



Carsten Schwemmer, M.A.

Computational Methods for the Social Sciences: Applications to the Study of Ethnic Minorities

Cumulative Dissertation

for obtaining the academic grade

Dr. rer. pol.

submitted to

University of Bamberg

Faculty for Social Sciences, Economics, and Business Administration

First advisor

Prof. Dr. Marc Helbling, University of Bamberg

Second advisor

Prof. Dr. Thomas Saalfeld, University of Bamberg

Additional member of the promotion committee

Prof. Dr. Kai Fischbach, University of Bamberg

Submitted in Bamberg on June 17, 2019

Successfully defended in Bamberg on September 20, 2019

Danksagung (Acknowledgments in German Language)

Diese Dissertationsschrift ist das Resultat meiner ersten Schritte auf der Reise durch die akademische Welt. Es war nicht immer klar, ob “Forscher zu werden” der richtige Weg für mich ist. Zu jeder Zeit, aber insbesondere in Phasen der Orientierungslosigkeit, hatte und habe ich das Glück von zahlreichen Menschen unterstützt zu werden. Bei allen möchte ich mich hiermit ganz herzlich bedanken. Ohne euch wäre diese Dissertationsschrift nie zu Ende geschrieben worden.

Zunächst möchte ich meinen Betreuern Marc Helbling und Thomas Saalfeld danken. Wenn ich durch die Tätigkeit als wissenschaftlicher Mitarbeiter eines gelernt habe, dann dass Zeit eine der wertvollsten Ressourcen ist. In dieser Hinsicht wart ihr nicht nur ausgezeichnete Betreuer, sondern auch die denkbar besten Chefs: Ihr habt mir ausreichend Zeit für meine Forschung gegeben und hattet gleichzeitig stets ein offenes Ohr wenn ich euren Rat gebraucht habe. Zudem habt ihr mich immer dabei unterstützt die nächsten Schritte meiner akademischen Reise vorzubereiten.

Ebenso bin ich dankbar für die Unterstützung zahlreicher Kolleginnen und Kollegen, die entweder selbst als Koautor/-innen an Teilen dieser Schrift beteiligt waren, oder wichtiges Feedback zu meinen Forschungsprojekten gegeben haben: Danke Michael Eberhardt, Jorge Fernandes, Kai Fischbach, Diana Fischer-Preßler, Lucas Geese, Sebastian Jungkunz, Menusch Khadjavi, Caroline Schultz, Stephan Simon, Jasper Tjaden, Oliver Wieczorek und Sandra Ziewiecki. Ich danke auch den Teilnehmer/-innen mehrerer Forschungskolloquien in Bamberg und der Graduiertenschule BAGSS für all die hilfreichen wissenschaftlichen Diskussionen.

Aus einigen Arbeitsbeziehungen sind über die Jahre hinweg Freundschaften entstanden. Ich möchte mich jedoch auch bei meinen Freund/-innen außerhalb der Wissenschaft, bei meiner Freundin Mareike und bei meiner Familie bedanken. Ihr habt mir auf unzählige Arten geholfen und mir emotionalen Rückhalt gegeben. Zuletzt gilt mein besonderer Dank meinen Eltern Manfred und Monika, die es mir ermöglicht haben, meinen Lebensweg nach eigenen Wünschen und Vorstellungen zu gestalten.

Contents

| | | |
|----------|---|------------|
| 1 | Preface | 1 |
| 1.1 | Substantive contributions to the study of ethnic minorities | 2 |
| 1.2 | About the application of computational methods | 16 |
| 1.3 | Concluding remarks | 39 |
| 2 | First Article: Ride with Me - Ethnic Discrimination, Social Mar- | |
| | kets, and the Sharing Economy | 51 |
| 3 | Second Article: MPs' principals and the substantive representa- | |
| | tion of disadvantaged immigrant groups | 94 |
| 4 | Third Article: Social Media Strategies of Right-Wing Movements | |
| | - The Radicalization of Pegida | 140 |

1 Preface

In this manuscript I introduce my contributions to the emerging academic discipline *Computational Social Science*. At the time of writing in 2019, scholars have already used this term for over a decade (Lazer et al. 2009), but the development of this field is still ongoing. At its core, computational social scientists, including myself, seek to provide new answers to important social science research questions. They draw on computational methods at the intersection of computer science and statistics. This interdisciplinary approach comes with many potential benefits, but also with challenges, both of which I try to address in this cumulative dissertation.

Naturally, the focus of computational social science research will lean stronger towards either of the involved disciplines. Trained as a sociologist, my research predominantly focuses on the application of computational methods for *social science* aspects rather than on the development of computational methods on its own merit. Or, to put it in the words of Andreas Jungherr, I am “taking the *social* in Computational Social Science seriously” (Jungherr 2018, p. 29). This dissertation deals with the study of ethnic minorities, a social science research field about the interactions between mainstream societies and minorities such as refugees. These dynamic interactions lead to the emergence of many societal problems, such as political mobilization with the aim to maintain power of majority members and exclude members of immigrant origin. The overarching question for this dissertation is: how can computational methods be applied to provide new insights for the study of ethnic minorities? The articles for this dissertation include findings from research across

three related and interconnected domains: ethnic discrimination in the sharing economy, political representation of ethnic minorities and collective action driven by xenophobia. In the first subsection of this preface, I will provide a summary of the substantial contributions to the study of ethnic minorities across these domains.

All of the included articles were submitted to international, peer-reviewed social science journals. At the time of writing, two of the three articles have already been published and one article is under review. Unsurprisingly, the corresponding journals predominantly focus on social science aspects rather than computational methods. This is strongly reflected in the content of all articles: details about many of the computational aspects either had to be moved to appendices or did not find a place at all. This makes it difficult to highlight the importance of my computational contributions, as topics like the development of research software or algorithms for working with textual data could not be discussed in depth. For this reason, I will use the second part of this preface to provide more insights into the computational methods which served as the backbone for this dissertation. At last, in the third part of this preface I will close with some concluding remarks about the present and the future of computational methods for social science research.

1.1 Substantive contributions to the study of ethnic minorities

This section provides an overview of the most important contributions to the study of ethnic minorities. The first article of this dissertation is related to discrimination of ethnic minorities (Tjaden, Schwemmer, and Khadjavi 2018). The second article

examines the political representation of ethnic minorities (Geese and Schwemmer 2019). The last article deals with xenophobic collective action affecting ethnic minorities (Schwemmer 2019b). These topics are connected to each other in several ways. To provide only one example, a stronger representation of ethnic minorities by political actors who act to fulfill their needs will make it harder for xenophobic movements to gain power and to lead the way for right-wing forces in the corresponding political system. Moreover, these topics are also connected in a methodological way: they share a lot of problems that make it difficult to conduct social science research. Analyzing phenomena such as ethnic discrimination, substantive representation and collective action requires the measurement of corresponding indicators in ways that fulfill standards of modern social science research. For instance, experimental research designs are often used to study ethnic discrimination. However, this approach tends to suffer from low external validity, that is the generalization of experimental research findings to real world scenarios. Likewise, using survey data to analyze attitudes towards ethnic minorities introduces other methodological issues, such as social desirability bias (Edwards 1957). In addition, studying (ethnic) minorities is difficult by definition, as it often comes with a low number of observations that can be analyzed. As demonstrated in this dissertation, using computational methods can help to overcome such methodological problems. I show that extracting and analyzing real world data, using computational models for working with unstructured data such as large text corpora and creating research software are efficient approaches for answering fundamental research questions for the study of ethnic minorities. In what follows, I will first discuss the substantive contributions of each article.

Discrimination of ethnic minorities

A large body of literature has consistently shown that discrimination of ethnic minorities is a persistent driver of inequalities across a multitude of domains (Bertrand and Mullainathan 2004; Pager and Shepherd 2008; Pager, Bonikowski, and Western 2009; Ahmed, Andersson, and Hammarstedt 2010; Lin and Lundquist 2013; Pedulla 2018). To name a few, ethnic minorities suffer from inequalities related to wages, education and employment. Many of these inequalities emerge from unequal treatment of minority groups in comparison to majority groups on markets like the housing market. Multiple studies have already been conducted to assess the role of ethnic discrimination in such markets (e.g. Pager and Shepherd 2008).

One of the more recent puzzles is the question to what extent discriminatory patterns observed for traditional markets are also apparent for digital markets that emerged in the last couple of years. Together with my co-authors, I join this effort in my first article to gain a better understanding about the magnitude of ethnic discrimination in these markets and the mechanisms behind it. In particular, we focus on discrimination in the sharing economy by studying a large online carpooling platform where drivers offer to share their rides with other people.

This setting helps to expand the view from discrimination studies, which predominantly focus on major cornerstones of life such as getting a job or buying a house. While such events have huge implications and lead to unequal treatment of minorities, they usually do not occur very often in a lifetime. In contrast, sharing a car with other people is a situation many people experience on a more regular basis.

Therefore, our study allows us to examine more subtle, everyday forms of unequal treatment that might otherwise go unnoticed.

Moreover, we argue that many studies neglected to consider that such markets are not only driven by economic, but also by social aspects. In our case, strangers agree to share very limited space in a car and spend a considerable amount of time together. In this context, our case selection and research design allow us to analyze both social and economic aspects in digital markets and their relation to ethnic discrimination.

In addition, the available literature predominantly draws on either observable studies, which are affected by omitted variable bias (see Heckman 1998), or on experiments to measure discrimination. In such experiments, researchers artificially construct advertisements or other primary resources for the corresponding markets. Outcome variations between experimental groups can then be analyzed to measure discriminatory behavior (e.g. Doleac and Stein 2013). With our research setting, we can address the disadvantages of both of these approaches by utilizing computational methods. We observe thousands of rides of real actors acting in a real market and therefore are able to analyze human behavior without artificially manipulating the marketplace. At the same time, we observe all characteristics that are visible to the customer, as can be seen in Figure 1.

The visual interface shows information about age, gender, user picture (if available), user rating, car, timing and stops of the ride, price, available seats and some preferences of the driver (e.g. smoking, music, talking). In the first article, we estimate the effect of drivers' perceived name origin on the demand for their offered rides













| | | |
|---|--|---|
|   | Mittwoch 11. November - 13:30 Uhr Rostock → Berlin  Rostock, Deutschland  Mollstraße 19, 10249 Berlin, Deutschland Fahrzeug: SKODA OCTAVIA Combi ★★ | 11 € pro Mitfahrer/in 2 Plätze frei |
|  Fortgeschrittene/r ★ 4.7 - 3 Bewertungen  | Mittwoch 11. November - 14:00 Uhr Rostock → Berlin  Rostock Hbf, Rostock  S+U Hermannstr. (Berlin), Neukölln Fahrzeug: VOLKSWAGEN PASSAT ★★★★★ | 12 € pro Mitfahrer/in 4 Plätze frei |
|  Aufsteiger/in ★ 5.0 - 1 Bewertung f 161 Freunde/innen  | Mittwoch 11. November - 15:10 Uhr Rostock → Berlin → Ansbach  Rostock Hbf, Rostock  Ankunft: Berlin (Bitte sprechen Sie die Details mit dem Fahrer/der Fahrerin ab.) Fahrzeug: SKODA OCTAVIA Combi ★★ | 11 € pro Mitfahrer/in 3 Plätze frei |

Figure 1: Screenshot of the German carpooling interface. Images, names and age of drivers are pixelated.

as indicated by the number of times customers clicked on the corresponding ride. To create a measure for the ethnic backgrounds of drivers, we conducted an online survey, in which participants were asked to categorize the names of drivers to distinguish their associated origin. In particular, the focus was on typically German names and names of Arab, Persian or Turkish origin. The former group is the largest, most recognizable immigrant community in Germany. More details on data acquisition and preparation for the carpooling analysis are available in the second part of this preface.

The rich information available for our case allows us to also get a better understanding of the mechanisms driving ethnic discrimination, where past research predominantly focused on theories of taste-based discrimination and statistical discrimination. The concept of taste-based discrimination defines discrimination as personal

prejudice or taste associated with certain groups (Becker 1971). For equally productive individuals in a market, some are preferred over others because of variations in taste, which can be formalized as a disutility function.

In contrast to taste-based theories, statistical discrimination is based upon beliefs and expectations rather than animus against certain groups. In our case, consumers would use the name as a proxy signal to infer the *true* value of the ride in economic, safety and social terms. In practice it is often difficult to clearly distinguish between taste-based and statistical discrimination mechanisms. In our study, we address this problem by assessing the role of information (e.g. about the driver’s rating) for discriminatory behavior. We find that as more information (e.g. a higher number of driver ratings) or a stronger quality signal (e.g. better driver rating) becomes available, differences in demand for German and minority drivers vanish. This finding provides evidence in favor of statistical discrimination.

In summary, these aspects of the first article contribute to the literature on ethnic discrimination. In the context of this article, it is not very surprising to find evidence for discrimination in general. As outlined above, the existence of ethnic discrimination in market environments has already been proven by a large number of scholars (e.g. Pager and Shepherd 2008). However, the application of computational methods produced new insights into mechanisms for discrimination mechanisms in markets related to everyday, social interactions. As an alternative to computational methods, findings from a (conventional) experimental research design would lack external validity and therefore also provide inferior estimates about the magnitude of ethnic discrimination. Likewise, manually collecting longitudinal data for several

thousand rides, which was important to capture enough observations with ethnic minorities, would not be feasible without relying on hundreds of human workers.

Political representation of ethnic minorities

Once potential mechanisms for the emergence of inequalities due to ethnic discrimination are identified, a possible solution is to introduce policies in order to counter-vail discrimination patterns. Policies might be introduced by providers of specific platforms (e.g. online markets), or by policymakers such as elected representatives in democratic systems. In such democratic systems, elected representatives are expected to act in the interest of the electorate. The second article of this dissertation examines the political representation of ethnic minorities and especially disadvantaged immigrant groups in the German Bundestag.

Proper representation of certain groups in democratic systems first can be understood in terms of socio-demographic attributes of elected political actors. According to census data, about one fourth of the people living in Germany have a migratory background and about six percent are immigrants (Statistisches Bundesamt 2017). From a normative point of view of, it is desirable for democratic systems (e.g. the German Bundestag) that these numbers are reflected in the proportion of elected politicians. The political science literature refers to this concept as *descriptive representation* (see e.g. Pitkin 1967; Mansbridge 1999; Dovi 2002). Unfortunately, the reality is far from this normative ideal case. In most democracies, residents with a migratory background are in fact politically underrepresented (see Alba and Foner 2015; Bloemraad, Graauw, and Hamlin 2015).

What makes the situation even more problematic is that descriptive representation of minority residents does not necessarily result in political behavior that reduces inequalities. While it can be expected that elected politicians who themselves identify as members of ethnic minority groups are able to better understand the needs and interests of such groups, membership is neither necessary nor sufficient on its own for addressing inequalities. Rather, the needs and interests of minority groups should also find more consideration in the activities of their representatives. This concept is commonly referred to as *substantive representation* (Dahl 1971). In the second article of this dissertation, we draw on principal-agent models of democratic representation to examine substantive representation in the 17th German Bundestag. To measure substantive representation, we rely on parliamentary written questions tabled by members of parliament. The article includes a detailed discussion about the advantages of using written questions in comparison to other approaches that have been used in the literature before (see also Martin 2011; Wüst 2014; Aydemir and Vliegenthart 2016; Fernandes, Leston-Bandeira, and Schwemmer 2017). The texts of written questions were extracted from official online archives of the Bundestag using Python programming scripts. Technical details about the procedure to acquire all questions tabled during the 17th Bundestag (about 20,000) from corresponding PDF files and to combine them with socio-demographic data will be given in the second part of this preface. The following two written questions provide examples for members of parliament who engage in substantive representation of minority groups:

“How does the government justify the Federal Office for Migration and Refugees recent announcement to cut the budget for integration courses in the light of the CDU, CSU and FDP’s coalition agreements’ plan to qualitatively and quantitatively upgrade those courses?”

Written question tabled by Aydan Özoğuz, SPD, May 7, 2010.

“How does the government want to ensure that the Federal Employment Office will bring residents with a migratory background into vocational training in similar proportions in their respective age groups as compared to Germans?”

Written question tabled by Mechthild Rawert, SPD, March 18, 2011.

For the procedure of identifying written questions related to substantive representation of immigrant groups, we developed a coding scheme that combined human coding with automated methods. In this manner, one important aspect was to filter out questions containing negative positions on the integration of immigrant-origin residents, for instance questions expressing reservations against the integration of immigrants or a multicultural society. To contribute to the literature on political representation, we examine to what extent members of parliament engage in more substantive representation of immigrant groups depending on several factors.

First, we examine geographic patterns of representation by analyzing whether members of parliament engage in more substantive representation with an increasing share of foreign nationals (a proxy for immigrant origin residents) in their districts. Figure 2 shows the percentages of residents with a migratory background across the states of Germany for 2017.

It can be seen that residents with a migratory background are not evenly distributed across the country, as for instance the state Baden-Württemberg has a much higher

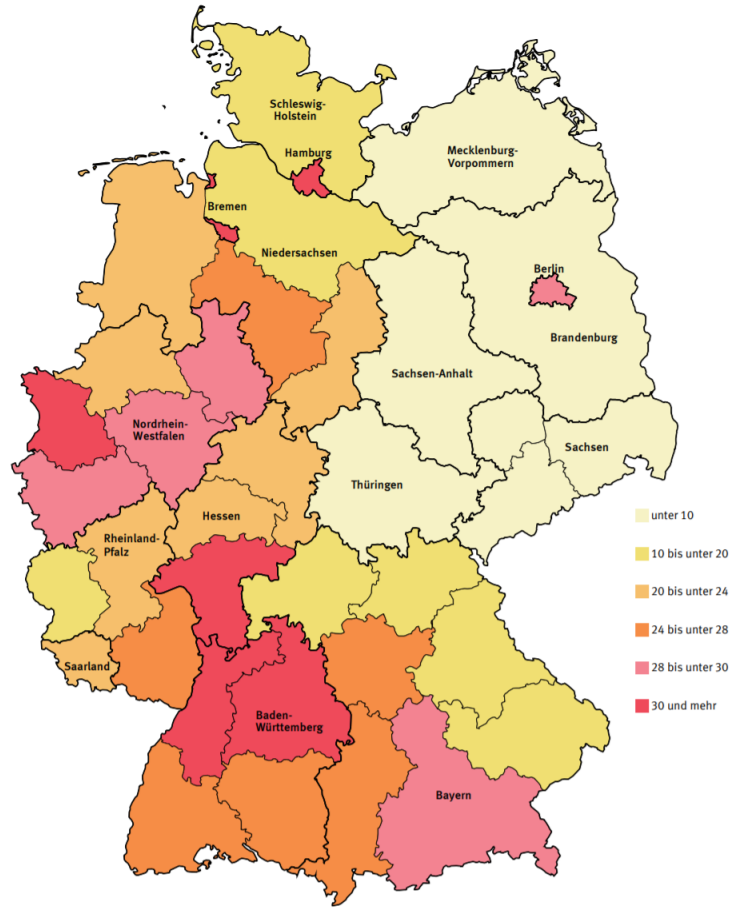


Figure 2: Percentage of residents in Germany with a migratory background by state. Source: Statistisches Bundesamt Report 2017, page 22.

share in comparison to Saxony (Sachsen). These differences are also apparent between districts within the German states. In our article, we show that higher geographical concentrations of foreign nationals in districts have a positive effect on the substantive representation behavior of corresponding members of parliament.

Furthermore, we examine the interaction of migration related committee memberships and party ideologies for matters of integration. In this context, we also analyze whether the electoral rules under which members of parliament came to power affect this interaction. For this, we differentiate between members of parliament elected in single-member plurality districts or multi-member districts (SMD tier) and those

elected under rules of closed-list proportional representation (PR tier). Our findings shed new light on the question whether electoral rules dominate the effects of legislative organization and candidate selection methods. Results suggest that, at least for the German case, electoral rules do not affect the engagement of members of parliament related to substantive representation of immigrants.

At last, we show that, in line with previous findings from the literature, members of parliament who are themselves of immigrant origin engage in more substantive representation in comparison to members of parliament without a migration background. The same is true for members of parliament who are members of migrant-related committees such as the committee for social affairs.

In summary, these findings from the second article of this dissertation contribute to the literature on political representation of ethnic minorities. The application of computational methods was crucial to obtain parliamentary written questions as a measure for substantive representation. As outlined in the second article, these questions are better indicators for the personal efforts of members of parliament in comparison to other legislative activities. This measurement approach in turn provided new insights for the political representation literature, suggesting that members of parliament remain responsive to the local demands of immigrant-origin citizens regardless of electoral rules.

Xenophobic collective action

While substantive representation is an example of political behavior for reducing unequal treatment of ethnic minorities, the third and last article of this dissertation

deals with a case where the opposite is true. In most Western-European countries radical right-wing and populist forces have increasingly gained influence in the last years (Arzheimer 2015). In Germany, the right-wing populist party *Alternative für Deutschland* became the third largest party in the Bundestag, which they first entered after the 2017 election. The party is associated with xenophobic and especially islamophobic attitudes and a harsh anti-immigrant agenda. For the establishment of the party, a grass roots movement called Pegida played an important role (W. J. Patzelt 2016). The movement is the focus of the third article and first caught public attention in 2014. Its supporters fear an increasing “alienation” of German culture and language by foreigners. In January 2015, a Pegida street rally attracted over 25,000 protesters. Although the public attention for the movement declined steadily soon after this peak, to this day most of its rallies are still joined by several hundred people.

In the third article, I analyze how the right-wing movement utilized social media to spread its xenophobic agenda and to mobilize supporters. Since the movement has been established, its administrators used Facebook as a platform for propaganda and mobilization, reaching over 100,000 likes within a few months (J. Patzelt W. K. 2016). The utilization of online platforms is in line with theories on social movements, which suggest that they are looking for ways to mobilize support for their cause and to acquire more resources (Opp 2009). The success of movements depends on factors such as common interests, shared identity, political power, supporter mobilization and resource availability (Tilly 1978; Harlow 2012). In this regard, social media platforms are a powerful tool for social movements, as they enable transna-

tional communication to reach a substantial amount of people . In addition, using social media platforms does not require a lot of resources to get started. Therefore, it is not surprising that previous studies already showed that several movements used social media platforms in the past and that their online activities can indeed affect on-site user mobilization (Budak and Watts 2015; Harlow 2012; Poell et al. 2016; Suh, Vasi, and Chang 2017).

