

Are Some Citizens More Equal than Others? Evidence from a Field Experiment on Discrimination by Public Bodies in Germany

Abstract

The equal treatment of all citizens by public institutions is a key element of democratic rule of law. Nevertheless, there is a growing number of media and scientific reports on unequal treatment by public administrations. The article examines the unequal treatment of citizens by sex and ethnic origin by means of a survey based field experiment in German local government. With the help of two vignettes and a randomized assignment of names, responses to fake citizen requests by local governments are analyzed for speed, quality and customer orientation. Our findings point to mixed results. While there is no evidence for general ethnic discrimination, a more differentiated analysis indicates patterns of ethnic discrimination conditioned by gender.

Key Words: Discrimination, Local Government, Experiment, Gender, Migration

This is the peer reviewed version of the following article: Are Some Citizens More Equal than Others? Evidence from a Field Experiment on Discrimination by Public Bodies in Germany, which has been published in final form at <https://doi.org/10.1111/puar.12439>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions. This article may not be enhanced, enriched or otherwise transformed into a derivative work, without express permission from Wiley or by statutory rights under applicable legislation. Copyright notices must not be removed, obscured or modified. The article must be linked to Wiley's version of record on Wiley Online Library and any embedding, framing or otherwise making available the article or pages thereof by third parties from platforms, services and websites other than Wiley Online Library must be prohibited.

Information on the authors

Dr. Stephan Grohs
Chair of Comparative Public Policy and Administration
Department of Politics and Management
University of Konstanz
P.O. Box D 91
D-78457 Konstanz (Germany)
Tel ++49 (0)7531 882756
Fax ++49 (0)7531 882381
Email: stephan.grohs@uni-konstanz.de

Dr. Christian Adam
Chair for Empirical Theories of Politics
Ludwig-Maximilians-University München
Geschwister-Scholl-Institute for Political Science
Oettingenstraße 67
80538 München(Germany)
Tel.: ++49 (0) 89 / 2180-9025
Email: christian.adam@gsi.uni-muenchen.de

Prof. Dr. Christoph Knill
Chair for Empirical Theories of Politics
Ludwig-Maximilians-University München
Geschwister-Scholl-Institute for Political Science
Oettingenstraße 67
80538 München
Tel.:+49 (0) 89 / 2180-9061
Email: christoph.knill@gsi.uni-muenchen.de

Are Some Citizens More Equal than Others? Evidence from a Field Experiment on Discrimination by Public Bodies in Germany

Introduction

One of the basic principles of good administrative practice is the equal treatment of all citizens. Especially in countries in which public administration is based on the strong adherence to legal rules (*Rechtsstaat* tradition), adherence to the principle of equality constitutes a central feature affecting the legality of administrative behavior. According to the general principle of equality stipulated in Article 3 of German Basic Law, for instance, public administration is committed to equal treatment irrespective of gender, descent, race, language, home country, creed, as well as religious or political convictions. Nevertheless, there are reports indicating potential violations of these requirements. In an expert study for the German Federal Anti-Discrimination Agency, over 20 per cent of the surveyed persons with an immigration background reported discrimination by public agencies and authorities – a much higher share than cases reported by those searching for an apartment on the private market for example (Sachverständigenrat 2012, 11). Similarly, public administration studies in the US context find substantial evidence of violations of non-discrimination rules both against citizens as well as in personnel recruitment (e.g. Cayer and Sigelman 1980; Kellough 1990; Llorens et al. 2007).

Hence, discrimination seems to be a ubiquitous problem in the public sector. This is surprising, as for public administration – in contrast to the private sector – there are no obvious benefits associated with institutional discrimination. For public institutions, we can to a much lesser extent assume a direct correlation between the assessment of individuals and

the achievement of organizational goals, as this might be the case for private sector employment or housing markets, for instance. For the performance of a public administration, it basically makes no difference whether to give information to residents or to foreigners, to deny a building permit, or to assign a penalty notice. Institutional discrimination should hence hardly be a feature of public organizations. If discrimination through public administration occurs, this should hence be the result of individual behavior of public officials rather than organizational features.

Yet, and in contrast to the US, there is still no systematic evidence on the prevalence of individual discrimination in the German public sector. While its law-led administrative tradition should prevent discriminatory behavior, the fact that Germany faces a history of ignorance of its de-facto role as an immigration country and the long-term denial of citizenship to immigrants might point into an opposite direction.

To address this question, we carried out a field experiment in which – similar to studies on ethnic discrimination in the labor and housing market - different names linked to ethnicity and gender) were used as a stimulus to test the discrimination hypothesis in public administration under realistic conditions. Using two different requests and a randomized allocation of names, responses to citizen requests in two policy areas were systematically recorded based on speed, quality and service orientation. Our findings point to mixed results. While there is no evidence for general ethnic discrimination, a more differentiated analysis indicates patterns of ethnic discrimination conditioned by gender.

After a short discussion of the situation of immigrants in the German administrative culture and an overview of the state-of-the-art, we present our experimental research design, and elaborate on our results.

Social Equity and Institutional Discrimination: Germany between rule of law and non-immigration legacies

Nondiscrimination is a central value of public administration. In the law-oriented administrative culture of continental Europe, nondiscrimination is primarily discussed as fundamental right. In the US and other Anglo-Saxon countries nondiscrimination also is related to equity as a normative foundation of democratic public administration.

