

IZA DP No. 4741

**Ethnic Discrimination in Germany's Labour Market:
A Field Experiment**

Leo Kaas
Christian Manger

February 2010

Ethnic Discrimination in Germany's Labour Market: A Field Experiment

Leo Kaas

*University of Konstanz
and IZA*

Christian Manger

University of Konstanz

Discussion Paper No. 4741
February 2010

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Ethnic Discrimination in Germany's Labour Market: A Field Experiment^{*}

This paper studies ethnic discrimination in Germany's labour market with a correspondence test. To each of 528 advertisements for student internships we send two similar applications, one with a Turkish-sounding and one with a German-sounding name. A German name raises the average probability of a callback by about 14 percent. Differential treatment is particularly strong and significant at smaller firms at which the applicant with the German name receives 24 percent more callbacks. Discrimination disappears when we restrict our sample to applications including reference letters which contain favourable information about the candidate's personality. We interpret this finding as evidence for statistical discrimination.

JEL Classification: C93, J71

Keywords: ethnic discrimination, hiring discrimination, correspondence test

Corresponding author:

Leo Kaas
Department of Economics
University of Konstanz
78457 Konstanz
Germany
E-mail: leo.kaas@uni-konstanz.de

^{*} We are grateful to Anja Ahtziger, Gerald Eisenkopf, Till Grossmass, Normann Lorenz, Anne Madden, and Thomas Hinz for helpful comments and remarks. We also thank Fabian Fink, Jutta Obenland and Christiane Ralf for research assistance.

1 Introduction

Discrimination against individuals with respect to ethnicity, gender, or religion has a wide impact on labour market outcomes, including job opportunities, promotions and earnings. The extent to which a society is plagued by discrimination is hard to measure, however. On the one hand, empirical studies based on field data can deliver measures for earnings inequality but they cannot unveil discriminatory practices in the hiring process, for example. Moreover, field data are not collected in a controlled environment, so that the researcher has typically much less information about worker characteristics than is available to the employing firm. Hence it is difficult to disentangle the effects of actual productivity differences from employer discrimination. On the other hand, laboratory experiments on discrimination can be conducted in fully controlled settings. What is measured there, however, is the behaviour of subjects in a sterile environment; how far the findings of such experiments extend to employer–worker interactions in real–world labour markets is unclear. Field experiments are a compromise between these approaches, combining the advantages of controlled experiments with a field context.¹ With regard to measuring hiring discrimination, the correspondence test method is a sensible way to measure the initial response of employers to varying characteristics of artificial applicants.

This paper describes a correspondence test conducted in the German labour market for student internships. We examine the hiring opportunities of individuals with a Turkish migration background. Germany has about 2.4 million persons with a Turkish ethnic background (2.9 percent of its population). Predominantly in the 1960s, migrants from Turkey came to Germany to enlarge its labour force. About forty years later, the children and grand–children of these workers, born and raised in Germany, represent a significant share of Germany’s workforce. Among other things, a full integration of these young women and men into the German labour market necessitates equal employment opportunities.

Our experiment generates a snapshot of ethnic discrimination in one particular subsection of the labour market. In particular, we send more than one thousand applications to firms that offer internships for students of economics and management

¹See Harrison and List (2004) for a survey on field experiments.

science. In practice such internships serve as an important prerequisites for access into regular jobs. Although they are not well paid, a student who has successfully completed an internship gains job experience and significantly improves his employment opportunities after graduation. Today, the completion of at least one internship is commonly expected and is often considered as a “foot in the door”.

To each of 528 job advertisements, we send two similar applications, one with a Turkish sounding name (“Fatih Yildiz” or “Serkan Sezer”) and one with a typical German name (“Dennis Langer” or “Tobias Hartmann”). Importantly, the name is the only distinguishing characteristic of the applicant with Turkish ethnical background. That is, all applicants have German citizenship and they were born and educated in Germany, and all of them specify “German” as their mother tongue. With that design, we are able to isolate the effect of ethnicity from possible language effects. We create two slightly different types of applications with similar grades, soft skills, and photographs. For every job vacancy, applicant names are randomly assigned to the two different applications. Furthermore, the amount of information provided by the students varies between the two different applications. In particular, one application type contains reference letters stating favourable information about the candidate’s personality traits such as conscientiousness and agreeableness. We use this variation to explore the effect of statistical discrimination versus taste-based discrimination.

The field experiment shows that an application with a German-sounding name is on average 14% more likely to receive a callback. Discrimination is more pronounced among smaller firms: firms with less than 50 employees give “Dennis” and “Tobias” about 24% more callbacks than “Fatih” and “Serkan”. We also find evidence that a reasonable fraction of the differential treatment can be attributed to statistical discrimination: while there is almost no difference in callback probabilities for the application that is equipped with personality information (37.4% with a German name vs. 36.9% with a Turkish name), the absence of such information in the other application gives rise to significant differences in callback probabilities (41.8% with a German and 32.5% with a Turkish name).

