

nccr —————>
on the move



Eva Zschirnt

Revisiting Ethics in Correspondence Testing



Working Paper #8
May, 2016

Eva Zschirnt (University of Neuchâtel)

Revisiting Ethics in Correspondence Testing

The nccr – on the move is the National Center of Competence in Research (NCCR) for migration and mobility studies, which has been operational since 2014. The overall scientific aim of the nccr – on the move is to offer a comprehensive and interdisciplinary framework for understanding the character of contemporary migration and mobility to Switzerland. It was established by the Swiss National Science Foundation with the objective of fostering an innovative and competitive field of research on migration and mobility in Switzerland. Managed from the University of Neuchâtel, the network consists of some 60 researchers from social sciences, economics and law, who collaborate in 19 projects that are based at the universities of Neuchâtel, Basel, Bern, Fribourg, Geneva, Lausanne, Lucerne, and Zurich.

The Working Papers Series is an online platform for academic debates by members and cooperating partners of the nccr – on the move. The authors are responsible for the analyses and arguments, which do not necessarily reflect those of the nccr – on the move.

nccr – on the move, University of Neuchâtel, Faubourg de l'Hôpital 106, 2000 Neuchâtel, Switzerland
Contact for the Working Paper Series: info@nccr-onthemove.ch

Abstract

Questions of research ethics always arise when planning a correspondence test to study discrimination in the market place. However, the issue is addressed relatively little in published correspondence tests with authors usually referring to the two seminal articles written in this field (i.e. Banton (1997) and Riach and Rich (2004)). Since then correspondence testing has become more widespread and the technique is increasingly relying on the internet to find and send applications. It is therefore necessary to revisit the question of ethics in correspondence testing. This paper addresses the ethical issues that researchers are facing in correspondence tests that study discrimination in hiring decisions in the labour market in particular. It provides a short overview on the development of research ethic guidelines. The main part of the paper focuses on the ethical issues that arise in correspondence testing, looking at questions of covert research, potential problems (regarding voluntary participation, informed consent, deception, entrapment of employers, employer's rights), possible solutions and technical challenges. Looking at specific country examples, decisions by ethical commissions and national legal frameworks are considered. These show that testing has to be renegotiated depending on the national context, and, in the case of Germany, legal implications of correspondence testing are discussed. The paper concludes that correspondence testing, if planned carefully and executed responsibly, does not violate research ethics in social sciences.

Keywords

Research Ethics, Correspondence Testing, Discrimination Research, Field Experiments

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 and Carolin Fischer for comments.

Contacts

Eva.zschirnt@unine.ch

Swiss Forum for Migration and Population Studies SFM

University of Neuchâtel, Faubourg de l'Hôpital 106, 2000 Neuchâtel

Contents

1.	Introduction	5
2.	The development of research ethics – from biomedical to social sciences	6
3.	Ethical issues in correspondence testing	7
3.1	Correspondence testing – an introduction to the technique	8
3.2	Justifications for the infringement of ethical principles in correspondence testing	9
3.2.1	Voluntary participation and informed consent	10
3.2.2	Deception	12
3.2.3	Entrapment of Employers	13
3.2.4	Employers' rights – avoiding harm and maintaining the confidentiality of records	14
3.3	Constructing contact details for fictitious applicants – legal challenges and open questions	15
4.	Managing ethical issues in correspondence testing	17
4.1.	The involvement of Research Ethics Committees	17
4.2	Obstacles in correspondence testing – country examples	18
4.2.1	Sweden and Norway	19
4.2.2	Germany	20
5.	Conclusion	22
	Bibliography	23

1. Introduction

The 20th century has seen a growing awareness of the question of ethics in research and researchers have increasingly become subject to adhering to codes of ethics. However, the development of these codes has been mostly based on biomedical research methods that social scientists were then subjected to as well. This dominance of biomedical sciences has affected social sciences worldwide. Today, most of the ethical regulations concerning social scientists have been adopted by professional associations, such as e.g. the British Sociological Association or the American Sociological Association, that exist next to individual institutional or national ethical regulations. However, in contrast to biomedical or social sciences, economists have so far not developed comparable research ethics codes.

Lying at the intersection between social sciences and economics, field experiments on discrimination in the market place always raise ethical questions. As Riach and Rich (2004) put it in one of the most frequently quoted articles on ethics in correspondence testing, “what is required [...] is a consideration of the ethical issues involved in the application of deceitful research procedures, originally developed in sociology and psychology, to the study of an economic institution” (p.459). Correspondence tests, that rely on two substantially equally qualified fictitious candidates to send applications to one job vacancy, are inherently deceitful research procedures, because potential employers do not know that the applications are fictitious and that they are part of a field experiment on discrimination in hiring decisions. Nevertheless, the issue of ethics in correspondence testing is addressed relatively little in published correspondence tests. The trend that researchers discuss their research methodology and procedures in detail but rarely address ethical dilemmas has also been observed (Burnham, Lutz, Grant, & Layton-Henry, 2008). In the particular case of correspondence testing, most articles just refer to the two seminal works on ethics in field experiment written by Banton (1997) and Riach and Rich (2004). However, since the publication of Riach and Rich’s article more than ten years have passed, correspondence testing has become more widespread and diverse (Zschirnt & Ruedin, 2016), and the technique increasingly uses the internet to find vacancies and send applications. It is therefore worth to revisit the question of ethics posed by Riach and Rich “Deceptive field experiments of discrimination: are they ethical?” to account for recent developments.