Accordingly, records of discrimination strike at the very heart of its normative roots. The so-called “formalistic impersonality” was already a core element of Weber’s theory of bureaucracy: “‘Sine ira et studio’, without hatred or passion, and hence without affection or enthusiasm. The dominant norms are concepts of straightforward duty without regard to the personal considerations. Everyone is subject to formal equality of treatment; that is everyone in the same empirical situation. This is the spirit in which the ideal official conducts his office” (Weber 1968, 225). The founding father of “New Public Administration”, H. George Frederickson declared equity as a central dimension of performance in administrative behavior along with efficiency, economy and effectiveness (Frederickson 1990). Social equity encompasses the criteria of fairness, justice and equality (Frederickson 2010: xiv), which exclude any kind of discrimination.

While the German debate concentrates on the juridical nature of discrimination and is based on the *Rechtsstaat* principle, the equity principle as a societal democratic value is much less fundamental. In contrast, the US debate is more coined by these latter cultural and societal values. Of course, anti-discrimination legislation is also very important in the US (Alba 2005). Nevertheless, cultural norms of fair access to economic and social opportunities, as expressed in the “American Creed” (Merton 1949), are more formative than law. Equity as the right of equal access to public bodies, irrespective of race, religion or ethnic origin, coin the understanding of administrative action – despite all evidence of a gap between creed and

conduct which was addressed as the "American dilemma" by Myrdal already in the 1940s (Myrdal 1944).

Interestingly, in the German case there is a conflict between the legally embodied equality principle and dominant cultural and societal norms. This can be traced to the fact that Germany became a target country for immigration rather late. The so-called "guest workers" attracted by the post-war boom formed the first generation of immigrants after the post-war resettlement in the 1960s and 1970s. These low skilled workers from Southern Europe and Turkey were deemed to leave the country after a certain period of time for which they were hired to support the booming economy with cheap labor. The aspiration of a temporary stay failed, however. Today the third generation of ancestors of these immigrants lives in Germany. Only in the 1990s, after the fall of the iron curtain and increased eastern European immigration, Germany slowly faced the reality of being a migratory country - now with a population of around 20.5 per cent with migratory background (Statistisches Bundesamt 2014). The largest immigrant population originates in Turkey (1,549 million by the end of 2013).

Politically, this learning process is marked by a late and incremental liberalization of labor migration as well as a reform of citizenship law: the dated *ius sanguinis* was replaced by *ius soli* in 2000 and a liberal immigration law in 2004. Especially these latter reforms were politically disputed. Immigrants faced severe riots still in the early 1990s, especially in the East German states, which were formerly part of the German Democratic Republic GDR. Despite much lower immigrant numbers, prejudices and violence against immigrants are still more common in this part of Germany.

Germany's late adoption of modern immigration policies might also affect the behavior of administrative officials; hence favoring discriminatory practices against non-German citizens. Lacking the tradition of the "American Creed" (Merton 1949), German administration is an

interesting test case, of how new liberal policies concerning migrant population affect the everyday behavior of public officials. In contrast to the US, Germany never faced demands for a "representative bureaucracy" (Groeneveld and Van De Walle 2010) or "diversity" (Ricucci 2002) until the turn of the century. On the contrary: Germany, like other European countries, is marked by a rather mono-ethnic civil service (OECD 2009).

Discrimination by public authorities: Methodological and theoretical approaches

When it comes to discriminatory behavior, which is defined here as an unequal treatment of persons or groups in the basis of ascriptive attributes (Pager and Shepherd 2008), the use of standard methodology of empirical public administration research reaches its limits. As a rule, empirical data base on standardized surveys and interview-based case studies, while researchers only seldom used administrative process data or participative observation (Raadschelders and Lee 2011). Controlled settings such as laboratory experiments are rarely used. Moreover, such data are typically gathered in highly sterile settings and observation effects cannot be ruled out.

One fundamental methodological problem with interview-based research on public administration is the tendency of the interviewees to regard the interview situation as a form of organizational self-portrayal and self-assurance. This potentially leads to euphemistic presentations of the represented organization and selective response behavior (Kuhlmann et al. 2008, 856). The tendency to "sugar-coat" facts is particularly relevant in normatively loaded contexts, in which the aspect of social desirability plays a central role (see Crowne and Marlowe 1964). This is obviously the case with discriminatory conduct.

Field experiments (Harrison and List 2004) link the advantages of a controlled experiment with realistic field settings and largely exclude observation effects. By simulating real

demands placed on the tested persons, their reactions to stimuli can be analyzed in a real setting. With regard to public administration, a number of artificially produced stimuli are conceivable, for example artificial requests and applications or even the artificial creation of problems. Such approaches indeed raise questions regarding the ethical boundaries of research (Riach and Rich 2004): field experiments of the kind presented in this article entail the deception of the units of analysis and are carried out without their consent. However, the artificial creation of requests to public authorities and the analysis of their responses, i.e. a so called “correspondence test”, enables us to gain extensive data material, which reflects everyday administrative behavior and is largely free of observation effects such as “social desirability” and artificial decision-making situations, as in the case in laboratory experiments.

Field experiments on discrimination have been primarily carried out for private labor markets (e.g. Bertrand and Mullainathan 2004; Riach and Rich 2004) and housing markets (Riach and Rich 2002; Yinger 1998). The various analyses for Germany show that immigrants are disadvantaged in comparison to citizens with German origin both in the labor as well as in the housing market (Kaas and Manger 2010; Diehl et al. 2013).

As for the public sector, no comparable studies exist up to now, except for a few analyses of discriminatory behavior of political decision-makers. For example, Butler and Brookman (2011) investigated the responsiveness of American state representatives to the requests of presumably white or black voters and identified discriminatory effects independent of party affiliation. In Robert Putnam’s classic “Making Democracy Work”, faked requests (although without randomized stimuli) were used to measure responsiveness (Putnam 1993, 73).