Our results can be compared with those from other field studies that explore ethnic discrimination in other countries.² Across these studies, the measured degree

²There are several field studies that test discrimination against other characteristics, e.g. Neu-

of differential treatment varies remarkably with the respective context. In the U.S. labour market, Bertrand and Mullainathan (2004) show that applications with White-sounding names receive 50% more callbacks for interviews than those with African-American-sounding names. They find that the racial gap is uniform across occupation, industry, and employer size. A similarly huge difference in callback rates is documented by Drydakis and Vlassis (2007) who analyze the labour market opportunities of Albanians in Greece. They find that Albanians not only face a 43% smaller chance of access to occupations, but they are also significantly less likely to be registered with insurance coverage. For Arabs in Sweden, Carlsson and Rooth (2007) find that every fourth employer discriminates against the minority. Wood et al. (2009) conduct a correspondence test in Britain, finding that there are considerable gaps in callbacks between whites and several different ethnical groups. For the German labour market, Goldberg et al. (1996) conducted various field experiments to analyze ethnic discrimination of migrants, also finding substantial differences in callback rates. However, the legal framework has changed since 1994 when their experiments were conducted. Further, Goldberg et al. analyze the situation of migrants, that is, workers that were born in Turkey and with Turkish mother tongue, whereas we focus on German citizens with a Turkish migrational background.

Compared to these other studies, the extent of discrimination in our experiment is comparatively small. There are at least two explanations for this phenomenon. First, our applications contain much more information than those in the studies cited above. In particular, in Germany it is common practice to submit not only a resume, but also copies of all school and university certificates; these certificates provide detailed hard-evidence information about various skills. Second, we focus on a high-skill segment of the labour market; it is unclear whether ethnic discrimination in Germany is stronger in other segments of the labour market.³ The rest of this

mark (1996), Goldin and Rouse (2000)) and Petit (2007) for gender, Banerjee et al. (2009) for caste and religious groups in India, Rooth (2007) for obesity and Weichselbaumer (2003) for sexual orientation.

³Carlsson and Rooth (2007) find large differences across occupations, with differences in callback rates varying from 10 percent (computer professionals) to over 100 percent (shop sales and cleaning). But even in their study, there are several high-skill segments of the labour market with much higher discrimination rates than in our study. Large callback differences across occupations are also observed by Wood et al. (2009).

paper is structured as follows. Section 2 describes the experimental design and Section 3 presents and discusses the results. Section 4 concludes.

2 Experimental Design

The Vacancies

The experiment focuses on a specific segment of the labour market, in particular the market for internships for students in economics and business. This restriction allows us to completely automate the application process by sending serial letters and to eliminate potential bias caused by individually written and adjusted applications. We also restrict our study to internships within Germany.

We only apply for internships with a duration ranging from 6 weeks to 6 months and consider all reasonable vacancies posted at large internet job sites (such as *monster.de* and *jobscout24.de*). The field experiment was conducted in two waves, the first one covering vacancies posted in December 2007 and January 2008, the second one considering positions posted in December 2008. Although all firms explicitly search for students in economics or business programmes, the internships are quite heterogeneous. This concerns employer characteristics (size, sector, location), the division within the firm (typically human resources, marketing, finance or controlling). Most of the vacancies are at firms with 500 or more employees. Large firms and banks are the most relevant employers for graduates of economics and business, and they are more likely to post their vacancies on large internet job sites. Further, there are only few vacancies from East Germany since most large corporations have their headquarters in West Germany.

Applications

All applicants are second-year students of age 21 or 22. Our applications are quite comprehensive compared to other field studies on hiring discrimination. In particular, each application contains a cover letter, a curriculum vitae, a high-school certificate and a certificate documenting university grades in the first year. In the German labour market, this amount of information is necessary to achieve a reasonable callback rate. In fact, most employers explicitly request copies of all these certificates. Omitting them would bias our results significantly since only quite

unattractive employers would respond to an application that contains only a resume.

We create two slightly different types of resumes, labeled type A and type B, such that we can send two application to each firm.⁴ Students of both types were born, raised and educated in West Germany, but in different regions, one in the state of North Rhine–Westphalia, the other one in Baden–Wuerttemberg. After graduating from school, both skip military service, work at a summer job and then attend different universities. Both aim at a bachelor’s degree in business economics. At the time of the application they are in their third semester and they are applying for an internship during their fourth semester. The school and university certificates document grades between “good” and “very good”, so that the students range in the top 25 percent of their peer groups.

Both types are fluent in English and they have basic knowledge of one further foreign language. Since applications are identical for applicants with a Turkish and a German name, the applicant with the Turkish name is a native German speaker and he also does not report any command of Turkish in his CV. Both types have reasonably developed computer skills. Moreover, both applicants state in there CV that they had two minor part–time jobs, but while type B provides two letters of reference from previous employers, type A does not add any related documents. The two reference letters contain positive statements about the student’s personality (affability, commitment, capacity for teamwork, conscientiousness). This variation in information is used to analyze the effects of statistical discrimination. In all other dimensions, applications of types A and B are rather similar; particularly, there are only minor deviations in individual school and university grades. Hence, the variation in information about personality is the decisive informational difference between these types.