In order to do so, the first part of the paper looks into the historical background that led to the establishment of research ethics and its evolution from biomedical sciences to social sciences. Section two then turns to the practical application of research ethics in the case of field experiments by giving a short introduction to the technique of correspondence testing and discusses the issue of covert research, before dealing with the ethical principles that are being challenged in correspondence testing. Finally, possible arrangements to respond to ethical dilemmas are discussed. The last part of the paper then portrays some country examples as case studies and shows how researchers addressed ethical and legal challenges in correspondence testing in the different national contexts.

2. The development of research ethics – from biomedical to social sciences

As Riach and Rich have pointed out, “psychologists and sociologists have a lengthy history of deceptive research activity in laboratory and social settings, and have consequently developed strict codes of research ethics” (2004, p. 459). Certain authors consider that the development of research ethic guidelines and ethics codes only really started with the 1947 Nuremberg Code following the crimes committed in the name of research by the Nazi regime (Dingwall, 2012). Yet other authors emphasise that the issue of research ethics had already been established on the agenda at the beginning of the 20th century (Dench, Iphofen, & Huws, 2004, p. 3). Prussia already introduced guidelines on medical experiments in 1900 which were reinforced in 1931 and several research ethics scandals took place well before the medical experiments of the Nazis, one of the earliest dating back to England in 1796 (Dingwall, 2012; Hunter, 2010). What all of these examples have in common is their predominant focus on medical or psychological research.

It is therefore not surprising that the field of research ethics has been strongly influenced by biomedical research. Israel (2015) provides a detailed overview over the development of research codes and guidelines in the field of bioethics. They range from the Nuremberg Code in 1947, the Declaration of Helsinki adopted by the World Medical Association in 1964, the 1979 Belmont Report by the US National Commission for the Protection of Human Subject of Biomedical and Behavioral Research, the International Ethical Guidelines for Biomedical Research Involving Human Subjects by the Council of International Organizations of Medical Sciences in 2002, to the Universal Declaration on Bioethics and Human Rights adopted by the UNESCO in 2005. Israel discusses these biomedical research ethics codes in detail, because

“These statements on biomedical research provide key, albeit contested, foundations for much current thinking and practices in research ethics and have an impact on the social sciences – either intellectually or through the dominance of the biomedical research model in shaping institutional ethical practice.” (Israel, 2015, p. 28)

He further emphasises that the biomedical ethic codes and guidelines were never “conceived with the methodologies and issues that concern social scientists in mind. Social scientists were rarely involved in their drafting, they were not consulted and they did not consent” (Israel, 2015, p. 41).¹

In a similar fashion to the development of biomedical research codes, “increased awareness of the importance of ethics in social science research has been driven in part by the outrage caused by particular studies” (Wassenaar, 2006, p. 62). The most known studies in the field of social sciences are probably the Milgram experiment in 1961-62, the Stanford Prison Experiment in 1971, or Humphrey’s Tearoom Trade in the mid-1960s, that clearly showed the ethical questions social scientist were dealing with in their research (Nakray, 2016, p. 14). Although, as Wassenaar pointed out while

¹ A critical approach to the development of research ethics in social sciences can be found in Dingwall (2012), aptly titled “How did we ever get into this mess? The rise of ethical regulation in the social sciences”.

“it could be argued that none of the above examples equal the nature and extent of the atrocities committed by the Nazi doctors, or the tragedy of the Tuskegee syphilis study, all of these studies involve ethical violations of one type or another” (2006, p. 62).

Furthermore, the involvement of researchers in “covert military and intelligence operation in South-East Asia in the 1960s” (Israel, 2015, p. 42) also led to the adoption of research ethics codes.

As Dench et al. (2004) point out, the development of research ethics covering numerous disciplines has been led by Anglo-Saxon and Scandinavian countries. Like Israel (2015), they refer to early codes of ethics developed by disciplinary associations, such as the Society for Applied Anthropology in 1948, the American Psychological Association in 1963, the British Sociological Association in 1968, the American Sociological Association in 1970, the American Anthropological Association in 1971, or the American Psychological Association in 1973. In particular in Anglo-Saxon and Scandinavian countries, obtaining ethical approval by formal research ethics board has become the norm and is often a prerequisite to obtain research funding. Countries on the European continent have so far been spared from this more institutionalised development of research ethics, as especially French, German and Italian researchers “retain and cherish traditions of professional autonomy, offering an alternative to the Anglo-Saxon movement towards a system of command and control” (Dingwall, 2012, p.4). Yet, the growing awareness of questions of research ethics and the increasingly international dimension of research seem to encourage a move to more institutionalised national or supranational approaches. One of these supranational approaches can be observed in the European Research Area, where the European Commission launched the RESPECT Project².

Since the methodology of correspondence testing tries to create laboratory like conditions to conduct an experiment of the labour market, developments in the fields of research ethics in the social sciences also influence researchers planning correspondence tests. Quite often research funding is now dependent on prior ethical approval. Since correspondence tests are usually first met with scepticism regarding their compliance with research ethics standards, a thorough preparation by the researchers is necessary, to meet the ethical questions that can arise.

