. Discrimination is therefore likely to be higher, if candidates had proceeded to the actual job interview. These so far largely unobserved biases are important as they could predict a higher cumulative discrimination rate when it comes to actual job offers at the end of the application process.

Even though correspondence tests target a very specific moment in the hiring process, the examples presented above show that there are several ways in which employers can resort to discriminatory treatment, both intentionally or unconsciously. While the binary outcome of an application (invited: yes/no) is the main focus of correspondence tests, the previous discussion has shown that subtle discrimination also occurs in response time or in the content of the responses. Going back to the definition by Van Laer and Janssens (2011), we can observe that the instances of subtle discrimination are not easily recognized and can be ambiguous – they become visible only when email responses are compared side by side, they are in many cases unintentional and unconscious, and they are entrenched in every day interaction, such as responding to an email. Yet, while these subtle forms of discrimination might be unintentional and unconscious, over time these small disadvantages can become cumulative. As Blank et al. (2004) have shown, cumulative effects of discriminatory treatment can occur “throughout the stages within a domain, across domains, across individual lifetimes, and even across generations” (Blank et al., 2004, p. 11). They argue that candidates who anticipate discriminatory treatment can be discouraged from even trying to obtain skills or apply for certain positions.

7. Conclusion

As this paper has shown, using data from a recent correspondence test in Switzerland, discrimination also occurs on the Swiss labour market, even if the call back ratio is quite low with 1.48 (sales) and 1.12 (electricians) on the occupation level, low compared to results from other correspondence tests, in as far as these numbers can be compared. As the timing and content of the messages analysed above has shown, even if cases are

reported as having an equal outcome in the correspondence test results, there can still be different treatment in the timing and the wording of responses. These are the cases in the ILO studies that were called “equal but different treatment” or in Hanson et al. (2011) “subtle discrimination”. Incidences of subtle discrimination in the hiring procedure can be considered “to represent a different dimension of discrimination that is more difficult to uncover” (Hanson et al., 2011, p. 283). As Pager et al. (2009) observed, “It was only through side-by-side comparisons of our testers’ experiences that patterns of subtle but consistent differential treatment were revealed” (p. 793). This was also observed in the Swiss results. While the evidence of correspondence tests is indeed “clear and convincing” it represents only the tip of the iceberg. Since most correspondence tests today are conducted by email, researchers have a lot of material at their disposal to analyse these instances of subtle discrimination in responses and it would be interesting to see more discussion of these materials. In this vein, Crabtree (2018) suggests an automated analysis of these instances of subtle discrimination for large-N correspondence tests.

As correspondence tests are often criticised for looking at only a very specific moment in the hiring process, that is, whether an applicant is invited for an interview or not, the material provided in emails or potentially also in mobile phone messages, can enlarge this point a little, by looking not only at the final decision but also at some of the underlying mechanisms. Coupled with work from social psychologists, the findings suggest that discrimination not only occurs in whether or not applicants are invited for an interview, but also in how long they have to wait for a reply or how they are treated by a potential employer. Many of the observed instances of subtle discrimination are probably unintentional and unconscious (e.g. misspelled names or different greetings), yet others

have shown a clear preferential taste for the majority candidate. While it is of course still possible that employers hire applicants that experienced subtle discrimination in the previous stage, expected discrimination could also prevent candidates from applying, for example in the wording of the vacancy. Taken together, the latter incidences of subtle discrimination only add to other observed workplace incivility (Krings et al., 2014) to show that discrimination can accumulate and that even seemingly minor incidences can build up in the long run.

Conflicts of interests

There is no potential conflict of interest.