Finally, all applications are completed by a type–specific photograph. While in many countries firms do not request or even oppose photographs in applications, they are still very common (and sometimes requested) in Germany. Omitting them would again bias the results. We select photographs that fit both a native German student as well as one with a Turkish migration background. Each resume type has its own unique photograph, while names are randomly assigned to the types. This

⁴Detailed CVs of both types are available upon request.

guarantees that the choice of the photograph has no overall effect on the callback rates of an individual name.

We systematically adjust certain details in the cover letter of every single application to match job-specific features relating to the sector of the firm or the division of the internship. This adjustment is performed automatically using serial letters. That is, we design for each type and each division a specific paragraph matching the basic requirements and interest for that division. For example, when applying for an internship in the human resource division, an applicant of type A would explain why he is interested in human resource management and he also states that he intends to pursue a master’s degree in human resource management after finishing his bachelor degree. Our approach to standardize serial letters reflects the trade-off between maximizing the callback rate by adjusting the application to the specific requirements of each post and generating unbiased letters (and also reducing our workload).

The Names

The first application that is supposed to be sent to each individual firm is randomly assigned a type (A or B) and a name (German or Turkish), while the second application then is assigned the complementary type and name.⁵ We choose “Dennis Langer” (first wave) and “Tobias Hartmann” (second wave) as names for the native German candidate. The first names as well as the surnames belong to the 30 most common ones for the birth years 1986 to 1988 in Germany. The name of the applicant from the ethnic minority is “Fatih Yildiz” in the first wave and “Serkan Sezer” in the second one. Both, first names and surnames, are common for male descendants of Turkish immigrants in Germany. It is also evident for every human resource manager to deduce the ethnic background from these names. We did not explicitly check for connotations of names regarding their social background, but we assured that the names do not contradict common sense, are very stereotype or exhibit other peculiarities (e.g. ruling out combinations between an Anatolian first name and a Kurdish surname). We also made sure that none of these names corresponds to a real person in Germany’s student web network (*studivz.de*).

⁵In particular, we simulate an urn model to determine the type and name of the application that is sent first. All four possible combinations of types and names are used equally often.

The application process

We create an individual e-mail address for each name and prepare mobile phones with name-specific numbers. However, we do not answer incoming calls directly, but firms are redirected to the voice mail where they are politely asked to leave their names and contact information. Additionally, we made arrangements such that answers by regular mail to both candidates and addresses were redirected to us. Thus firms could contact the applicants via mail, e-mail and phone.

When applying, all documents (cover letters, CVs, additional documents such as certificates) are automatically merged to one single pdf file and are e-mailed to the firm. After sending the first application we waited two days before sending the second one. A reasonable fraction of firms (especially larger ones that expect to receive many applications) do not accept applications by e-mail, but instead require the applicants to complete several pre-defined online forms. In these cases, the forms are filled out with the respective applicant's information and our documents were attached as pdf files whenever possible.

After applying for the vacancy we registered callbacks in the subsequent four months. A callback is defined as any action of a firm that signals interest in the respective resume, including offers for interviews, direct job offers and leaving contact information on the voice mail. In contrast, automatic responses confirming the receipt of our applications are not considered as callbacks, as well as written requests for additional information which were answered whenever possible. For every reaction of a firm, be it a callback, rejecting the applicant, or a request for more information, we collect the date and the type of reaction. Within 24 hours of that reaction, we politely withdraw the candidate's further interest in the position.

3 Results

The application process at each of the 528 firms can have several outcomes, summarized in Table 1. Either the firm shows no positive reaction towards any of the candidates (column 1), or at least one applicant receives a callback (column 2). In the latter case, either both receive a callback (column 3), or the firm prefers one of the students, either the one with the German (column 4) or with the Turkish name (column 5). Column (6) calculates net discrimination as the difference in call-

backs between applications with a German and with a Turkish name, expressed as a percentage of those observations where at least one candidate received a callback. This definition of net discrimination treats those cases where no candidate receives a callback as a non-observation. Riach and Rich (2002) discuss whether a negative answer (or no answer at all) for both candidates should be considered as equal treatment or as a non-observation. On the one hand, if a firm rejects both applicants (or does not even send an answer) this could be considered equal treatment; that is, somebody reviewed both applications and found them not suitable for the job. On the other hand, it is also conceivable that the firm was not even considering the applications, for instance because the vacancy has already been filled.

We conduct a standard χ^2 test of the hypothesis that the two possible outcomes of unequal treatment (that is, columns (4) and (5)) are equally likely.⁶ If the application with the German name is preferred significantly more often than the Turkish one, the H_0 of equal treatment is rejected. As this test considers only observations with differential treatment, observations of firms that either decline both applicants or callback both applicants are irrelevant.

