

## DISCRIMINATION

*Methodological controversies and sociological perspectives on future research***Abstract**

Field experiments have proved productive in documenting the continuing relevance of race and ethnicity in the shaping of opportunities in modern societies. However, methodological questions have been raised concerning the existence of unobservable variables and the subsequent danger of drawing inferences on a wrong basis. In this article, we address these debates and discuss a research strategy which to a large degree exceeds the challenges posed in the literature. By supplementing a correspondence test study with in-depth interviews with employers who were subjected to the fictitious job applicants in the experiment, we get access to the extent of discrimination in the labour market as well as to the mechanisms involved in discriminatory hiring practices. The design has several advantages compared to 'single-method' approaches and provides a more substantial understanding of the processes leading to ethnic inequality in the labour market.

**Keywords**

Ethnic discrimination • field experiment • correspondence testing • labour market • mixed methods

Arnfinn H. Midtbøen<sup>1</sup>, Jon Rogstad<sup>2</sup>

<sup>1</sup> Institute for Social Research, Oslo

<sup>2</sup> Fafo - Institute for Labour and Social Research

## 1 Introduction

Ethnic discrimination remains a controversial topic in modern societies. Insofar as ethnicity constitutes an obstacle to full participation in the educational system, the labour market and other societal arenas, the normative principle of equality of opportunity widely held by most governments is being challenged. Despite the fact that ethnic inequalities for decades have been a matter of great concern, however, two questions continue to create debate among both scholars and politicians: What is the actual *extent* of discrimination in different spheres of social life? And what are the *causes* of discriminatory treatment?

The first question is primarily related to how and by which methods discrimination should be measured (Banton 1994; Bovenkerk 1992). While traditional methods such as surveys and regression analysis of large data sets have been perceived as insufficient for measuring discrimination directly, there has been a growing interest for experimental methods among social scientists studying discrimination in different market situations (Pager & Shepherd 2008; Riach & Rich 2002; Blank, Dabady & Citro 2004). By creating field

experiments that test access to opportunity for ethnic minorities compared to equally qualified persons from the majority population in, for instance, the labour market, it is possible to measure the causal effect of ethnicity and draw statistically significant conclusions about the extent of discrimination (Quillian 2006; Pager 2007).

Although field experiments have proven productive in documenting the continuing relevance of ethnicity in the shaping of opportunities in modern societies, the evidence they offer does not shed much light on the second question, i.e., the underlying causes of discrimination. Whether discrimination is caused by racist motives, statistical uncertainty or stereotypes unconsciously working in disfavour of ethnic minorities, is widely debated in the discrimination literature (e.g., Pager & Karafin 2009; Petersen 2008; Reskin 2000; Reskin 2008). An important question is thus how to develop a research strategy which simultaneously addresses both the extent of discrimination and the mechanisms involved in discriminatory practices.

The aim of this article is to discuss the methodological possibilities and limitations of the field experiment tradition. We argue that combining a correspondence study with in-depth employer interviews represents a methodological innovation for research on

\* E-mail: ahm@socialresearch.no, jon.rogstad@fafo.no

discrimination. The use of mixed methods is obviously not new to sociological research and is in itself not entirely new in the field experiment tradition either (see Carlsson & Rooth 2007; Pager & Quillian 2005; Pager, Western & Bonikowski 2009). Drawing on empirical illustrations from a research project addressing the extent and causes of ethnic discrimination in the Norwegian labour market, however, we show that this specific research design has important advantages when the aim is to gain insight into the nature of discrimination. The field experiment makes it possible to draw statistically significant conclusions on the extent of ethnic discrimination. By interviewing a selection of the exact same employers who participated in the initial testing, we are furthermore able to explore the various reasons why employers act as they do and the contexts surrounding their hiring decisions. The crucial point is that the qualitative interviews take place in a situation where we as researchers already know the outcome of the specific tests, allowing us to link the interviews to the employers' behaviour in actual employment processes.<sup>1</sup>

We begin by briefly discussing traditional methods for the study of discrimination, highlighting their insufficiencies in producing direct measures of the extent of discriminatory practices. We then discuss the audit method more specifically, both by clarifying its importance in documenting the continuing significance of ethnic discrimination and by accounting for important critical perspectives (e.g., Heckman 1998; Heckman & Siegelman 1993). Critique of the audit method has resulted in several refinements of the testing technique, including a more frequent use of correspondence testing in recent years (see Bertrand & Mullainathan 2004; Bursell 2007; Carlsson & Rooth 2007; Jackson 2009; Kaas & Manger 2012). Although such refinements answer some of the main objections, the most serious critique of the audit method – that unobservable variables within firms may influence the outcomes of the audit study without the researcher's awareness – is not accommodated by a simple change of test technique. Supplementing the findings of the field experiment with data from qualitative interviews largely solves this problem, because the interviews provide a broader contextual frame of interpretation and a more empirically based understanding of the mechanisms leading to discriminatory hiring decisions. Thus, this design has clear advantages over single-method approaches in that it offers a more complete picture of the complex ways in which discrimination takes place in modern societies.

## 2 Traditional methods for measuring discrimination

Discrimination has been a central field of study for social scientists for decades, resulting in a variety of methods and techniques for measurement. Traditional approaches include studying the *experiences* individuals may have with discriminatory treatment, which best can be investigated through the use of surveys or qualitative interviews; a focus on the role of *potential discriminators* and their attitudes

and prejudices, which usually also are addressed by surveys and interviews; and lastly, studies of *ethnic differentials* in, for instance, labour market outcomes, which are most commonly examined by statistical analysis (Pager & Shepherd 2008; Blank, Dabady & Citro 2004). Which method is the most suitable depends on the question posed: A focus on people's experiences highlights central aspects of everyday life; studies of potential discriminators may provide insights into the way individuals in power positions make their decisions; and statistical analyses are of indisputable importance in providing large-scale pictures of ethnic inequalities.