Table 1:

	Number of jobs [1]	None invited [2]	Both invited [3]	Only majority invited [4]	Only minority invited [5]	Net discrimination rate % [6]	% call back majority [7]	% call back minority [8]	Ratio [9]	Percentage difference [10]	Chi-square test [11]
<i>Sales assistants</i>	136	83	19	24	10	26.42	31.62	21.32	1.48	10.30	0.0164**
Germany	45	24	7	11	3	38.10	40.00	22.22	1.80	17.78	0.0325**
Kosovo	46	28	8	7	3	22.22	32.61	23.91	1.36	8.70	0.2059
Turkey	45	31	4	6	4	14.29	22.22	17.78	1.25	4.44	0.5271
<i>Electrician</i>	136	56	64	12	4	10.00	55.88	50.00	1.12	5.88	0.0455**
Germany	45	16	23	3	3	0.00	57.78	57.78	1.00	0.00	1.0
Kosovo	46	16	23	6	1	16.67	63.04	52.17	1.21	10.87	0.0588*
Turkey	45	24	18	3	0	14.29	46.67	40.00	1.17	6.67	0.0833*
Total	272	139	83	36	14	16.54	43.75	35.66	1.23	8.09	0.0019***

Table 1: Correspondence test results from the Swiss (German) labour market (chi square test: significant at * 10%, ** 5% level).

Figure 1:

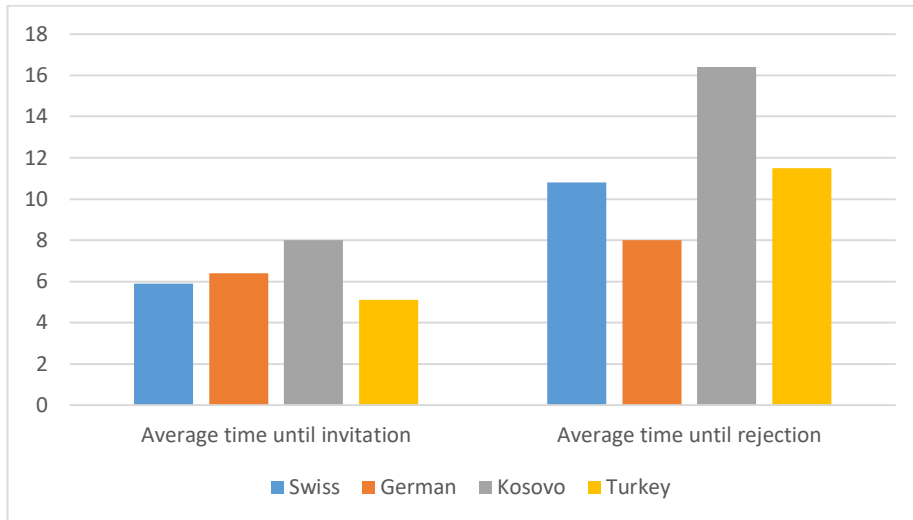


Figure 1: Working days until response

Data on which Figure 1 is based:

Working days until invitation or rejection by ethnicity

	Swiss	German	Kosovo	Turkey
Average time until invitation	5,9	6,4	8,0	5,1
Max time until invitation	61,0	21,0	87,0	15,0
Average time until rejection	10,8	8,0	16,4	11,5
Max time until rejection	56,0	29,0	56,0	43,0

Table 2:

Cases with at least one email response	194
Only majority contacted	26
Only minority contacted	23
Both contacted	145
- Exactly the same text	100
- Almost the same or very similar texts	12
- Different texts	33
o Equal outcome	16
o Different outcome	17

Table 2: Email responses

Table 3:

	Majority	Minority
First and last names used in greeting	0	2
Misspelled name	1	4
No name in greeting	3	5
Wrong gender	1	2
Total	5	13