The conventional approach used in these analyses is the variation of the treatment variable by names, which function as the main stimulus. The other traits of the person were held constant. Thus, it is assumed that the simple signal of the name, which is associated with ascriptive

features, is the trigger of unequal treatment effects. These are generally explained by economic discrimination theories (Pager and Shepherd 2008). On the one hand, this pertains to statistic discrimination (Arrow 1973), which presumes that actors discriminate on rational grounds by following the assumption that according to statistic regularities certain features trigger effects, which are perceived as undesirable. On the other hand, “taste-based discrimination” (Becker 1957) involves discrimination based on personal aversions and prejudices against people, for example depending on their gender or ethnic affiliation.

Notwithstanding the advantages of field experiments, some of their benefits may be lost when, as is common, outcomes are not defined for some sample members. For analyzing service response quality, for example, the outcome is undefined if the request remains unanswered. This problem of truncation by death, can lead to systematic biases that need to be accounted for when interpreting results.

Research design and methods

In order to test for discrimination based on perceived origin in different contexts and to determine whether gender has a mediating effect on how citizens are treated, we applied a 2 x 2 x 2 experimental design including three factors with two values each. Ethnic background and gender represent the factors of main interest. Both stimuli were conveyed by names, as we selected names that could clearly be identified to be male or female names and sounded *typically* German or Turkish.

The variation between a German and Turkish background is useful for several reasons. First, immigrants of Turkish origin constitute the largest migrant population in Germany (BAMF 2013, 159). It is thus particularly relevant to identify potential cases of discrimination against this group. Second, the Turkish immigrant group is among the groups that experience the highest level of prejudice according to existing research on the German case. In particular,

sociological integration research shows that the Turkish population group is less integrated into society compared to immigrants of other nationalities and are perceived "as much more 'alien' than other immigrant groups" (Diehl et al. 2013, 1680). Third, Turkish names are easy to recognize as such and therefore provide a clear stimulus. The specific names were selected with consideration to sociological name research (Humpert and Schneiderheinze 2000).¹

Furthermore, our experiment captures two different types of citizen request. In a first request, a potential tourist inquires about possibilities for parking his or her mobile home while travelling through the city. In the second request, a person who will move to the town in the near future inquires about child-care options and leisure time facilities for a six-member family (see appendix). The topics were selected in order to cover two policy areas that would potentially impose different types of burdens on the city. The first email pertains to a temporally limited issue, which concerns public order, in particular. The second email involves a long-term issue, which could potentially pose demands on the city to provide services and result in a financial burden.² With regard to the above mentioned issue of research ethics, we selected a compromise, that keeps the degree of deception relatively low and aims to minimize the work burden and consequences for the analyzed public authorities. This entails ensuring strict anonymity, as no public authorities should be put in an embarrassing situation.³

For each citizen request, individual email addresses were created, that consisted of the given first name and a family name. The main stimuli were thus easily recognizable not only through the signature but also through the indicative email address. Two emails were sent with the names Ayşe Karagöz, Mehmet Karagöz, Eberhard Schuster and Dorothee Schuster in June and July 2013 in intervals of three weeks to the central email address of all municipal administrations (e.g. info@stadtxyz.de) of German cities with more than 25,000 inhabitants (excluding the three city-states Berlin, Hamburg and Bremen; n=501).⁴

We consciously confronted every city with two emails so that every city had to answer each type of citizen request once. This way we could increase the number of responses per request (two answers per city) while avoiding suspicion about the potential ‘fake’ nature of the requests on the side of city administrations likely to arise when a city would have been confronted with the same request coming from different senders. By combining the factors gender (male vs. female), ethnicity (German vs. Turkish), and request (child-care vs. mobile home), the analysis represents a 2 x 2 x 2 design resulting in 8 different vignettes used within citizen requests.

The responses to the emails were coded according to three sub-categories: response speed, response quality (completeness) and service orientation. After an initial review of all responses, the basic structure of the coding was adjusted and subdivided accordingly. In order to ensure a high degree of congruence, the responses were coded from several coders independently. Differing assessments were discussed and brought in line in order assure for consistency and thus a uniform coding with similar cases.

The first sub-category *response speed* attributes the public authorities a score resulting from the days between initial contact and the response. Public authorities were granted a duration of 15 days maximum (Monday to Friday) to process our requests. To rule out distortions in the processing time, all requests were sent out more or less simultaneously on weekends. The score in the sub-category response speed resulted from the difference between the maximum duration to process the request and number of days it took to answer the request. Public authorities, which already reacted to our request on the first day subsequently, received the maximum score of 15 points. A score of 0, by contrast, was given to authorities that did not respond to our request within a period of 15 days.

For the assessment of *completeness*, the responses were coded according to their informational content and the presentation of information. All requests comprised two

thematic blocks with different sub-questions. If all sub-questions of both thematic blocks were answered, a score of 4 points was given. Congruently, the comprehensibility and preparation of the responses were rated with two points respectively. The maximum score for the sub-category *response quality* thus amounted to 8 points.

In the final sub-category of *service orientation* we gave 7 points in the best case. We gave up to three points with regard to the thoroughness of the response, a friendly and courteous tone of the response, and the mentioning of additional contact persons for further questions. If all questions were answered in one single email, we rated the response with two points. If we did not receive such an encompassing response but instead two or more response emails, we rated the response as less service-oriented. Finally, we gave one additional point, if public authorities provided information and additional material beyond the content of the request. The three sub-indexes were merged into an additive total index (max. 30 points together) (see table 1).