However, despite an increasing availability of studies about the social media use of movements, we still don't know how exactly they utilize such platforms and what strategies they use to mobilize supporters. The following translated post by Pegida's administrators, which relates to the appearance of a former Pegida member in a German television show, demonstrates that they are well aware of the effects of links and hashtags on information diffusion mechanisms on the platform:

“Thanks Kathrin! You took our view very well and held your ground against the constantly interrupting, aggressive and arrogant CDU politician Spahn. Next time together with Rene or Lutz! This was only the first round which was clearly won by you! #DresdenShowsHowToDoIt PS: All the stupid comments on some watch-site - for which we do not want to provide reach with links or hashtags - obviously show how they boil with rage because of Kathrin's confident performance. Beforehand, they predicted a big disaster. Well, once again a proof that do-gooders just don't have a clue about anything.”

Facebook posts by Pegida administrators, created on January 19, 2015.

One of the most important aspects for information diffusion on social media platforms such as Facebook is user activity. Liking, commenting and sharing content of Pegida posts on Facebook affects how fast and to whom right-wing propaganda can spread on the platform (see Rieder et al. 2015 on Facebook algorithms). In the third article of this dissertation, I examine what factors influence the activity on Pegida's Facebook page. I argue that in order to gain a better understanding of

the social media usage of right-wing movements, it is important not to study social media in isolation, but rather to examine the interplay of social media with public activities and a movement's salience in the media. In order to do so, I apply a variety of computational methods (see the second part of this preface) to analyze and compare data from Pegida's Facebook page with the occurrence of exogenous shocks like terrorist attacks and Pegida's salience in the public sphere. This not only allows to shed light on variations in user activity, but also on changes in the topics Pegida issued in their Facebook posts.

Results of my analysis show that Pegida can not simply affect user activity on Facebook by posting more content. Although the administrators created more and more posts during the observed time period, the activity on the platform is mostly determined by changes in the public attention that Pegida receives and the content of its posts. Over time, the movement increasingly created more xenophobic material, which attracted more users than other themes like posts about demonstrations. Pegida resorted to more and more radical mobilization methods, underlining the responsibility of social media platforms to successfully detect and remove obnoxious content. Findings of this work also suggest a possible reinforcement process between the strategies of right-wing movements and the reactions of the audience: more radical posts lead to more user reactions and more reaction will eventually lead to more radicalized posts. This in turn results in less mobilization from the public, since more radical methods do not appeal to an audience with moderate ideology.

In summary, while a number of studies have already shown that right-wing movements use social media platforms for mobilization purposes, we could not learn from

these studies how strategic mobilization efforts are related to temporal dynamics and the public attention received by social movements. With the third article, I contribute to the literature on xenophobic collective action by examining these questions.

1.2 About the application of computational methods

The substantive contributions to the study of ethnic minorities outlined above were achieved by applying a variety of computational methods. Before I am going to outline the most important computational aspects with some examples, the question arises how *computational methods* can be defined from the viewpoint of social scientists. After all, in line with other disciplines, social scientists already use computers for their research (both quantitative and qualitative) since decades. If we all use computers to run our analysis, what would we consider as computational methods?

What are computational methods?

Lazer et al. were one of the first research groups to describe the field of Computational Social Science. They write (2009, 722f):

“In short, a computational social science is emerging that leverages the capacity to collect and analyze data with an unprecedented breadth and depth and scale.”

First, it is important to note that they mention both the collection as well as the analysis of data. Afterwards, what they describe as “unprecedented breadth, depth and scale” of data is similar to how the related term *big data* is often described in

the literature. A common definition is termed in relation to the *three V's*: Volume (a large amount of data), Velocity (data availability at rapid speed) and Variety (data in many forms, such as text, audio and video). Related to that, Ward and Baker (2013) conducted a survey of big data definitions. They concluded that all definitions of big data used in the literature mention the importance of at least one of the following aspects:

- **size**: the volume of the datasets
- **complexity**: the structure, behaviour and permutations of the datasets
- **technologies**: the tools and techniques used to process sizable or complex datasets

What can be taken from these studies is that computational methods are first and foremost methods for collecting, storing, processing, and analyzing data. These methods for the most part require experience with programming languages such as R or Python. Especially for social scientists, I think it is crucial to stress at this point that the size of the data is a possible, but not a necessary reason for the need of such programming skills. In this context, Riebling (2018) argues that challenges in applying computational methods more often lie in exogenous processes of data generation, which researchers can not control, and in working with complex data structures. As outlined by Salganik (2017, 18f), some researchers do indeed process large amounts of data while expressing their excitement about it:

“[Our] corpus contains over 500 billion words, in English (361 billion), French (45 billion), Spanish (45 billion), German (37 billion), Chinese (13 billion), Russian (35 billion), and Hebrew (2 billion). The oldest works were published in the 1500s. The early decades are represented by only a few books per year, comprising several hundred thousand words. By 1800, the corpus grows to 98 million words per year; by 1900, 1.8 billion; and by 2000, 11 billion. The corpus cannot be read by a human. If you tried to read only English-language entries from the year 2000 alone, at the reasonable pace of 200 words/min, without interruptions for food or sleep, it would take 80 years. The sequence of letters is 1000 times longer than the human genome: If you wrote it out in a straight line, it would reach to the Moon and back 10 times over.”

Michel et al. 2011

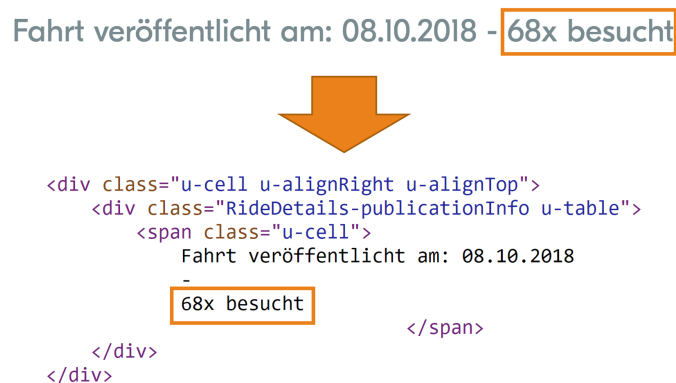
In order to answer some social science questions, for instance related to the study of very rare events, a large amount of data may be necessary. However, in general I highly doubt that the number of times data sequences reach to the moon and back correlates with the usefulness of the corresponding data for social science research. In fact, many questions social scientists might be interested in, including those examined in this dissertation, actually do not require very large amounts of data. With one exception discussed below, all the data used for the three research projects could be processed and analyzed on a single computer.

Examples of applying computational methods

In the context of this dissertation, computational methods were predominantly required to collect and process data from sources such as web pages, PDF files and programming interfaces. A major contribution of computational methods in all three articles of this dissertation is about processing data and reshaping it such that it can be used for analysis. In the second and third article, computational methods were also required for analyzing textual data.

Computational methods for data scraping

In the first article related to the discrimination of ethnic minorities, I draw on observable data in form of offered rides by drivers in Germany (see Figure 1). The process of extracting and preparing data from the carpooling platform required an extensive computational framework. First, programmatic procedures have been used to combine data from the carpooling provider's application programming interface (API) with additional data scraped from websites for each ride. This was necessary because the access to the provider's API did not include ride-specific information such as the main independent variable: consumer demand as measured by clicks on a ride. While web scraping techniques were used to a varying degree in all three articles, discussing the many challenges of automated web scraping in depth is beyond the scope of this preface (see Munzert et al. 2014 for an introduction to the topic). To only provide one example from the first article, Figure 3 shows a (simplified) concept of extracting the number of clicks on a ride from the source code of the carpooling website:



```
Fahrt veröffentlicht am: 08.10.2018 - 68x besucht
```



```
<div class="u-cell u-alignRight u-alignTop">
  <div class="RideDetails-publicationInfo u-table">
    <span class="u-cell">
      Fahrt veröffentlicht am: 08.10.2018
      -
      68x besucht
    </span>
  </div>
</div>
```

Figure 3: Web scraping example for carpooling data.

The top of the figure shows what users were able to see when they visited a ride-

specific side on the carpooling platform. The bottom of the figure shows the structure of the HTML code containing the information of interest. Finding this corresponding piece of code for each ride-specific site of our sample of initially 47,000 rides required an equal number of calls to the corresponding server for extracting the data. However, in order to be able to observe rides that were offered on short notice, this procedure had to be repeated several times per day. This was the only instance in this dissertation where the use of one single computer was insufficient, as the latency introduced by sending requests and receiving data from the server for thousands of rides was too high. For this reason, the extraction procedure was handled by distributing tasks over a cluster of computers. When it comes to distributed computing, an important question is whether the computational tasks of interest are dependant on each other.

In the carpooling case, there were no dependencies between different rides, which can be described as an *embarrassingly parallel* task (Herlihy and Shavit 2011, 14 ff.). The sample of rides could therefore be divided according to different routes from departure to arrival cities. Each computer in the cluster could then execute the scraping procedure for the corresponding routes and rides several times per day. In every iteration, trips detected in earlier steps were identified and merged subsequently. In addition, geographical data for federal states of all observed locations have been accessed. The whole procedure can be formalized in pseudo code as follows:

```

for each route  $r$  in routes do

    connect to carpooling API and retrieve all trips  $t \in r$ 

    for each  $t$  in  $r$  do

        extract information  $t_i$ 

        identify all trip locations  $l_1 \dots l_n$ 

        for each  $l$  in  $l_1 \dots l_n$  do

            extract federal state  $l_f$  from Google Geocoding API

            append  $l_f$  to  $t_i$ 

        end for

        identify trip specific url  $u$  in  $t_i$ 

        scrape html data for  $u$  and extract additional information  $u_i$ 

        append  $u_i$  to  $t_i$ 

    end for

    if  $t$  already in sample  $s$  then

        merge  $t$  with  $s_t$ 

    else

        append  $t$  to  $s$ 

    end if

end for

```

This procedure created a longitudinal sample of rides and reconstructs the platform from a consumer's point of view. This means that every observation in the resulting dataset contains information about a ride and the corresponding driver, given that a user had searched for according departure and arrival locations.

If the required sample for the first article would have only required a small number of observations, human labour could have been used for data collection instead of the presented computational architecture. In fact, for only a handful of rides the time required to manually collect the data several times per day would have been lower than the time that I spent to develop the computational architecture. However, a large sample was necessary in order to identify a sufficient number of (perceived) ethnic minority drivers and human labour does not scale nearly as good as computational architecture. I consider the significant reduction in resource costs as one of the most important contributions of computational methods for social science research.

Computational methods for unstructured data formats

In the first article, computational methods were used to create a dataset ready for statistical analysis. For the second article about substantive representation of disadvantaged immigrant groups, computational methods were used to retrieve and process data, but in part also for analysis. Data for the second article have been obtained within the project *Pathways to Power: The Political Representation of Citizens of Immigrant Origin in Seven European Democracies*. In this project, several research teams collected data such as socio-demographic information for members of parliament and macro-level data, for instance the share of foreign nationals in districts for several countries and legislatures. I was responsible for collecting parliamentary recordings, in particular the questions for written answers used as a measure for substantive representation in the second article. For some countries such as the United Kingdom, data for parliamentary recordings can be extracted

from APIs in structured formats.

In the case of the German Bundestag, to this date, parliamentary recordings are only provided as PDF files stored on an online server. For the 17th German Bundestag, about 89,000 files were available, but only 202 contained written questions. These were at first identified in an automated fashion, after which they could be extracted with web scraping techniques. Unfortunately, PDF files are a complicated format to work: they are often not very well structured and can not only include text, but also tabular data, images or even videos. Processing the data at first requires to convert the binary PDF to raw text. Afterwards, in order to extract the data of interest (in this case written questions) from raw text files, several pattern detection procedures, called regular expressions (see Sipser 2012, 63ff), had to be combined with other computational methods.

Starting at page 24 of this manuscript, a document from the 17th Bundestag which contains written questions is included. This example shows page 1 and page 27 of the corresponding document, which contains questions received by the government on August 19, 2013 (Bundestag 2013). Within one legislative period, the majority of documents from the Bundestag containing written questions are structured in a similar way, but not always exactly as shown in this example. Irregularities across documents can arise for instance due to missing parenthesis around party or due to the absence of line breaks after paragraphs. Programming code used to extract the questions therefore needed to be written in a way such that it detects irregularities, corrects them if possible and otherwise stores the corresponding documents for manual inspection.

Schriftliche Fragen

mit den in der Woche vom 19. August 2013
eingegangenen Antworten der Bundesregierung

Verzeichnis der Fragenden

| <i>Abgeordnete</i> | <i>Nummer der Frage</i> | <i>Abgeordnete</i> | <i>Nummer der Frage</i> |
|--|-----------------------------|---|-----------------------------|
| Aken, Jan van (DIE LINKE.) | 37 | Kotting-Uhl, Sylvia (BÜNDNIS 90/DIE GRÜNEN) | 77 |
| Bätzing-Lichtenthäler, Sabine (SPD) | 27 | Krumwiede, Agnes (BÜNDNIS 90/DIE GRÜNEN) | 1 |
| Dr. Bartels, Hans-Peter (SPD) | 50 | Liebing, Ingbert (CDU/CSU) | 78 |
| Bartol, Sören (SPD) | 57 | Lösekrug-Möller, Gabriele (SPD) | 44, 45 |
| Behm, Cornelia (BÜNDNIS 90/DIE GRÜNEN) | 48 | Maisch, Nicole (BÜNDNIS 90/DIE GRÜNEN) | 49 |
| Birkwald, Matthias W. (DIE LINKE.) | 28, 29 | Mattheis, Hilde (SPD) | 46 |
| Brähmig, Klaus (CDU/CSU) | 58, 59, 60 | Movassat, Niema (DIE LINKE.) | 79, 80 |
| Ehrmann, Siegmund (SPD) | 2, 3 | Müntefering, Franz (SPD) | 32, 33, 34 |
| Ernst, Klaus (DIE LINKE.) | 42, 43 | Dr. Mützenich, Rolf (SPD) | 10, 11 |
| Fograscher, Gabriele (SPD) | 16, 17 | Nahles, Andrea (SPD) | 53 |
| Groth, Annette (DIE LINKE.) | 4, 5, 6, 7 | Ostendorff, Friedrich (BÜNDNIS 90/DIE GRÜNEN) | 71, 72, 73, 74 |
| Hacker, Hans-Joachim (SPD) | 61, 62 | Roth, Claudia (Augsburg) (BÜNDNIS 90/DIE GRÜNEN) | 40 |
| Hagemann, Klaus (SPD) | 63 | Schäfer, Axel (Bochum) (SPD) | 20, 21 |
| Hellmich, Wolfgang (SPD) | 51 | Schäfer, Paul (Köln) (DIE LINKE.) | 12, 13, 14, 15 |
| Herzog, Gustav (SPD) | 64, 65 | Schäffler, Frank (FDP) | 35, 75 |
| Dr. Höll, Barbara (DIE LINKE.) | 30, 31 | Schmidt, Ulla (Aachen) (SPD) | 54, 55, 56 |
| Hoppe, Thilo (BÜNDNIS 90/DIE GRÜNEN) | 76 | Stüber, Sabine (DIE LINKE.) | 22, 23, 24 |
| Hunko, Andrej (DIE LINKE.) | 52, 66 | Dr. Tackmann, Kirsten (DIE LINKE.) | 25, 47 |
| Jelpke, Ulla (DIE LINKE.) | 18 | Tiefensee, Wolfgang (SPD) | 41 |
| Dr. Jüttner, Egon (CDU/CSU) | 38, 67 | Winkler, Josef Philip (BÜNDNIS 90/DIE GRÜNEN) | 26 |
| Keul, Katja (BÜNDNIS 90/DIE GRÜNEN) | 8, 39 | Ziegler, Dagmar (SPD) | 36 |
| Kipping, Katja (DIE LINKE.) | 68 | | |
| Koenigs, Tom (BÜNDNIS 90/DIE GRÜNEN) | 9 | | |
| Dr. Kofler, Bärbel (SPD) | 69, 70 | | |
| Korte, Jan (DIE LINKE.) | 19 | | |

nicht vorrangig auf das Alter, sondern vielmehr auf die besondere Situation Langzeitarbeitsloser abstellen. Die Überlegungen zu einem künftigen Programm befinden sich noch im Planungsstadium.

Die Anwendung der Steuerungslogik des Bundesprogramms in der Regelförderung nach dem SGB II wäre weitaus komplexer, als es im Bundesprogramm selbst der Fall ist. Das BMAS prüft derzeit Ansatzpunkte, wie eine Verknüpfung von Zielsteuerung und Ressourcenverteilung realisiert werden kann.

- | | |
|---|--|
| 45. Abgeordnete Gabriele Lösekrug-Möller (SPD) | Gedenkt die Bundesregierung, zukünftig in der Arbeitsförderung mehr auf Dienstleistung zu setzen, um durch einen verbesserten Personal- bzw. Betreuungsschlüssel bessere Ergebnisse zu erzielen? |
|---|--|

**Antwort des Staatssekretärs Gerd Hoofe
vom 22. August 2013**

Die Träger vor Ort bestimmen das Nähere über die Organisation und die Art der Leistungserbringung im Jobcenter; im Rahmen der Trägerversammlung wird über die Betreuungsschlüssel beraten und das örtliche Arbeitsmarkt- und Integrationsprogramm abgestimmt.

- | | |
|---|---|
| 46. Abgeordnete Hilde Mattheis (SPD) | Mit welchem Ergebnis hat die Bundesregierung ihr Prüfvorhaben umgesetzt, das im Entwurf des Vierten Armuts- und Reichtumsberichts (vom 17. September 2012) dahingehend formuliert war zu prüfen, „ob und wie über die Progression in der Einkommensteuer hinaus privater Reichtum für die nachhaltige Finanzierung öffentlicher Aufgaben herangezogen werden kann“ (S. XLII des Entwurfs), und im endgültigen Bericht lautete zu prüfen, „wie weiteres persönliches und finanzielles freiwilliges Engagement Vermögender in Deutschland für das Gemeinwohl eingeworben werden kann“ (S. XLVIII des Berichts), und wann ist mit der Veröffentlichung der Prüfung zu rechnen? |
|---|---|

**Antwort des Staatssekretärs Gerd Hoofe
vom 23. August 2013**

Das Thema des freiwilligen sozialen Engagements Vermögender war im Vierten Armuts- und Reichtumsbericht ein Schwerpunkt im Rahmen der Reichtumsberichterstattung. Privates Engagement baut nicht zuletzt dort Brücken, wo der Staat weniger flexibel, kreativ und zielgenau agieren könnte. Die Bundesregierung ermunterte deshalb im Bericht ausdrücklich zu mehr freiwilligem sozialem Engagement. Dieses ersetzt freilich nicht staatliches Handeln, sondern ergänzt dieses sinnvoll.

What follows is an iterative process where code is updated, refined and reapplied to process all documents. The following function (listing 1), which is written in the programming language Python, is one of many functions that were used to process the data for the second article.

In the beginning of the function (lines 5-17), multiple regular expressions are compiled to be used in the bottom part of the function. For each text document in the list of input files, the function extracts meta data like a document identifier and the document date from the corresponding first page. The remaining pages are then searched for instances of written questions, which in turn are processed and at last converted to a structured data type. Throughout the function, several exception blocks were inserted to document and print out potential errors, for instance for detecting and extracting the name of the politician who tabled the question (lines 34-38).

Listing 1: Python function for parsing written questions

```
1 import re
2 def parse_questions(drucksachen):
3     all_questions = {}
4     # find written questions
5     start_q = re.compile(r'[0-9]+\.\sAbgeordnete[r]{0,1}')
6     # replace line breaks
7     newlines = re.compile(r'\n\x0c.*?\n\n.*?\n\n.*?\n\n')
8     # split questions
9     split_question = re.compile('(.*)\n\n', re.DOTALL)
10    # detect mp name
11    mpname = re.compile(r'\n(.*?)\s*(?=\n)', re.DOTALL)
12    # detect mp party
13    mpparty = re.compile(r'\n(.*?)\s*', re.DOTALL)
```

```

14     # detect id
15     id_ = re.compile(r'Wahlperiode\n\n(.*)\n')
16     # detect date
17     date = re.compile(r'Wahlperiode\n\n.*?\n(.*)\n\n')
18
19     for document in drucksachen: # iterate over all documents
20         try:
21             metasplitter = document.split('1')[0]
22             identifier = re.search(id_, metasplitter).group(1)
23             docdate = re.search(date, metasplitter).group(1).replace('_', '')
24         except AttributeError:
25             print('Metaerror', metasplitter[:200], '=====\n')
26
27         docsplitter = document.split('1')[1]
28         parts = re.split(start_q, docsplitter)[1:]
29
30         for p in parts: # process single questions
31             fixed = re.sub(newlines, '', p)
32             splitted = re.split(split_question, fixed)
33             splitted = [i for i in splitted if len(i) > 1]
34             try:
35                 name = re.search(mpname, splitted[0]).group().replace('\n', '_')
36                 name = re.search(r'^(.*)$', name).group(1)
37             except AttributeError:
38                 print(p, '\nNameError=====\n')
39             try:
40                 party = re.search(mpparty, splitted[0]).group().replace('\n', '_')
41                 party = party.strip('().')
42             except AttributeError:
43                 print(identifier, p, '\nPartyError=====\n')
44             question = splitted[1].replace('\n', '_')
45             if len(question) < 10: # ignore empty entries
46                 pass
47             else:
48                 try:
49                     answer = splitted[2].replace('\n', '_')

```

```

50         except IndexError:
51             answer = -99
52             # add data to structured dictionary
53             helpdic = {'name': name, 'party': party}
54             if name not in all_questions:
55                 all_questions[name] = {'info': {}, 'texts': []}
56                 all_questions[name]['info'] = helpdic
57             qdic = {'date': docdate, 'druck_id': identifier,
58                    'question': question, 'answer': answer}
59             all_questions[name]['texts'].append(qdic)
60
61     return all_questions

```

From mathematical concepts to computer programs

After all written questions (about 20,000) were converted to a rectangular data format in sufficient quality, they needed to be matched with socio-demographic data about the members of parliament from the Pathways project. I mention this because it required another aspect of computational methods that was used throughout this dissertation: the conversion from mathematical concepts to program code. One simple yet important example is the Levenshtein distance (Levenshtein 1966). Given two sequences of characters (strings), for instance the name strings “Hilde Mattheis” and “Hilde Mattheus”, the Levenshtein distance returns the minimum number of operations required to convert one string into the other.