Table 3: Problems with names

References

- Aigner, D. J., & Cain, G. G. (1977). Statistical theories of discrimination in labor markets. *Industrial and Labor Relations Review*, 175-187.
- Arriijn, P., Feld, S., & Nayer, A. (1998). *Discrimination in access to employment on grounds of foreign origin: the case of Belgium*. Geneva: International Labour Office, Conditions of Work Branch.
- Arrow, K. J. (1973). The Theory of Discrimination. In O. Ashonfelter & A. Rees (Eds.), *Discrimination in Labor Markets* (pp. 15-42). Princeton: Princeton University Press.
- Attström, K. (2007). *Discrimination against native Swedes of immigrant origin in access to employment*. Geneva: International Labor Organization.
- Becker, G. (1957). *The Economics of Discrimination* (2nd ed.). Chicago: University of Chicago Press.
- Bendick, M. (1996). *Discrimination Against Racial-ethnic Minorities in Access to Employment in the United States: Empirical Findings from Situation Testing*: Employment Department, International Labour Office.
- Bendick, M., Jackson, C. W., Reinoso, V. A., & Hodges, L. E. (1991). Discrimination against Latino job applicants: A controlled experiment. *Human Resource Management*, 30(4), 469-484.
- Bendick, M., Rodriguez, R. E., & Jayaraman, S. (2010). Employment discrimination in upscale restaurants: Evidence from matched pair testing. *The Social Science Journal*, 47(4), 802-818. doi:<http://dx.doi.org/10.1016/j.soscij.2010.04.001>
- Bertrand, M., & Duflo, E. (2017). Field Experiments on Discrimination *Handbook of Economic Field Experiments* (Vol. 1, pp. 309-393): Elsevier.
- Blank, R. M., Dabady, M., & Citro, C. F. (2004). *Measuring racial discrimination*. Washington, DC: National Academies Press.
- Bovenkerk, F., Gras, M., Ramsoedh, D., Dankoor, M., & Havelaar, A. (1995). *Discrimination against migrant workers and ethnic minorities in access to employment in the Netherlands*. Retrieved from Geneva:
- Cediey, E., & Foroni, F. (2008). *Discrimination in access to employment on grounds of foreign origin in France: A national survey of discrimination based on the testing methodology of the International Labour Office* (Vol. 85E). Geneva: International Labor Organisation (ILO).
- Crabtree, C. (2018). An introduction to conducting email audit studies *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 103-117): Springer.
- Daniel, W. W. (1968). *Racial discrimination in England: based on the PEP report* (Vol. 1084). Middlesex: Penguin.
- Darity, W. A., & Mason, P. L. (1998). Evidence on discrimination in employment: Codes of color, codes of gender. *Journal of Economic Perspectives*, 12(2), 63-90.
- de Prada, M. A., Actis, W., Pereda, C., & Molina, R. P. (1995). *Labour market discrimination against migrant workers in Spain* (Vol. 9). Geneva: International Labour Office, Employment Department.

- Drydakis, N., & Vlassis, M. (2010). Ethnic Discrimination in the Greek Labour market: Occupational Access, Insurance Coverage and Wage Offers. *The Manchester School*, 78(3), 201-218. doi:10.1111/j.1467-9957.2009.02132.x
- Fibbi, R., Kaya, B., & Piguët, E. (2003). *Le passeport ou le diplôme ? Etude des discriminations à l'embauche des jeunes issus de la migration*. Neuchâtel: Forum suisse pour l'étude des migrations et de la population.
- Fix, M., & Struyk, R. J. (Eds.). (1993). *Clear and Convincing Evidence: Measurement of Discrimination in America*. Washington, D.C.: The Urban Institute Press.
- Gaddis, S. M. (2014). Discrimination in the Credential Society: An Audit Study of Race and College Selectivity in the Labor Market. *Social Forces*, 1-29.
- Ghumman, S., & Ryan, A. M. (2013). Not welcome here: Discrimination towards women who wear the Muslim headscarf. *Human Relations*, 66(5), 671-698.
- Goldberg, A., Mourinho, D., & Kulke, U. (1995). *Labour market discrimination against foreign workers in Germany*: International Labour Office, Employment Department.
- Hanson, A., Hawley, Z., & Taylor, A. (2011). Subtle discrimination in the rental housing market: Evidence from e-mail correspondence with landlords. *Journal of Housing Economics*, 20(4), 276-284.
- Heckman, J. J., & Siegelman, P. (1993). The Urban Institute Audit Studies: Their Methods and Findings. In M. Fix & R. Struyk (Eds.), *Clear and Convincing Evidence: Measurements of Discrimination in America* (pp. 187-258). Washington, DC: The Urban institute Press.
- Helbling, M. (2011). WHY SWISS-GERMANS DISLIKE GERMANS. *European Societies*, 13(1), 5-27. doi:10.1080/14616696.2010.533784
- Kaas, L., & Manger, C. (2012). Ethnic Discrimination in Germany's Labour Market: A Field Experiment. *German Economic Review*, 13(1), 1-20. doi:10.1111/j.1468-0475.2011.00538.x
- Krings, F., Johnston, C., Binggeli, S., & Maggiori, C. (2014). Selective incivility: Immigrant groups experience subtle workplace discrimination at different rates. *Cultural Diversity and Ethnic Minority Psychology*, 20(4), 491-498. doi:10.1037/a0035436
- Lodder, L., McFarland, S., & White, D. (2003). Racial Preference and Suburban Employment Opportunities. *Chicago: Legal Assistance Foundation of Chicago*.
- Longchamp, C., Imfeld, M., Tschöpe, S., Müller, M., Rochat, P., & Schwab, J. (2014). *Verbreitung und Entwicklung von Rassismus, Fremdenfeindlichkeit, Muslimfeindlichkeit und Judenfeindlichkeit - Schlussbericht zur Studie "Zusammenleben in der Schweiz 2010-2014"*. Retrieved from Bern: <http://www.gfsbern.ch/de-ch/Detail/zusammenleben-in-der-schweiz-2010-2014>
- Neumark, D. (2018). Experimental research on labor market discrimination. *Journal of Economic Literature*, 56(3), 799-866.
- Pager, D., Bonikowski, B., & Western, B. (2009). Discrimination in a Low-Wage Labor Market: A Field Experiment. *American Sociological Review*, 74(5), 777-799. doi:10.2307/27736094