Each of these methods thus represents different ways to acquire knowledge of important aspects of the discrimination complex. Yet, they all have obvious limitations if the purpose is to establish knowledge of the actual extent of discrimination. Studies of perceived discrimination may result in both over- and underreported levels of discrimination due to the subjective character of experiences. Asking individuals in power positions what they would have done in hypothetical situations does not have to correspond with what they actually do in practice. Lastly, the investigation of ethnic differentials in large data sets only provides an indirect measure of discrimination because what is actually measured is the *residual gap* between different groups in an outcome (Midtbøen & Rogstad 2008; Pager & Shepherd 2008; Rogstad 2001). These limitations of traditional methods for measuring discrimination have paved the way for the increasing use of experimental approaches, usually referred to as *audit testing* in the literature.<sup>2</sup>

## 3 Audit testing

Audit testing was developed simultaneously in Britain and the US in the late 1960's. In the British case, the first study (Daniel 1968) was conducted to assess the extent of discrimination in areas that were not covered by the Race Relation Act of 1965, such as labour and housing markets (Riach & Rich 2002: 491). Similarly, in the US audit testing was introduced in conjunction with legal progress. As the economist Kenneth J. Arrow would put it later, discrimination in American society before the Civil Rights Act of 1964 was 'a fact too evident for detection, and too gross for aggravation' (Arrow 1998: 92). After the bill was adopted the most obvious examples of discrimination, for instance in advertising vacancies, disappeared. The pertinent question was thus whether it also implied less *de facto* discrimination or simply a change in expression, from an open to a more covert and subtle character.

Audit testing was introduced as a relevant method for investigating whether such a change had occurred. When investigating the prevalence of discrimination in employments, the method requires two individuals, one with a name and a physiological appearance of a person from the majority population, the other with a name and appearance signalling an ethnic or racial minority background, to apply for the same job. Because the two individuals are given the same

qualifications (i.e., identical education levels and work experience), and have the same age, the same sex, etc., systematically differential treatment of the minority applicant is interpreted as evidence of discrimination.

Compared to traditional methods of measuring discrimination, the audit method has several advantages. First, it documents the extent of discrimination based on the *actual treatment* of identical applicants, not on what employers say that they do, which is the case when using qualitative interviews or survey questionnaires. Second, the discrimination rate is a result of a *direct measure* of discriminatory treatment, not a reflection of indirect indications. Since the experimental design gives the researcher control over all the variables involved, it is possible to isolate an 'ethnic' variable and determine the causal effect of ethnicity on, for instance, access to the labour market. As such, and through studies from several countries over four decades, the audit method has proved important in documenting the continuing significance of racial and ethnic discrimination in modern societies (see reviews in Pager & Shepherd 2008; Pager 2007; Riach & Rich 2002).

## 4 Critique of the audit method

The audit test method has been an important contribution to the research on discrimination, traditionally in the Anglo-American world but in recent years also in a wider European context (e.g., Allasino *et al.* 2004; Attström 2007; Cediey & Foroni 2008; Zegers de Beijl 1999). However, the method has also been criticised on several occasions. On the one hand, objections are made with respect to research ethics, as field experiments by definition represent a violation of the widely held principle of informed consent when participating in research (Banton 1997; Midtbøen & Rogstad 2008; Riach & Rich 2004; Rogstad 1996). On the other hand, important questions are raised to what field experiments are actually measuring – i.e., the validity of the method – which is the topic addressed in this article.<sup>3</sup> For an understanding of the methodological debate concerning audit testing it is worth giving a few examples from the early American research, which had a tendency to overstate the implications of its findings. We will then discuss the most important critique of the method, before turning to some different ways in which these may be met.

### 4.1 Devoted social scientists

Many readers of the audit literature may have been struck by the confidence of some of its practitioners. It might be the experimental character of the method, the opportunity to determine causal effects, and the apparently easily interpreted results which have created a research tradition marked by a certain presumptuousness on behalf of the conclusions drawn. An anthology published by The Urban Institute in Washington D.C., which summarises much of the early

research using audit testing, is a good example of the type of presumptuousness we have in mind. The anthology carries the illustrative title *Clear and Convincing Evidence* (Fix & Struyk 1993) and leaves little room to doubt the existence of a considerable amount of racial discrimination in American society or that the audit method documents this fact in a convincing way. The editors are also certain that the findings have one particular type of policy implication, namely the strengthening of the content and the enforcement of the civil rights law.

Another example is a symposium on racial discrimination, arranged at Princeton University in 1997 and later published in *Journal of Economic Perspectives*. At this symposium, three researchers accounted for the extent of racial discrimination in different marked arenas in the US, i.e., the labour market (Darity & Mason 1998), the consumer market (Yinger 1998) and the credit market (Ladd 1998). All three found comprehensive discrimination in their respective fields. As Darity & Mason write, there is 'a substantial racial [...] disparity in the American economy. [...] discriminatory treatment within the labor market is a major cause of this inequality. The evidence is ubiquitous' (Darity & Mason 1998: 63).