The Levenshtein distance between two strings a and b can be written as

$$\text{lev}_{a,b}(i, j) = \begin{cases} \max(i, j) & \text{if } \min(i, j) = 0, \\ \min \begin{cases} \text{lev}_{a,b}(i-1, j) + 1 \\ \text{lev}_{a,b}(i, j-1) + 1 \\ \text{lev}_{a,b}(i-1, j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

where $1_{(a_i \neq b_j)}$ is equal to 0 when $a_i = b_j$ and equal to 1 otherwise, and $\text{lev}_{a,b}(i, j)$ is the distance between the first i characters of a and the first j characters of b . The three operations refer to the deletion, insertion, or replacement of a single character. The following function (listing 2) provides an example for a matrix-based Python implementation of the Levenshtein distance:

Listing 2: Python implementation of the Levenshtein distance

```

1 def levenshtein(a, b):
2     rows = len(a) + 1 # number of rows
3     cols = len(b) + 1 # number of columns
4     matrix = [[0 for i in range(cols)] for i in range(rows)]
5
6     for i in range(1, rows):
7         matrix[i][0] = i
8     for j in range(1, cols):
9         matrix[0][j] = j
10    for col in range(1, cols):
11        for row in range(1, rows):
12            if a[row-1] == b[col-1]: # test equality
13                cost = 0
14            else:

```



```

15         cost = 1
16         matrix[row][col] = min(matrix[row-1][col] + 1, # deletion
17                                matrix[row][col-1] + 1, # insertion
18                                matrix[row-1][col-1] + cost) # replacement
19
20     return matrix[row][col]

```

Given two input strings $a = \text{"Hilde Mattheis"}$ and $b = \text{"Hilde Mattheus"}$, the function returns 2 as result, as this is the number of operations required to convert string a to b : the deletion of the first “l” and the replacement of the “u” with an “i” in string b . This concept proved to be very useful for semi-automated matching of the several hundred names of members of parliament in the Bundestag documents with the names from the Pathways data. Similar concepts were also required for the third article to identify duplicate (or almost identical) news reports about Pegida.

Computational methods for data analysis

Until now, all examples I provided were related to getting and processing data ready for analysis. In the second and third article of this dissertation, computational methods were also applied in form of machine learning techniques (e.g. implementations by Pedregosa et al. 2011) for analyzing textual data. Although the application of such techniques is often associated with computer science, they are in principle quite familiar to tools commonly used by social scientists, such as regression models. A major difference between the disciplines is that computer scientists and social scientists are interested in different parameters. Computer scientists predominantly focus on \hat{y} , that is the best possible prediction results for an outcome of interest.

In comparison, social scientists are more often interested in $\hat{\beta}$ estimates, that is the coefficients for some explanatory variables to gain an understanding about the relation between these variables and an outcome of interest (see Mullainathan and Spiess 2017 for a more elaborate comparison). Switching the focus to \hat{y} for applying machine learning techniques is in my opinion not particularly difficult for social scientists who are already trained in the use of statistical models. For the sake of brevity, I will therefore only provide two short summaries for the use of supervised machine learning techniques in this dissertation, which were predominantly used in articles two and three.

In the second article, written questions were identified that are relevant for the substantive representation of immigrants. After labeling a subset of written questions manually (see the appendix of Article 2), several classification models were trained to predict the labels of all 20,000 questions in our corpus by learning from the labels in our hand-coded subset. For our training data, the best machine learning models achieved the same predictive performance in comparison to a dictionary approach. However, for predictions of questions in our entire corpus, a dictionary look-up produced better results, indicating a potential overfit of the machine learning models. For this reason, we ultimately used a dictionary look-up instead. Although the machine learning approach was discarded, it nevertheless was helpful for refining our dictionary for substantive representation. Over all, this procedure was vital to reduce the costs for the second article, as labeling over 20,000 questions would have required a very long time and/or a substantial amount of financial resources. In the third article, news reports about Pegida were analyzed to understand the time-

dependent salience of the movement. A machine learning model was trained on time stamps of news reports in order to find the most important terms for correctly predicting several time intervals. In conjunction with aggregated daily counts for the number of news articles related to Pegida, this provides a measure of context-enriched issue salience. This approach was also important to understand how media content about Pegida changed over time and which public events were mentioned in articles about the movement.

After briefly discussing machine learning applications, I would like to focus on a final computational aspect the development of research software. In the context of this dissertation, I created software applications that were primarily used for data analysis. To this date, the majority of social scientists conducting some kind of quantitative analysis rely on proprietary closed-source software such as SPSS or Stata. These programs are specialized for the analysis of data coming in rectangular, spread-sheet like formats. These software solutions come with a number of benefits. For example, they are considered easy to use, they offer a graphical user interface and their functionality is well documented. However, relying on proprietary closed-source software also comes with severe limitations. To name a few, SPSS or Stata are not available for free. Depending on the corresponding licence, using them might require to spend a significant amount of money. Moreover, their source code is not available to the public, which effectively turns them into black boxes. For instance, although Stata offers a promising range of methods for estimating both basic and highly specific statistical models, users have to trust the company with regards to their functionality and are not able to look behind the wheels. Not being able to

inspect algorithms and their inner workings is a severe problem for core principles of scientific research (see Heiberger and J. R. Riebling 2016, 6f and Trilling and Jonkman 2018, 7f).

As of lately, an increasing number of academics - including social scientists - is using open-source programming languages, such as Python or R, in supplement to or as a replacement for proprietary software (Lindeløv 2019). Both Python and R were used extensively for the research projects of this dissertation. Over the last years, they have improved in a number of ways. For example, the programming language R is becoming more and more beginner friendly. This is especially helpful for social scientists who usually have little to no knowledge about computer programming. In addition, thanks to many users who create R packages to expand the functionality of the language, R now also covers the majority of analyses that would be of interest for social scientists. Most importantly, R is an open-source language, meaning every line of code can be inspected by the user and R can be downloaded and used for free. I consider the development of research software an important aspect of computational methods for social science. In the context of this dissertation, I developed two R packages that proved to be very useful for research collaborations and content analysis.

The first package was developed in the context of the aforementioned project Pathways. The *pathways* package (Schwemmer 2019a) contains all the datasets for parliamentary recording that I collected via automated procedure such as described above for the German Bundestag. In addition, the package includes a graphical interface to explore these datasets, which is illustrated in Figure 4.

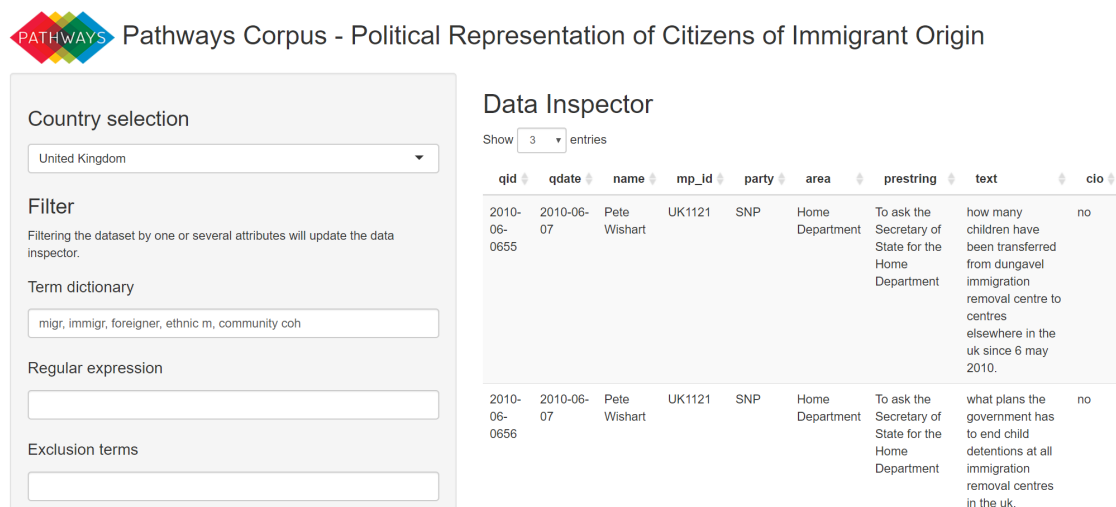


Figure 4: Graphical interface of the software package *pathways*

The software interface allows researchers to quickly inspect country-specific datasets, filter them based upon multiple attributes and create simple visualizations. The example in Figure 4 shows written questions tabled by members of parliament in the United Kingdom. The corpus is filtered on the left-hand side by a couple of strings that could for instance be used to identify the substantive representation of minorities. This proved to be very helpful for the development of the dictionary that was ultimately used to categorize written questions related to immigrant groups for the second article. Moreover, as the Pathways project covered data in several languages, such as German, English, French and Greek, the graphical interface proved useful for collaborative research purposes. Native speakers of certain languages were able to use the tool for inspecting the data and generating dictionaries without the need to write a single line of R code.

The second example for research software that I developed supports researchers working with topic models for analyzing textual data. As outlined in more detail in

the third article, topic models are statistical models for reducing the dimensionality of textual data. They were used in the third article of this dissertation to first gain an understanding of the content of Facebook posts by Pegida. The recently introduced variant of topic models I used is named *Structural Topic Model* (M. E. Roberts et al. 2014). This extension of conventional topic models then allowed in a second step to examine changes in the topics covered in Pegida posts over time, which was substantial for the contributions of the third article. Although the approach of topic modeling is a quantitative approach, model selection and validation of topic model results can be quite labor intensive, as it requires qualitative inspection of many documents and terms. In order to make these tasks easier for researchers, I developed the R package *stminsights* (Schwemmer 2018). It was used in the third article and several other research projects (e.g. Schwemmer and Ziewiecki 2018; Fischer-Prefler, Schwemmer, and Fischbach 2019; Rodriguez and Storer 2019). An illustration of the main interface is shown in Figure 5.

The package enables interactive validation, interpretation and visualization of one or several structural topic models. It also includes a range of utility functions for converting outputs of the vanilla R package for structural topic models (M. Roberts, Stewart, and Tingley 2015) to modern R frameworks for data analysis (Wickham 2016). At the time of writing, *stminsights* has been downloaded over 4000 times from the R package repository (CRAN).

Writing software for research purposes comes with many benefits, but also with challenges. From the viewpoint of a social scientist, an important factor for soft-



Figure 5: Interface of research software *stminsights*

ware development that needs to be considered is time. There are many reasons why writing software is time intensive. Hadley Wickham wrote several books about related content with focus on programming and software development with R (see Wickham 2014; Wickham 2015). Unlike research papers, which usually do not require to invest additional time once they are (finally) published, software needs to be maintained on a regular basis. Packages need to be updated for various reasons, for instance for adding new functionality or fixing errors. In this context, another aspect - which I believe social scientists often do not consider - are package dependencies. Developing software usually builds upon existing software in order to use resources efficiently and not reinvent the wheel. To provide one final example for challenges of applying computational methods, Figure 6 shows dependencies for *stminsights*. These dependencies are illustrated as a network, where nodes correspond

For example, *stminsights* depends on *stm* (level 1) and this package in turn relies on code from *glmnets* (level 2). Even from this simplified illustration it becomes apparent that software dependencies can be complex. Maintaining packages therefore requires the need to monitor changes to the code base of dependencies on a regular basis. Likewise, software should be written in a way that updates will not break working code from earlier versions of other packages.

Despite such hurdles for software development, I would advocate for more social scientists to engage in this endeavor. With an increasing availability of data relevant for answering social science questions, the need for specialized tools will grow simultaneously. Social scientists have specific needs for their research software and one possible way to address those needs is to develop software by themselves. For the third article, the use of topic models for analyzing Facebook posts by Pegida administrators was important to understand the social media strategies of the right-wing movement.

To find a model that is suited for this research task, computer scientists would for the most part rely on optimizing some performance metric, for instance word predictions for out-of-sample documents. However, such procedures do not necessarily result in finding a model that is also useful for social scientists. This turned out to be true for the third article as well. For this reason, I developed *stminsights* to assist researchers in qualitatively examining documents and output from topic models in order to find the best model for substantive insights. While using the application for the third article, I discovered the Pegida post quoted in section 1.1, which showed that the movement is well aware of the effects of links and hashtags on information diffusion

mechanisms on Facebook. This was an important qualitative finding for this article, which provided further evidence for the strategic use of Facebook by Pegida.

In summary, the examples provided for data scraping, working with unstructured data formats, methods for data analysis and the creation of research software demonstrate how computational methods can be utilized to conduct innovative studies of ethnic minorities.

1.3 Concluding remarks

With this preface, I outlined the most important contributions of this dissertation for the study of ethnic minorities. It has been shown that computational methods are helpful for answering fundamental questions about interactions between mainstream societies and minorities across three interconnected domains: ethnic discrimination, political representation of minorities and collective action driven by xenophobia. Regarding ethnic discrimination, computational methods helped to gain novel insights into mechanisms of subtle, everyday forms of unequal treatment in social markets. Such methods were also crucial to enhance our knowledge on substantive representation of immigrant groups in Germany, showing for instance that members of parliament respond to local concentrations of immigrant voters. At last, the application of computational methods shed new light on social media strategies of right-wing movements such as Pegida, which resorted to increasingly radical mobilization methods over time, underlining the responsibility of online platforms to detect and remove obnoxious content.

In the course of this preface, I discussed the potential, but also the challenges that

come with applying such methods. Regarding the question why computational methods are becoming increasingly important, social scientists have used data for several decades to learn about the social world. In this context, social scientists have also become experts in understanding the limitations of their data, which was predominantly based upon interviewing people in surveys. We are well aware of phenomena such as the social desirability bias (Edwards 1957) and came up with survey and questionnaire techniques to reduce such biases (Presser and Stinson 1998; Hanmer, Banks, and White 2014; Porst 2014). What has changed is the increasing availability of large amounts of data that comes in many forms, which introduces new technical challenges for social scientists.

Data from online markets or social media platforms allows us to observe human behavior and can help to answer important questions that can not be resolved with survey data. However, such data not only requires computational methods to be useful for social scientist, but also comes with other challenges. Several data sources are simply not available for researchers (Salganik 2017, 27f). For example, Facebook shut down its API access for public pages in 2017, which basically makes it impossible to conduct quantitative analyses without violating the terms of service (see Schwemmer, Bolle, and Seeberg 2018 for a related interview). For that reason, scholars are speaking of a *Post-API-Age* (Freelon 2018) and some researchers are even suing governments to regain access to important data sources (Wilson and Misllove 2017). Moreover, the underlying data generating process of black boxes that we use to retrieve data, like the Twitter sample endpoint, is for the most part unknown. This makes it difficult to draw reasonable inferences (Grimmer 2015), especially be-

cause researchers have shown that such data sources can be manipulated (Pfeffer, Mayer, and Morstatter 2018).

Such difficulties are further amplified when social scientists and computer scientists conduct interdisciplinary research. Both disciplines are fundamentally interested in different phenomena and treat data in different ways, as described by the following quote:

[C]omputer scientists may be interested in finding the needle in the haystack (such as [...] the right web page to display from a search), but social scientists are more commonly interested in characterizing the haystack. Certainly, individual document classifications, when available, provide additional information to social scientists, since they enable one to aggregate in unanticipated ways, serve as variables in regression-type analyses, and help guide deeper qualitative inquiries into the nature of specific documents. But they do not usually [...] constitute the ultimate quantities of interest.

Hopkins and King 2010, 230f

Research in the field of computational social science therefore is not just about the application of computational methods in a black-box fashion, as if computational social science were simply computer science plus social data (Wallach 2018). Rather, it is crucial to connect the disciplines in a way that translates fundamental questions of computer science to social science and vice versa. I consider it an important contribution of my work to bridge the gap between disciplines in the field of computational social science. The example of my software packages on the one hand demonstrate technical and methodological contributions to the field. On the other hand, my research software is designed in a way that, besides overcoming technical hurdles and working with complex data structures, it can be used to conduct fundamental social science research questions without sophisticated programming knowledge. Recently,

computational social scholars have acknowledged the usefulness of this *translation service* and used one of my applications to examine how quantitative topic models can be used for qualitative research (Rodriguez and Storer 2019).

Besides the need to bridge the gap between disciplines, how is the field of computational social science supposed to move forward? An increasing number of scholars, including myself, agrees that an important step forward is to combine custom-made data, for instance from surveys, with ready-made data, for example from online platforms (Salganik 2017, p. 355). This allows to combine the advantages of both worlds to overcome limitations. For this reason, I believe that the need for computational methods in the social sciences will continuously increase in the next years. It is interesting to note that in 2009, Lazer et al. (2009, 722f) already discussed the importance of training scholars in the application of computational methods. In this context, they raised the question of how this training might look like in the future:

The emergence of a computational social science shares with other nascent interdisciplinary fields [...] the need to develop a paradigm for training new scholars. [...] In the long run, the question will be whether academia should nurture computational social scientists, or teams of computationally literate social scientists and socially literate computer scientists.

Lazer et al. 2009, 230f

At the time of writing in 2019, most scholars would probably agree that the question of how computational methods should find their way into education curricula is still unanswered. An increasing number of higher education institutes is offering modules or even entire degrees related to computational methods for social scientists. It is possible that, another ten years from now, the field of computational methods will be taught in the majority of (quantitatively oriented) social science degrees, or even

that what we today consider as basics of empirical social research will include computational methods. Unfortunately, at the time of writing, computational methods for the social sciences are still far away from being part of conventional social science programs. With this dissertation, I provided several examples for why this should change. It will be my goal for the upcoming years to accelerate this change.

References

- Ahmed, A. M., L. Andersson, and M. Hammarstedt (2010). “Can Discrimination in the Housing Market Be Reduced by Increasing the Information about the Applicants?” In: *Land Economics* 86.1, pp. 79–90. URL: <http://le.uwpress.org/cgi/doi/10.3368/le.86.1.79>.
- Alba, Richard and Nancy Foner (2015). *Strangers no more: Immigration and the challenges of integration in North America and Western Europe*. Princeton University Press.
- Arzheimer, Kai (2015). “The AfD: Finally a Successful Right-Wing Populist Eurosceptic Party for Germany?” In: *West European Politics* 38.3, pp. 535–556.
- Aydemir, Nermin and Rens Vliegthart (2016). “‘Minority Representatives’ in the Netherlands: Supporting, Silencing or Suppressing?: Table 1”. In: *Parliamentary Affairs* 69.1, pp. 73–92. URL: <https://academic.oup.com/pa/article-lookup/doi/10.1093/pa/gsv009>.
- Becker, Gary S (1971). *The economics of discrimination*. University of Chicago press.
- Bertrand, Marianne and Sendhil Mullainathan (2004). “Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination”. In: *American Economic Review* 94.4, pp. 991–1013. URL: <http://pubs.aeaweb.org/doi/10.1257/0002828042002561>.
- Bloemraad, Irene, Els de Graauw, and Rebecca Hamlin (2015). “Immigrants in the Media: Civic Visibility in the USA and Canada”. In: *Journal of Ethnic and Migration Studies* 41.6, pp. 874–896.
- Budak, Ceren and Duncan Watts (2015). “Dissecting the Spirit of Gezi: Influence vs. Selection in the Occupy Gezi Movement”. In: *Sociological Science* 2, pp. 370–397. URL: <https://www.sociologicalscience.com/articles-v2-18-370>.
- Bundestag (2013). *Drucksache 17/14617*. URL: <http://dipbt.bundestag.de:80/dip21/btd/17/146/1714617.pdf>.
- Dahl, Robert A. (1971). *Polyarchy: Participation and Opposition*. New Haven: Yale University.