3. Ethical issues in correspondence testing

Field experiments on discrimination in hiring that take place in the market place, i.e. both the in-person audits as well as the written correspondence tests, pose, by the nature of their research methodology, a problem from a research ethics point of view, especially concerning the responsibilities of the researcher towards the research participants. Most research ethics guidelines and books focus on these responsibilities, namely the voluntary participation of research subjects, their participation based on informed consent, the confidential and anonymous use of data and the protection of research participants from harm or distress (e.g. Dench et al., 2004; Israel, 2015). Yet, so far, field experiments are the most reliable and most direct methodology to measure discrimination in entering the labour market, even though ethically the methodology is violating

² RESPECT Project: Professional and Ethical Codes for Technology-related Socio-Economic Research. One aspect of the RESPECT Project was the development of “An EU Code of Ethics”, by Sally Dench, Ron Iphofen and Ursula Huws. Yet this code of ethics is still voluntary.

several core principles of social science research ethics, in particular the principles of informed consent and voluntary participation. However, there are only two articles, by Banton (1997) and Riach and Rich (2004), that explicitly deal with the question of research ethics in this deceptive technique and “at times this omission has been detrimental to the technique’s application” (Riach & Rich, 2004, p. 457).

3.1 Correspondence testing – an introduction to the technique

Since the 1960s field experiments have been used to study the phenomenon of discrimination in the labour market, making use of both in-person audit methods as well as written correspondence testing. 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 a real-life vacancy and present the potential employer with two substantially equal and thus interchangeable candidates, that differ only in the characteristic to be studied, e.g. membership of a particular ethnic minority group. As Midtbøen and Rogstad pointed out, this “gives the researcher complete control over the experiment” (2012, p. 206) since the researcher constructs the application material and, furthermore, randomization of the applications “provides a strong opportunity to draw causal inferences” (p. 207). Moreover, the fact that correspondence testing enables researchers to test for discrimination in a greater variety of jobs with varying skill and qualification levels, not only those where in-person walk-in applications are common, allows for a better representativeness of the results.

Correspondence tests are very carefully planned experiments and most published correspondence tests contain a very detailed section on the research design. Looking at the testing conducted by Wood et al. (2009) in the UK as one example, they describe in detail:

- the number of application used for each application set,
- the signals chosen to convey the racial identity of the applicants,
- the locations chosen in which the testing was conducted,
- the choice of occupations included in the experiment,
- how job adverts for the study were identified (i.e. newspapers or job-search websites),
- the development of applications focusing on the competitive quality of the applications, their plausibility and comparability,
- how employers could contact potential applicants by creating physical addresses, email addresses and telephone numbers for all applicants,
- the time of testing and the sample size achieved, distributed by occupations, locations, market sectors, size of organisation, gender, source of the advertisement, the type of application (CV, an application form, or a web-based form), and the mode in which the application was submitted.

These careful considerations before the actual testing takes place show that correspondence test require a lot of time and detailed preparation before they can be conducted. Yet, once this labour intensive preparatory work has been done, Bendick and Nunes (2012, p. 238) point out, that in giving the researcher complete control, correspondence tests are an “innovative research technique

[...] that offers laboratory-like controlled conditions in quasi-experiments in real-world hiring situations”.

Closely linked to the thorough preparatory work, many authors refer to the ethical aspects that correspondence studies are challenged with, but only very few look at the question of research ethics in correspondence testing in more detail. Most studies only refer to Bovenkerk (1992), who briefly considers ethical questions in his manual on testing discrimination in natural experiments, the two articles focusing on ethical issues in correspondence testing written by Banton (1997) and Riach and Rich (2004), as well as a short section on ethics by Pager (2007). Does this mean, that a general 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? Could it even be the case, that some researchers omit the discussion of ethical issues altogether? In any case, there is a need to look at the question of ethics in correspondence testing again, since the last article focusing specifically on ethics in correspondence testing (Riach & Rich, 2004) is already more than ten years old and does not include information on the use of modern technology, i.e. emails and online applications or the use of mobile phones, or developments observed in the recent waves of correspondence testing³.

3.2 Justifications for the infringement of ethical principles in correspondence testing

By the nature of their research design, in which an employer is not aware that an experiment is being conducted, correspondence tests are an example of covert research, which goes against the principles of voluntary and informed consent that are promoted as key principles in research ethics guidelines across disciplines. This becomes apparent in the definition provided in the *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)

Similarly the British Sociological Association argues that while “there are serious ethical and legal issues in the use of covert research, [...] 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 that it may violate the privacy of its research subjects, making it a method that should only be used as a last resort if it is not possible to obtain the data needed using other research methods. The *EU Code of Ethics for Socio-Economic Research* also focuses on the question if the deception used in covert research is acceptable if the researcher has no other possibility to obtain information:

³ Of the 43 correspondence tests conducted between 1990 and 2015 included in the meta-analysis by Zschirnt and Ruedin (2016), 32 studies were carried out after 2004.

“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, which distorts the research findings. They also point out the example of “mystery shopping”⁴ as an area where covert research is frequently used to evaluate the quality of a service, an experiment that is – in its setup – similar to the idea of field experiments in hiring decisions (2004, p. 33). In a second example they mention a study conducted by Calvey (2000) using a covert participant observation on nightclub bouncers and who reasoned that “gaining access and the analytic richness of the data collected [...] would have otherwise been seriously diminished” (in Dench et al., 2004, p. 34). Dench et al. even refer to field experiments in the labour market explicitly saying that

“For example, 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, apparently justified in situations in which information of a similar quality and richness cannot be achieved using other research methodologies. It is therefore worth dealing with the problems most often cited by opponents of covert research that are infringing ethical principles of social sciences research, and to provide a reasoning why infringing these ethical principles can be justified. Throughout this section the correspondence testing methodology serves as a background for the discussion.