- Pager, D., & Shepherd, H. (2008). The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets. *Annual Review of Sociology*, 34(1), 181-209.
- Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review*, 62(4), 659-661.
- Quillian, L., Pager, D., Hexel, O., & Midtbøen, A. H. (2017). Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. *Proceedings of the National Academy of Sciences*, 114(41), 10870-10875. doi:10.1073/pnas.1706255114
- Rapp, C. (2015). More diversity, less tolerance? The effect of type of cultural diversity on the erosion of tolerance in Swiss municipalities. *Ethnic and Racial Studies*, 38(10), 1779-1797. doi:10.1080/01419870.2015.1015582
- Raymann, U. (2003). *Meinungen und Einstellungen gegenüber AusländerInnen in der Schweiz*. Retrieved from Zürich:
- Reskin, B. F. (2003). Including Mechanisms in Our Models of Ascriptive Inequality: 2002 Presidential Address. *American Sociological Review*, 68(1), 1-21. doi:10.2307/3088900
- Riach, P. A., & Rich, J. (2002). Field Experiments of Discrimination in the Market Place. *The Economic Journal*, 112(483), F480-F518. doi:10.2307/798458
- 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).
- Rowe, M. P. (1990). Barriers to equality: The power of subtle discrimination to maintain unequal opportunity. *Employee Responsibilities and Rights Journal*, 3(2), 153-163.
- Ruedin, D., D'Amato, G., Wichmann, N., & Pecoraro, M. (2013). *Migrationsängste der Schweizer Bevölkerung*. Retrieved from Neuchâtel:
- Van Laer, K., & Janssens, M. (2011). Ethnic minority professionals' experiences with subtle discrimination in the workplace. *Human Relations*, 64(9), 1203-1227.
- Weichselbaumer, D. (2016). Discrimination Against Migrant Job Applicants in Austria: An Experimental Study. *German Economic Review*, 18(2), 237-265. doi:10.1111/geer.12104
- Zschirnt, E. (2016). Measuring Hiring Discrimination – A History of Field Experiments in Discrimination Research. *NCCR Working Papers*, 7, 1-32.
- Zschirnt, E., & Ruedin, D. (2016). Ethnic Discrimination in Hiring Decisions: A Meta-Analysis of Correspondence Tests 1990–2015. *Journal of Ethnic and Migration Studies*, 42(7), 1115-1134. doi:10.1080/1369183X.2015.1133279