- Doleac, Jennifer L. and Luke C.D. Stein (2013). “The Visible Hand: Race and Online Market Outcomes”. In: *The Economic Journal* 123.572, F469–F492. URL: <https://academic.oup.com/ej/article/123/572/F469/5080452>.
- Dovi, Suzanne (2002). “Preferable descriptive representatives: Will just any woman, black, or latino do?” In: *American Political Science Review* 96.4, pp. 729–743.
- Edwards, Allen L (1957). *The social desirability variable in personality assessment and research*. Dryden Press.
- Fernandes, Jorge M, Cristina Leston-Bandeira, and Carsten Schwemmer (2017). “Election proximity and representation focus in party-constrained environments”. In: *Party Politics*.
- Fischer-Preßler, Diana, Carsten Schwemmer, and Kai Fischbach (2019). “Collective sense-making in times of crisis: Connecting terror management theory with twitter reactions to the Berlin terrorist attack”. In: *Computers in Human Behavior*. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0747563219301876>.
- Freelon, Deen (2018). “Computational Research in the Post-API Age”. In: *Political Communication*, pp. 1–4. URL: <https://www.tandfonline.com/doi/full/10.1080/10584609.2018.1477506>.
- Geese, Lucas and Carsten Schwemmer (2019). “MPs’ principals and the substantive representation of disadvantaged immigrant groups”. In: *West European Politics* 42.4, pp. 681–704. URL: <https://www.tandfonline.com/doi/full/10.1080/01402382.2018.1560196>.
- Grimmer, Justin (2015). “We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together”. In: *PS: Political Science & Politics* 48.01, pp. 80–83. URL: http://www.journals.cambridge.org/abstract_S1049096514001784.
- Hanmer, Michael J., Antoine J. Banks, and Ismail K. White (2014). “Experiments to Reduce the Over-Reporting of Voting: A Pipeline to the Truth”. In: *Political Analysis* 22.1, pp. 130–141. URL: https://www.cambridge.org/core/product/identifier/S1047198700013644/type/journal_article.
- Harlow, Summer (2012). “Social media and social movements: Facebook and an online Guatemalan justice movement that moved offline”. In: *New Media & So-*

- ciety* 14.2, pp. 225–243. URL: <http://journals.sagepub.com/doi/10.1177/1461444811410408>.
- Heckman, James J (1998). “Detecting Discrimination”. In: *Journal of Economic Perspectives* 12.2, pp. 101–116. URL: <http://pubs.aeaweb.org/doi/10.1257/jep.12.2.101>.
- Heiberger, Raphael H. and Jan R. Riebling (2016). “Installing computational social science: Facing the challenges of new information and communication technologies in social science”. In: *Methodological Innovations* 9, p. 205979911562276. URL: <http://journals.sagepub.com/doi/10.1177/2059799115622763>.
- Herlihy, Maurice and Nir Shavit (2011). *The art of multiprocessor programming*. Morgan Kaufmann.
- Hopkins, Daniel J. and Gary King (2010). “A Method of Automated Nonparametric Content Analysis for Social Science”. In: *American Journal of Political Science* 54.1, pp. 229–247. URL: <http://doi.wiley.com/10.1111/j.1540-5907.2009.00428.x>.
- Jungherr, Andreas (2018). “Normalizing Digital Trace Data”. In: *Digital Discussions - How Big Data Informs Political Communication*. Ed. by Natalie Jomini Stroud and Shannon C. McGregor. Routledge. URL: <https://www.taylorfrancis.com/books/9781351209427>.
- Lazer, D. et al. (2009). “Social Science: Computational Social Science”. In: *Science* 323.5915, pp. 721–723. URL: <http://www.sciencemag.org/cgi/doi/10.1126/science.1167742>.
- Levenshtein, Vladimir I (1966). “Binary codes capable of correcting deletions, insertions, and reversals”. In: *Soviet physics doklady*, pp. 707–710.
- Lin, Ken-Hou and Jennifer Lundquist (2013). “Mate Selection in Cyberspace: The Intersection of Race, Gender, and Education”. In: *American Journal of Sociology* 119.1, pp. 183–215. URL: <https://www.journals.uchicago.edu/doi/10.1086/673129>.
- Lindeløv, Jonas (2019). *SPSS is dying. It's time to change*. URL: <https://lindeloev.net/spss-is-dying/>.

- Mansbridge, Jane (1999). "Should Blacks Represent Blacks and Women Represent Women? A Contingent "Yes"". In: *The Journal of Politics* 61.03, p. 628.
- Martin, Shane (2011). "Parliamentary Questions, the Behaviour of Legislators, and the Function of Legislatures: An Introduction". In: *The Journal of Legislative Studies* 17.3, pp. 259–270. URL: <http://www.tandfonline.com/doi/abs/10.1080/13572334.2011.595120>.
- Michel, J.-B. et al. (2011). "Quantitative Analysis of Culture Using Millions of Digitized Books". In: *Science* 331.6014, pp. 176–182. URL: <http://www.sciencemag.org/cgi/doi/10.1126/science.1199644>.
- Mullainathan, Sendhil and Jann Spiess (2017). "Machine Learning: An Applied Econometric Approach". In: *Journal of Economic Perspectives* 31.2, pp. 87–106. URL: <http://pubs.aeaweb.org/doi/10.1257/jep.31.2.87>.
- Munzert, Simon et al. (2014). *Automated Data Collection with R*. Chichester, UK: John Wiley & Sons, Ltd. URL: <http://doi.wiley.com/10.1002/9781118834732>.
- Opp, Karl-Dieter (2009). *Theories of political protest and social movements: A multidisciplinary introduction, critique, and synthesis*. Routledge.
- Pager, Devah, Bart Bonikowski, and Bruce Western (2009). "Discrimination in a Low-Wage Labor Market". In: *American Sociological Review* 74.5, pp. 777–799. URL: <http://journals.sagepub.com/doi/10.1177/000312240907400505>.
- Pager, Devah and Hana Shepherd (2008). "The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets". In: *Annual Review of Sociology* 34.1, pp. 181–209. URL: <http://www.annualreviews.org/doi/10.1146/annurev.soc.33.040406.131740>.
- Patzelt W; Klose, Joachim (2016). *PEGIDA. Warnsignale aus Dresden*. Social coherence studies 3. Dresden: Thelem.
- Patzelt, Werner J. (2016). *"Rassisten, Extremisten, Vulgärdemokraten!" Hat sich PEGIDA radikalisiert?* Dresden. URL: <https://www.docdroid.net/M5uwYZS/pegida-studie-januar-2016-finale-ppt.pdf.html>.
- Pedregosa, Fabian et al. (2011). "Scikit-learn: Machine Learning in {P}ython". In: *Journal of Machine Learning Research* 12.Oct, pp. 2825–2830.

- Pedulla, David S (2018). “How Race and Unemployment Shape Labor Market Opportunities: Additive, Amplified, or Muted Effects?” In: *Social Forces* 96.4, pp. 1477–1506. URL: <https://academic.oup.com/sf/article/96/4/1477/4938480>.
- Pfeffer, Jürgen, Katja Mayer, and Fred Morstatter (2018). “Tampering with Twitter’s Sample API”. In: *EPJ Data Science* 7.1, p. 50. URL: <https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-018-0178-0>.
- Pitkin, Hanna Fenichel (1967). *The Concept of Representation*. Berkeley: University of California Press.
- Poell, Thomas et al. (2016). “Protest leadership in the age of social media”. In: *Information Communication and Society* 19.7, pp. 994–1014.
- Porst, Rolf (2014). *Fragebogen*. Wiesbaden: Springer Fachmedien Wiesbaden. URL: <http://link.springer.com/10.1007/978-3-658-02118-4>.
- Presser, Stanley and Linda Stinson (1998). “Data Collection Mode and Social Desirability Bias in Self-Reported Religious Attendance”. In: *American Sociological Review* 63.1, p. 137. URL: <http://www.jstor.org/stable/2657486?origin=crossref>.
- Riebling, Jan (2018). “The Medium Data Problem in Social Science”. In: *Computational Social Science in the Age of Big Data. Concepts, Methodologies, Tools, and Applications*. Ed. by Cathleen Stuetzner, Martin Egger, and Welker Marc. Neue Schri. Halem Verlag, pp. 76–100.
- Rieder, Bernhard et al. (2015). “Data critique and analytical opportunities for very large Facebook Pages: Lessons learned from exploring “We are all Khaled Said””. In: *Big Data & Society* 2.2, p. 205395171561498. URL: <http://journals.sagepub.com/doi/10.1177/2053951715614980>.
- Roberts, Margaret E et al. (2014). “Structural Topic Models for Open-Ended Survey Responses Despite broad use of surveys and survey and”. In: *American Journal of Political Science* 58.4, pp. 1064–1082.
- Roberts, Margaret, Brandon Stewart, and Dustin Tingley (2015). *stm: R Package for Structural Topic Models*. URL: <http://www.structuraltopicmodel.com>.
- Rodriguez, Maria Y. and Heather Storer (2019). “A computational social science perspective on qualitative data exploration: Using topic models for the descriptive

- analysis of social media data*". In: *Journal of Technology in Human Services*, pp. 1–32. URL: <https://www.tandfonline.com/doi/full/10.1080/15228835.2019.1616350>.
- Salganik, Matthew (2017). *Bit by Bit: Social Research in the Digital Age*. Open Revie. Princeton, NJ: Princeton University Press, p. 448.
- Schwemmer, Carsten (2018). *stminsights. A 'Shiny' Application for Inspecting Structural Topic Models*. URL: <https://cschwem2er.github.io/stminsights/>.
- (2019a). *pathways: A 'Shiny' Application for Exploring the Pathways Corpus*. URL: <https://cschwem2er.github.io/pathways/>.
- (2019b). “Social Media Strategies of Right-Wing Movements - The Radicalization of Pegida”. URL: <https://osf.io/preprints/socarxiv/js73z/>.
- Schwemmer, Carsten, Anja Bolle, and David Seeberg (2018). *Facebook macht dicht - Datenskandal beeinträchtigt Sozialforschung: Interview mit Carsten Schwemmer*. URL: <https://detektor.fm/digital/datenskandal-und-wissenschaft>.
- Schwemmer, Carsten and Sandra Ziewiecki (2018). “Social Media Sellout: The Increasing Role of Product Promotion on YouTube”. In: *Social Media + Society* 4.3, p. 205630511878672. URL: <http://journals.sagepub.com/doi/10.1177/2056305118786720>.
- Sipser, Michael (2012). *Introduction to the Theory of Computation*. Cengage Learning.
- Statistisches Bundesamt (2017). *Bevölkerung mit Migrationshintergrund - Ergebnisse des Mikrozensus 2017*. Tech. rep.
- Suh, Chan S., Ion Bogdan Vasi, and Paul Y. Chang (2017). “How social media matter: Repression and the diffusion of the Occupy Wall Street movement”. In: *Social Science Research* 65, pp. 282–293.
- Tilly, Charles (1978). *From Mobilization to Revolution*. Tech. rep. Addison-Wesley, pp. 1–20.
- Tjaden, Jasper Dag, Carsten Schwemmer, and Menusch Khadjavi (2018). “Ride with Me—Ethnic Discrimination, Social Markets, and the Sharing Economy”. In: *European Sociological Review* 34.4, pp. 418–432. URL: <https://academic.oup.com/esr/article/34/4/418/5048414>.

- Trilling, Damian and Jeroen G. F. Jonkman (2018). “Scaling up Content Analysis”. In: *Communication Methods and Measures* 12.2-3, pp. 158–174. URL: <https://www.tandfonline.com/doi/full/10.1080/19312458.2018.1447655>.
- Wallach, Hanna (2018). “Computational social science is not equal to computer science plus social data”. In: *Communications of the ACM* 61.3, pp. 42–44. URL: <http://dl.acm.org/citation.cfm?doid=3190347.3132698>.
- Ward, Jonathan Stuart and Adam Barker (2013). “Undefined By Data: A Survey of Big Data Definitions”. In: URL: <http://arxiv.org/abs/1309.5821>.
- Wickham, Hadley (2014). *Advanced r*. Chapman and Hall/CRC.
- (2015). *R packages: organize, test, document, and share your code*. " O'Reilly Media, Inc."
 - (2016). *tidyverse: Easily Install and Load 'Tidyverse' Packages*. URL: <https://cran.r-project.org/package=tidyverse>.
- Wilson, Christo and Alan Mislove (2017). *We're suing the federal government to be free to do our research*. URL: <http://theconversation.com/were-suing-the-federal-government-to-be-free-to-do-our-research-74676>.
- Wüst, Andreas M. (2014). “A Lasting Impact? On the Legislative Activities of Immigrant-origin Parliamentarians in Germany”. In: *The Journal of Legislative Studies* 20.4, pp. 495–515. URL: <http://www.tandfonline.com/doi/abs/10.1080/13572334.2014.907601>.

2 First Article: Ride with Me - Ethnic Discrimination, Social Markets, and the Sharing Economy

This manuscript was accepted for publication at the journal *European Sociological Review* in June 2018. It is available online and in print.

Jasper Dag Tjaden, Carsten Schwemmer, and Menusch Khadjavi (2018). “Ride with Me—Ethnic Discrimination, Social Markets, and the Sharing Economy”. In: *European Sociological Review* 34.4, pp. 418–432. URL: <https://academic.oup.com/esr/article/34/4/418/5048414>

Ride with Me - Ethnic Discrimination, Social Markets and the Sharing Economy*

Jasper Dag Tjaden¹, Carsten Schwemmer² & Menusch Khadjavi³

[Published in European Sociological Review: <https://doi.org/10.1093/esr/jcy024>]

Abstract

We study ethnic discrimination in the sharing economy using the example of online carpooling marketplaces. Based on a unique dataset of 16,624 real rides from Germany, we estimate the effects of drivers' perceived name origins on the demand for rides. The results show sizable ethnic discrimination – a discriminatory price premium of about 32% of the average market price. Further analyses suggest that additional information about actors in this market decreases the magnitude of ethnic discrimination. Our findings broaden the perspective of ethnic discrimination by shedding light on subtle, everyday forms of discrimination in social markets; inform ongoing discussions about ways to address discrimination in an era in which markets gradually move online; and respond to increasingly recognized limitations of experimental approaches to study discrimination.

Keywords

Ethnic discrimination; sharing economy, statistical discrimination; online markets; computational social science

* We would like to thank the three anonymous reviewers for their time and constructive feedback. We are also grateful for comments on previous versions from Uri Gneezy, Ruud Koopmans, Jack DeWaard, Marc Helbling, Sebastian Wenz and Cornelia Kristen.

¹ Corresponding author: Global Migration Data Analysis Centre, International Organisation for Migration, Taubenstraße 20-22, 10117 Berlin, Germany

² Social Sciences, Economics, and Business Administration Faculty, Chair of Political Sociology, University of Bamberg, Germany

³ Kiel Institute for the World Economy and Department of Economics, Christian Albrechts-University Kiel, Germany

1. Introduction

Decades of social science research provide evidence of ethnic and racial discrimination in various areas of society and in numerous countries (e.g. Pager, 2007; Pager & Shepherd, 2008; Rich, 2014). Despite a long history of policy responses and the introduction of anti-discrimination legislation designed to attenuate ethnic and racial disparities, discrimination appears to persist (Pager et al., 2009). Discrimination studies continue to enjoy attention as discrimination is seen as one of the key mechanisms for explaining enduring economic and social inequality in society. Online markets offer a new perspective on the diverse settings in which ethnic discrimination can occur and provide new channels to test assumptions about why and how members of ethnic or racial groups are being discriminated against (e.g. Zussman, 2013; Edelman et al., 2017).

We join this effort by examining the extent and the causes of ethnic discrimination in Europe's largest online carpooling market. We compile a new dataset of 16,624 carpooling rides offered in Germany by programmatically collecting ride information from an online platform. We estimate the effect of drivers' perceived name origin on the demand for their offered rides (clicks on offer). In order to group names into perceived ethnic backgrounds, we conducted a separate online survey with 1,577 participants who rated a total of 1,381 unique first names to distinguish the associated origin of drivers. Participants distinguished between typically German names and names with an Arab, Turkish or Persian origin. The latter group is the largest and most recognizable immigrant community in Germany.ⁱ Previous studies found that this particular group is disproportionately affected by discrimination (e.g. Blommaert et al., 2014; Diel et al., 2013).

In carpooling markets, private individuals use online platforms to offer seats in their car for a particular ride. Carpooling websites have become serious competitors for conventional bus and train providers across Europe, in particular in low-budget segments of the transportation

market.ⁱⁱ Carpooling offers mid- to long distance rides from city to city rather than short taxi rides within cities (like services such as Uber).

Carpooling platforms are a compelling application for ethnic discrimination studies for several reasons. First, carpooling is not only an economic market where riders select drivers with the best economic value (e.g. price per distance). Carpooling is also a social market. The decision to acquire the service is linked to spending one-off time with a stranger, i.e. the driver. The element of face-to-face personal interaction in a non-professional setting distinguishes carpooling platforms from labor or consumer good markets where ethnic discrimination has been studied (e.g. Doleac & Stein, 2013; Ayres et al., 2015, Ewens et al., 2014). Carpooling may thus help to draw attention to ethnic discrimination in social situations and reveal subtle, everyday forms of discrimination that may otherwise go unnoticed.

Second, online markets such as carpooling are ideally suited to isolate ethnic effects. We are able to observe all characteristics that are visible to the customer including the driver rating, experience, the car et cetera. This setup allows us to overcome issues related to both experimental designs – because we do not introduce an artificial treatment that would otherwise not occur in this wayⁱⁱⁱ – and many observational studies which may suffer from omitted variable bias.^{iv} For example, audit studies have often been criticized for introducing additional unobserved factors such as demeanor and socioeconomic background that may ‘pollute’ the treatment. Our analysis controls for every signal available to the consumer.

Third, we exploit variation in the information about drivers to test assumptions about the mechanisms driving ethnic discrimination. Additional information about the driver – such as the rating and experience – could work as a trust signal for consumers (e.g. Abrahao et al., 2017), however, previous evidence is unclear about whether additional information offsets ethnic discrimination (Ahmed et al., 2010; Nunley et al., 2011).

Fourth, we provide tentative evidence that ethnic discrimination effects are not driven by social class bias – a common dilemma given that many ethnic groups in Germany are overrepresented among lower social classes.

Our results indicate large ethnic discrimination effects. Controlling for all observable information, drivers with an Arab/Turkish/Persian sounding name attract significantly less interest in their offers (fewer clicks on the offer) than drivers with typically German names. To achieve the same demand compared to a driver with a typical German name, the average driver with an Arab/Turkish/Persian sounding name would have to offer the ride at 32% less than the price for an average ride. Group differences cannot be explained by any other observable characteristic associated with the driver or the offered ride and are robust against a series of robustness checks.

Consumers appear to use the name as a proxy signal to infer the ‘true’ value of the ride in economic, safety and social terms. When rich information about the driver is available (i.e. high rating, profile picture), ethnic discrimination decreases, as consumers rely less on the name.

Our findings have important implications for policy. First, ethnic discrimination occurs in social online market platforms. This expands the view from traditional discrimination studies in the labor market and housing to more subtle, everyday forms of unequal treatment. The results draw attention to other sectors with stronger social interaction elements, including the service and care sector or group environments such as membership in clubs, associations and interest groups. Second, insights into the mechanisms of discrimination can be the starting point for policy design aimed at reducing disparities (e.g. Guryan & Charles, 2013; Nunley et al., 2011). Our results suggest that providing more relevant context information about market actors may be a powerful strategy to reduce discrimination effects. As such, our results inform the discussion around the need for anti-discrimination efforts in markets that increasingly operate online. While consumers and service providers are often protected against discrimination in

traditional, offline markets (for instance hiring, housing, hospitality and consumer goods), similar provisions do not exist in online markets (see Edelman et al., 2017) and are difficult to prosecute.

2. Evidence and Mechanisms of Ethnic Discrimination

Ethnic and racial discrimination can be defined as differential treatment that leads to unequal outcomes based entirely on ascribed features such as race, ethnic background, name origin, foreign appearance etc. (Blank et al., 2004).

Recent reviews document discrimination effects in employment, housing, credit and commodity markets in many countries (Pager & Shepherd, 2008; Rich, 2014). The strongest evidence for ethnic discrimination is based on studies employing experimental designs (Rich, 2014). These studies show that racial or ethnic groups often are – *ceteris paribus* – disadvantaged in terms of access to labor market (interview invitations, call back rates, wage offers, treatment in interviews) and the housing market (renting, buying or selling apartments and houses).

More recent studies have made advances in two ways: first, they have broadened the application of discrimination studies to other markets (Bryson & Chevalier, 2015; Doleac & Stein, 2013; Edelman et al., 2017; Gneezy et al., 2012; Nunley et al., 2011; Zussman, 2013). Second, researchers have fine-tuned experiments to test hypotheses about why discrimination occurs opposed to whether it occurs (Gneezy et al. 2012; Guryan & Charles, 2013).

In terms of relevant mechanisms, much of the literature across the various domains traditionally attempts to discern whether discrimination stems primarily from taste-based discrimination (racial animus/ prejudice) or from statistical discrimination (asymmetric information).

In the case of taste-based discrimination (Becker, 1971), the driver of discrimination is a negative disposition towards certain groups. In our case, an individual may suffer ‘disutility’ resulting from contact with a specific ethnic group. As such, taste-based discrimination relies on the presence of prejudice. Prejudice in return can loosely be defined as an affective, mostly unfavorable feeling toward a person or group member based solely on their group membership.

In the case of statistical discrimination, differential treatment based on race and ethnic background arises from incomplete or asymmetric information about the productivity of actors (Arrow, 1973; Phelps, 1972).^v When limited information about a product or an individual is available, agents rely on observable group characteristics (such as ethnic group, race) to make inferences about the individual. Another class of statistical discrimination models focuses on the reliability of the information that employers have about individual productivity (Aigner & Cain, 1977; Altonji & Blank, 1999). At the core of both of these strands of statistical discrimination is the notion that a lack of information leads the employer to treat individuals as members of groups (Guryan & Charles, 2013).^{vi}

Past evidence on the dominant form of ethnic discrimination remained inconclusive, as support for either mechanism varies considerably across studies (Ewens et al., 2014). However, recent (experimental) studies point to the importance of statistical discrimination (rather than tastes) for explaining why discrimination persists (Altonji & Pierret, 2001; Bryson & Chevalier, 2015; Ewens et al., 2014; ; List, 2004; Zussman, 2013). Growing evidence in favor of statistical discrimination may be good news for policy makers given that information asymmetries may more easily be addressed than deep-rooted prejudice. In practice, statistical and taste-based discrimination are difficult to isolate in experimental and non-experimental study designs. Similar to taste-based discrimination, statistical discrimination relies on the concept of stereotypes in the form of certain beliefs associated with a group. It is not clear where

stereotypes originate and whether they are related or congruent with the concept of prejudice. This complicates interpreting evidence for one or the other.

Our study avoids a framing of statistical vs. taste-based discrimination as mutually exclusive mechanisms. We examine the role of information on its own merit while we are aware that previous research has interpreted information effects as indicative of statistical discrimination (e.g. Nunley et al., 2011). Rather than discerning the origins of ethnic discrimination, our primary aim with this study is to examine how additional information affects ethnic discrimination levels. In addition, responding to critiques of experimental studies (e.g. Heckman 1998), we take advantage of rich observational data covering real interactions occurring in a real market.

3. Ethnic Discrimination in Online Carpooling

Our study joins the effort of leveraging online markets – in our case, the largest German online carpooling market – for the study of ethnic discrimination. Carpooling markets match drivers that offer available seats in their car to riders that look for affordable one-off transport between cities. Riders can search rides by departure/ arrival town and date. Besides the place, day and time of departure, the price for a seat and a number of other ride-specific characteristics, carpooling offers show the first names of the drivers. We estimate the effect of an Arab/Persian/Turkish sounding first name on the demand of offered rides as measured by clicks. Arab/Turkish/Persian sounding names are associated with the largest and most recognizable immigrant community in Germany (mostly descendants of low skilled guest workers that arrived since the 1960s). Previous studies have highlighted that members of the Arab/Turkish/Persian community appear disproportionately affected by discrimination (Blommaert et al., 2014; Diehl et al., 2013) in Europe.