[Table 1 here]

Results

Due to the randomized allocation of the respective stimuli we can assume that no systematic differences exist between the different groups except for the stimulus itself. It is thus possible to determine the effect of the respective stimulus without controlling for potentially confounded variables. This implies that we can rely on comparing group means to assess whether a certain stimulus has an effect.

The frequency distribution of the respective stimulus is summarized in table 2. Overall, 498 requests regarding child-care facilities and 493 requests regarding mobile home travel were

sent out. On both topics, we contacted all 501 German cities with more than 25,000 inhabitants. However, due to technical problems – the name of the email address did not coincide with the name of the enquiring person – eleven requests had to be excluded from the analysis.

[Table 2 here]

Discrimination by ethnic origin

The quality of received responses is assessed from multiple angles. To begin with, we assess whether (1) a response was provided at all (response). To further assess the quality of responses that were actually received, we capture (2) how quickly the city responded (speed), (3) how completely the respective inquiries were answered (completeness), and (4) how service-oriented the respective answer was (service). These aspects were aggregated into a total score.

[Table 3 here]

Generally, we find only little evidence for discrimination based on whether citizens' name indicates a Turkish ethnic background. Neither the probability to receive a response, nor the total index score hints towards substantial differences in the treatment of citizens with Turkish and German names. The only result that suggests discrimination relates to the completeness of answers. Specifically, in case a response was received at all, the responses

sent to citizens with Turkish names were on average less complete than the answers to citizens with German names. This result is still present in the aggregate score across both types of request. While the lower average level of completeness is present in both requests individually, difference in means is only large enough in the mobile home request to conclude that the difference is substantial based on conventionally accepted levels of statistical significance. Interestingly, the results suggest that the level of service orientation of mobile home requests received by citizens with Turkish names was even higher than for citizens with German names. However, this result is susceptible to the way in which we handle the truncation by death problem discussed above. The result only holds for the answers received. If we include the requests that did not receive an answer with a score of zero for service orientation, the difference between citizens with German and Turkish names in terms of average service orientation disappears. In contrast, the significant difference in the average level of completeness is robust to this adjustment. In sum, the difference-in-means analysis indicates that if the answers to citizen requests reflect any discrimination at all, then it is mostly with respect to the completeness of these answers. Yet, one has to keep in mind that this difference in means of 4.58 versus 4.33 is not very large. After all, these means are calculated on the basis of ordinal scales and the median level of completeness for both groups (Turkish and German names) is 4. Yet, despite the ordinal character, the difference in means should not be dismissed. It does in any case reflect that citizens with Turkish names received less complete answers more frequently than citizens with German names.

Discrimination by gender

We find more evidence for discrimination by gender than for discrimination based on the ethnicity indicated by citizens' names. Yet, the resulting patterns are somewhat ambiguous. The results do not suggest a clear discrimination of one particular gender. To begin with, 81% of male but only 77% of female citizens received a response at all. Female citizens thus

had a statically significant lower probability of receiving an answer to their request than male citizens. At the same time, table 4 suggests that if an answer was received, the average answer to male citizens was less complete than the average answer to female citizens. This result holds across both types of requests. Yet, this result is only indicative for the level of service orientation. If we include cases in which requests remained unanswered with a score of zero on completeness the gender specific difference on completeness disappears. Since female citizens received no answer more often than male citizens, the average completeness score for female citizens decreases sufficiently through this handling of the truncation by death problem that the significant difference in means disappears. In contrast, the differences in the average level of service orientation reflected by answers to male citizens are robust. We find different levels of service orientation independent of our treatment of truncation by death.

[Table 4 here]

Interestingly, however, the differences in service orientation vary by request. For mobile home requests, female citizens received a higher average level of service orientation. In contrast, male citizens received answers with a higher average level of service orientation to child-care requests. Only because the difference is particularly pronounced for the mobile home request, for which male citizens received answers of lower service orientation, the aggregate score for both requests tilts towards a discrimination against male citizens. Overall, however, these differences across different aspects of the quality of answers are not large enough to be reflected in the total index score. Taking all different aspects of quality into account, we find no evidence for discrimination by gender.

Interaction effects of ethnic origin and gender

While the lack of substantial differences in the total quality scores is good news for the German municipal administrations, the following paragraphs further explore the empirical patterns of differences we found in terms of completeness and service orientation of answers received. The results presented above suggest service orientation and completeness not only differed by ethnicity but also by gender as well as by request. Carrying out an analysis of variance (ANOVA) for each type of request allows us to further assess how name and gender interact to create group specific effects.

This approach does not yield much additional insight in terms of patterns of answers' completeness. Completeness only seems to vary for the mobile home request and depends on name only without a strong indication for any further group specific effects. In terms of completeness, it does not seem to matter whether one has a female or male Turkish name.

In stark contrast to this result for answers' completeness, we find evidence for substantial group differences in the level of service orientation. Even more so, these differences are relevant for both types of requests. By carrying out post-hoc tests to the ANOVA in the form of pairwise comparisons, we can tell for which citizen groups these significant differences exist. The following tables present the results including the ranking of the different groups in terms of the average level of service orientation the answers reflected.

[Table 5 here]

Table 5 ranks the different groups of citizens according to the average level of service orientation they received if they did receive an answer to their request. Most strikingly, the results show that service orientation crucially depends on whether citizens with a Turkish name are male or female. While the female Turkish name received by far the best level of service orientation, the male Turkish request received the lowest level of service orientation of all groups. All group differences are significant except for the difference between the groups with the lowest score on service orientation; i.e. male German and male Turkish requests. This difference of 0.28 service orientation points only slightly misses a level of conventionally accepted statistical significance, however. These results are robust even when we account for potential biases through truncation by death.