The first row shows the aggregate results of the field study. Out of the 258 firms that accepted at least one application, 29.1% contacted only the German, and 19.0% only the Turkish candidate, while 51.9% contacted both. This corresponds to a callback rate of 34.7% for the Turkish student and of 39.6% for the German student. In other words, while the German candidate has to write 15 applications to obtain 6 callbacks, the Turkish candidate must send 17 applications for the same number of callbacks. This difference is significant at the 5%-level, but it is remarkably small compared to studies on employment discrimination of ethnic minorities in other countries, such as Albanians in Greece (Drydakis and Vlassis 2007), Arabs in Sweden (Carlsson and Rooth 2007) or Afro-Americans in the U.S. (Bertrand and Mullainathan 2004); see Section 4 for further discussion.

Regarding the effect of firm characteristics on discrimination, we note that only four (8.9%) out of the 45 firms with less than 50 employees made an offer only to the Turkish candidate, while 12 (26.7%) preferred the German over the Turkish

⁶The result is the same as the χ^2 of a McNemar test which considers the null hypothesis that a dichotomous and paired outcome variable (i.e. the reaction dummies to German and Turkish candidates) have the same distributions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	no callback	at least one	both	only G	only T	net discr.	χ^2 test
all firms (528)	51.14 (270)	48.86 (258)	51.94 (134)	29.07 (75)	18.99 (49)	10.08 (26)	5.45**
Firm Size							
large (> 500) (379)	51.19 (194)	48.81 (185)	49.19 (91)	29.73 (55)	21.08 (39)	8.65 (16)	2.72*
medium (59)	52.54 (31)	47.46 (28)	50.00 (14)	28.57 (8)	21.43 (6)	7.14 (2)	0.29
small (< 50) (90)	50.00 (45)	50.00 (45)	64.44 (29)	26.67 (12)	8.89 (4)	17.78 (8)	4.00**
Location							
South (292)	48.29 (141)	51.71 (151)	52.32 (79)	29.14 (44)	18.54 (28)	10.60 (16)	3.56*
East (40)	62.50 (25)	37.50 (15)	40.00 (6)	40.00 (6)	20.00 (3)	20.00 (3)	1.00
other (196)	53.06 (104)	46.94 (92)	53.26 (49)	27.17 (25)	19.57 (18)	7.60 (7)	1.14
Division							
Marketing (173)	43.35 (75)	56.65 (98)	54.08 (53)	29.59 (29)	16.33 (16)	13.26 (13)	3.76*
Controlling (91)	52.75 (48)	47.25 (43)	46.51 (20)	25.58 (11)	27.91 (12)	-2.33 (-1)	0.04
Finance (84)	48.81 (41)	51.19 (43)	46.51 (20)	32.56 (14)	20.93 (9)	11.63 (5)	1.09
Human Res. (122)	59.84 (73)	40.16 (49)	51.02 (25)	32.65 (16)	16.33 (8)	16.33 (8)	2.67
Consulting (32)	65.63 (21)	34.37 (11)	72.73 (8)	9.09 (1)	18.18 (2)	9.09 (-1)	0.33
other (26)	46.15 (12)	53.85 (14)	57.14 (8)	28.57 (4)	14.29 (2)	14.29 (2)	0.67
Industry							
Fin. Services (105)	41.90 (44)	58.10 (61)	40.98 (25)	34.43 (21)	24.59 (15)	14.75 (6)	1.00
Consulting (98)	56.12 (55)	43.88 (43)	69.77 (30)	16.28 (7)	13.95 (6)	2.33 (1)	0.08
Manufacturing (150)	55.33 (83)	44.67 (67)	56.72 (38)	26.87 (18)	16.42 (11)	10.45 (7)	1.69
IT & Telecom (31)	58.06 (18)	41.94 (13)	69.23 (9)	23.08 (3)	7.69 (1)	15.38 (2)	1.00
Public Services (11)	63.64 (7)	36.36 (4)	25.00 (1)	50.00 (2)	25.00 (1)	25.00 (1)	0.33
other (133)	47.37 (63)	52.63 (70)	44.29 (31)	34.29 (24)	21.43 (15)	12.86 (9)	2.08

Notes: This table shows the distribution of the firm responses, absolute numbers are in parentheses. Column (1) reports the fraction of firms that gave none of the candidates a callback, so the remainder in column (2) contacted at least one applicant. Firms that gave both candidates a positive reaction, column (3), are considered as equal treatment, while the rest preferred either the candidate with the German or the one with the Turkish name, columns (4) and (5). Net discrimination is calculated as (6)=(4)-(5). Column (7) contains the χ^2 for equality between (4) and (5) (H_0 : Turkish and German candidates are equally likely to receive a callback at any of the paired observations). * denotes significance at the 10%-level, ** significance at the 5%-level.

Table 1: Callbacks conditional on names and firm characteristics.

applicant. We would assume that smaller firms with fewer vacancies have a less standardized recruitment process. This leaves more scope for individual preferences of the human resource manager to influence hiring decisions. Indeed, discrimination is less prominent in larger firms, presumably since their recruitment processes follow pre-defined rules.