We interpret group differences in clicks net off all observable characteristics of the driver and the offered ride as evidence for ethnic discrimination.

In this brief section, we formulate three different possible outcomes to our research question of whether and why ethnic discrimination exists in carpooling markets: (1) no distinct discrimination effects, (2) discrimination effects *unaffected by information* and (3) discrimination effects *sensitive to the information provided*.

Regarding (1), we propose that it is plausible to expect no distinct ethnic discrimination effect owing to the particular context of carpooling and our study design. First, carpooling consumers are on average younger than the general population (Destatis, 2017). Second, rides provide transport between urban centers which suggest that the customers are also more likely to live in urban areas. Third, sharing a ride with a stranger already requires a certain level of trust. Fourth, online market platforms have been shown to reduce information asymmetries associated with productivity and correct biases against certain groups (Agrawal et al., 2013). Accordingly, it would *not* be surprising to find *no* significant discrimination effects given that we control for all observable information about the ride and the driver. In this light any effect that we may find is likely conservatively small compared to ethnic discrimination in other contexts and with other sub-populations in the German society.

Regarding (2), in our setting, traditional taste-based ethnic discrimination approaches suggest that potential consumers discriminate against drivers with a foreign name, because they simply wish to avoid contact with a member of a specific ethnic group. In carpooling, this means that we would expect discrimination effects regardless of variation in other information about the driver. Compared to commodity markets, taste-based ethnic discrimination may be more pronounced in carpooling as the customer is spending several hours with someone from another ethnic group in a narrow space (a car). In this case, simply the fact that the driver is associated with another ethnic group should lead to unequal treatment regardless of other observable

characteristics of the ride or the driver. In other words, we would expect that variation in the information provided about the ‘quality’ of the driver should not affect ethnic discrimination.

Regarding (3), in our setting, statistical discrimination approaches commonly assume that potential riders use the name of a driver as a signal to infer the ‘true value’ of the ride. Following this approach, the value of the ride depends on the degree of provided information about a ride rather than exclusively on the name origin of the driver. One advantage of our large dataset is that we can test for different stereotypes. We hypothesize that a negative ethnic effect on clicks could generally be driven by three different sets of considerations: 1) price and comfort relative to distance, 2) personal safety and 3) the social value.

Based on a narrow economic perspective, consumers simply click on the ride that offers the cheapest price relative to the distance travelled. Other factors may include the car quality as an indication how fast and comfortable the ride will be. In our setting, we control for the distance of the ride, the price and the car comfort.

Other consumers may choose an offer based on how secure they perceive the ride. Security has to be inferred from other available information as there is no objective indicator of security and safety on the platform. We assume that consumers use the name of the driver as two signals for perceived safety of a ride. First, it is a common stereotype that ‘migrants’, especially males, commit more crimes (e.g. Fitzgerald et al., 2011, Trager et al., 2014).^{vii} The other common safety-related stereotype could be that foreigners drive less safely because traffic regulations are less strict or less enforced in their origin countries. To the best of our knowledge there is virtually no reliable comparable data to prove or disprove this stereotype, but surveys suggest that the stereotype exists.^{viii}

Lastly, the value of a particular ride (and as a result, demand for that ride) may be driven by the desire for pleasant social interaction. We know from previous research that certain ethnic

groups are disadvantaged, for example, in flat sharing markets – a market where choices include social interaction (e.g. Przepiorka, 2011). Consumers may click on those offers that suggest the most enjoyable time during the ride. Sharing a ride means sharing private space as car-poolers sit in close proximity. Again, the name of the driver could be a proxy for language. Pleasant conversation is less likely if the driver speaks a different language and possibly listens to ‘foreign’ music. We estimate an interaction effect with music and dialog preferences to test this assumption. Studies on online dating have shown that clear ethnic/racial preferences exist that commonly disadvantage minorities (Jakobsson & Lindholm, 2014, Lin & Lundquist, 2013; Robnett & Feliciano, 2011). Similar to those markets, consumers in carpooling markets may be driven by homophily preferences, i.e. looking to meet drivers who are most like them (see McPherson et al., 2001). Again, the name would signal greater social distance given that the large majority of consumers are Germans.

As there is no direct indicator for safety or sociability, consumers have to rely on other available information, including the name. We argue that the user rating, number of ratings and the driver experience are suitable aggregate proxies for both categories. A bad user rating or low experience suggests that the ride may be less safe and less pleasant. Similar to studies that attempt to test statistical discrimination, we will interact the ethnic indicator with other indicators about productivity signals, in this case, the user rating and experience (Blommaert et al., 2014; Ewens et al., 2014; Nunley et al., 2011). Similar to Nunley et al. (2011), we argue that consumers’ relative weight on beliefs regarding the trustworthiness of drivers with an ethnically distinct name diminishes as other pertinent information about credibility becomes available. The scarcer other information about the ‘true’ safety and ‘fun’ of the ride, the more consumers rely on stereotypes regarding the perceived name origin. In our design, the user rating is based on experience of the driver (how many rides he or she has offered in the past) and the customer satisfaction. The user rating is a strong signal about the trustworthiness of provided information online and the ‘true’ productivity of the ride. We use additional proxies

of safety and sociability to test the effect of additional information including the profile picture (homophily, trust), talking and music preferences (sociability) and gender (safety).^{ix}

4. Data & Methods





















In this section, we first present details of the data collection process and then elaborate on the empirical methods.

4.1 Data Collection

We compile a new dataset with the aim to achieve a meaningful balance between internal and external validity of discrimination effects. Using one of the largest online carpooling platforms in Germany, we compile a dataset of 16,624 observations (i.e. rides) in Germany that were listed online between 16 July 2015 and 27 July 2015.^x According to the provider, the platform offered 250,000 rides in 2013 and 2014. The platform has 30 million members in 22 countries. According to the company's website, 10 million users use the website every quarter. Based on access to an Application Programming Interface, in short API, we collected information on all observable information on the offered rides and the drivers. The visual interface (see Figure 1) shows information about age, gender, user picture if available, user rating, car, timing and stops of the ride, price, available seats and some preferences of the driver (smoking, music, talking).

Figure 1: User Interface of Online Carpooling Platform

Entfernung: 233 km - Dauer: 2Std.18Min.
Ordnen nach

| | | |
|--|---|---|
|  19 Jahre Aufsteiger/in ★ 4.0 - 2 Bewertungen  | Morgen - 21:00 Uhr Rostock → Berlin  Rostock  Berlin Fahrzeug: OPEL CORSA ★★ | 11 € pro Mitfahrer/in 3 Plätze frei |
|  21 Jahre Fortgeschrittene/r ★ 4.7 - 3 Bewertungen  | Mittwoch 11. November - 13:30 Uhr Rostock → Berlin  Rostock, Deutschland  Mollstraße 19, 10249 Berlin, Deutschland Fahrzeug: SKODA OCTAVIA Combi ★★★ | 11 € pro Mitfahrer/in 2 Plätze frei |
|  25 Jahre Fortgeschrittene/r ★ 4.7 - 3 Bewertungen  | Mittwoch 11. November - 14:00 Uhr Rostock → Berlin  Rostock Hbf, Rostock  S+U Hermannstr. (Berlin), Neukölln Fahrzeug: VOLKSWAGEN PASSAT ★★★★★ | 12 € pro Mitfahrer/in 4 Plätze frei |
|  28 Jahre Aufsteiger/in ★ 5.0 - 1 Bewertung 161 Freunde/innen  | Mittwoch 11. November - 15:10 Uhr Rostock → Berlin → Ansbach  Rostock Hbf, Rostock  Ankunft: Berlin (Bitte sprechen Sie die Details mit dem Fahrer/der Fahrerin ab.) Fahrzeug: SKODA OCTAVIA Combi ★★ | 11 € pro Mitfahrer/in 3 Plätze frei |
|  30 Jahre Aufsteiger/in ★ 5.0 - 2 Bewertungen 237 Freunde/innen  | Mittwoch 11. November - 17:00 Uhr Rostock → Berlin → Bernau Bei Berlin  Rostock, Deutschland  Ankunft: Berlin, Deutschland (Bitte sprechen Sie die Details mit dem Fahrer/der Fahrerin ab.) Fahrzeug: VOLKSWAGEN GOLF V ★★ | 11 € pro Mitfahrer/in 3 Plätze frei |

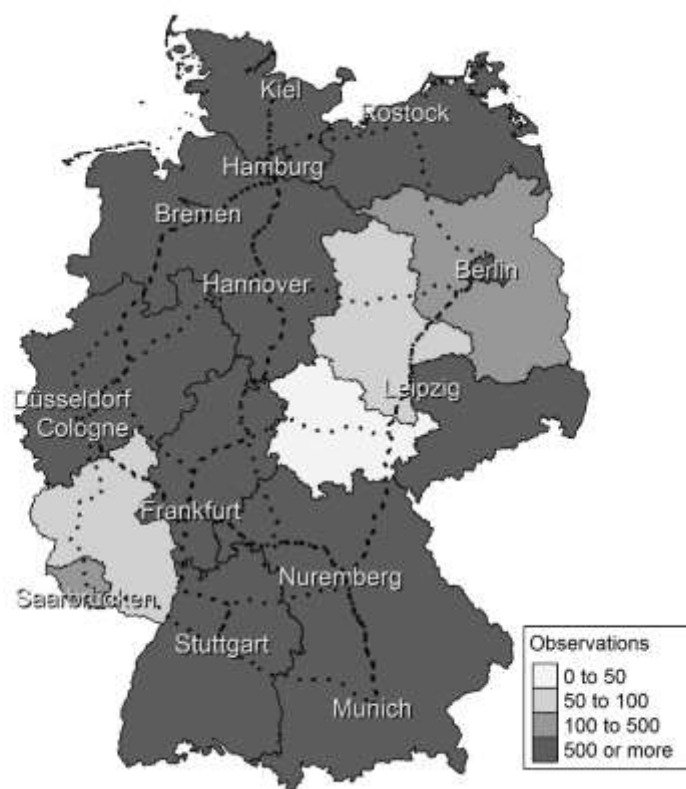
Source: Carpooling Data Germany 2015 (compiled by authors). Images, names and age of drivers pixelated. See main text for description in English.

Rides shown in Figure 1 are offered by drivers traveling from Rostock to Berlin. For instance, the first driver offered a seat for his ride at the price of 11 Euros. He has two positive user ratings from earlier interactions on the market. Furthermore, he prefers to talk during the trip

and does not mind riders to travel with their pets. In contrast to the other ride offers in Figure 1, the driver did not upload a picture.

Due to volume and restrictions from the provider, not all rides in the market could be collected. As a result, we selected routes between the largest cities in all 16 German states (Länder). Carpooling is more common between urban centers due to supply and demand for rides. Our strategy allowed us to approximate a balanced geographical representation of observed rides for different regions in Germany. As a second step, we included additional large cities in regions with larger populations, particularly regions with higher shares of ethnic minority residents. Oversampling of urban areas was necessary to ensure a sufficient sample of drivers with foreign-sounding names (see Figure 2).

Figure 2: Sampled rides in online carpooling market



Source: Carpooling Data Germany 2015 (compiled by authors)

The data was collected in two steps: First, we pulled data from the API and, second, we used the programming language Python to automatically access the website four times a day at equal intervals between 16 July 2015 and 27 July 2015. Accessing the website in addition to pulling API data was necessary for retrieving information about our main dependent variable, the number of clicks that each ride in our sample received.

4.2 Measurements

We assess group differences regarding demand for offered rides on the carpooling platform. Demand is measured by clicks. We regress the maximum number of clicks that a ride received until departure using a negative binomial regression (see Figure B1 in the Appendix for distribution of clicks).^{xi} In the analysis, the number of clicks is adjusted for the number of days that the offer was displayed online until departure.

We restrict our sample in several ways: first, we drop rides that depart after the end of our observation period (right-censoring). This is necessary to obtain an accurate measure of maximum clicks before departure. Second, we restrict the sample to rides that were uploaded no sooner than seven days before we began the data collection. One week is an appropriate time window given that most rides are uploaded a few days before departure. Third, we drop cross-border rides, as consumers are most likely not German and clicks are inflated as affected rides are also listed in carpooling platforms for neighboring countries, where the provider also operates. Fourth, we limit our sample to routes (e.g. departure city: Munich – arrival city: Berlin) that have more than one offered ride per day and have at least one driver with an Arab/Turkish/Persian name.^{xii} This step is important to ensure that we can observe a counterfactual, i.e. consumers cannot discriminate against drivers from another ethnic group if there are none. It is important to note that all our models additionally control for route and volume (number of offered rides per route and day). The final sample for the analysis of clicks includes 16,624 rides, including 528 rides with an Arab/Turkish/Persian driver.

As our main independent variable regarding discrimination, we use the first name of the driver to infer whether the name is ‘typically Arab/Turkish/Persian’ or ‘typically German’. Names signal membership to a particular ethnic group (regardless of whether the signal is true) and ‘ignite’ potential stereotypes (e.g. Bertrand & Mullainathan, 2004, Booth et al., 2012).

Driven by concerns about the objectiveness and reliability of name ratings, we conducted a large online survey in which respondents were asked to rate driver names that we extracted from our carpooling sample. In total, 1,577 student raters participated in the survey. The origin of 1,381 unique first names were on average rated by 20 student raters (SD=4.6).^{xiii} As carpooling riders are younger than the national average (Destatis, 2017), students represent a reasonable approximation of typical riders. Table 1 shows the most frequent names by perceived name origin.

Table 1: Most frequent name origins with high origin certainty

| | Male names | | Female names | |
|----|---------------------------|-----------|---------------------------|-----------|
| # | Arab/ Persian/ Turkish | German | Arab/ Persian/ Turkish | German |
| 1 | Ali | Thomas | Sanam | Julia |
| 2 | Mohammed | Christian | Halime | Sarah |
| 3 | Süleyman | Daniel | Sahar | Johanna |
| 4 | Seref | Martin | Hülya | Lisa |
| 5 | Mohamed | Michael | Taman | Anna |
| 6 | Kadir | Alexander | Büsra | Katharina |
| 7 | Serdar | Andreas | Dersimgül | Grit |
| 8 | Ismail | Sebastian | Güllü | Maria |
| 9 | Mustafa | Markus | Gülten | Laura |
| 10 | Cem | Jens | Husna | Anne |
| 11 | Osman | Peter | Nasrin | Franziska |
| 12 | Salman | Tobias | Nesrin | Lena |
| 13 | Yusuf | Christoph | Senem | Stefanie |
| 14 | Amir | Matthias | Sinem | Alexandra |
| 15 | Ercan | Stefan | Özlem | Anja |
| 16 | Mehdi | Chris | Hasiba | Annika |
| 17 | Oguz | Robert | Cigdem | Nadine |
| 18 | Rami | Jan | Elif | Sandra |
| 19 | Ahmad | Volker | Fatemeh | Miriam |
| 20 | Ersin | Friedrich | Gülcan | Carolin |

Note: Most frequent driver names by name origin based on online survey ratings (N=1,577 survey participants; 20 ratings per name on average). Note that only 30 unique female Arab/Persian/Turkish names were available in the sample.

For the analysis, we use an 80% cut-off for determining an Arab/Turkish/Persian name. That is, the driver is considered to have an Arab/Turkish/Persian name if four out of five raters (i.e. 16 out of 20 raters per name on average) considered the name to be typically “Arab, Turkish or Persian”. We also report results for the continuous measure of Arab/Turkish/Persian name origin variable in percentage points (see Figure 4 below). We grouped Arab/Turkish/Persian sounding first names together because they are difficult to distinguish for the average resident in Germany and are commonly associated as being from the same world region.^{xiv} Members of this broad group are associated with the largest and most recognizable immigrant community in Germany (mostly descendants of low skilled guest workers that arrived since the 1960s). Previous studies have highlighted that members of the Arab/Turkish/Persian community appear disproportionately affected by discrimination (Blommaert et al., 2014; Diehl et al., 2013).

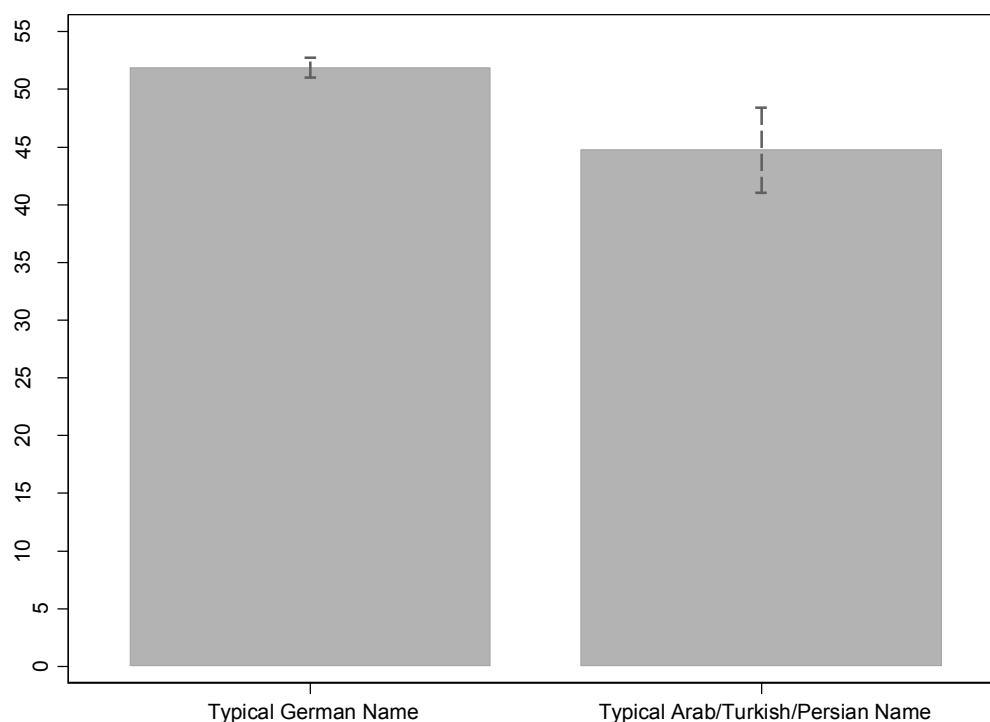
For the analysis of mechanisms, we exploit variation in information about each offered ride. We estimate the interaction effect of an Arab/Turkish/Persian first name with the user rating associated with each offered ride. The rating score is an average of past riders’ general evaluation of their ride with the respective driver who offered it. Past riders can rate the driver retrospectively. As a result, the rating is a strong signal of safety, sociability and overall trustworthiness. To test the sociability argument in particular, we estimate an interaction of the name with the profile picture and the music and dialog preference. Drivers that provide a profile picture as personal information likely increase their trustworthiness. The profile picture may also function as a proxy for sociability perceptions. The ‘talking preference’ indicates whether or not the driver is interested in talking during the ride which we use as one additional measurement for the sociability argument. Finally, we estimate an interaction with name and gender. We assume that negative stereotypes regarding the safety of rides with Arab/Turkish/Persian drivers largely apply to male drivers.

Controls include all the information that is observable to consumers, including information about the offered ride (route, time, distance) and the driver such as age and gender (see Table A2 and Table A3 in the Appendix for a full description and distribution of all model variables). Regardless of ride and driver information, the clicks on offer may simply be driven by the size of the potential user population, which varies considerably across the sampled cities and regions in Germany. For this reason, we control for the demand side using a route identifier for all routes in our sample.

5. Results

The analysis confirms substantial discrimination effects in Germany's online carpooling market. Drivers with Arab/Turkish/Persian names attract less demand (measured in clicks) than drivers with typical German names for the same ride. Controlling for all observable characteristics of the ride and the driver that are visible to consumers, we find that drivers with an Arab/Turkish/Persian sounding name obtain on average 7 clicks less than a driver with a typical German name (significant at $p < 0.01$, see Figure 3 and Table A3 the Appendix). 7 clicks represent approximately 13% of the average number of clicks per offered ride in the sample (51 clicks). In a separate step, we calculate the average discriminatory price premium, i.e. the average willingness to pay to avoid riding with an Arab/Turkish/Persian driver. Dividing the name coefficient by the price coefficient indicates that Arab/Turkish/Persian drivers would have to offer their rides on average 4.20 € cheaper than German drivers to achieve the same number of clicks. This accounts for 32% of the average price of an average ride in our sample. This discriminatory price premium increases to 34% when setting covariates to different values, for example, a male, thirty-year-old Arab/Turkish/Persian driver with little experience offering a ride over 300 kilometers in a comfortable car on a Sunday afternoon.