The point here is not to say that the audit method in general or these research projects in particular are problematic *per se*, but that researchers using this method should, like every researcher, interpret their findings with caution. When Yinger (1998: 36) states that audit testing documents the *causes* of discrimination or Darity and Mason (1998: 76) claim that discrimination in hiring, uncovered by testing, provides a convincing explanation for the ethnic differentials in *income*, in our opinion their conclusions are not to be regarded as clear and convincing evidence, but as generalisations which are not sustained by the actual findings.<sup>4</sup>

These critical remarks on some examples of early audit testing research are simply meant to emphasise what the test approach is actually measuring: The likelihood of landing a job for an individual with a racial or ethnic minority background compared to an individual from the majority population. However, in the methodological debate concerning the audit test technique, mainly originating from Heckman's contributions (Heckman 1998; Heckman & Siegelman 1993), other and more substantial reasons for questioning the technique have also been proposed.

### 4.2 Heckman's critique of the audit method

Heckman's first visit to the audit test tradition was in an article in the early 1990s anthology, *Clear and Convincing Evidence* (Fix & Struyk 1993). He and his colleague Peter Siegelman had been invited to comment on the research accomplished at The Urban Institute (Heckman & Siegelman 1993). In their response, they question the researcher's ability to construct identical test persons in every relevant aspect, arguing that differences between the candidates could bias the results in ways hidden from the researcher. Furthermore,

they point to the risk that the minority applicant may be motivated to document the existence of discrimination by 'performing worse' – resulting in a biased measure of discrimination. At the same time, the authors acknowledge that the audit test approach may offer valuable insight into the extent of discrimination, concluding that 'the audit pair methodology, properly augmented, promises to shed much new light on the nature of racial and ethnic disparities in labor market outcomes' (Heckman & Siegelman 1993: 227).

A far more direct critique of the scientific potential of the audit method was aired at the Princeton symposium on racial discrimination in 1997. Heckman was invited to comment on the three articles mentioned above and he states that the conclusions – in his words that the 'American society is riddled with racism and discrimination by employers' (Heckman 1998: 101) – by no means are made probable by the evidence proposed. In his view, the documentation of ethnic discrimination provided by the audit method is far less convincing than it is presented to be, and he argues that the results are 'fragile to alternative assumptions about unobservable variables and the way the labor markets work' (Heckman 1998: 102).<sup>5</sup>

Heckman has three main objections to the audit test method. First, he questions the *representativeness* of the research using the test technique. Early research has primarily investigated the occurrence of discrimination in low skilled occupations. This makes the foundation for generalisation vulnerable, partly because the testing is only implemented in a small section of the labour market and partly due to the fact that a large amount of jobs in the American labour market are found through social networks. Since audit testing leans on job openings advertised, such studies have a tendency to be highly unrepresentative.

Second, Heckman points to the way the discrimination rate is *calculated*. In the majority of research projects using audit testing, the net discrimination rate is estimated on the basis of the situations where at least one of the candidates gets a positive response.<sup>6</sup> Heckman, however, argues that the discrimination rate has to be calculated on the basis of all situations where the two applicants are treated symmetrically – i.e., the sum of situations where both got a job and both were rejected. Considering that in the majority of tests neither of the applicants get a job offer (and hence are registered as non-observations in the literature), Heckman's proposed way of calculating discrimination dramatically changes the results: By re-analysing three audit studies accomplished by The Urban Institute and using his own calculation technique, Heckman shows that the evidence suggests only a slight preference in favour of the majority applicant (Heckman 1998: 105).<sup>7</sup>

The third objection concerns the existence of *unobservable variables*. The audit method is supposed to control for every relevant variable that might affect the outcome in hiring decisions by making the two applicants equally qualified, and hence measuring the causal effect of ethnicity on the probability to get a job. According to Heckman, however, researchers using audits for investigating discrimination in the labour market usually have limited knowledge

of the decisive determinants of productivity within firms. This insufficient institutional knowledge makes it unlikely that every relevant factor actually is perfectly matched in each test, raising the danger of biased results in favour of one applicant at the expense of the other. This may have two different consequences for the conclusions drawn: If the unobservable variables are in favour of the majority applicant, the results will suggest that discrimination occurs when it actually does not. On the other hand, and depending on the level of qualifications chosen for the applicants, the results may suggest equal treatment in contexts where unequal treatment is the norm, and hence that 'the method could fail to detect discrimination when it does exist' (Heckman 1998: 111).

In other words, the main argument for using the audit test approach – the possibility of measuring the causal effect of ethnicity in the access to employment – is disqualified, and, according to Heckman, the method therefore lacks embedment in social reality. His solution to these problems is to leave audit testing as a scientific method, focusing only on statistical analysis of large data sets. In Heckman's view, employers' discriminatory hiring practices are no longer relevant in explaining racial inequality in American society. The differences in income and employment are instead explained by pre-market factors – i.e., differences in skills that are due to inequality of family environments, schools and neighbourhoods – and not discrimination. This, according to Heckman, should in turn have implications for the direction of social policy.

## 5 From audits to correspondence testing: A first refinement of the experimental approach

Heckman's critique raises serious objections to the use of audit testing. In recent years, the presumptuousness that characterised parts of the audit testing research also seem to have evaporated. Overstated generalisations on the basis of test results are no longer the norm, and most researchers now seem to specify what is actually being measured. In addition, the critique has led to a change of test technique, where several researchers have switched to the correspondence testing method instead, leaving the audit method behind (e.g. Bertrand & Mullainathan 2004; Bursell 2007; Carlsson 2009; Carlsson & Rooth 2007; Jackson 2009; Kaas & Manger 2012).