However, differences with respect to other firm characteristics are rather limited. We note that there are twice as many firms in East Germany⁷ that favour the candidate with the German name; however, there are both too few vacancies as well as too low callback rates to give rise to significant results. If we consider jobs in different divisions, differential treatment is weakly significant for jobs in the marketing department. But discrimination is not particularly strong in this division; instead there are simply many internships in marketing divisions, which implies that only the large sample is responsible for significant discrimination here. On the other hand, differences in callback probabilities are remarkably strong for internships in human resource departments, where the number of employers which favour the German candidate is twice as large as the number of employers preferring the Turkish one. Strong discrimination in human resource departments could be considered to support taste-biased discrimination according to Becker (1957): managers refuse to hire minority workers either because they do not wish to work with them personally (due to their own discriminatory taste), or because they fear that discriminatory tastes among co-workers or outsiders impairs productive efficiency in a division with high exposures to outsiders and intense team interactions. However, there are too few callbacks to provide statistically significant evidence for this hypothesis.

Table 2 shows the callback rates for the different names conditional on the resume type. For applications of type B which include reference letters containing information regarding the candidate's personality, Turkish and German applicants achieve almost identical callback rates. However, for applications of type A (without personality information), the minority student receives only for 32.5% of his applications a callback, while the German student is successful in 41.8% of his applications. The difference in the number of callbacks between German and Turkish students of type

⁷“East Germany” is defined as the states Berlin, Brandenburg, Mecklenburg–Western Pomerania, Saxony, Saxony–Anhalt and Thuringia. “South Germany” includes the states Baden–Wuerttemberg, Bavaria, Hesse, Rhineland–Palatinate, and Saarland.

A is significant at the 5%-level, while it is not significant for students of type B. We cautiously interpret this as evidence for statistical discrimination (Arrow 1973): the difference in callbacks decreases with the provision of information about the applicant’s character. Note however that “information” and any other type characteristics are perfectly correlated. Hence the observation that callback differences are smaller at applications of type B could, in principle, also be due to other characteristics of a type B application. However, it is not true that type B is generally more attractive: in contrast, German applicants of type B receive *fewer* callbacks than German applicants of type A. More importantly, as we have argued above, the other differences between type A and type B applicants are rather minor and they should not be expected to generate the large observed difference in discrimination rates.⁸ It is worth to note that there is no evidence that firms from different regions favour a certain type, although both types come from different regions within Germany.

callback rate	type A	type B
German name	41.8% (110 out of 263 applications)	37.4% (99 out of 265 applications)
Turkish name	32.5% (86 out of 265 applications)	36.9% (97 out of 263 applications)

Notes: Applications of type B contain two reference letters with information about the applicant’s personality, those of type A do not.

Table 2: Callback rates for different types.

Probit Estimation

In order to disentangle the effects of different employer and worker characteristics, we conduct several probit estimations with the callback dummy as the dependent variable. Table 3 summarizes the various configurations. Columns (1) to (5) show regression outputs using the full sample. In the basic model (1) we regress the callback only on a constant and a dummy for the Turkish name and find a weakly

⁸Although the universities of both types are comparable in their academic standards within Germany, the university of type A offers a more business-oriented program which could explain why such a candidate is a bit more attractive for most employers. Nonetheless, it is rather implausible that this feature explains the strong difference in callback rates between ethnical types.

significant (at the 10 percent level) negative effect of the Turkish name on the callback probability. The interpretation of the result is that a Turkish name reduces the probability of a callback on average by about 14%. This effect is robust to adding a dummy for the location of the firm and the workers's type in (2), a dummy for firms with less than 50 employees in (3) and dummies for the sector of the firm as well as for the division of the internship in (4).⁹ The coefficient of the dummy for South Germany is (weakly) significantly positive, suggesting that it is easier to find an internship in regions with less unemployment and more job openings. There is no evidence that one of the two application types is more successful than the other in general. That is, the type dummy is always small and insignificant. However, once we add an interaction term between the name (ethnicity) and the type dummies in (5), this term catches all the disadvantage of the Turkish name, and the coefficient of the name dummy becomes small and insignificant. Again this confirms the conclusion from Table 2 that the difference in callbacks occurs predominantly at applicants of type A where personality information is not provided. The reduction in information seems to hurt applicants with a Turkish name. Again we cautiously interpret this finding as evidence in support of statistical discrimination.

Other means of discrimination

Discrimination can manifest itself in several ways, not only in different callback rates. If a firm has no interest in any of the applicants, but the applicant with the German name receives a polite message declining him, while an applicant with a Turkish name is just ignored, we would consider this as discriminatory treatment, even though it takes no direct impact on the job search outcome.

Table 4 provides a first snapshot. The most noticeable difference in treatment of the two applicants takes place at firms that callback one candidate and do not respond to the second one; indeed, 28 firms showed interest in the applicant with the German name and ignored the Turkish applicant, while only 12 firms contacted the Turkish and ignored the German one. We would consider this as the strongest form of discrimination: the firm has a vacant post, it shows interest in the German

⁹We tested several other variables but found all of them (and their interactions with the ethnicity dummy) insignificant. This includes a "first application" dummy, a dummy on "top employers" (obtained from a student survey), and a dummy for employers located in a large city region.