Figure 3: Predicted Number of Clicks on Offer Ride by Name Origin of the Driver

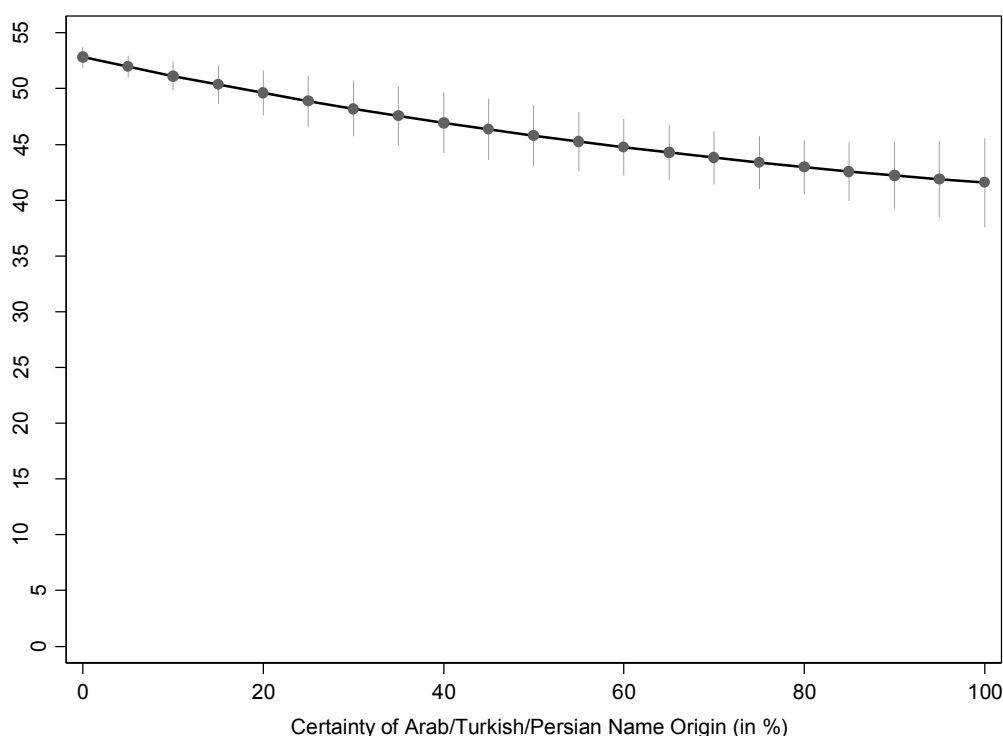


Note: Carpooling Data Germany 2015 (compiled by authors). Predicted clicks are adjusted (see full model in Table A3 in the Appendix). N= 16,624. Group differences are statistically different ($p < 0.01$).

Figure 3 uses the 80% cut-off to determine drivers with an Arab/Turkish/Persian name. Figure 4 reports the result for the continuous measure of Arab/Turkish/Persian name origin (the percentage of survey respondents who rated the name to be typically Arab/Turkish/Persian). Disparities between Arab/Turkish/Persian and German drivers increase with the degree of certainty that the name is associated with an Arab/Turkish/Persian background (see Figure 4).

Figure 4: Predicted Number of Clicks on Offer by Certainty of Arab/Turkish/Persian Name

Origin

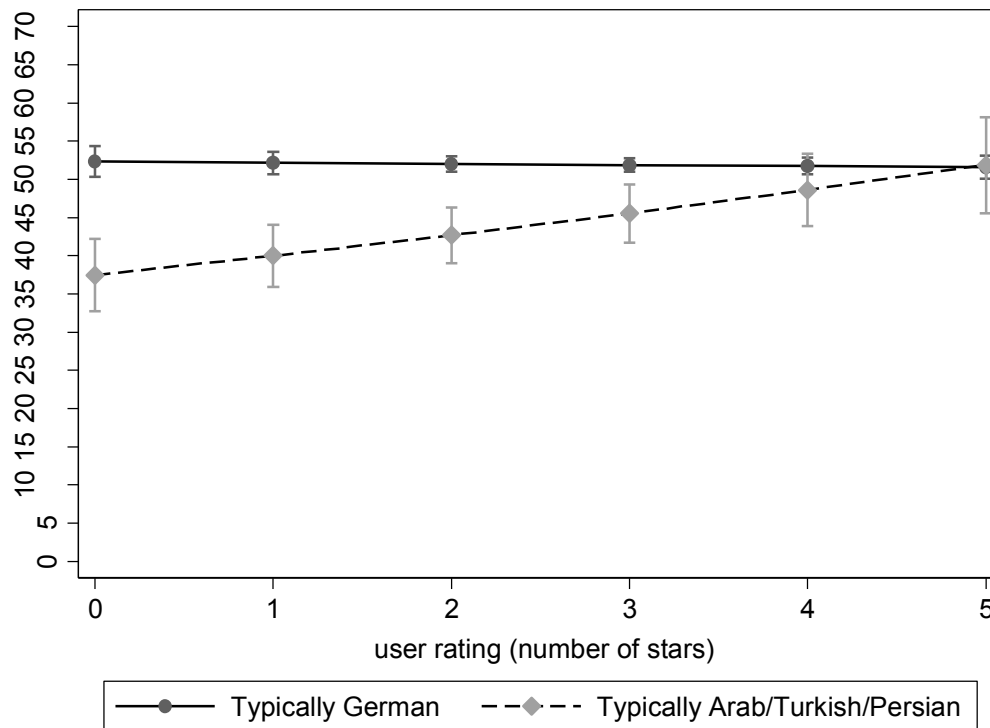


Note: Carpooling Data Germany 2015 (compiled by authors). Predicted clicks are adjusted (see Table A3 in the Appendix). N= 16,624. Effect of continuous name measure is statistically significant ($p < 0.01$). 95% confidence interval.

In the second part of the analysis, we turn to the role of information for ethnic discrimination. Our tests suggest that disparities between groups depend on the variation of information about individual drivers.

Figure 5 shows that drivers with an Arab/Turkish/Persian name are disadvantaged against German drivers when they have no or low ratings. Disparities appear to vanish when both drivers have equally high user ratings. The interaction effect is statistically significant at $p < 0.01$ (see Table A3). Similar results for the number of user ratings and the driver experience corroborate these findings (see Figure B2 and Figure B3 in the Appendix).

Figure 5: Predicted Number of Clicks on Offered Ride by Name Origin of the Driver and the User Rating



Note: Carpooling Data Germany 2015 (compiled by authors). Predicted clicks are adjusted for all observable characteristics (see Table A3 in the Appendix). N= 16,624. 95% confidence interval.

The findings for the user rating, number of ratings and experience suggest that consumers place less weight on the name of the driver when other relevant information is available. This is consistent with the statistical discrimination hypothesis as the name of the driver may be used as one more source of information about the perceived ‘value’ of the ride.

We conduct a number of additional indirect tests to approximate different motives (as discussed in section 2.1). There may be at least two possible sources of stereotypes. First, consumers may be concerned with safety since carpooling entails sharing a ride with a stranger. Given stereotypes regarding crime and driving styles among ‘foreigners’, we hypothesized that statistical discrimination may be driven by safety concerns. Our results provide indirect evidence that this is the case. First, ethnic discrimination effects are larger for males compared

to females (see Figure B4). This is consistent with the assumption that stereotypes regarding crime and driving styles apply particularly to male foreigners (Trager et al., 2014). Consumers may generally feel less safe with a male driver with an Arab/Turkish/Persian sounding name compared to a male German driver. Female drivers with an Arab/Turkish/Persian sounding name may not be disadvantaged compared to female German drivers given that the crime stereotype largely applies to males. However, sample size limitations regarding female drivers with foreign names (N=31) do not allow us to infer that females are not subject to discrimination. The effect is smaller but not zero (see Figure B4 and Table A4 in the Appendix).

Second, Arab/Turkish/Persian drivers without a profile picture are much more disadvantaged than drivers from the same ethnic group with a profile picture (see Figure B5).^{xv} We interpret the profile picture to be a trust enhancing measure. Ethnic stereotypes regarding safety may simply have more room to engage imagination when users do not know what the driver looks like.

Third, we suspected that consumers may select rides based on sociability considerations. Drivers with an Arab/Turkish/Persian name could be discriminated against when consumers assume that ‘foreign’ drivers may not speak the language or do not share similar music tastes which could make the joint ride less enjoyable. In fact, our findings show that Arab/Turkish/Persian drivers are less disadvantaged when they have indicated a preference for talking during the ride (see Figure B6). We speculate that consumers interpret a talking preference for foreigners as a sign of good German language skills. In contrast, a preference against talking may simply be interpreted as a potential language barrier. This could explain why the positive effect of talking preference on clicks is considerably larger for Arab/Turkish/Persian drivers compared to German drivers.

The results for music preference could be interpreted in a similar vein (see Figure B7). Our findings show that Arab/Turkish/Persian drivers are more disadvantaged when they indicate a

preference for music during the ride. Again, consumers may infer that a music preference implies a lack of willingness to talk which, in turn, could be perceived as a language barrier. In addition, consumers may assume that drivers with an Arab/Turkish/Persian name might want to listen to ethnic music which could reduce the enjoyment of the ride for German consumers who may be less likely to share similar tastes.

In sum, our results document substantial ethnic discrimination in Germany's carpooling market. Our findings highlight the power of information about drivers. The more useful information is provided about drivers with foreign-sounding names, the less likely they are to be discriminated. We provided tentative evidence that safety and sociability considerations may drive this information effect. Our results are robust against a series of checks including different samples, variable operationalization, estimators and potential social class bias (see the appendix for details).

One important potential source of bias is perceived social class given that Arab and Turkish migrant communities are overrepresented among lower social classes in Germany. We compare the effects of typically Arab/Turkish/Persian names with the effect of typically Anglo-Saxon names (i.e. Steven, Justin, Kevin). Studies in the German context have shown that Anglo-Saxon names in Germany are associated with low social class (e.g. Kaiser, 2010). The effect of such names, however, is not statistically significant and not negative. As such, we provide tentative findings that the ethnic penalty appears to be robust against social class bias.

6. Summary & Discussion

Recent ethnic discrimination studies increasingly make use of online market data to better understand when and why ethnic discrimination occurs. We aim to contribute to this effort with a novel application of ethnic discrimination in Europe's largest online carpooling market (i.e.

Germany). We argue that there are four aspects that make our study a valuable contribution to existing research. *First*, carpooling is a social market that involves one-off, face-to-face interaction in a non-professional setting. This unique setting broadens the scope of ethnic discrimination research to more subtle, diverse and everyday interactions where ethnic minorities may face unequal treatment. *Second*, the social element of this market and the fact that we are able to measure all relevant observable characteristics allows us to test assumptions regarding the mechanisms driving ethnic discrimination. We focus in particular on the effects of information. *Third*, the advantage of our application is that we observe real actors making real decisions in real markets while being able to control all relevant factors that may influence consumer choice. Using observational data while holding all relevant confounders constant is a key advantage of our study and responds to the critique of experimental methods that might create rare and artificial situations (e.g. Heckman 1998). *Fourth*, we can provide tentative analysis to disentangle social from ethnic cues which is a limitation of many ethnic discrimination studies.

We find evidence of substantial ethnic discrimination in Germany's carpooling market. Drivers with Arab/Turkish/Persian sounding names obtain – *ceteris paribus* – less demand (on average 13% fewer clicks) compared to German drivers. The average Arab/Turkish/Persian driver in our analysis would have to offer his ride 4.20 € cheaper than the average German driver to achieve the same demand, a discriminatory price premium that is equivalent to 32% of the price for an average ride. This finding is robust against a broad range of checks.

Discrimination of drivers with Arab/Turkish/Persian names persists despite the relatively young and urban consumer composition in this particular market. Therefore, the estimated discrimination effect may be conservative compared to other everyday social interactions with ethnic minorities in the German society. Our findings are consistent with other recent studies that show ethnic/racial discrimination effects in other online consumer markets (Ayres et al.,

2011; Blommaert et al., 2014; Doleac & Stein, 2013; Edelman et al., 2017; Przepiorka, 2011; Robnett & Feliciano, 2011; Zussman, 2013). Discrimination against individuals with Arab, Persian and Turkish sounding names is consistent with findings in other studies across Europe (Blommaert et al., 2014; Gaddis & Ghoshal, 2015; Rich, 2014).

One main result of our study is that ethnic disparities decrease depending on the level of relevant information that is available about the service provider (the driver). Higher user ratings, a higher number of ratings and information on driver experience decrease ethnic discrimination. In fact, ethnic disparities seem to disappear entirely for the highest rated drivers. This shows that stereotypes regarding particular ethnic groups become more salient and active when other information that could signal trust is scarce. In other words, discrimination is more pervasive in information-scarce environments. Consumers appear to use the name origin as a signal for other relevant information that is otherwise not available. Previous studies have argued that such information effects are consistent with statistical discrimination. We are cautious to make strong judgments on the relative importance of statistical and taste-based ethnic discrimination in general, as we do not have the tools to adequately isolate taste-based discrimination. Regardless of relative importance, the strong effects of information, including user ratings, deserves attention in its own merit and support some previous evidence that suggests that information can ameliorate discrimination effects (Abraham et al., 2017; Nunley et al., 2011; Ahmed et al., 2010).

Common statistical discrimination assumes that discriminatory behavior is based on stereotypes which are commonly difficult to capture empirically. Our data allowed us to provide a number of indirect tests of underlying stereotypes that might drive discrimination. Unlike conventional studies in the area of employment, for example, stereotypes in carpooling do not (only) revolve around low productivity, low educational achievement or work ethics. Our analysis suggests that safety and sociability considerations apply. The results show that customers may have –

ceteris paribus – less trust in a driver with a foreign name and that foreign drivers may signal a lower social value, for example, because they may speak a different language and prefer different music during the ride. Tentative evidence suggests that these effects are not driven by social class bias.

Our findings have implications for policy. The results underscore the importance of a general discussion about anti-discrimination legislation in the internet age. It is possible that due to one-to-one communication in online markets, discrimination goes largely undetected and unsanctioned. Moreover, our findings suggest that the type and level of information provided matters for the degree of discrimination, providing a useful leverage point for policy makers. The magnitude of ethnic discrimination *decreases* with an *increase* of available context information about individual actors. Growing evidence in support of statistical discrimination is good news for policy makers (as compared to taste-basted discrimination) as information is often more malleable to policy than deep-rooted prejudice. In cases where adding context information is not possible, another strategy is to remove the information or signal (ethnic cue) that induces unequal treatment of some users, i.e. the name. Without the name it is harder to assign (ethnic) group membership and thus, harder for stereotypes to be activated. Which approaches are most effective depends on the context and remains an empirical question. It is clear, however, that online markets are increasingly under pressure to find solutions. The online apartment sharing platform Airbnb – for example – has recently introduced changes to growing evidence of discrimination on their platform (see Edelman et al., 2017).^{xvi} The startup adopted new non-discrimination policies and systems to address user complaints. Airbnb is now promising to allow guests to book without prior approval or screening by the host and to reduce the prominence of pictures on guests' profiles in favor of more 'objective', reputation-enhancing information.

Our study also faces certain limitations. Unfortunately, information about the consumers of the rides is not available. We were also not able to match more disaggregated regional population statistics. This information would allow us to disaggregate effects by location and look at how population attitudes may correlate with discriminatory behavior. Future comparative research is also needed to assess discrimination varies across national contexts and different ethnic groups. In addition, more research is needed to study the effect of gender in online market discrimination. Our results suggest that ethnic penalties are smaller for women compared to men, however, our sample size of female drivers was too small to explore gender-specific processes in more detail.

In summary, our results have illustrated the power of a name and the information associated with it. We find that the name is used as a proxy for the trustworthiness of actors in a social market environment. Foreign-sounding drivers are trusted less than German drivers when information is scarce. When more information is provided about both drivers, discrimination decreases to levels that are statistically undistinguishable from zero. In other words, when little is known about the quality of a ride, drivers with typical German names enjoy a certain ‘blind trust premium’ that cannot be explained by any relevant quality indicator. Unique to our social market scenario, we were able to provide indication that discrimination in social markets is based on assumptions regarding safety and the social value of spending time with a member of another ethnic group in terms of language barriers and tastes (e.g. music preference). Our findings highlight the role of ethnic discrimination in subtle, everyday social interactions between ethnic groups and the powerful role of information to influence discrimination.

References

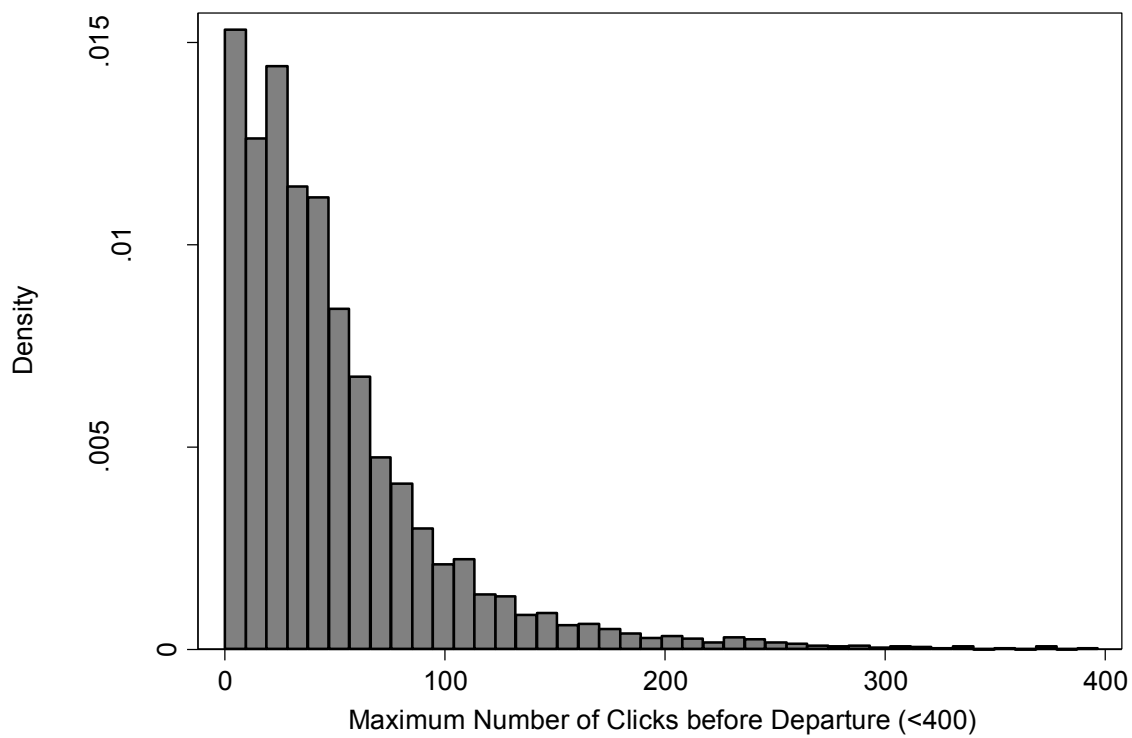
- Abrahao, B., Parigi, P., Gupta, A., & Cook, K. S. (2017). Reputation offsets trust judgments based on social biases among Airbnb users. *Proceedings of the National Academy of Sciences*, 114 (37), 9848–9853
- Ahmed, A. M., Andersson, L., & Hammarstedt, M. (2010). Can discrimination in the housing market be reduced by increasing the information about the applicants?. *Land Economics*, 86(1), 79-90.
- Al Ramiah, Ananthi, and Miles Hewstone (2013). Intergroup Contact as a Tool for Reducing, Resolving, and Preventing Intergroup Conflict: Evidence, Limitations, and Potential, *American Psychologist*, 68(7), 527.
- Agrawal, Ajay, Lacetera, Nicola, and Elizabeth Lyons. (2013). Does Information Help or Hinder Job Applicants from Less Developed Countries in Online Markets? *National Bureau of Economic Research* No. w18720.
- Aigner, Dennis, and Glen Cain. (1977). Statistical Theories of Discrimination in Labor Markets. *Industrial and Labor Relations Review*, 30(2), 175-187.
- Altonji, Joseph, and Charles Pierret. (2001). Employer Commitment and Statistical Discrimination. *The Quarterly Journal of Economics* 116, 313-350
- Altonji, Joseph, and Rebecca Blank. (1999). Race and Gender in the Labor Market. *Handbook of Labor Economics* 3, 3143-3259.
- Anderson, Lisa, Fryer, Roland and Charles Holt. (2006). Discrimination: Experimental Evidence from Psychology and Economics. Pp. 97-118 in *Handbook on the Economics of Discrimination* edited by William Rodgers. Edward Elgar Publishing.
- Arrow, Kenneth. (1973). The Theory of Discrimination. Pp. 3–33 in *Discrimination in Labor Markets*, edited by Orley Ashenfelter and Albert Rees. Princeton University Press.
- Ayres, Ian, Banaji, Mahzarin and Christine Jolls. (2015). Race Effects on eBay. *The RAND Journal of Economics* 46 (4): 891-917.
- Becker, Gary. (1971). *The Economics of Discrimination*. Chicago: University of Chicago Press.
- Bertrand, Marianne, and Sendhil Mullainathan. (2004). Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination, *The American Economic Review*, 94 (4), 991-1013.
- Blank, Rebecca, Dabady, Marilyn, and Constance Citro. (2004). Measuring Racial Discrimination: National Research Council Panel on Methods for Assessing Discrimination. National Academy of Sciences.
- Blommaert, Lieselotte, Coenders, Marcel, and Frank van Tubergen. (2014). Discrimination of Arabic-named Applicants in the Netherlands: An Internet-based Field Experiment Examining Different Phases in Online Recruitment Procedures, *Social forces* 92 (3), 957-982.
- Booth, Allison, Leigh, Andrew, and Elena Varganova. (2012). Does Ethnic Discrimination Vary Across Minority Groups? Evidence from a Field Experiment. *Oxford Bulletin of Economics and Statistics* 74(4), 547-573.
- Bryson, Alex, and Arnaud Chevalier. (2015). Is There a Taste for Racial Discrimination Amongst Employers? *Labour Economics*, 34, 51-63.