While audit testing requires that actual individuals are sent in person to entry-level jobs acting as job seekers, correspondence testing does not involve individual test persons; instead, it corresponds with the employers through written applications. This difference meets several of Heckman's objections. First, it gives the researcher complete control over the experiment. The use of test persons, as Heckman has pointed out, introduces the risk of bias due to different performances, first impressions, etc. In correspondence testing, this bias is eliminated because the researcher controls the content of the applications. Unintended bias from the researcher is also avoided

by randomly assigning the names to applications each time they are sent to job openings, making correspondence testing a randomised experiment which provides a strong opportunity to draw causal inferences. Because correspondence testing avoids interaction between individuals, however, what is measured is job interview offers – so-called *call-back rates* – and not actual job offers. This is obviously a limitation. On the other hand, a comprehensive meta-study of the field experiment tradition show that about 90 per cent of the discrimination occurs in the stage of the job seeking process where job interview offers are made, making this limitation less decisive (Riach & Rich 2002: 494).

Second, shifting from audits to correspondence testing speaks to the question of representativeness. Not using test persons makes it possible to apply for more job openings and hence get more data, compared to the expensive and time consuming audit technique. Furthermore, the researcher is free to investigate the extent of discrimination not only in low-wage labour markets but also in occupations requiring an academic degree. Obviously, very high-skilled labour markets still represent a problem. Occupations requiring highly specialised qualifications are hard to apply for due to the difficulties in making reliable applications. Neither should the relevance of social networks in these labour markets segments be underestimated (e.g., Loury 1998; Petersen, Saporta & Seidel 2000). Particularly high-skilled labour markets are, however, not Heckman's main concern. His critique is aimed at audit studies which were 'conducted for hiring in entry level jobs in certain low skill occupations using overqualified college students during summer vacations' (Heckman 1998: 104). Although there are also limits to the generalisability of research using correspondence testing, this technique clearly allows for a wider range of occupations and qualification levels which, to a large degree, meets the problem of representativeness.<sup>8</sup>

## 6 Addressing the causes of discrimination: A second refinement

Heckman's critique of the audit method has proven productive in raising awareness of the potential pitfalls of field experiments in the social sciences and in increasing the utilisation of correspondence testing in discrimination research. Neither audit nor correspondence testing can, however, explain why and under what circumstances discrimination occurs. The question of causes is closely related to the last point in Heckman's critique, namely that the existence of unobservable variables – e.g., the reasons why employers act as they do – raises the danger of inferring the wrong conclusions from the outcome of a field experiment. Although obviously important, this point has been strangely overlooked in the literature (but see Neumark 2010 for an interesting discussion). In this last section of the article, we will thus present and discuss a research strategy which simultaneously addresses both the extent of discrimination and the mechanisms

involved in discriminatory practice; a dual objective which to some extent solves this problem in the field experiment tradition.

### 6.1 The advantage of combining methods

In a research project investigating the occurrence of ethnic discrimination in the Norwegian labour market (Midtbøen & Rogstad 2012), we coupled correspondence testing with in-depth interviews with a selection of the employers who had participated in the initial testing. 1800 written résumés and application letters were sent to job openings in 12 occupational categories in Oslo and the surrounding area. The different occupations were chosen to cover a relatively broad picture of the Norwegian labour market, varying from qualified occupations like economists, high school teachers and IT-professionals, to less qualified occupations such as drivers, warehouse workers and auxiliary nurses, in both the private and the public sector. We operated with two pairs of applicants, one male pair and one female pair, and for each pair there were two résumés with identical educational qualifications, length of work experience, etc., making the name of the applicants – one with a typical Norwegian name and the other with a name indicating Pakistani background<sup>9</sup> – the only difference between the two. As soon as one or both of the applicants were contacted by an employer, the outcome of the test was registered and the job offers were politely refused. After a while, a selection of employers who had contacted one or both of the applicants received a letter informing them about the project and their participation, and encouraging them to contact us for an interview.

Letters were sent to 163 employers and 42 in-depth interviews were conducted, providing a response rate of approximately 25 per cent. 95 letters were sent to employers who had invited both candidates for a job interview, 64 letters to employers who invited only the candidate with a majority background and 4 letters were sent to employers that invited only the minority applicant. Among the final respondents there were employers from both the private and public sector and from occupations where both the male and the female pair applied. The interviews were conducted on the basis of tests where both applicants were invited for a job interview (23 interviews), where only the majority applicant was invited (17 interviews) and where only the minority applicant was invited (2 interviews). The distribution of responses among the employers who contacted us suggests that the research does not suffer from a direct selection bias – which might have been expected.

Although the combination of correspondence testing and qualitative interviews represent a novel research strategy in the study of discrimination, mixed method approaches in the field experiment traditions are not unique. One example is Pager and Quillian (2005), who matched an audit study with a telephone survey of the same employers and documented severe discrepancies between what employers say and what is actually done in practice. Another example

is Pager and colleagues (2009) who supplemented the findings of an audit study of a low-wage labour market with qualitative material from field notes the test persons had written down after interacting with employers in job interviews. Both of these approaches resulted in sophisticated analyses of contemporary patterns of discrimination and are important contributions to the field experiment tradition.