Callback	(1)	(2)	(3)	(4)	(5)
constant	-0.26*** (0.06)	-0.34*** (0.08)	-0.36*** (0.08)	-0.04 (0.22)	-0.1 (0.22)
Turkish name	-0.13* (0.08)	-0.13* (0.08)	-0.13* (0.08)	-0.13* (0.08)	0.003 (0.11)
Southern Germany		0.13* (0.08)	0.14* (0.08)	0.14* (0.08)	0.13 (0.08)
type A		-0.0001 (0.08)	-0.0001 (0.08)	0.002 (0.08)	0.14 (0.11)
small firm			0.13 (0.10)	0.14 (0.12)	0.14 (0.12)
sector dummies				yes	yes
division dummies				yes	yes
Turk*Type A					-0.27* (0.16)
Observations	1056	1056	1056	1056	1056

Notes: Each column represents a probit regression with the callback dummy as dependent variable. Robust standard errors are in parentheses. * denotes significance at the 10%-level, ** significance at the 5%-level, and *** at the 1%-level. Several other interactions have been tested and not found significant.

Table 3: Probit regression with callback dummy as dependent variable.

candidate and does not even answer the Turkish one. The difference in politeness can also be illustrated by the following comparison: 74.3% of the employers that show no interest in the German applicant nevertheless send him a polite rejection message, while an applicant with a Turkish name receives such a message only from 71.3% of the firms. However, this difference is not statistically significant.

German / Turkish	callback	rejection	no reaction	Σ
callback	134	47	28	209
rejection	37	179	21	237
no reaction	12	20	50	82
Σ	183	246	99	528

Table 4: Differential treatment in answers.

As another type of discrimination, we check how long applicants with different names have to wait for the firm’s decision, and what determines that waiting period. Most firms react only a few workdays after we sent the application (see Table 5). A callback is, on average, received after 11 workdays, while a rejection takes on average 17 workdays. An applicant with a Turkish name has to wait slightly longer for a

callback (11.3 workdays with a Turkish and 10.7 workdays with a German name), but this difference is not statistically significant. For rejections, the difference is even smaller (17.39 workdays with a Turkish and 17.36 workdays with a German name). There are two peculiarities, however. First, small firms react faster in general. Second, the difference in reaction times between German and Turkish applicants is larger at smaller firms. For example, to decline a Turkish applicant, a small firm needs 2.3 workdays more than to decline a German applicant. We would assume that small firms are faster in general because there is often just one decision maker, whereas larger firms are more likely to have standardized recruitment processes where applications have to go through many hands.

	callback			rejection		
	German name	Turkish name	average	German name	Turkish name	average
small	6.1	7.0	6.5	9.9	12.2	11.0
medium	7.8	7.5	7.7	19.6	19.0	19.3
large	12.4	13.0	12.7	18.1	17.8	17.9

Table 5: Average reaction time in workdays.

We also conduct a multinomial logit with the outcomes 1 (callback), 2 (applicant receives a rejection) versus 0 (no reaction) for every application and every workday, beginning at the application day until the workday of the firm’s reaction. We complemented the controls of the previous probit estimation by the number of workdays that have passed without a reaction (t and t^2) and by an interaction term between t and the firm-size dummy (see Table 6). The main results of the probit estimate can also be observed in the multinomial logit: applicants with a Turkish name are less likely to receive a callback, and firms in South Germany are more likely to give a positive response. In contrast, there are no such effects for rejections. That is, applications from the Turkish applicant do not receive significantly less (or more) rejections than candidates with a German name. Table 6 also confirms that large firms take more time to answer an application. We also checked other interaction terms (in particular those involving the name dummy), but found all of them to be insignificant.

To complement this survival analysis, we show in Figure 1 the distribution and

reaction	1	2
constant	-2.59*** (0.16)	-3.48*** (0.17)
t	-0.11*** (0.14)	-0.03*** (0.008)
t^2	-0.00008 (0.0001)	-0.00025*** (0.00008)
Turkish name	-0.187* (0.103)	-0.03 (0.09)
largefirm	-0.76*** (0.16)	0.036 (0.156)
$t \cdot$ largefirm	0.068*** (0.015)	0.019*** (0.007)
typeA	-0.018 (0.102)	0.022 (0.092)
South	0.187* (0.104)	0.038 (0.093)
Sample	full	full
Observations	1056	1056

Notes: Multinomial logit with the outcomes 1: callback, 2: rejection (vs 0: no reaction at workday t). Standard errors are in parentheses. * denotes significance at the 10%-level, ** significance at the 5%-level, and *** at the 1%-level. Several other interactions have been tested and not found significant.

Table 6: Multinomial Logit.

survival functions for our total sample. We do not show the differences between German and Turkish candidates since they are rather tiny, especially for the callback distribution. The figure shows that the probability to receive an answer (callback or rejection) is significantly declining over time and is nearly zero after about 30–40 workdays. Before that, the probability declines about linearly for callbacks, but non-linearly for rejections. This pattern explains why the coefficient for t^2 in column 2 of Table 6 is significantly negative.