- Diehl, Claudia, Andorfer, Veronika, Khoudja, Yassine, 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(10), 1679-1697.
- Doleac, Jennifer, and Luke Stein. (2013). The Visible Hand: Race and Online Market Outcomes, *The Economic Journal*, 123(572), 469-492.
- Edelman, Benjamin, Luca, Michael, and Dan Svirsky. (2017). Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment, *American Economic Journal: Applied Economics*, 9(2), 1-22.
- Ewens, Michael, Tomlin, Bryan and Liang Choon Wang. (2014). Statistical Discrimination or Prejudice? A Large Sample Field Experiment, *Review of Economics and Statistics*, 96(1), 119-134.
- Fitzgerald, Jennifer, Curtis, Amber, and Catherine L. Corliss. (2011). Anxious Publics - Worries About Crime and Immigration, *Comparative Political Studies*, 45 (4), 477 – 506.
- Gaddis, Michael, and Raj Ghoshal. (2015). Arab American Housing Discrimination, Ethnic Competition, and the Contact Hypothesis. *The ANNALS of the American Academy of Political and Social Science*, 660(1), 282-299.
- Gneezy, Uri, List, John, and Michael Price. (2012). Toward an Understanding of Why People Discriminate: Evidence From a Series of Natural Field Experiments, *National Bureau of Economic Research* No. w17855
- Guryan, Jonathan, and Kerwin Charles. (2013). Taste-based or Statistical Discrimination: The Economics of Discrimination Returns to its Roots, *The Economic Journal*, 123(572), 417-432.
- Heckman, James. (1998). Detecting Discrimination, *The Journal of Economic Perspectives*, 12(2): 101-116.
- Lin, Ken-Hou, and Jennifer Lundquist. (2013). Mate Selection in Cyberspace: The Intersection of Race, Gender, and Education, *American Journal of Sociology*, 119(1), 183-215.
- Jakobsson, Niklas, and Henrik Lindholm. (2014). Ethnic Preferences in Internet Dating: A Field Experiment. *Marriage & Family Review* 50(4), 307-317.
- Kaiser, Astrid. (2010). Vornamen: Nomen est omen? Vorerwartungen und Vorurteile in der Grundschule. *Schulverwaltung. Zeitschrift für Schulleitung und Schulaufsicht*, 21(2), 58-59.
- Kalter, Frank. (2006). In Search of an Explanation for the Specific Labor Market Disadvantages of Second Generation Turkish Migrant Children. *Zeitschrift für Soziologie*, 35(2), 144-160.
- List, John. 2004. The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field, *The Quarterly Journal of Economics*, 119(1), 49-89.
- McPherson, Miller, Smith-Lovin, Lynn, and James M. Cook. (2001). Birds of a Feather: Homophily in Social Networks, *Annual Review of Sociology*, 27(1), 415-444.
- Neumark, David. (2012). Detecting Discrimination in Audit and Correspondence Studies, *Journal of Human Resources* 47(4), 1128-1157.
- Nunley, John, Owens, Mark, and Stephen Howard. (2011). The Effects of Information and Competition on Racial Discrimination: Evidence from a Field Experiment, *Journal of Economic Behavior & Organization*, 80(3), 670-679.

- 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.
- Pager, Devah., Bonikowski, Bart, and Bruce Western (2009). Discrimination in a Low-wage Labor Market: A Field Experiment, *American Sociological Review*, 74(5), 777-799.
- Phelps, Edmund (1972). The Statistical Theory of Racism and Sexism, *The American Economic Review*, 62(4), 659-661.
- Przepiorka, Wojtek. (2011). Ethnic Discrimination and Signals of Trustworthiness in an Online Market: Evidence from two Field Experiments, *Zeitschrift für Soziologie*, 40(2), 132-141.
- Rich, Judith. (2014). What Do Field Experiments of Discrimination in Markets Tell Us? A Meta Analysis of Studies Conducted Since 2000, *IZA Discussion Paper*, 8584(1), 1-68.
- Robnett, Belinda, and Cynthia Feliciano. (2011). Patterns of Racial-Ethnic Exclusion by Internet Daters, *Social Forces* 89(3), 807-828.
- Trager, Glenn and Charis E. Kubrin. (2014). Complicating the Immigration-Crime Nexus: Theorizing the Role of Gender in the Relationship between Immigration and Crime. Pp. 527-548 in Rosemary Gartner and Bill McCarthy (Eds.), *The Oxford Handbook of Gender, Sex, and Crime*. New York: Oxford University Press.
- Zussman, Asaf. (2013). Ethnic Discrimination: Lessons from the Israeli Online Market for Used Cars, *The Economic Journal*, 123(572), 433-468.

APPENDIX (in the order of appearance)

Figure B1: Distribution of Clicks



Source: Carpooling Data Germany 2015 (compiled by authors) (N=16,624, clicks above 400 clicks excluded for visualization purposes)

Table A1: Operationalization of model variables

| Variable | Description | Operationalization (see also Table A3) |
|-----------------------------|---|--|
| Dependent Variables | | |
| Clicks | The maximum number of clicks an offered ride received before departure | Continuous |
| Independent Variable | | |
| Name origin | Rated origin of the driver's name | Categorical: Arab/Turkish/ Persian vs. German (Continuous scale in Figure 4) |
| Controls | | |
| Time of day | The time during a day when the ride departs | Categorical: night, morning, midday, afternoon, evening |
| Day of week | The day of the week when the ride departs | Categorical: Monday - Sunday |
| Distance in km | The distance between departure city and arrival city | Continuous |
| Price in euro | Price to be paid for one seat on the ride | Continuous |
| Gender | Gender of the driver | Categorical: Female, male |
| Age | Age of the driver | Continuous (in years) |
| Smoking preference | Smoking preference of the driver | Categorical: yes, no |
| Music preference | Music preference of the driver | Categorical: yes, no |
| Dialog preference | Dialog preference of the driver | Categorical: yes, no, maybe |
| Rating | Rating of the driver by previous customers | Categorical: 0-5 stars |
| Experience | Experience of the driver based on the number of offered rides in the past | Categorical: 0 – no experience to 4 – high experience |
| Picture | Availability of a profile picture for the driver | Categorical: yes, no |
| Comfort | Comfort of the ride conditional of the type of car | Categorical: simple/normal, comfortable, luxury, score not available |
| Auxiliary Variables | | |
| Route ID | Control for routes between cities | Dummy variable |
| Ride Volume | Control for number of rides offered per route | Continuous variable |
| Log days until departure | Log number of days before departure (i.e. the time the offer was online before departure) | Continuous variable |

Table A2: Summary Statistics of Clicks Model (see Figure 3 and Table A3)

| | German | | | | Arab/Turkish/Persian | | | | T-test | |
|-----------------------------------|--------|-------|-----|------|----------------------|-------|-----|-----|-----------|---------|
| | mean | sd | min | max | mean | sd | min | max | b | T |
| Maximum Number of Clicks on Offer | 51.4 | 53.7 | 0 | 629 | 45.8 | 49.4 | 0 | 331 | 5.53* | (2.50) |
| Traffic volume | 16.1 | 11.9 | 2 | 58 | 12.4 | 9.8 | 2 | 58 | 3.67*** | (8.29) |
| Days online until departure | 3.6 | 2.7 | 1 | 11 | 4.3 | 3.2 | 1 | 11 | -0.78*** | (-5.51) |
| Female | 0.3 | 0.4 | 0 | 1 | 0.1 | 0.2 | 0 | 1 | 0.21*** | (18.76) |
| Age* | 31.3 | 9.7 | 18 | 101 | 31.3 | 7.0 | 18 | 54 | -0.07 | (-0.24) |
| Number of Ratings | 5.3 | 11.8 | 0 | 154 | 4.3 | 10.1 | 0 | 79 | 1.03* | (2.28) |
| User Rating | 2.9 | 2.3 | 0 | 5 | 2.6 | 2.3 | 0 | 5 | 0.36*** | (3.58) |
| Experience | 1.1 | 1.2 | 0 | 4 | 0.9 | 1.1 | 0 | 4 | 0.20*** | (3.89) |
| Profile picture | 0.4 | 0.5 | 0 | 1 | 0.4 | 0.5 | 0 | 1 | 0.01 | (0.27) |
| Smoking preference | 0.0 | 0.2 | 0 | 1 | 0.1 | 0.3 | 0 | 1 | 0.06*** | (4.68) |
| Music preference | 0.4 | 0.5 | 0 | 1 | 0.4 | 0.5 | 0 | 1 | -0.06*** | (-4.68) |
| Talking preference | | | | | | | | | -0.01 | (-0.56) |
| maybe | 0.9 | 0.3 | 0 | 1 | 0.8 | 0.4 | 0 | 1 | 0.01 | (0.56) |
| yes | 0.1 | 0.3 | 0 | 1 | 0.2 | 0.4 | 0 | 1 | 0.10*** | (5.56) |
| no | 0.0 | 0.2 | 0 | 1 | 0.1 | 0.2 | 0 | 1 | -0.08*** | (-4.95) |
| Car comfort | | | | | | | | | -0.02* | (-2.02) |
| simple/normal | 0.5 | 0.5 | 0 | 1 | 0.3 | 0.5 | 0 | 1 | 0.18*** | (8.58) |
| comfortable | 0.2 | 0.4 | 0 | 1 | 0.2 | 0.4 | 0 | 1 | 0.01 | (0.42) |
| luxury | 0.0 | 0.2 | 0 | 1 | 0.1 | 0.3 | 0 | 1 | -0.04*** | (-3.54) |
| n/a | 0.2 | 0.4 | 0 | 1 | 0.4 | 0.5 | 0 | 1 | -0.14*** | (-6.59) |
| Nighttime | 0.0 | 0.2 | 0 | 1 | 0.1 | 0.3 | 0 | 1 | -0.07*** | (-4.99) |
| Morning | 0.2 | 0.4 | 0 | 1 | 0.1 | 0.3 | 0 | 1 | 0.06*** | (3.83) |
| Midday | 0.2 | 0.4 | 0 | 1 | 0.2 | 0.4 | 0 | 1 | 0.04* | (2.37) |
| Afternoon | 0.4 | 0.5 | 0 | 1 | 0.4 | 0.5 | 0 | 1 | 0.00 | (0.08) |
| Evening | 0.1 | 0.4 | 0 | 1 | 0.2 | 0.4 | 0 | 1 | -0.03 | (-1.88) |
| Sunday | 0.3 | 0.4 | 0 | 1 | 0.2 | 0.4 | 0 | 1 | 0.03 | (1.35) |
| Monday | 0.1 | 0.3 | 0 | 1 | 0.1 | 0.3 | 0 | 1 | -0.01 | (-0.66) |
| Tuesday | 0.1 | 0.2 | 0 | 1 | 0.1 | 0.2 | 0 | 1 | -0.01 | (-1.19) |
| Wednesday | 0.1 | 0.2 | 0 | 1 | 0.0 | 0.2 | 0 | 1 | 0.01 | (1.22) |
| Thursday | 0.1 | 0.3 | 0 | 1 | 0.2 | 0.4 | 0 | 1 | -0.04* | (-2.36) |
| Friday | 0.3 | 0.4 | 0 | 1 | 0.2 | 0.4 | 0 | 1 | 0.03 | (1.73) |
| Saturday | 0.1 | 0.3 | 0 | 1 | 0.1 | 0.3 | 0 | 1 | -0.01 | (-0.60) |
| Distance in km | 260.9 | 154.2 | 36 | 1178 | 290.0 | 168.8 | 39 | 653 | -29.11*** | (-3.87) |
| Price in Euro | 13.3 | 7.9 | 1 | 58 | 14.9 | 8.8 | 1 | 40 | -1.61*** | (-4.13) |
| N | 16107 | | | | 517 | | | | | |

* We decided to keep age outliers in the estimation model. Drivers that reported an age over 80 years old represent 0.004% of the sample. Excluding outliers from the estimation does not change the results

Table A3: The effect of name origin on clicks – all model coefficients (Average Marginal Effects)

| | | Empty Model | Full Model | Full model including interaction between rating and name origin (see Fig 5) |
|----------------------------|--------------|---------------------|---------------------|---|
| | | AME | AME | Coef. |
| Arab/Turkish/Persian name | | -7.725*** (1.87) | -7.178*** (1.92) | -0.334*** (0.06) |
| Number of rides per route | | 0.0613 (0.04) | -0.0973 (0.05) | -0.00194* (0.00) |
| Log days until departure | | -3.838*** (0.53) | -3.668*** (0.54) | -0.0714*** (0.01) |
| Time of day (ref. night) | morning | | -3.044 (2.13) | -0.0494 (0.04) |
| | mid-day | | -2.408 (2.10) | -0.0384 (0.04) |
| | afternoon | | -2.889 (2.17) | -0.0463 (0.04) |
| | evening | | -9.319*** (2.10) | -0.178*** (0.04) |
| Day of week (ref. Sunday) | Monday | | -2.752 (1.62) | -0.0533 (0.03) |
| | Tuesday | | -11.02*** (1.78) | -0.219*** (0.04) |
| | Wednesday | | -7.369*** (1.85) | -0.143*** (0.04) |
| | Thursday | | -10.17*** (1.33) | -0.204*** (0.03) |
| | Friday | | -2.425* (1.10) | -0.0452* (0.02) |
| | Saturday | | -6.907*** (1.45) | -0.132*** (0.03) |
| Distance in km | | | 0.277*** (0.02) | 0.00537*** (0.00) |
| Price in euro | | | -1.695*** (0.19) | -0.0330*** (0.00) |
| Female | | | 4.365*** (0.93) | 0.0825*** (0.02) |
| Age | | | -0.00684 (0.04) | -0.000107 (0.00) |
| Number of user ratings | | | 0.118** (0.04) | 0.00227** (0.00) |
| Rating | | | -0.0301 (0.31) | -0.00280 (0.01) |
| Experience (ref. Newcomer) | Intermediate | | -0.891 (1.53) | -0.0169 (0.03) |
| | Experienced | | 1.026 (1.80) | 0.0200 (0.03) |
| | Expert | | 0.970 (1.88) | 0.0213 (0.04) |
| | Ambassador | | 6.204* (2.85) | 0.119* (0.05) |
| Profile picture available | | | 1.113 (0.85) | 0.0195 (0.02) |

| | | | |
|---|--------------------|----------|---------------------|
| Smoking preference (ref. No/ Maybe) | | | |
| Yes | 2.830 (1.96) | | 0.0488 (0.04) |
| Music (Maybe/No) | | | |
| Yes | 1.025 (0.86) | | 0.0206 (0.02) |
| Dialog (ref. Maybe) | | | |
| Yes | 5.451*** (1.48) | | 0.0984*** (0.03) |
| No | 1.032 (1.96) | | 0.0205 (0.04) |
| Car comfort (ref. simple/normal) | | | |
| Comfortable | -1.374 (0.97) | | -0.0269 (0.02) |
| Luxury | 1.999 (2.08) | | 0.0366 (0.04) |
| n/a | -2.463* (1.00) | | -0.0475* (0.02) |
| Arab/Turkish/Persian name x user rating | | | 0.0678*** (0.02) |
| Observations | 16624 | 16624 | 16624 |
| <i>AIC</i> | 163749.4 | 163297.5 | 163286.0 |
| <i>BIC</i> | 164521.2 | 164285.5 | 164281.7 |

Note: AMEs based on Negative Binomial Regression model, standard errors in parentheses. Coefficients for auxiliary variables (i.e. route id) not reported. *** p<0.01; ** p< 0.05; * p< 0.1. Source: Carpooling Data Germany 2015 (compiled by authors)

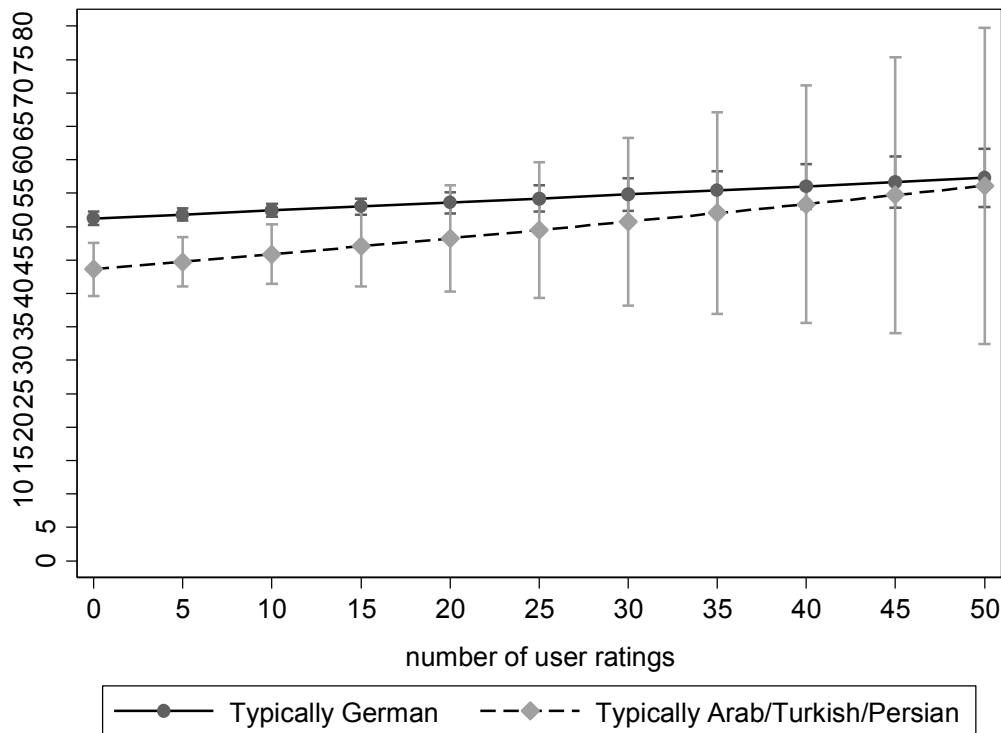
Table A4: Sub-group effects by sex, region and profile picture.

| | Men only | Women only | East Germany only | West Germany only | With Profile Picture | Without Profile Picture |
|---|---------------------|------------------|-------------------------|-------------------------|----------------------------|-------------------------------|
| Arab/Turkish/Persian (ref. German names) | -7.113*** (2.02) | -4.287 (8.12) | -4.088 (4.97) | -7.820*** (2.15) | -3.826 (3.21) | -7.625** (2.51) |
| Observations | 12299 | 4325 | 3872 | 12752 | 7396 | 9228 |
| Arab/Turk./Pers. (N) | 486 | 31 | 52 | 465 | 227 | 290 |
| <i>AIC</i> | 120583.3 | 42771.91 | 36561.04 | 126713.7 | 73114.31 | 90226.97 |
| <i>BIC</i> | 121517.9 | 43542.95 | 36855.33 | 127555.9 | 73970.99 | 91103.96 |

Standard errors in parentheses

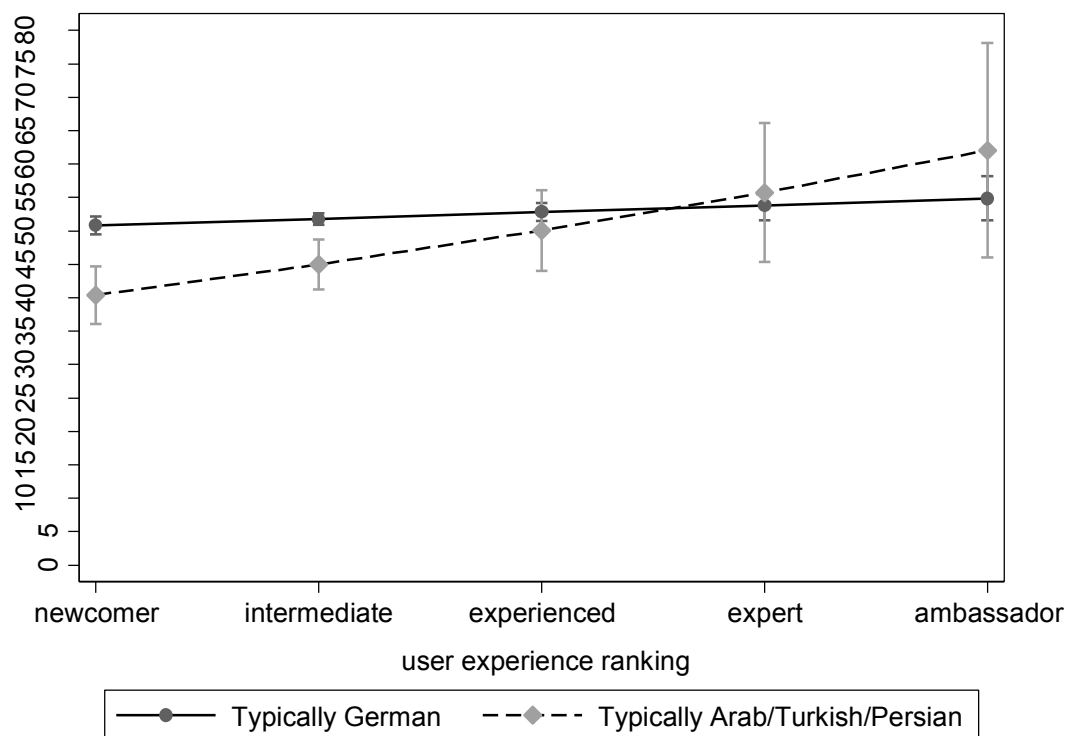
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure B2: Predicted Number of Clicks on Offered Ride by User Name Origin and Number of User Ratings



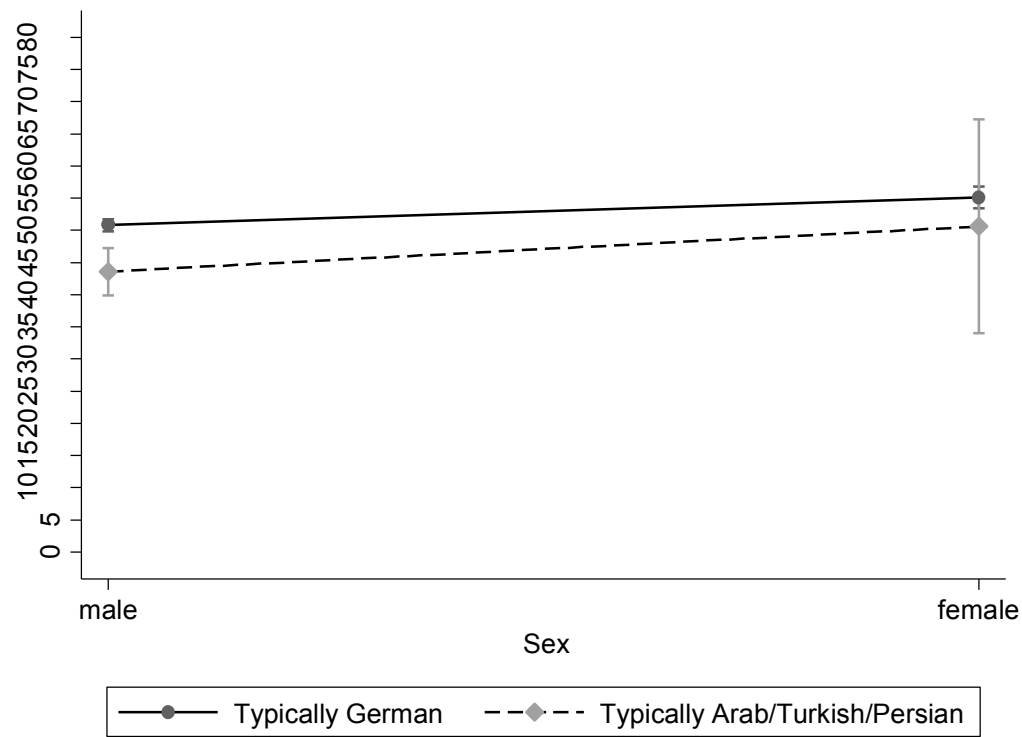
Note: Carpooling Data Germany 2015 (compiled by authors). Predicted clicks are adjusted (see Table A3 in the Appendix). N= 16,624. 95% confidence interval.