Despite these different strategies, recent research projects have not been able to provide a direct linkage between the outcomes of field experiments, the context of employment processes and the ways in which employers account for their recruitment practices. This is a general limitation in the discrimination literature due to the obvious importance of the role of the employers in the shaping of employment opportunities (see e.g. Jackson 2009; Moss & Tilly 2001; Waldinger & Lichter 2003; Kirschenman & Neckerman 1991; Nilsson 2006). When exploring the causes of ethnic discrimination in the Norwegian labour market, we thus ask the employers themselves to interpret and consider their actions and the amount of discrimination that we document through the field experiment. By encouraging the employers to reflect upon the dilemmas that occur when they are evaluating a set of job applications, we get closer to the uncertainty and stereotyping processes that characterise hiring decisions. Getting insight in the context of the employment processes also sheds light on the variety of reasons why employers act as they do, resulting in a more complex analysis than one could make from the outcomes of the field experiment alone.

## 6.2 Complicating empirical illustrations

Interpreting the outcomes of audits and correspondence studies is usually considered a more or less easy task: To put it simply, employers favouring the majority applicant are expected to display some sort of conscious or unconscious racial bias; employers who practise symmetrical treatment of the two applicants are seen as colour blind and unbiased; and the few employers actually favouring the minority applicant at the expense of the majority candidate are expected to represent firms with some kind of affirmative action policy. However, such straight-forward assumptions about the relationship between cause and effect in field experiments may be challenged by analysing two empirical examples from our interviews with Norwegian employers who have been through the initial testing stage.

The first example is an interview with a manager of a public nursing home outside of Oslo where only the majority applicant received a call-back. At the interview, the employer was at first surprised that the test had ended up that way, due to his self-understanding as tolerant of ethnic diversity and because the nursing home already had an ethnically heterogeneous workforce. Reflecting upon the result, he mentioned that he could have acted the way he did because the firm had a hiring policy which aimed at avoiding the dominance of any particular ethnic group – to ensure that Norwegian remained the firm's main language.

The second example represents a rather rare incident in our material because it is one of few examples in which the only applicant who received a call-back was the candidate with a minority background. These interviewees, the assistant director and a section manager of a private insurance company in Oslo city centre, were also surprised when confronted with the result of the test – but for the opposite reason: Rather than having an affirmative action policy, the company had no explicit intention of increasing the number of ethnic minority employees even though only one of the 50 employees was of foreign descent. Instead, the assistant director expressed the widely held view that job applicants with a minority background have to have better qualifications than majority applicants to be considered of equal value for the company.

The two examples are interesting because they challenge the immediate interpretation of the results known from earlier field experiments. To be sure, the manager of the public nursing home ascribed a lack of knowledge of the Norwegian language merely due to the foreign sounding name of the application: There was nothing in the résumé or application letter from the minority candidate which indicated that he was a threat to the language at the firm; on the contrary, an important aim of the field experiment was to measure the extent in which children of immigrants – i.e., individuals who have grown up in Norway and will have achieved linguistic fluency in Norwegian when they enter the labour market – are experiencing employment discrimination. However, the heterogeneous workforce at the nursing home gave no consistent indications of discrimination on the part of the manager, or that the next minority applicant would be treated the same way.

The case of the insurance company is even more interesting. Although we expected the few companies which favoured the minority applicants to run an explicit affirmative action policy, the interviewees were among the most clear spoken in our data material in saying that minority applicants have to prove themselves *more* productive than their majority peers to be considered as equal candidates for a job. That the minority candidate was preferred in this case thus might have been a coincidence, but later in the interview the assistant director of the insurance company also mentioned that due to a recent change in the ethnic composition of their clientele they had discussed the relevance of hiring a multilingual candidate for this position, which might explain this rather contradictory finding. In any case, it should be apparent that insight into the contexts in which hiring decisions are made complicates the initial inferences one would draw from the outcomes of a field experiment.

## 7 Revisiting the critique of the field experiment tradition

What are the methodological implications of these empirical illustrations? Heckman concludes that the potential biasing effects of unobservable variables at the firm level should lead to the

abolishment of field experiments as a method for studying the extent of discrimination in different market contexts. The two examples discussed above indeed suggest that factors unobservable for the field experiment researcher should not be neglected. Firms and industries differ in terms of the ethnic composition of the workforce, the experiences employers have with managing ethnically diverse workplaces, and they differ in terms of organisational policies and recruitment strategies. These factors are just some of many which may affect the hiring decisions employers make, and ignoring them may bias our understanding of the dynamics of discrimination.

It is, however, far from recognising the importance of contextual knowledge of firm characteristics and recruitment strategies in interpreting the findings from field experiments to suggest that this entire tradition of discrimination research is without scientific value. Heckmans critique of the audit method does ignore the overwhelming fact that for 40 years sociologists and economists have documented that ethnic and racial discrimination continue to shape the access to opportunities for young men and women across a wide range of countries. This research cannot be overlooked, but it is time to address some of its methodological limitations. One important way of doing this is to combine field experiments with qualitative methods because crucial knowledge of the context of employments and the mechanisms involved in discriminatory practices cannot be accessed without some sort of qualitative approach.