3.2.1 Voluntary participation and informed consent

One of the basic principles underlying research across disciplines, is that “potential research subjects should be given the opportunity to refuse to participate in the research” and it is argued that “Voluntary participation in research is basic human right” (Dench et al., 2004, p. 56). This strong emphasis on voluntary participation can be traced back to the aforementioned medical experiments conducted by the Nazi regime. While the Nuremberg Code of 1947 mentioned ten principles which researchers should abide to, “the Code gave considerable emphasis to the voluntary and informed consent of people competent to make decisions. Indeed, it was underpinned by the concept of voluntary consent” (Israel, 2015, p. 27). The first paragraph of the Nuremberg Code states that “The voluntary consent of the human subject is absolutely essential.” (Nuernberg Military Tribunals, 1949, p. 181) However, it lies in the nature of covert research that “respondents don’t know it is happening and hence are unable to decide whether to participate or not” (Dench et al., 2004, p. 61).

Closely related to, often overlapping with the principle of voluntary participation and also based in the Nuremberg Code, the principle of informed consent is seen as one of the most fundamental

⁴ Mystery shopping is one example of covert research that is “routinely used in market research as a way of evaluating the quality of service delivery and, in some cases, the honesty of employees” (Dench et al., 2004, p. 33).

principles in the literature on research ethics and ethical guidelines frequently point out the importance of obtaining the informed consent of research participants (e.g. Dench et al., 2004, p. 64). As the National Committee for Research Ethics in Norway (NESH) stated, “the consent requirement is intended to prevent invasions of personal integrity” (2006, p. 13). This, however, is not possible in covert research, which research ethics institutions and codes have also acknowledged. NESH for example recognizes that there are exceptions in which research can be conducted without the consent of the participants:

“In certain cases, participant’s freedom and self-determination can be respected even though consent has not been obtained beforehand. Although informed consent is the general rule also in projects in which the participants do not participate actively, exceptions from the requirement regarding informed consent 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.” (2006, p. 14)

Looking at the example of correspondence testing, Bovenkerk already recognized that “While situation testing may be a superior method to discover how people in a position of taking decisions really behave, failure to inform those who are being studied is inconsistent with an important norm in research: the right of the participants to provide informed consent” (1992, p. 33). He lists three reasons that in his opinion justify breaking this principle: first, hiring decisions are not a private matter and discrimination in this field 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 only normal hiring decisions that are observed and researchers “do not lure employers in to a situation in which they are enticed to deviate from their normal course of action” (Bovenkerk, 1992, pp. 33-34).

This problem has also been addressed by other researchers conducting correspondence tests and it has been argued that breaking the principle of informed consent “is a crucial feature of this type of research, as informing participants would invalidate the experiment” (Blommaert, Coenders, & van Tubergen, 2014, p. 964). In her correspondence test in Sweden, Bursell also discusses this issue and refers to Swedish law, according to which “research without the participant’s informed consent can still be performed if the research i) is of high societal importance *and* ii) if there is no other way of getting at the information *and* iii) if the research is of high quality” (Bursell, 2007, p. 9). Looking at the US, Pager also refers to the legal provisions, according to which

“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 she explicitly refers to audit studies of discrimination, these conditions are also met in correspondence tests of discrimination. Thus, there seems to be a consensus that covert research is permissible under certain (strict) provisions.

The British Sociological Association's Statement of Ethical Practices proposes that in covert research, which has been conducted without the prior informed consent of the participants, researchers should still obtain this consent post-hoc (British Sociological Association, 2002, p. 4). However, in the case of correspondence testing this is debatable. As Pager (2007) pointed out "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). While Midtbøen (2014b) decided to contact his research participants who had replied in one way or another to the fictitious candidates to conduct interviews with them, the German research team of the Expert Council of German Foundations on Integration and Migration (SVR) decided not to inform their research subjects. They argued that informing the participants after the testing had taken place would not improve the chances of minority applicants in the future and that it might pose a problem for further research if the technique of correspondence testing became too well-known. Furthermore, following Pager's argument, they claimed that obtaining post-hoc consent could potentially prove problematic for the employees responsible for the hiring decisions and might cause conflicts within the enterprise. Thus, in order to limit the potential damages incurred by individuals, it can be argued against seeking post-hoc consent. Furthermore, correspondence testing conducted for research does not look at the individual behaviour of research participants, but focuses on aggregated behaviour in hiring decisions. This focus on using anonymised aggregated data guarantees the protection of the participants' information.

3.2.2 Deception

Another ethically questionable feature of correspondence testing is the reliance on the deception of the research subjects, since fictitious applicants pretend to be real candidates for a vacancy. The deceptive nature of field experiments has been the focus of Riach and Rich's (2004) article where it featured in the title "Deceptive field experiment" on the ethics of field experiments in discrimination research.

Riach and Rich 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 use the example of the American legal system where US Courts endorsed the use of testing in housing discrimination cases. The courts endorsed the use of testing in these cases, and argued that deception is seen as regrettable but unavoidable:

"It is surely regrettable that testers must mislead commercial landlords and home-owners as to their real intentions to rent or buy housing. Nonetheless, 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 both benefits unbiased landlords by quickly dispelling false claims of discrimination and is a major resource in society's continuing struggle to

eliminate the subtle but deadly poison of racial discrimination.” (Boggs, Sellers and Bendick, 1993, p. 366-367, in Riach and Rich, 2004, p.458)

In a similar way 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, in Riach & Rich, 2004, p. 460). The deceptive nature of field experiments is thus seen as necessary to obtain information about the socially harmful practice of discrimination.