[Table 6 here]

Interestingly, for the child-care request, the ranking of average level of service orientation is reversed (table 6). In this case the group of citizens with a male Turkish name received the best level of service orientation, while the group with a female Turkish name received the lowest average level of service orientation. Again, the gender difference seems to matter greatly for citizens with a Turkish name. In contrast, there is no significant difference between male and female citizens with German names in this case. These results only change slightly when accounting for potential biases through truncation by death. Again, when non-answers are included with a score of zero on service orientation, the group ranking remains the same. Yet, the differences between average scores decrease so that only the differences between the best and the worst category remain significant. This gives further support to the

finding that differences between citizens with Turkish names but different genders are particularly pronounced.

In sum, the level of service orientation varies greatly for citizens with Turkish names but different genders. In contrast, gender does not make much difference for citizens with German names. Furthermore, whether service orientation is particularly high for citizens with female Turkish names crucially depends on the type of request. The question is, of course, how to make sense of these results. Yet, before we engage in a discussion of potential causes, further assessment of these patterns is in order.

So far, the analysis only focused on gender, ethnic origin indicated by the name, and the interaction between these two factors. Other potentially confounded variables were left out of the picture. To test whether these preliminary results indicating positive discrimination hold when other potentially confounded factors are taken into account, we conduct a regression analysis.

Specifically, we focus on two further factors that might influence the level of service orientation besides having a male or female Turkish name. First of all, we control for whether the citizen making a request and the public official answering the request are of the same gender. Secondly, we take the anonymity of answers into account. Whenever public officials do not identify themselves by name in an answer to the citizen request and simply respond in the name of the city or the particular city department, it is more difficult to hold them accountable for bad answers to citizen requests. It is simply more difficult to find out who exactly gave the particular answer. Consequently, anonymity might influence the level of service orientation.

Since the dependent variable of service orientation is a count variable, i.e. its values are only non-negative integers, simple OLS regression is inappropriate. Instead, count data regression models should be used. Since testing for overdispersion indicates that the data is

equidispersed, a Poisson regression model is used (Winkelmann, 2008; Cameron and Trivedi, 2013).

Table 7 presents the results of the regression analysis. Model 1 simply restates the finding from above. The significant interaction terms (request X male Turkish name; request X female Turkish name) confirm the results from above: the effect of a male (and female) Turkish name on the level of service orientation depends on the type of request. The positive coefficient for the group with female Turkish names indicates that this group received a higher level of service orientation than Turkish males for the mobile home request (i.e., when the type of request = 0). Models 2 and 3 add the control variables discussed above. Model 3 shows that while the fact that questioner and respondent are of the same gender enhances the level of service orientation, adding this variable does not change the request specific effect of having a male or female Turkish name. Whether answers were sent anonymously or not did not seem to make a difference to the level of service orientation at all. The request specific effect of gender for Turkish names is thus robust for these control variables. This further supports the finding that the differential treatment results from aspects that are specific to each request.

[Table 7 here]

To further assess whether the positive and/or negative discrimination in terms of service orientation of male and female citizens with Turkish names depends on the gender of the responding public official, we conduct an additional regression analysis for each type of request. The results are presented in tables 8 and 9. For the child-care request, model 1

simply restates the central finding from above: citizens with female Turkish names receive a lower level of service orientation (negative coefficient) and citizens with male Turkish names receive a higher level of service orientation (positive coefficient). This coefficient does not change when controlling for whether a female public official responded to the request (model 2). In addition, even more importantly, model 3 provides no evidence for a significant interaction between the gender of the respondent and the gender of the citizen with Turkish name. The gender specific patterns for citizens with Turkish names in the context of the mobile home request thus seem to be independent of the respondents' gender.

[here Table 8]

The same holds true for the mobile home request. While model 2 generally suggests that answers provided by female public officials were of a higher level of service orientation, the effect of having a male or female Turkish name does not depend on the respondents' gender. The positive coefficient for the variable female Turkish name in model 3 indicates that male respondents were more service oriented towards this group than to other citizen groups. Yet, this also holds for female public officials as indicated by the non-significant interaction term. The effect of a female or male Turkish name on the level of service orientation does not depend on the respondents' gender.

[Table 9 here]

In sum, the analysis suggests that the request specific effects of Turkish names on the level of service orientation are either due to variables that this study was unable to observe or caused by the characteristics of the particular requests.

Discussion and conclusions

The starting point of this paper was striking misfit between legally embedded principles of administrative equality and non-discrimination and rising societal claims of discriminatory administrative practice in Germany. To address the extent of individual discrimination in the German public administration, we conducted a field experiment that involved all German cities over 25,000 inhabitants. The aim of this analysis was to investigate whether and to what extent discrimination by gender and ethnic origin impacts the responsiveness of public administrations.

Our findings generally reveal that contrary to the private sector, German public administration does not appear to systematically discriminate between people based on gender or ethnic origin. Thus, it lives up to Weberian ideals to a very large extent. A more detailed analysis of interaction effects between gender and ethnic origin, however, shows rather striking patterns of positive and negative discrimination with regard to service orientation. While for the mobile home request the female Turkish name received by far the best level, the male Turkish request received the lowest level of service orientation. For the child-care request, by contrast, we observe exactly the opposite picture. In short, there is discrimination by ethnic origin, but this discrimination is conditioned by gender.