The role of the business cycle

Our field study also permits to explore the impact of the business cycle. We conducted the study in two waves, the first one in Winter 2007/08 and the second one in Winter 2008/09. The macroeconomic situation changed substantially between these two dates. In particular, the accelerating financial crisis strongly affected the real economy in late 2008. By then, most research institutes published forecasts for

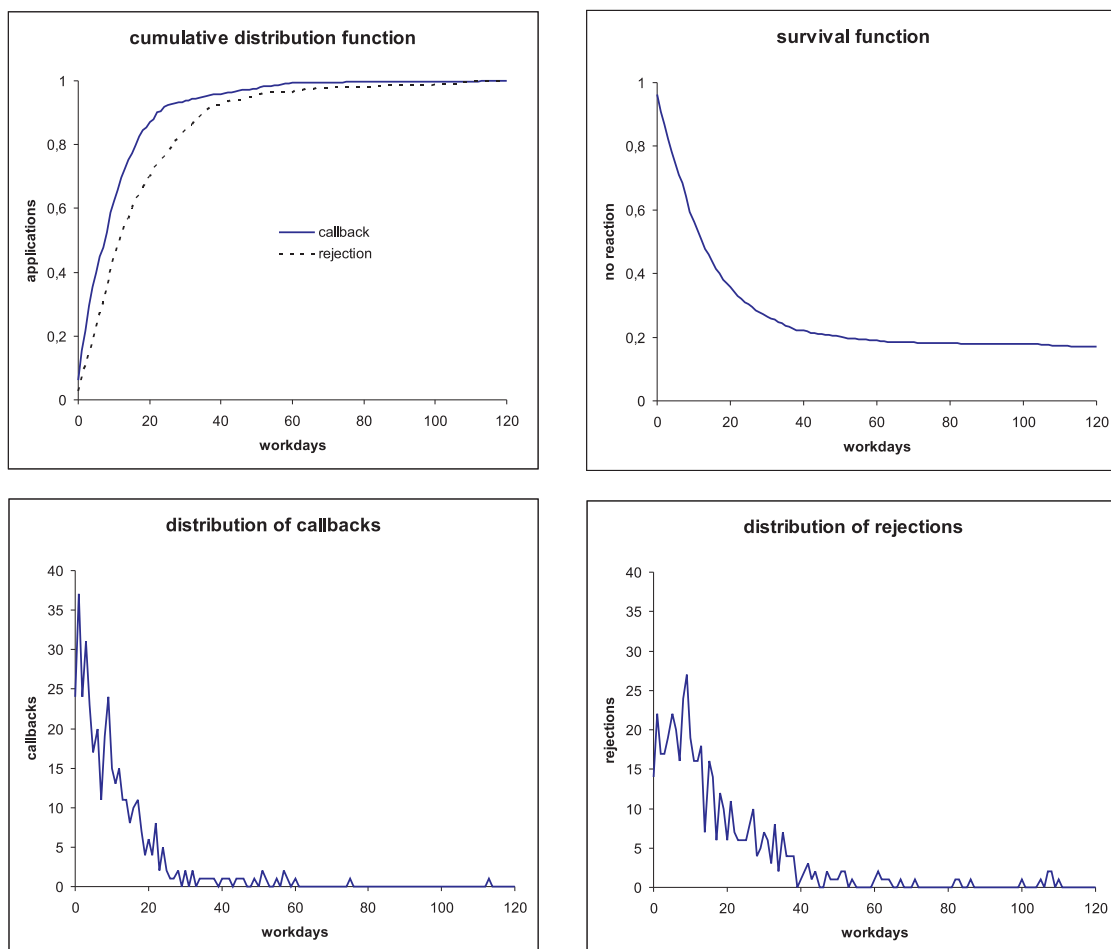


Figure 1: Response time of firms for callbacks and rejections.

Germany's GDP growth in 2009 at -3 percent or less. Although the German labour market remained remarkably robust during the course of 2009, we would expect that firms adjusted their hiring behaviour in January 2009 relative to the year before. However, the aggregate callback rates (as well as the name-specific ones) were relatively stable over time (38.8% in 2007/08 versus 35.4% in 2009). It seems that two effects offset each other: as global demand decreases, firms should be expected to hire fewer workers, but they could also try to replace regular employees with cheaper interns.