Riach and Rich then turn towards the deception of the research subjects, i.e. in correspondence testing the employer. 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” (2004, p. 461) since “the labour and the real estate markets are notorious for their deceptive and discriminatory activity” (p.462). They conclude that deceiving the research subjects is justified, 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, 2004, p. 463).

Thus, the fact that testing is accepted as proof of discrimination by courts in a number of countries and the fact that markets are regarded as inherently dishonest shows that deception is accepted as an unavoidable feature of the technique to combat the social problems continued discrimination might cause.

3.2.3 Entrapment of Employers

Bovenkerk also addressed the question if researchers could be held liable for the entrapment of their research subjects if the experiment encourages them to behave in a manner that could break the law. According to him this “concern is ill-conceived as discriminating employers break the legal rules probably more than the researcher does” (1992, p. 34). Furthermore, and more important to him, researchers just observe normal hiring practices, they do not trap the employer into any action he would not have taken under different circumstances.

Yet, the fact that an employer might be pushed towards behaving illegally was one of the reasons that two Swedish research proposals were subjected to a thorough ethical assessment. Rolf Nygren, a law professor who prepared the report for the Swedish Council for Social Research’s committee on research ethics, concluded that the proposals should be rejected. He argued that “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).

Both Pager (2007) and Banton (1997) refer to decisions of US courts, including the US Supreme Court, that have confirmed the legal standing of testers and, thus “broadening 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). Banton also argued that the strongest argument for correspondence testing is the “resemblance to an accepted method for gathering evidence for the enforcement of anti-discrimination law” (1997, p. 416). Furthermore, referring to undercover operations used by the police, e.g. to investigate drug-dealing or prostitution, he argues that it might be in the public interest if “officials encourage the commission of offences in order to reduce the incidence of offending” (1997, p. 418).

3.2.4 Employers’ rights – avoiding harm and maintaining the confidentiality of records

A second risk to employers discussed by Pager (2007) is the loss of time employers incur by assessing fictitious applications. 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 (interviewing candidates not interested in actual employment)” (Gingrich, 1998). Researchers acknowledge that this assessment of additional fictitious applications poses a burden on the employers’ time (e.g. Pager & Western, 2012) and most research designs in correspondence testing limit the burden that is being placed on one employer. Usually a company is only considered once, even if more vacancies fitting the requirements of the researchers are published during the testing time, and the number of fictitious applications is limited, usually to two or three applications per company. 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 et al., 2009).

Furthermore, it is worth addressing concerns about breaching the privacy of employers. This is the first concern addressed by Bovenkerk who claims “that in this case 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. Riach and Rich cite a Urban Institute publication supporting Bovenkerk’s argument “that privacy is not a legitimate expectation where public and commercial acts, in the form of advertising vacancies, are involved, and where there is public regulation proscribing discriminatory activities” (Fix, Galster and Struyk (1993) in Riach & Rich, 2004, p. 459).

Finally, concerns regarding the reputation of enterprises and possible negative effects of being part of a correspondence test are also met. 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 and that it is analysed aggregately to avoid inferences about individual employers. Furthermore, researchers are not interested in accusing individual employers of discriminatory behaviour, but in reporting trends in discrimination patterns in a society.

While the concerns about the ethical questions in correspondence testing that were addressed in this section, i.e. the voluntary participation and informed consent, the issue of deception, the possible

entrapment of employers, or causing a potential harm to employers, are valid, a thorough preparation of the correspondence test can alleviate these concerns. Using correspondence tests is justified, because it is not possible to observe authentic hiring decisions when participants are informed, labour markets are faced with an inherent dishonesty, the social damage caused by discriminatory treatment is substantial, and potential damages to employers are almost non-existent. Furthermore, employers are not coerced into acting differently from their normal decision making procedures and potential harm is minimised by only looking at aggregated data and discrimination patterns in the society instead of seeking to prosecute discriminating employers.

3.3 Constructing contact details for fictitious applicants – legal challenges and open questions

Next to the ethical questions, planning a correspondence test also involves numerous technical challenges, which in some situations, can also have ethical or legal implications, in particular the use of email addresses for the fictitious candidates, setting up phone numbers or the use of online application forms. These issues have, so far, not been addressed in detail in articles on the methodology of correspondence testing. A further challenge in the construction of the application materials is the use of photographs which are commonly attached in applications in German speaking countries. While this is not an ethical or legal question, it is still worth mentioning.

One of the most important elements in the application is the section on contact details for the applicant. Almost every recent correspondence test provided a valid email-address, a mobile phone number and usually a street address for the applicant. Each of these ways to contact an applicant provides its own challenge.

Looking at the correspondence tests included in the meta-analysis by Zschirnt and Ruedin (2016), the most frequently used **email** providers are gmail.com, Hotmail.com, and yahoo.com. While Google's Terms of Service are very vague when it comes to who can open a google account and only states "Don't misuse our services" (Google, 2014), Yahoo clearly spells out "Your Registration Obligations" 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 and are not a person barred from receiving the Yahoo Services under the laws of the United States or other applicable jurisdiction. You also agree to (a) provide true, accurate, current and complete information about yourself [...] and (b) maintain and promptly update the Registration Data to keep it true, accurate, current and complete. If you provide any information that is untrue, inaccurate, not current or incomplete, or Yahoo has reasonable grounds to suspect that such information is untrue, inaccurate, not current or incomplete, Yahoo has the right to suspend or terminate your account and refuse any and all current or future use of the Yahoo Services (or any oration thereof)." (Yahoo, 2012)

Similarly to Yahoo, Microsoft – who offers Hotmail or Live email addresses, stipulates in its terms of services that "You agree not to use any false, inaccurate or misleading information when signing

up for your Microsoft account or Skype account” (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, researchers have to provide “false” information to create an email account for a fictitious applicant. Thus it is impossible to “provide true, accurate, current and complete information” as requested by Yahoo. While Google’s Terms of Services seem to be the vaguest and do not specifically define who is allowed to open a google account and which conditions have to be fulfilled, another potential problem arises. 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” as mentioned above.