This methodological stance also has theoretical implications: Recent studies have borrowed insight from social psychology (in particular Fiske 1998), arguing that unconsciously working stereotypes may explain why racial and ethnic minorities continue to experience unequal access to opportunity (see e.g., Bertrand, Chugh & Mullainathan 2005; Reskin 2000; Reskin 2008; Quillian 2008; Rooth 2007; Quillian 2006; Quillian & Pager 2010). We agree that stereotypes are important causes of discrimination, and hope that the adoption of perspectives from social psychology may offer a way out of the ever-lasting discussion of motive-based explanations well known from the economics literature. Used in relation to field experiment research, however, these theoretical discussions tend to assume that all discriminatory outcomes may be explained by the same factor (i.e., stereotypes), while the causes of equal treatment and 'reverse' discrimination (a call-back only for the minority applicant) are seldom addressed. The implication of this tendency is not just that the situations where employers are 'caught in the act' of discrimination are the only cases worth theorising. Leaving other outcomes unexplained and the possibility for mediating contextual factors providing more or less room for discrimination unaddressed, what started as a methodological problem (the limitations of field experiments) has consequences for the theoretical contributions to the research on discrimination and ethnic inequality. In scratching the surface of the outcome numbers we are reminded that the processes in which ethnicity come to matter are highly complex and that the interpretation of the findings from field experiments should be treated with caution. While the results of research using only

the testing technique are unable to clarify more than the fact that discrimination in employments exists, supplementing field experiments with qualitative approaches complicate this one-sided picture of the discrimination complex, paving the way for new insights into the dynamics of contemporary discrimination.

## 8 Conclusion

In this article, we have argued that a combination of correspondence testing and in-depth interviews with employers represents an innovative way to address questions of the extent and causes of ethnic discrimination in the labour market. Using only the audit or correspondence technique may result in conclusions based on insufficient knowledge of the context of employment and thereby increasing the danger of drawing inferences on a wrong basis. Supplementing the testing with qualitative interviews, on the other hand, enables a more substantial understanding of the processes leading to ethnic inequality in the labour market, thus offering both methodological and theoretical advantages over single-method approaches to the study of discrimination.

Why is it important to address the extent and causes of ethnic discrimination? Whether or not discrimination is an important factor explaining ethnic differentials in the labour market is still a debated question, and the answer has significant implications for the development of social policy. Field experiments usually document that discrimination in the labour market exists, providing a more or less explicit recommendation to strengthen civil rights law. On the basis of a critical dissection of the audit method, however, Heckman (1998) argues that racial differences in income and employment in the US are due to the different skills racial groups bring to the market, and not to discrimination *within* the market. According to him, the important tasks are skill promotion through the improvement of schools and neighbourhoods.

This discrepancy in the implications of research results illustrates the importance of understanding the mechanisms creating ethnic differentials in the labour market. Such an understanding includes a focus on the employers, who, through their role as gatekeepers, determine access to opportunities in the labour market. Despite the legal protection of ethnic minorities in hiring processes, employers have the autonomy to determine who is hired and who is not, and they have a wide range of legitimate arguments to justify their decisions. Hence it is clear that an understanding of why, when and under what conditions employers discriminate is of unquestionable relevance for the legal and political means to prevent such discrimination.

## Acknowledgments

The authors are grateful for valuable comments from Grete Brochmann, John Wrench, and two anonymous referees, as well as

from the research group Equality, inclusion and migration at Institute for Social Research, Oslo, and participants at a seminar held at Institut National d'Etudes Démographiques, Paris, on July 5.-6. 2010.

**Arnfinn H. Midtbøen** holds an MA in sociology from the University of Oslo and is a PhD Candidate at Institute for Social Research in Oslo, Norway. He has contributed to several academic journals in Norway, writing on the differences between Norway, Denmark and Sweden on the fields of nationality legislation, citizenship and integration policies, and on dilemmas concerning racism and ethnic discrimination in the labour market. His current PhD project concerns the nature and causes of employment discrimination of ethnic minorities and the political and judicial dilemmas following such discrimination.

**Dr. Jon Rogstad** is Research Director at Fafo–Institute for Labour and Social Research. Rogstad has written several books and articles concerning discrimination, social exclusion, and labor market participation as well as political mobilization, social movements and transnational engagement. Among his recent publications are 'Antiracism and social movements: The importance of critical events' *Journal of Ethnic Migration Studies*, 2012 (with Carl Endre Espeland), and 'The Art of Articulation. The Mediation of Political Engagement in Multicultural Networks', *Social Movement Studies*, 2011 (with Viggo Vestel).

## Notes

1. The project was submitted to The National Ethical Committee for Research in The Social Sciences and Humanities (NESH) in Norway in September 2008 and received approval in March 2009. To minimise the ethical challenges involved in the design, particularly the potential negative consequences for the employers that were confronted with the outcome of their decisions, we avoided any normative 'blaming' of the employers' decisions. Furthermore, to participate in the interviews the employers themselves had to respond to a written invitation, providing an informed consent in the second part of the study. The original application to NESH, the subsequent discussion and the final letter of approval can be found in Rogstad and Midtbøen (2009).
2. Audit testing is equivalent to *situation testing*, which has been a more widely used term in European research, but differs from *correspondence testing* which will become apparent later on in this article. More general concepts including all the different testing techniques are 'discrimination testing' or 'field experiments'.
3. By leaving the ethical challenges involved in experimental research at this point, we do not suggest that such challenges are unimportant. In our opinion, the problems concerning the ethical dimension of field experiments have been too poorly addressed in the literature. There are reasons to question a frequent use

of such methods unless the research design allows for a more substantial understanding of the mechanisms leading to ethnic inequality. Field experiments conducted without being supplemented with other methods, we argue, do not have the ability to shed light on much more than the fact that discrimination exists and the relevance of each new project should consequently be ethically scrutinised. See Midtbøen and Rogstad (2008) for an elaboration of this argument and Rogstad and Midtbøen (2009) for a detailed account of the process leading to the approval of this current project by NESH.