4 Conclusions

We conducted a correspondence test, sending more than a thousand applications with randomly assigned German and Turkish names to firms advertising student internships. The difference in callbacks is significant but, compared to similar studies for other countries, relatively small. Several explanations can account for this result. First, we focus on a specific high-skill segment of the labour market. If competition for qualified students is intense, discriminating firms cannot survive the “war for talents” and are driven out of the market. On the other hand, in labour market segments with an excess supply of qualified workers, discrimination should be stronger as firms can choose their favourite candidate among a large number of applicants. Moreover, in our experiment the student with the Turkish name is a German citizen with a migration background. He was born and raised in Germany, went to school in Germany and is now studying in Germany. We focus on these second and third generations of immigrants as they represent the largest ethnic minority in the German labour market. The fact that all applicants are observationally equivalent (except their name) permits us to isolate the name effect from any language effects. Conducting a similar study with non-German citizens with a mother tongue other than German should be expected to produce a larger difference in callback rates. Both applications were endowed with very good grades and interesting enough CVs so as to guarantee a reasonably high callback rate. It is conceivable that net discrimination is substantially larger for candidates with mediocre grades; for such applications negative stereotypes of human resource managers (such as “students with Turkish migration background are under-performing”) become potentially more important. On the other hand, the rather good grades can be inconsistent with employers’ expectations about a Turkish applicant. Hence such an observation is more salient and takes a stronger impact on impression formation than it would do for a German candidate (see e.g. Sherman, Stroessner, Loftus, and Deguzmani (1997)). We would further expect that most firms with a standardized recruitment procedure use a threshold strategy. Any candidate that fulfills certain minimal criteria receives a callback, and these criteria apply equally for all candidates. Since larger firms have more often a standardized recruitment process, they also discriminate less than smaller firms.

This study gives only a first insight into the extent of ethnical discrimination in Germany’s labour market. There are several further questions that should be explored in further research. First, measuring hiring discrimination in different segments of the labour market would provide more information on the effects of sector and firm characteristics. Second, varying the quality of the applications would also permit to measure group-specific returns to skills, as in Bertrand and Mullainathan (2004). Third, our experiment shows that provision of information about personality reduces the extent of discrimination. In many countries it is uncommon to use references at early stages of the recruitment process.¹⁰ Our result suggests that such conventions can potentially backfire on minority employees. Future experiments, also those conducted in the laboratory, should further illuminate the role of information about personality on recruitment decisions.

References

- ARROW, K. J. (1973): “The Theory of Discrimination,” in *Discrimination in Labor Markets*, ed. by O. Ashenfelter, and A. Rees. Princeton University Press.
- BANERJEE, A., M. BERTRAND, S. DATTA, AND S. MULLAINATHAN (2009): “Labor Market Discrimination in Delhi: Evidence from a Field Experiment,” *Journal of Comparative Economics*, 37, 14–27.
- BECKER, G. (1957): *The Economics of Discrimination*. The University of Chicago Press.
- BERTRAND, M., AND S. MULLAINATHAN (2004): “Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination,” *American Economic Review*, 94, 991–1013.
- CARLSSON, M., AND D.-O. ROTH (2007): “Evidence of Ethnic Discrimination in the Swedish Labor Market Using Experimental Data,” *Labour Economics*, 14, 716–729.

¹⁰In the UK, the “Employment Practices Data Protection Code” states that employers should only carry out pre-employment vetting (e.g. reference letters) on an applicant at a late stage in the recruitment process.

- DRYDAKIS, N., AND M. G. VLASSIS (2007): “Ethnic Discrimination in the Greek Labour Market: Occupational Access, Insurance Coverage, and Wage Offers,” Working Paper No. 0715, University of Crete, Department of Economics.
- GOLDBERG, A., D. MOURINHO, AND U. KULKE (1996): “Labour Market Discrimination against Foreign Workers in Germany,” International Migration Papers 7, International Labour Office, Geneva.
- GOLDIN, C., AND C. ROUSE (2000): “Orchestrating Impartiality: The Impact of Blind Auditions on Female Musicians,” *American Economic Review*, 90, 715–741.
- HARRISON, G. W., AND J. A. LIST (2004): “Field Experiments,” *Journal of Economic Literature*, 42, 1009–1055.
- NEUMARK, D. (1996): “Sex Discrimination in Restaurant Hiring: An Audit Study,” *Quarterly Journal of Economics*, 111, 915–942.
- PETIT, P. (2007): “The Effects of Age and Family Constraints on Gender Hiring Discrimination: a Field Experiment in the French Financial Sector,” *Labour Economics*, 14, 371–391.
- RIACH, P. A., AND J. RICH (2002): “Field Experiments of Discrimination in the Market Place,” *The Economic Journal*, 112, F480–F518.
- ROOTH, D.-O. (2007): “Evidence of Unequal Treatment in Hiring against Obese Applicants: A Field Experiment,” IZA Working Paper No. 2775.
- SHERMAN, J., S. STROESSNER, S. LOFTUS, AND G. DEGUZMANI (1997): “Stereotype suppression and recognition memory for stereotypical and nonstereotypical information,” *Social Cognition*, 15, 205–215.
- WEICHELBAUMER, D. (2003): “Sexual Orientation Discrimination in Hiring,” *Labour Economics*, 10, 629–642.
- WOOD, M., J. HALES, S. PURDON, T. SEJERSEN, AND O. HAYLLAR (2009): “A Test for Racial Discrimination in Recruitment Practice in British Cities,” Department for Work and Pensions Research Report No. 607.