So far, all published correspondence tests I am aware of have used free and frequently used email providers such as Gmail, Yahoo or Hotmail and, to my knowledge, there have never been any legal repercussions. During a conference on labour market discrimination in 2015, David Neumark reported that for a research project conducted in California, he and his team were able to set up their own email provider⁵. While this would avoid violating the terms of services of the major free email providers, the question of online impersonation still remains.

Similarly, **online application forms** can pose a problem for researchers planning a correspondence tests. Frequently these forms require user to tick a box in the end confirming that all the information they have provided is correct and true. As with the email addresses discussed above it is, however, not possible to provide employers with this confirmation without openly lying. To my knowledge there are only two correspondence tests that made use of online application forms, which are Wood et al. (2009) and Midtbøen (2014a).

The next part of the contact detail that might become problematic are **street addresses** which, for the sake of completeness, are included in almost all applications. Researchers argue that it is

⁵ “Conference on Discrimination and Labour Market Research”, August 25-26, 2015 in Kalmar (Sweden). Keynote Lecture by David Neumark “Is It Harder for Older Workers to Find Jobs? New and Improved Evidence from a Field Experiment”.

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. 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. Apparently UK employers sometimes carry out background checks, including credit checks, and it was therefore decided to create fictitious addresses for the applicants. Thus, 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, 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.

The final element of the contact details is made up by the **mobile phone** number that is provided by each applicant. Here almost all researchers use the same approach: mobile phone numbers, including voice mail boxes were set up for different fictitious applicants, using either real 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 available phone numbers, depending on the gender and the ethnicity of the fictitious applicant. In all studies the voicemail messages were either standard voice mail messages by the phone provider or recorded without any discernible accent. One of the challenges of using mobile phones 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.

4. Managing ethical issues in correspondence testing

Since the focus of this paper has so far been put on the predominantly 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 turns to recent examples of correspondence tests. It discusses how ethical committees responded to the proposed research designs and provides country examples. While most recent correspondence tests acknowledge the question of research ethics without going into further details, some researchers specifically refer to the ethics bodies in their countries, and the conditions they placed on the researchers in order to approve the research design.

4.1 The involvement of Research Ethics Committees

In the case of Norway, Midtøen reports that the National Committees for Research Ethics in Norway (NESH) approved the research design given that it met three 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 and no specific information about them should be

registered. 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. “By meeting these conditions, according to NESH, the potential societal value of the research findings would exceed the ethical problems related to field experiment research” (Arnfinn H Midtbøen, 2013, p. 53).

Before conducting their correspondence test in Germany the research team of the Expert Council of German Foundations on Integration and Migration (SVR) also submitted their research design to the ethical commission of the German Sociological Society and the Association of Sociologists in Germany. Here, too, the research design was judged to be unproblematic both from a data protection and an ethical point of view. It was argued that the aggregated analysis of the data would not allow inferences about individual research subject, and that the fact that the applications were made using fictitious applications did not infringe any personal rights (Schneider et al., 2014, p. 16).

In their study on the Netherlands, Blommaert et al. pointed out that they “minimized possible inconveniences to employers or genuine applicants by responding to positive reactions quickly and politely” (2014, p. 966). Similarly, Wood et al. reported that their internal ethics committee decided that “the burden for employers of considering an additional three applications while engaged in a public recruitment process was minimal, and that a speedy response to decline offers of interviews would minimise problems” (2009, p. 2).

Using these 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 criteria were met. The most important being:

- Inconveniences to employers should be kept to a minimum. Testing takes place at an early stage of the hiring process and invitations to interviews are declined swiftly.
- The number of fictitious applications sent to each employer should be kept low.
- Confidentiality and the privacy of research subjects are guaranteed.
- Data should be analysed in an aggregated form to avoid inferences being made about individual research subjects.
- Further steps in the research to follow-up on the results obtained in the testing should again follow the principles of ethical research in the social sciences.

Given the surge in numbers of correspondence tests conducted in OECD countries in the last years, provided researchers presented a thoughtful and detailed research design, correspondence tests have been accepted as a common methodology in studying discrimination in the labour market by ethical committees in all countries in which correspondence tests have been conducted.

4.2 Obstacles in correspondence testing – country examples

Unfortunately, in-depths information on the ethical approval process is only a small number of cases, namely, Sweden, Norway, and Germany. The case of Sweden is of interest, since it was the first country where the ILO field experiments did not receive ethical clearance during the first wave of studies and could only be conducted about ten years later. The decisions in Sweden further had

implications on researchers in Norway, which is briefly discussed as a second example. The most extensive information on the preparation of a correspondence test for research, however, comes from Germany, where not only the ethical questions were discussed, but where three comprehensive legal expertise were conducted before the experiment took place to eliminate any legal repercussions for the researchers.