4. Yinger (1998) does acknowledge, however, that the causes of discrimination may be multifaceted and that more research is needed to further explore the mechanisms involved in discriminatory practices in the market place. Still, he leaves the impression that field experiments can assess this situation, which we argue they cannot.
5. Heckman limits the discussion to audit testing used to investigate ethnic discrimination in the labour market, and we will concentrate on this area in the subsequent sections.
6. To put it more technically, the net discrimination rate is estimated by dividing the difference between the positive responses where only the majority candidate was preferred and the positive responses where only the minority applicant was preferred, with the sum of all positive responses. This is the most frequently used measure for discrimination in the literature.
7. The correct way to calculate the discrimination rate is an ongoing debate in the literature and has been most discussed problem in the aftermath of Heckman's critique. In this article, we will concentrate on his other objections to the testing approach, but see Pager (2003) and Pager and colleagues (2009) for excellent examples of which results from field experiments may be presented in easy interpretable ways which at the same time meet Heckmans' objections.
8. Switching from audits to correspondence testing also has ethical consequences: Completing the full employment process by which employers have to make a decision as to whether the fictitious in-person applicants should get a job offer or not, as is the case when conducting an audit study, is far more challenging than in correspondence testing, where actual interaction between individuals does not take place at all and were the process is finished at an earlier stage. The ethical problems attached to the audit method are, however, not of Heckman's concern.
9. Immigrants of Pakistani heritage comprise one of the largest immigrant communities in Norway and were among the first of the 'new' immigrants to arrive in the late 1960s (Brochmann & Kjeldstadli 2008). In 2009, the children of Pakistani immigrants furthermore constituted 34 per cent of the so-called 'second generation' in Norway (Olsen 2011), making this group suitable for a field experiment.

References

- Allasino, E, Reyneri, E, Venturini, A & Zincone, G 2004, *Labour market discrimination against migrant workers in Italy*, International Labour Office, Geneva.
- Arrow, KJ 1998, 'What has economics to say about racial discrimination?', *Journal of Economic Perspectives*, vol. 12, no. 2, pp. 91–100.
- Attström, K 2007, *Discrimination against native swedes of immigrant origin in access to employment. A research study on Stockholm, Malmö and Gothenburg, utilizing the ILO discrimination practice testing approach*, International Labour Office, Genève.
- Banton, M 1994, *Discrimination*, Open University Press, Buckingham.
- Banton, M 1997, 'The ethics of practice-testing', *New Community*, vol. 23, no. 3, pp. 413–420.
- Bertrand, M, Chugh, D & Mullainathan, S 2005, 'Implicit discrimination', *The American Economic Review*, vol. 95, no. 2, pp. 94–98.
- Bertrand, M & Mullainathan, S 2004, 'Are Emily and Greg Morge employable than Lakisha and Jamal? A field experiment on labor market discrimination', *The American Economic Review*, vol. 90, no. 4, pp. 991–1013.
- Blank, RM, Dabady, M & Citro, CF (eds) 2004, *Measuring racial discrimination. Panel on methods for assessing discrimination*, National Research Council, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, National Academies Press, Washington, DC.
- Bovenkerk, F 1992, *Testing discrimination in natural experiments. A manual for international comparative research on discrimination on the grounds of "race" and ethnic origin*, International Labour Office, Genève.
- Brochmann, G & Kjeldstadli, K 2008, *A history of immigration. The case of Norway 900–2000*, Universitetsforlaget, Oslo.
- Bursell, M 2007, 'What's in a name? A field experiment test for the existence of ethnic discrimination in the hiring process', *Working Paper 2007:7*, p. Stockholm University Linnaeus Centre for Integration Studies.
- Carlsson, M 2009, *Essays on discrimination in hiring*, Växjö University, Ph. D Thesis. Department of Economics and Statistics.
- Carlsson, M & Rooth, DO 2007, 'Evidence of ethnic discrimination in the Swedish labor market using experimental data', *Labour Economics*, vol. 14, pp. 716–729.
- Cediey, E & Feroni, 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*, International Labour Office, Geneva.
- Daniel, W 1968, *Racial discrimination in England*, Penguin Books, Middlesex.
- Darity, WA & Mason, PL 1998, 'Evidence on discrimination in employment: codes of color, codes of gender', *Journal of Economic Perspectives*, vol. 12, no. 2, pp. 63–90.
- Fiske, S 1998, 'Stereotyping, prejudice, and discrimination', in *The handbook of social psychology*, 4 edn, vol. 2, eds D Gilbert, S Fiske & G Lindzey, McGraw Hill, New York, pp. 357–411.
- Fix, M & Struyk, RJ (eds) 1993, *Clear and convincing evidence: measurement of discrimination in America*, The Urban Institute Press, Washington D. C.
- Heckman, J 1998, 'Detecting discrimination', *Journal of Economic Perspectives*, vol. 12, no. 2, pp. 101–116.
- Heckman, J & Siegelman, P 1993, 'The urban institute audit studies: their methods and findings', in *Clear and convincing evidence: measurement of discrimination in America*, eds M Fix & RJ Struyk, The Urban Institute Press, Washington D. C.
- Jackson, M 2009, 'Disadvantaged through discrimination? The role of employers in social stratification', *The British Journal of Sociology*, vol. 60, no. 4, pp. 669–692.
- Kaas, L & Manger, C 2012, 'Ethnic discrimination in Germany's labour market: a field experiment', *German Economic Review*, vol. 13, no. 1, pp. 1–20.
- Kirschenman, J & Neckerman, KM 1991, "We'd love to hire them, but...": the meaning of race to employers', in *The urban underclass*, eds C Jencks & PE Peterson, Brookings Institution, Washington.
- Ladd, HF 1998, 'Evidence on discrimination in mortgage lending', *Journal of Economic Perspectives*, vol. 12, no. 2, pp. 41–62.
- Loury, GC 1998, 'Discrimination in the post-civil rights era: beyond market interactions', *Journal of Economic Perspectives*, vol. 12, no. 2, pp. 117–126.
- Midtbøen, AH & Rogstad, J 2008, 'Diskriminerings art, omfang og årsaker', *Søkelys på arbeidslivet*, vol. 25, no. 3, pp. 417–429.
- Midtbøen, AH & Rogstad, J 2012, *Diskriminerings omfang og årsaker. Etniske minoriteters tilgang til norsk arbeidsliv*. Institutt for samfunnsforskning, ISF-rapport 2012:001, Oslo.
- Moss, P & Tilly, C 2001, *Stories employers tell. Race, skill, and hiring in America*, Russell Sage Foundation, New York.
- Neumark, D 2010, *Detecting discrimination in audit and correspondence testing in Working Paper 16448*, National Bureau of Economic Research, Cambridge.
- Nilsson, A 2006, *Gränsvakter: Tankestilar och sortering vid rekrytering av personal i sex kommuner*, CEIFO, Stockholm University, Doctoral dissertation. Stockholm.
- Olsen, B 2011, *Unge med innvandrerbakgrunn i arbeid og utdanning 2009*, Statistisk sentralbyrå, Oslo/Kongsvinger.
- Pager, D 2003, 'The mark of a criminal record', *American Journal of Sociology*, vol. 108, no. 5, pp. 937–975.
- Pager, D 2007, 'The use of field experiments for studies of employment discrimination: contributions, critiques, and directions for the future', *The ANNALS of the American Academy of Political and Social Science*, vol. 609, no. 1, pp. 104–133.