What might be potential explanations for this pattern? Our two requests seem to cater to different stereotypes regarding the role of men and women that are very widespread in German society, as recent survey results show (IfD 2013). Stereotypically, women are thought to be responsible for children and the organization of everyday family life. In

contrast, men are stereotypically thought to be responsible for the organization of special family events such as vacations; particularly when these involve the operation of mobile homes. Consequently, certain requests can be seen to be stereotype conform while others are not. Specifically, a request for child-care placement by a male citizen can be seen to be stereotype non-conform. The same can be argued about a mobile home request by a female citizen.

Stereotypes about these gender-specific roles can be assumed to be stronger among the German public when they concern Turkish families with a stereotypically strong orientation in Muslim faith and paternalistic family structures (Blohm and Wasmer 2008; Diehl 2013). Ethnic background might thus interact with the gender specific stereotypes. As a result, for the mobile home request a citizen with a female Turkish name can be assumed to maximize the level of stereotype nonconformity. Similarly, a citizen with a male Turkish name asking about child-care facilities can also be assumed to create a higher level of stereotype nonconformity than the other groups. Our results suggest that positive discrimination in the form of better service orientation occurs wherever stereotype nonconformity is maximized. In turn, negative discrimination in terms of substantially lower levels of service orientation seems to occur wherever requests are completely stereotype conform. These considerations, however, need further elaboration and testing in future research.

Endnotes

1. Besides the largest possible stimulus, we also attempted to select names for which no real persons could be clearly identified by relevant electronic searches (Google; electronic telephone book). In other words, we aimed to ensure that there were no or too many “hits” for the respective names. We did so in order to protect non-participating bearers of the names from unwarranted requests by the municipal authorities.
2. The procedure was presented to the Ethics Commission of the University of Konstanz, which deemed it as unproblematic. If a central email address did not exist or was not visible, the requests were sent by means of the contact form on the internet sites of the municipal administrations.
3. If a central email address did not exist or was not visible, the requests were sent by means of the contact form on the web sites of the municipal administrations.
4. While constructing vignettes, we additionally had to consider that it was not possible to refer to local circumstances and that the requests also had to be plausible for non-residents of the towns and cities, because otherwise the register of residents would have revealed that the requests were faked.

References

- Alba, Richard. 2005. "Bright vs. blurred boundaries: Second-generation assimilation and exclusion in France, Germany, and the United States." *Ethnic and Racial Studies* 28, 20-49.
- Arrow, Kenneth J. 1973. "The Theory of Discrimination." In *Discrimination in Labor Markets*, edited by Orley Ashenfelter, and Albert Rees, 3–33. Princeton: Princeton University Press.
- Atzmüller, Christiane, and Peter M. Steiner. 2010. "Experimental Vignette Studies in Survey Research." *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 6:128-138.
- Becker, Gary S. 1957. *The Economics of Discrimination*. University of Chicago Press, Chicago.
- Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *The American Economic Review* 94: 991-1013.
- Blohm, Michael, and Martina Wasmer. 2008. „Einstellungen und Kontakte zu Ausländern.“ In *Datenreport 2008*, edited by Statistisches Bundesamt Deutschland, 208-214. Bonn: Statistisches Bundesamt.
- Bundesamt für Migration und Flüchtlinge (BAMF). 2013. *Migrationsbericht 2011*. Berlin: Bundesministerium des Inneren.
- Butler, Daniel M, and David E. Brookman. 2011. "Do Politicians Racially Discriminate Against Constituents? A Field Experiment on State Legislators." *American Journal of Political Science* 55 (3):463-477.
- Cameron, A. Colin, and Pravin K. Trivedi. 2013. *Regression Analysis of Count Data. 2nd Edition*. Cambridge: Cambridge University Press.

- Cayer, Joseph, and Lee Sigelman. 1980. "Minorities and Women in State and Local Government: 1973-1975." *Public Administration Review* 40:443-450.
- Crowne, D. P., and D. Marlowe. 1964. *The Approval Motive*. New York: John Wiley & Sons.
- Diehl, Claudia, Veronika Andorfer, Yassine Khoudja, and Karolin Krause. 2013. „Not in my kitchen? Ethnic discrimination and discrimination intentions in shared housing among university students in Germany." *Journal of Ethnic and Migration Studies* 39:1679-1697.
- Frederickson, H. George. 1990. "Public Administration and Social Equity." *Public Administration Review* 50:228-237.
- Frederickson, H. George. 2010. *Social Equity and Public Administration. Origins, Developments, and Applications*. Armonk: Sharpe.
- Groeneveld, Sandra, and Steven Van de Walle. 2010. "A Contingency Approach to Representative Bureaucracy: Power, Equal Opportunity and Diversity." *International Review of Administrative Sciences* 76:239-258.
- Harrison, Glenn W., and John A. List. 2004. "Field Experiments." *Journal of Economic Literature* 42:1009-1055.
- Humpert, Andreas, and Klaus Schneiderheinze. 2000. „Stichprobenziehung für telefonische Zuwandererbefragungen – Einsatzmöglichkeiten der Namensforschung.“ *ZUMA-Nachrichten* 47:36-64.
- Institut für Demoskopie (IfD) (2013). *Der Mann 2013: Arbeits- und Lebenswelten - Wunsch und Wirklichkeit*. Allensbach: Institut für Demoskopie.
- Kaas, Leo, and Christian Manger. 2010. "Ethnic Discrimination in Germany's Labour Market: A Field Experiment." *IZA Discussion Paper No. 4741*.
- Kellough, J. Edward. 1990. "Integration in the Public Workforce: Determinants of Minority and Female Employment in Federal Agencies." *Public Administration Review* 50:557-566.