4.2.1 Sweden and Norway

The first country in which a correspondence test was stopped by an ethics commission was Sweden. In the cadre of the ILO Project on labour market discrimination in the 1990s, Swedish researchers submitted two proposals to carry out a research using the correspondence test design as outlined by Bovenkerk (1992), yet their research proposals were not approved by the Swedish ethics board.

The grounds of objection to the research proposals focused in particular on “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). While a first assessor did not express any doubts, the ethical committee decided to carefully examine the ethical dimension of the research design. This analysis was conducted by one of its members, the law professor Rolf Nygren. Nygren concluded that while the research might be in the public interest, the potential consequences for people found guilty of discriminatory behaviour were too big. He claimed that “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, Nygren addressed potential issues of liability for the researcher or funding organisations and finally concluded his report with the recommendation that neither researcher should be funded to conduct a correspondence test as the Swedish contribution to the ILO project.

Yet, the position of the Swedish Ethics Board seemed to change around 2005, with three correspondence tests being published in Sweden in 2007. This might be due to the adoption of the new “Prohibition of Discrimination Act” in 2003 that also implemented the EU directives of 2000 against discrimination.

The first correspondence test was carried out by Carlsson and Rooth (2007) who conducted a field experiment on male Middle Eastern named applicants in the Swedish labour market between May 2005 and February 2006. Their paper, however, does not provide any information on the ethical aspects of the research, but refers readers to Riach and Rich (2004). The second research project by Bursell was approved by the responsible ethical vetting board in February 2006 and, in a very brief section discussing ethics, she claimed that it was “according to [her] knowledge, the only field experiment testing ethnic discrimination in Sweden that has been approved by the board” (Bursell, 2007, p. 9). Following the approval, Bursell conducted her field experiment on male and female applicants with Arab names or foreign names that were perceived as Arabic or stemming from the Horn of Africa region from March 2006 to September 2007. The third field experiment was conducted in 2006 by Attström (2007) and finally offered a Swedish contribution to the ILO project started in the 1990s, yet, as in other ILO studies the issue of ethics was not discussed. Attström used the ILO methodology that also included applications by phone and in person and focused on second generation immigrants of Middle Eastern origin. All three studies conducted in Sweden found discrimination against the minority applicants.

As Midtbøen (2013) points out, the rejection of the Swedish ILO projects also influenced Norwegian researchers:

“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 changed its view on the methodology and approved several research projects using correspondence testing, researchers in Norway also applied for funding and ethical approval to conduct a field experiment on the Norwegian labour market. Following a first enquiry to the National Committee for Research Ethics in the Social Sciences and Humanities (NESH) a report was prepared by Rogstad and Midtbøen that was then discussed in a meeting with NESH. In March 2009, the committee decided to approve the research project (Arnfinn H Midtbøen, 2013).

4.2.2 Germany

In Germany so far four correspondence tests on ethnic discrimination in hiring decisions have been conducted, the first one by Goldberg et al. (1995) on male Turkish applicants as part of the ILO Project. They, however, do not mention ethical issues at all in their publication. It took more than fifteen years until the second testing in Germany was carried out by Kaas and Manger (2012). Like Carlsson and Rooth (2007) in Sweden, they do not discuss the ethical issues inherent in correspondence testing, but also refer readers to Riach and Rich (2004).

This omission to discuss the ethical aspects of correspondence testing changed with the 2014 publication of the Expert Council of German Foundations on Integration and Migration’s (SVR) report, which addressed discrimination of Turkish named applicants in the apprenticeship market. Their report includes a short section that refers to the ethical and legal challenges researchers are faced with in correspondence testing (Schneider et al., 2014, p. 16). They emphasise that the research design was approved by the ethical committees of the German Sociological Society as well as the German Association of Sociologists. However, the research team went even further than obtaining only 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 the research design proposed by their research team in preparation 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.

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 as regulated in paragraph 22 of the German General Law on Equal Treatment (2010). This second expertise, however, excludes testing for research purposes.

In particular, Klose and Kühn argue that testing does not fulfil the elements of hiring fraud crime, which would only be applicable once an employment contract is signed (2009). They also address the possibility that research subjects might suffer from financial repercussions, but conclude that a loss of working time or a loss of income due to fraud cannot be claimed (2010, p. 26). Furthermore, Klose and Kühn focus on the question of forgery of documents. As it is common to include detailed information about an applicant in an application package in Germany, including high school or university certificates, these need to be constructed by the researchers. Applications missing these certificates will be regarded as incomplete and probably be discarded immediately. Klose and Kühn explain that including certificates is possible under certain conditions:

“A copy or a collage is, on its own, not a certificate. According to §267 of the Penal Code it is therefore not punishable, to make a photocopy of a real certificate, change details in this copy, make another photocopy of the modified document, to hide the manipulation, if this photocopy can then be recognised as a copy and is not used as an original certificate” (Klose & Kühn, 2009, p. 40, own translation).

Klose and Kühn conclude that testing is an indispensable instrument to study discrimination and that the results obtained using this methodology, if they adhere to strict methodological requirements, can also be used as circumstantial evidence in discrimination cases. The use of testing is therefore ethically and legally justified (2009, p. 49).

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 preparation of the correspondence test planned by the Expert Council of German Foundations on Integration and Migration. It is therefore the most relevant publication to be considered here as it addresses legal problems that might arise under both criminal and civil law. Regarding the 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. 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 the civil law, Kühn et al. argue that claims for liability of the researcher due to the work 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 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.