- Pager, D & Karafin, D 2009, 'Bayesian bigot? Statistical discrimination, stereotypes, and employer decision making', *The ANNALS of the American Academy of Political and Social Science*, vol. 621, pp. 70–93.
- Pager, D & Quillian, L 2005, 'Walking the talk? What employers say versus what they do', *American Sociological Review*, vol. 70, no. 3, pp. 355–380.
- Pager, D & Shepherd, H 2008, 'The sociology of discrimination: racial discrimination in employment, housing, credit and consumer markets', *Annual Review of Sociology*, vol. 34, pp. 181–209.
- Pager, D, Western, B & Bonikowski, B 2009, 'Discrimination in a low-wage labor market: a field experiment', *American Sociological Review*, vol. 74, pp. 777–799.
- Petersen, T 2008, 'Discrimination: conscious or nonconscious?', in *Social stratification: class, race and gender in sociological perspective*, ed. DB Grusky, Westview Press, Boulder, pp. 780–785.
- Petersen, T, Saporta, I & Seidel, M-DL 2000, 'Offering a job: meritocracy and social networks', *American Journal of Sociology*, vol. 106, no. 3, pp. 763–816.
- Quillian, L 2006, 'New approaches to understanding racial prejudice and discrimination', *Annual Review of Sociology*, vol. 32, pp. 299–328.
- Quillian, L 2008, 'Does unconscious racism exist?', *Social Psychology Quarterly*, vol. 71, no. 1, pp. 6–11.
- Quillian, L & Pager, D 2010, 'Estimating risk: stereotype amplification and the perceived risk of criminal victimization', *Social Psychology Quarterly*, vol. 73, no. 1, pp. 79–104.
- Reskin, BF 2000, 'Proximate causes of employment discrimination', *Contemporary Sociology*, vol. 29, no. 2, pp. 319–328.
- Reskin, BF 2008, 'Rethinking employment discrimination and its remedies', in *Social stratification: class, race and gender in sociological perspective*, ed. DB Grusky, Westview Press, Boulder, pp. 770–779.
- Riach, P & Rich, J 2002, 'Field experiments of discrimination in the market place', *The Economic Journal*, vol. 112, no. 483, pp. 480–518.
- Riach, P & Rich, J 2004, 'Deceptive field experiments of discrimination: are they ethical?', *Kyklos*, vol. 57, no. 3, pp. 457–470.
- Rogstad, J 1996, 'Etikk og metode—noen problemer i studiet av diskriminering', *Sosiologisk tidsskrift*, vol. 4, no. 3, pp. 219–228.
- Rogstad, J 2001, *Sist blant likemenn? Synlige minoriteter på arbeidsmarkedet*, Unipax, Oslo.
- Rogstad, J & Midtbøen, AH 2009, *Fra symptom til årsak. Metodiske utfordringer og forskningsetiske dilemmaer ved bruk av tester i studiet av diskriminering*, Institutt for samfunnsforskning., ISF-rapport 2009:002., Oslo.
- Rooth, D-O 2007, *Implicit discrimination in hiring: real world evidence*, in IZA DP#2764, IZA.
- Waldinger, R & Lichter, MI 2003, *How the other half works: immigration and the social organization of labor*, University of California Press, Berkeley.
- Yinger, J 1998, 'Evidence on discrimination in consumer markets', *Journal of Economic Perspectives*, vol. 12, no. 2, pp. 23–40.
- Zegers de Beijl, R (ed.) 1999, *Documenting discrimination against migrant workers in the labour market: a comparative study of four European countries*, International Labour Office, Geneva.