- Kuhlmann, Sabine, Jörg Bogumil, and Stephan Grohs. 2008. "Evaluating Administrative Modernization in German Local Governments: Success or Failure of the "New Steering Model"?" *Public Administration Review* 68 (5):851-863.
- Llorens, Jared J., Jeffery B. Wenger, and Edward Kellough. 2007. "Choosing Public Sector Employment: The Impact of Wages on the Representation of Women and Minorities in State Bureaucracies." *Journal of Public Administration Research and Theory* 18:397-413.
- Merton, Robert K. 1949. "Discrimination and the American creed." In: *Discrimination and National Welfare*, edited by Richard M. MacIver, 99–126. New York: Harper and Brothers.
- Myrdal, Gunnar. 1944. *An American Dilemma. The Negro Problem and Modern Democracy*. New York: Harper and Brothers Publishers.
- Pager, Devah, and Hana Shepherd. 2008. "The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets." *Annual Review of Sociology* 34:181-209.
- Putnam, Robert D. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton: Princeton University Press.
- Raadschelders, Jos, and Kwang-Hoon Lee. 2011. "Trends in the Study of Public Administration: Empirical and Qualitative Observations from *Public Administration Review*, 2000-2009." *Public Administration Review*, 71:19–33.
- Riach, Peter A., and Judith Rich. 2002. "Field Experiments of Discrimination in the Market Place." *The Economic Journal* 112:F480-F518.
- Riach, Peter A., and Judith Rich. 2004. "Deceptive Field Experiments of Discrimination: Are they Ethical?" *Kyklos* 57:457-470.
- Riccucci, Norma. 2002. *Managing Diversity in Public Sector Workforces*. Boulder: Westview Press.

- Sachverständigenrat deutscher Stiftungen für Integration und Migration (SVR). 2012. *Benachteiligungserfahrungen von Personen mit und ohne Migrationshintergrund im Ost-West-Vergleich. Expertise für die Antidiskriminierungsstelle des Bundes*. Berlin: Antidiskriminierungsstelle des Bundes.
- Statistisches Bundesamt (2014). *Mikrozensus 2013. Pressemitteilung Nr. 402 vom 14.11.2014*. Wiesbaden: Statistisches Bundesamt.
- Weber, Max. 1968. *Economy and Society. Vol. 1*. Berkley: University of California Press.
- Winkelmann, Rainer. 2008. *Econometric Analysis of Count Data. Fifth Edition*. Heidelberg, New York: Springer.
- Yinger, John. 1998. "Testing for Discrimination in Housing and Related Markets." In *A National Report Card on Discrimination in America: The Role of Testing*, edited by Michael Fix, and Margery Austin Turner, 27-46. Washington DC: The Urban Institute Press.

Appendix:

1. The vignettes

Vignette 1: Mobile home

Dear city administration XY,

My family and I are planning a city tour through Germany with our new mobile home. In the past, we have had good experiences with parking spaces that are free of charge or camping grounds near the city. Before our trip, we would therefore kindly request you to provide us information on the following topics:

- Is there an explicitly marked free parking space for mobile homes in your city? If so, where is it located and what facilities does it have?
- How many nights maximum may one stay there?
- If your city does not offer any free parking spaces for mobile homes, can you give us a suggestion where we can alternatively camp with our mobile home as close as possible to the city center?
- Finally, could you indicate two or three places of interest that we definitely should not miss or a place where we can gather extensive information on the history of the city?

Kind regards,

Vignette 2: Family with many children

Dear Ladies and Gentlemen,

At the end of the year my family and I will be moving to your beautiful city for career reasons.

In order for us to get off to a good start, I would be very thankful if you could answer a few questions. My wife and I have four children (2, 4, 7, 11) and we are still looking for a child-care facility for our two youngest children. Could you please give us some information on the public child-care facilities in your city? Are there free places as of December and what

expenses can we plan on. Are there any special price reductions for families with many children?

Furthermore, I would be delighted if you could provide me some information on leisure activities for our oldest children. Are there special offers for public services such as the music school or the open air swimming pool for families with many children?

Kind regards,

Table 1: Coding scheme

Sub-index	Incorporated items	Scores	Max. score
Response speed	Duration in work days	Per work day -1 point	15
Completeness	Responses to all sub-questions (1) / (2) Comprehensibility and preparation of the response (1) / (2)	2 x 2 points each	8
Service orientation	Mention contact persons; (offer) to send additional information (3) Completeness of the response (4) Perception of the friendliness of the public administration / service (5)	1 point each Up 2 points Up to 3 points	7
Total		Total	30

Table 2: Distribution of requests by stimulus

	German		Turkish		Total
	female	Male	Female	male	
Mobile home	124	125	124	120	493
Child-care	124	125	124	125	498
Total	248	250	248	245	991

Table 3: Name of the inquiring person

		Name						Difference (absolute)	p-value (t-test)
		German			Turkish				
		n	x	sd	N	x	sd		
Together	Response	498	.79	.40	493	.78	.41	.01	.292
	Speed	396	13.73	2.27	385	13.58	2.27	.15	.173
	Completeness	396	4.58	2.44	385	4.33	2.24	.25	.067*
	Service	396	2.58	1.15	385	2.67	1.34	.09	.166
	Total	498	16.61	9.11	493	16.14	9.11	.47	.209
Child-care	Response	249	.79	.41	249	.78	.42	.01	.373
	Speed	196	13.31	2.61	193	13.17	2.53	.14	.295
	Completeness	196	4.08	2.49	193	3.98	2.17	.10	.341
	Service	196	2.33	1.10	193	2.33	1.20	.01	.483
	Total	249	15.50	8.78	249	15.14	8.77	.36	.323
Mobile home	Response	249	.80	.40	244	.78	.41	.02	.327
	Speed	200	14.15	1.79	192	14.00	1.90	.15	.203
	Completeness	200	5.08	2.27	192	4.69	2.27	.39	.044**
	Service	200	2.83	.08	192	3.01	1.39	.18	.086*
	Total	249	17.72	9.32	244	17.16	9.35	.55	.255

Note: ***/**/* indicate levels of statistical significance at the 10% / 5% / 1% level; p-values result from one-sided t-tests.