The fourth and most recent correspondence test conducted in Germany examined the discrimination of Turkish women, testing in particular the effect of wearing a headscarf on the picture enclosed in the application (Weichselbaumer, 2015). There is no mentioning of any ethical issues in this paper.

These 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 and found not be an obstacle to conducting a correspondence test. While a similar wealth of information on ethical and legal preparation work was not available for other countries, the examples of Sweden and Norway 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.

5. Conclusion

The spill-over of research ethics from biomedical sciences to social sciences has also affected correspondence testing on the labour market. While economists do not have to adhere to research ethics guidelines, the fact that correspondence testing lies at the intersection between economics and social sciences, makes the discussion of ethical questions a fundamental part of the research design. Since the methodology is deceptive by design, some major concerns have been voiced regarding not adhering to the principles of voluntary participation and informed consent, the use of deception, the potential entrapment of employers, or the question of employers' rights.

As the previous discussion of ethical issues concerning correspondence testing has shown, the questions and concerns invoked deserve a careful deliberation and any research project planning to conduct a correspondence test should address these issues carefully. The increasing use of correspondence testing in recent years across OECD countries, as well as the acceptance of evidence obtained by testings by numerous courts in Europe and the US indicate, that the use of this covert research methodology has been deemed ethically acceptable to obtain knowledge about the extent of discrimination. The potential harm that research participants that are not aware of their role in their experiment might suffer, has to be weighed against the societal interest of knowing the extent of discrimination in any society studied. This knowledge can then be used as a basis to combat discrimination in hiring decisions.

It has, however, been emphasised that these field experiment require a high level of planning and need to be of high quality. Field experiment are currently deemed to be the best suited methodology to measure discrimination and no other approach leads to similar results. Studying discrimination and obtaining a “rigorous and realistic measurement of discrimination is fundamental to understanding and addressing persistent barriers to employment facing members of stigmatized groups” (Pager, 2007, p. 78). Given the increased diversification of modern societies, such barriers can prove detrimental to social cohesion.

References

- Attström, K. (2007). *Discrimination against native Swedes of immigrant origin in access to employment*. Geneva: International Labor Organization.
- 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
- 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
- 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 15.01.2016, from <http://www.britsoc.co.uk/media/27107/StatementofEthicalPractice.pdf>
- Burnham, P., Lutz, K. G., Grant, W., & Layton-Henry, Z. (2008). *Research methods in politics* (2nd ed.). Basingstoke: Palgrave Macmillan.
- Bursell, M. (2007). » What's in a name?-A field experiment test for the existence of ethnic discrimination in the hiring process «, The Stockholm University Linnaeus Center for Integration Studies (SULCIS): Working Paper 2007: 7.
- Calvey, D. (2000). Getting on the door and staying there: A covert participant observational study of bouncers. *Danger in the Field: Risk and Ethics in Social Research*. New York: Routledge.
- 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>
- Dench, S., Iphofen, R., & Huws, U. (2004). An EU Code of Ethics for Socio-Economic Research: The Institute for Employment Studies.
- 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.

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.

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/gingrich.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 18.02.2016, 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.

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*.

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?".

Microsoft. (2015). Microsoft Services Agreement (Published June 4, 2015). Retrieved 18.02.2016, from <https://www.microsoft.com/en-us/servicesagreement/>

Midtbøen, A. H. (2013). *Determining Discrimination* University of Oslo.

Midtbøen, A. H. (2014a). 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

Midtbøen, A. H. (2014b). 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., & Rogstad, J. (2012). Discrimination. Methodological controversies and sociological perspectives on future research. *Nordic Journal of Migration Research*, 2(3), 203-212.

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 15.01.2015, from <https://www.etikkom.no/globalassets/documents/publikasjoner-som-pdf/guidelines-for-research-ethics-in-the-social-sciences-law-and-the-humanities-2006.pdf>

Nuernberg Military Tribunals. (1949). *Trials of War Criminals Before the Nuernberg Military Tribunals Under Control Council Law No. 10* (Vol. 2). Washington, DC: U.S. Government Printing Office.

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.

RESPECT Code of Practice - Introduction. (n.d.). Retrieved 07.03.2016, from <http://www.respectproject.org/code/index.php?id=files/18/index.html>

RESPECT for research ethics: summary. (n.d.). Retrieved 07.03.2016, from <http://www.respectproject.org/ethics/files/12/ethics.html>

RESPECT project - Context: why a project like RESPECT is needed. (n.d.). Retrieved 07.03.2016, from <http://www.respectproject.org/main/context.phpfiles/16/context.html>

Riach, P. A., & Rich, J. (2004). Deceptive field experiments of discrimination: are they ethical? *Kyklos*, 57(3), 457-470.

Schneider, J., Yemane, R., & Weinmann, M. (2014). Diskriminierung am Ausbildungsmarkt: Ausmass, Ursachen und Handlungsperspektiven Berlin Forschungsbereich beim Sachverständigenrat deutscher Stiftungen für Integration und Migration (SVR).

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.

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). Beyond the veil: Discrimination against female migrants wearing a headscarf in Germany. *University of Linz Working Paper*.

Wood, M., Hales, J., Purdon, S., Sejersen, T., & Hayllar, O. (2009). A Test for Racial Discrimination in Recruitment Practice in British Cities. Norwich: Department for Work and Pensions.

Yahoo. (2012). Yahoo Terms of Service (Last updated March 16, 2012). Retrieved 18.02.2016, from <https://policies.yahoo.com/us/en/yahoo/terms/utos/>

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