Table 4: Gender of the inquiring person

		Gender						Difference (absolute)	p-value (t-test)
		Male			Female				
		n	x	sd	N	x	sd		
Together	Response	494	.81	.39	497	.77	.42	.04	.066*
	Speed	399	13.69	2.25	382	13.62	2.30	.07	.342
	Completeness	399	4.23	2.30	382	4.70	2.37	.46	.003***
	Service	399	2.36	1.09	382	2.90	1.34	.54	.000***
	Total	494	16.38	8.68	497	16.37	9.52	.01	.495
Child-care	Response	249	.79	0.41	249	.78	.42	.01	.373
	Speed	196	13.31	0.42	193	13.16	2.71	.15	.282
	Completeness	196	3.81	2.27	193	4.25	2.39	.45	.029**
	Service	196	2.56	1.18	193	2.10	1.08	.46	.000***
	Total	249	15.49	8.77	249	15.16	8.79	.33	.338
Mobile home	Response	245	.83	.38	248	.76	.43	.07	.033**
	Speed	203	14.05	2.00	189	14.10	1.66	.04	.413
	Completeness	203	4.65	2.26	189	5.15	2.28	.50	.014**
	Service	203	2.17	.96	189	3.72	1.06	1.55	.000***
	Total	245	17.29	8.51	248	17.60	10.08	.31	.358

Note: ***/**/* indicate levels of statistical significance at the 10% / 5% / 1% level; p-values result from one-sided t-tests.

Table 5: Pairwise comparison of group specific differences in service orientation for the mobile home request

Rank	Name	Median	Mean	std	Difference	Difference	Difference
1	Turkish; female	4	4.01	.93	.58***	_____	
2	German; female	3	3.43	1.10			
3	German; male	2	2.3	.90	_____	1.13***	_____
4	Turkish; male	2	2.02	1.00			

Note: *** indicate levels of statistical significance at the 1% level; these results are robust for different post-hoc tests (i.e. the Sidak method, Scheffé's method, and the Bonferroni procedure)

Table 6: Pairwise comparison of group specific differences in service orientation for the child-care request

Rank	Name	Median	Mean	std	Difference	Difference	Difference
1	Turkish; male	3	2.86	1.02	.47**	_____	
2	German; female	2	2.39	.92			
3	German; male	2	2.26	1.25		.14	_____
4	Turkish; female	2	1.80	1.14			.46**

Note: ** indicate levels of statistical significance at the 5% level; these results are robust for different post-hoc tests (i.e. the Sidak method, Scheffé's method, and the Bonferroni procedure)

Table 7: Poisson regression results

	(1)	(2)	(3)
female Turkish name [0=no; 1=yes]	.33*** (.06)	.32*** (.06)	.32*** (.06)
male Turkish name [0=no; 1=yes]	-.33*** (.08)	-.32*** (.08)	-.32*** (.08)
request X male Turkish name	.51*** (.11)	.56*** (.11)	.56*** (.11)
request X female Turkish name	-.57*** (.11)	-.61*** (.11)	-.61*** (.11)
request [0=mobile home request; 1=child-care request]	-.22*** (.06)	-.20*** (.06)	-.20*** (.06)
same gender [0=no; 1=yes]		.10** (.04)	.09* (.04)
Anonymous [0=no; 1=yes]			-.11 (.17)
Constant	.82*** (.0420)	.99*** (.04)	1.00*** (.05)
Observations	991	781	781

Note: Coefficients are unstandardized Poisson regression coefficients. Standard errors in parentheses. Note: */**/** indicate levels of statistical significance at the 10% / 5% / 1% level.

Table 8: Poisson regression results for the child-care request

	(1)	(2)	(3)
female Turkish name	-.25*** (.09)	-.25*** (.09)	-.30 (.20)
male Turkish name	.18** (.07)	.19** (.08)	.19 (.15)
female respondent		-.00 (.07)	-.01 (.10)
female respondent X female Turkish name			.06 (.22)
female respondent X male Turkish name			.00 (.17)
Constant	.61*** (.05)	.86*** (.07)	.87*** (.08)
Observations	498	380	380

Note: Coefficients are unstandardized Poisson regression coefficients. Standard errors in parentheses. Note: */**/** indicate levels of statistical significance at the 10% / 5% / 1% level.

Table 9: Poisson regression results for the mobile home request

	(1)	(2)	(3)
female Turkish name	.33*** (.06)	.35*** (.06)	.28* (.16)
male Turkish name	-.33*** (.08)	-.33*** (.08)	-.34* (.18)
female respondent		.13* (.08)	.09 (.12)
female respondent X female Turkish name			.08 (.18)
female respondent X male Turkish name			.02 (.21)
Constant	.82*** (.04)	.93*** (.08)	.96*** (.12)
Observations	493	377	377

Note: Coefficients are unstandardized Poisson regression coefficients. Standard errors in parentheses. Note: */**/** indicate levels of statistical significance at the 10% / 5% / 1% level.