Figure B3: Predicted Number of Clicks on Offered Ride by User Name Origin and Driver Experience



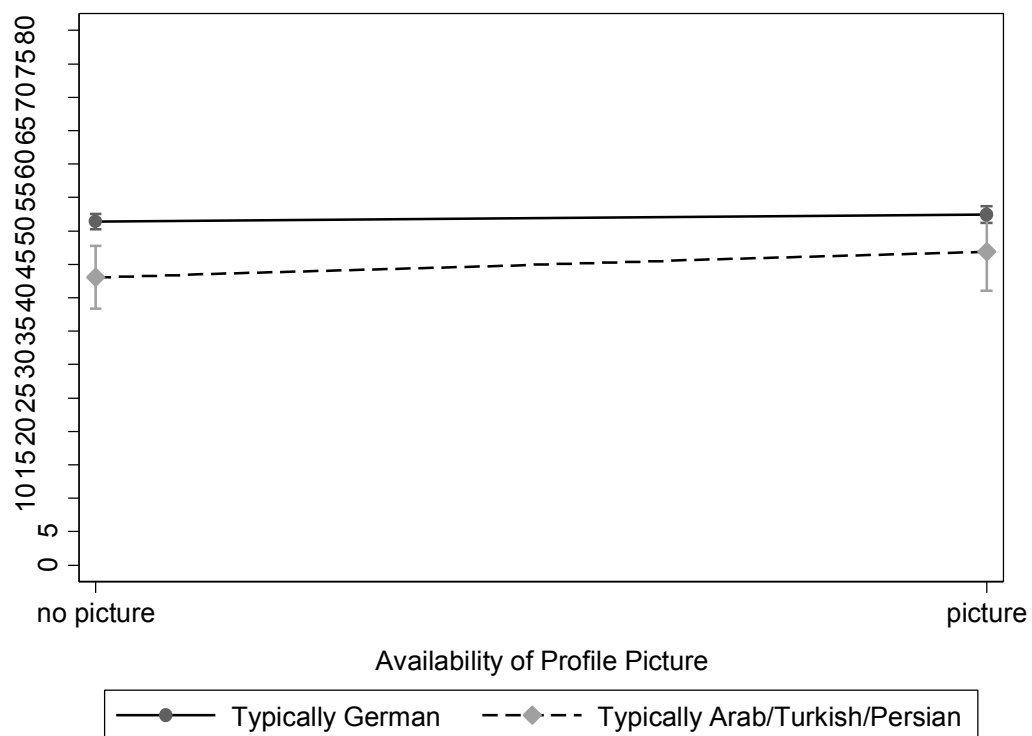
Note: Carpooling Data Germany 2015 (compiled by authors). Predicted clicks are adjusted (see Table A3 in the Appendix). N= 16,624. 95% confidence interval.

Figure B4: Predicted Number of Clicks on Offered Ride by Name Origin and Sex



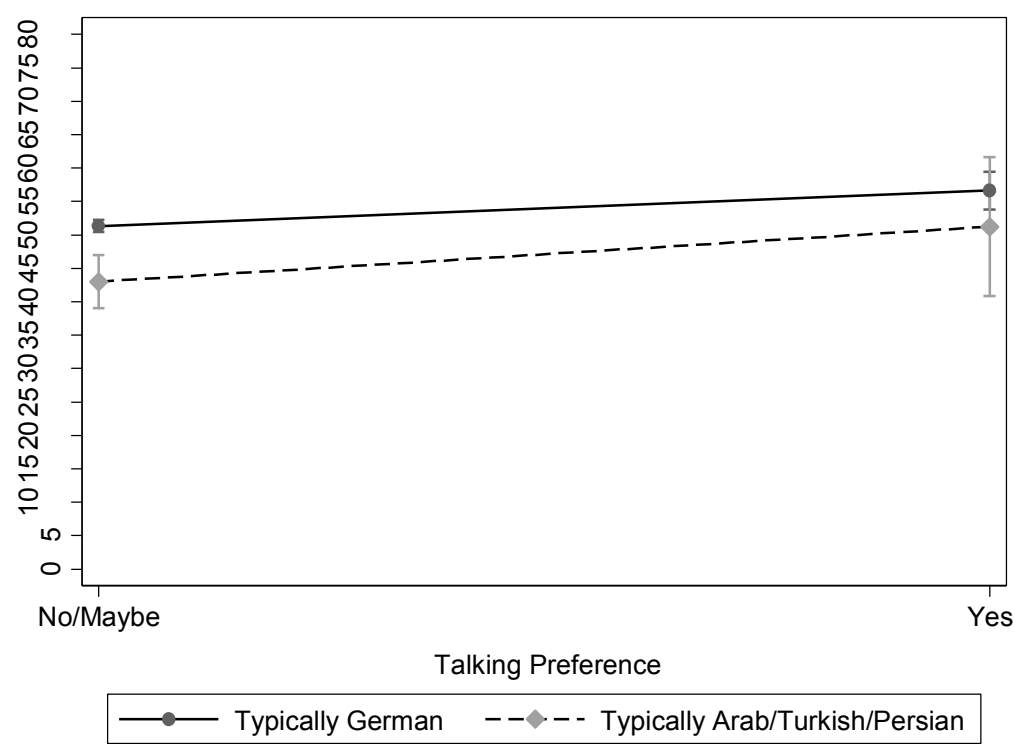
Note: Carpooling Data Germany 2015 (compiled by authors). Predicted clicks are adjusted (see Table A3 in the Annex). N= 16,624. 95% confidence interval.

Figure B5: Predicted Number of Clicks on Offered Ride by Name Origin and Profile Picture



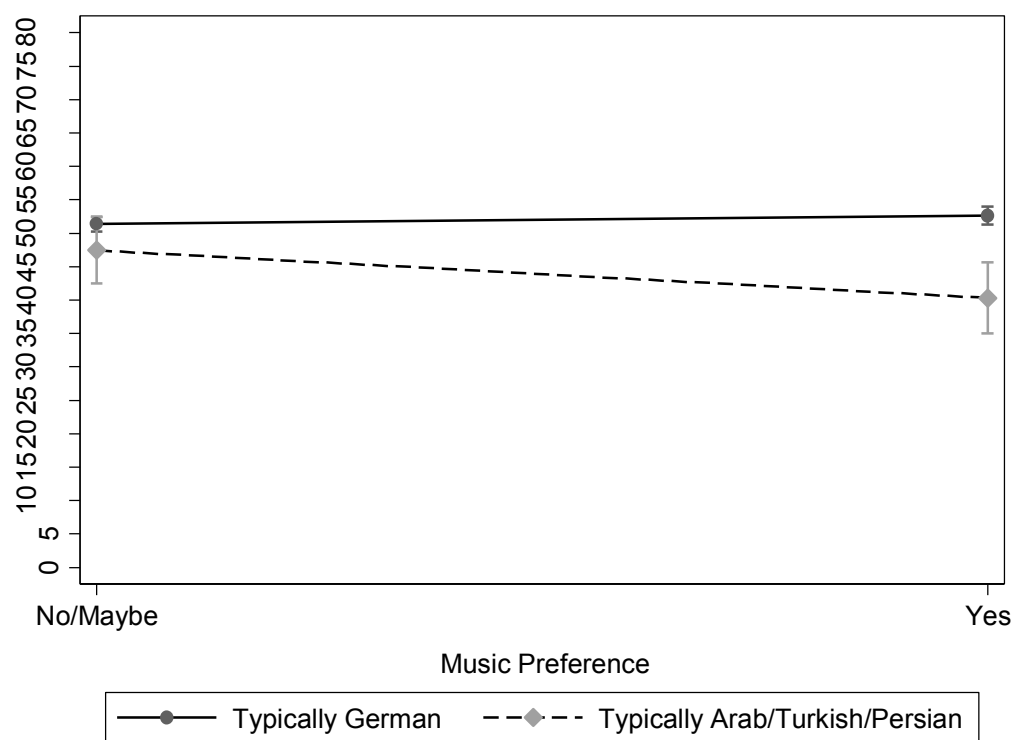
Note: Carpooling Data Germany 2015 (compiled by authors). Predicted clicks are adjusted (see Table A3 in the Annex). N= 16,624. 95% confidence interval.

Figure B6: Predicted Number of Clicks by Name Origin and Talking Preference



Note: Carpooling Data Germany 2015 (compiled by authors). Predicted clicks are adjusted (see Table A3 in the Annex). N= 16,624. 95% confidence interval.

Figure B7: Predicted Number of Clicks by Name Origin and Music Preference



Note:

Carpooling Data Germany 2015 (compiled by authors). Predicted clicks are adjusted (see Table A3 in the Annex). N= 16,624. 95% confidence interval.

Robustness checks

As referenced throughout the paper, we conducted a series of robustness checks to assess the sensitivity of our results:

First, to assess the robustness of our price and distance information, we varied how price and distance enter the model. Including both variables separately or including a separate ‘price per km’ measure does not change the results.

Second, we applied propensity score matching to assess the robustness of the key ethnic name effects. Nearest neighbor matching yields slightly larger ethnic penalties and confirms our regression results.

Table A5: Propensity Score Matching

| Nearest neighbor matching | Coeff. | AI Robust SE | z | p-value | 95% Confidence Interval | |
|---|---------------|---------------------|----------|----------------|--------------------------------|----------|
| Average Treatment Effect | -10.15991 | 2.594111 | -3.92 | 0.000 | -15.2442 | -5.07554 |
| Average Treatment Effect on the Treated | -8.577434 | 2.407805 | -3.56 | 0.000 | -13.2966 | -3.85822 |

Third, we investigate the effect of name origin for different sub-groups including by gender, region and availability of profile picture (see Table A4 in the Appendix).

Fourth, we attempt to disentangle ethnic from social discrimination effects by comparing the effect of drivers with an Arab/Turkish/Persian name and drivers with an ‘Anglo-Saxon name’. Several studies have shown that Anglo-Saxon names (i.e. Steven, Justin, Kevin) signal low social class in Germany (Kaiser, 2010). Table A6 presents the results for ethnic and social discrimination. We do not find any significant effects for the first names indicated by Kaiser (2010).

Table A6: Ethnic vs. Social Cues

| Name origin | AME | SE | z | p-value | 95% conf. Interval | |
|---------------------------|------------|-----------|----------|----------------|---------------------------|--------|
| Arab/Turkish/Persian name | -7.178 | 1.915 | -3.75 | 0.000 | -10.932 | -3.423 |
| Anglo-Saxon name | 2.753 | 2.968 | 0.93 | 0.354 | -3.063 | 8.569 |

ENDNOTES

ⁱ The exact origin of Arab, Persian and Turkish names is difficult to distinguish for a lay person. However, member of all three groups in Germany are commonly associated to be of the same migrant group with assumed cultural similarities.

ⁱⁱ See article in the German newspaper *die Welt* entitled „BlaBlaCar und Co. vor diesen hippen Mitfahrdiensten zittert die Bahn“ (i.e. these are the carpooling services that the train companies are afraid of). Accessible at <https://www.welt.de/wirtschaft/article129721188/Vor-diesen-hippen-Mitfahrdiensten-zittert-die-Bahn.html>.

ⁱⁱⁱ As a pre-test, we uploaded a limited number of artificial rides on the route Munich to Cologne, varying profiles by the name origin only (using ‘Mehmet’ and ‘Serkan’ as typical Turkish first names and ‘Johannes’ and ‘Tobias’ as typical German first names). The pre-test indicated large discrimination effects which further strengthened our objective to collect real market data.

^{iv} See Heckman (1998) and Neumark (2012) for discussions on the limitations of audit and correspondence studies.

^v Other forms of discrimination include implicit, unintentional biases (e.g. Anderson, Fryer, & Holt, 2006). We will focus our discussion on statistical and taste-based discrimination as the decision to car-pool (our application) involves conscious weighing of numerous alternatives (other competing rides) and evaluation of several characteristics (location, price, timing, age, experience, rating et cetera).

^{vi} Specific applications of statistical discrimination approaches may not be able to explain average group disadvantages when group stereotypes are, in fact, correct. However, individual members of the respective group can still be subject to discrimination (e.g. Kalter, 2006).

^{vii} See news articles referring to the issue of crimes rates and foreigners in Germany: <http://www.strafrecht-wi.de/auslaenderkriminalitaet/>; <http://www.bpb.de/politik/innenpolitik/innere-sicherheit/76639/auslaenderkriminalitaet?p=all>;

<http://www.spiegel.de/lebenundlernen/schule/kriminalitaet-von-migranten-laut-gutachten-nicht-hoher-a-983536.html>

^{viii} See news report covering a survey on the reputation of car drivers in different European countries: <https://www.welt.de/motor/news/article108612704/Europaweite-Umfrage.html>

^{ix} Stereotypes regarding high crime rates for ethnic minorities largely affect males (Trager et al. 2014).

^x We would like to thank the provider for supporting academic research by allowing access to this data.

^{xi} Goodness of Fit tests revealed that negative binomial regression is superior to other count models. Robustness checks reveal similar results for the OLS estimator.

^{xii} There is generally less demand for carpooling in rural areas. Ethnic minority drivers are also less often offer rides on rural routes. Including rural routes would hence bias the average number of clicks for German drivers downward.

^{xiii} Raters were paid for their efforts and consisted mostly of students from laboratory pools at the University of (...) and the University of (...) (all in Germany). We thereby guaranteed that no rater participated more than once. The raters did not receive information on the aim of the study to avoid demand effects. More technical information available upon request.

^{xiv} To support this fact, we collected additional data from 38 respondents in a university lab setting. Each respondent was asked to allocate a particular origin to the most frequent names in the Arab/Turkish/Persian group. On average, neither of the three origins was chosen with more than 50% certainty. In comparison, the German origin was allocated to typical German names with a certainty of over 90%.

^{xv} We also estimated the baseline model based on a subsample of rides with and without a profile picture (total N = 7,664; Arab/Turkish/Persian = 233). The effect of an Arab/Turkish/Persian name is smaller and not significant.

^{xvi} See <http://blog.airbnb.com/an-update-on-the-airbnb-anti-discrimination-review>; <https://www.theguardian.com/technology/2016/jul/20/airbnb-hires-eric-holder-racial-discrimination-bias>

3 Second Article: MPs' principals and the substantive representation of disadvantaged immigrant groups

This manuscript was accepted for publication at the journal *West European Politics* in June 2018. At the time of writing it is available online, but not yet in print.

Lucas Geese and Carsten Schwemmer (2019). "MPs' principals and the substantive representation of disadvantaged immigrant groups". In: *West European Politics* 42.4, pp. 681–704. URL: <https://www.tandfonline.com/doi/full/10.1080/01402382.2018.1560196>

MPs' Principals and the Substantive Representation of Disadvantaged Immigrant Groups

This is an accepted Manuscript of an article published by Taylor & Francis in West European Politics, available online: <https://doi.org/10.1080/01402382.2018.1560196>.

Lucas Geese^a and Carsten Schwemmer^b

^a Faculty for Social Sciences, Economics, and Business Administration, University of Bamberg, Bamberg, Germany, Lucas.Geese@uni-bamberg.de, phone: +49(0)951-863-3010 (Corresponding author); ^b Faculty for Social Sciences, Economics, and Business Administration, University of Bamberg, Bamberg, Germany, Carsten.Schwemmer@uni-bamberg.de, phone: +49(0)951-863- 2736

Funding

This work was supported by the German Research Foundation (DFG) under Grant SA 2160/3-1 (principal investigator: Thomas Saalfeld)

Acknowledgments

Earlier versions of this article were presented at the ABC Conference 2016 in Bamberg and at the 'Anxieties of Democracy' workshop 2017 in Mainz. We thank Thomas Saalfeld, Marc Helbling, Jorge M. Fernandes, Henning Bergmann, Javier Martínez Cantó, Simon Fink, Daniel Gillion, Margret Hornsteiner, Stefanie John, Ira Katznelson, Caroline Schultz and two anonymous reviewers for helpful comments and suggestions. We also thank Magdalena Stiegler, Elena Maier, David Beck, Johannes Geiger and Emanuel Slany for research assistance, and Joanna MacLeod for proofreading. Data have been obtained within the project 'Pathways to Power: The Political Representation of Citizens of Immigrant Origin in Seven European Democracies (PATHWAYS)'. This project was funded by the ANR (France), DFG (Germany), ESRC (United Kingdom) and NWO (Netherlands) under the Open Research Area (ORA+) framework. The PATHWAYS consortium is formed by the University of Amsterdam (Professor Jean Tillie), the University of Bamberg (Professor Thomas Saalfeld), the University of Leicester (Professor Laura Morales) and the CEVIPOF-Sciences Po Paris (Professor Manlio Cinalli)

Abstract

This article provides an alternative understanding of the substantive representation of immigrant-origin citizens compared to previous work in the ‘politics of presence’ tradition. Rather than assuming that the representational activities of members of parliaments (MPs) are underpinned by intrinsic motivations, it highlights extrinsic motives. Drawing on principal-agent theory, the article conceptualises MPs as delegates who are to act on behalf of their main principals, constituents and party bodies. This approach permits the rigorous analysis of the impact of electoral rules, candidate selection methods and legislative organisation on substantive representation. Based on an analysis of more than 20,000 written parliamentary questions tabled in the 17th German Bundestag (2009-13), empirical findings suggest that electoral rules do not influence the relationship between MPs and their principals in relation to the substantive representation of disadvantaged immigrant groups, however, results indicate that candidate selection methods as well as powerful parliamentary party group leaderships do.

Keywords

Substantive representation; Immigrant-origin citizens; Parliamentary questions; Electoral rules, Candidate selection, Legislative organisation

Introduction

The normative ideal of democratic representation (e.g. Dahl 1971) suggests that as more immigrant-origin residents acquire citizenship and thus the right to vote, their interests should find more consideration in the parliamentary activities of members of parliament (MPs). Indeed, the relevance of this ideal should not be underestimated, given that immigrants and their descendants remain socially and economically disadvantaged in most Western democracies of immigration (Alba and Foner 2015). Consequently, political scientists are called for to examine the mechanisms underlying the substantive representation of disadvantaged immigrant groups.

Conceptually, substantive representation refers to whether MPs ‘act in the interest of’ citizens, while descriptive representation refers to whether MPs’ sociodemographic features ‘stand for’ a certain group of citizens (Pitkin 1967). Despite this conceptual differentiation, however, normative claims of a connection between the two concepts of representation (Mansbridge 1999; Phillips 1995) has inspired the lion’s share of previous research on immigrants’ substantive representation (e.g. Aydemir and Vliegthart 2016; Saalfeld 2011; Saalfeld and Bischof 2013; Wüst 2014a). Doubtlessly, this literature has advanced our understanding considerably, confirming by and large a link between the descriptive and substantive representation of immigrant-origin citizens. Nevertheless, it is no secret that immigrant-origin citizens remain descriptively *underrepresented* in Western European parliaments (Alba and Foner 2015; Bird *et al.* 2011; Bloemraad and Schönwälder 2013). Therefore, this group of citizens has to rely to a considerable extent on the level of substantive representation provided by *native* MPs. There is, however, a dearth of research on immigrants’ substantive representation unrelated to MPs’ own national or ethnic backgrounds.

Consequently, rather than relying on the assumption of intrinsically motivated ‘descriptive’ representatives, we think in this paper of MPs as agents in a principal-agent relationship, with local voters as well as political party bodies inside and outside parliament being the most important principals (Carey 2009; Mitchell 2000; Müller 2000). Speaking on behalf of immigrants and their descendants is understood as being part of MPs’ strategies to please the demands of their principals. Yet, the incentive to please the demands of one principal at the expense of another one is a function of the institutional environment. On one hand, MPs favour the demands of a centralised party body under party-centred electoral rules, a centralised candidate selection method and due to powerful parliamentary party groups (PPGs). On the other, they are ‘pulled’ towards local demands by candidate-centred electoral rules and a localised candidate selection method (Carey and Shugart 1995; Gallagher 1988; Strøm 1997).

The question arises what happens when the incentives encoded in these institutional features conflict (Martin 2014; Preece 2014). Do MPs remain responsive to the local demands of immigrant-origin citizens when a decentralised candidate selection method clashes with party-centred electoral rules? Do they remain responsive to the demands of the PPG leadership when electoral rules are candidate-centred? To examine these questions, we turn to a case study of MPs’ legislative behaviour in the German Bundestag, a complex institutional context combining mixed electoral rules with a localised candidate selection method and tightly organised PPGs. Here, MPs are ‘pulled’ by their principals’ demands in different directions, thus providing researchers the opportunity to better disentangle the effects of institutional variables while holding country-specific context fixed. Empirically, this study is based on a dataset of all MPs serving in the 17th Bundestag (2009-13), combined with a semi-automated content analysis of more than 20,000 of their parliamentary questions (PQs) for written answer.

Quantitative analyses of this dataset suggest that principals’ demands are important determinants of the substantive representation of disadvantaged immigrant groups in MPs’ PQs.

However, our findings provide little support that different electoral rules moderate MPs' attentiveness towards the demands of their principals. A localised candidate selection on the one hand and powerful PPG leaderships on the other, by contrast, are found to be more consequential for the substantive representation of disadvantaged immigrant groups.

Institutional Context and the Substantive Representation of Disadvantaged Immigrant Groups

A major controversy in political science is the question of whether MPs should be conceptualised as *trustees*, who act based on their own conscience, or as *delegates*, who act based on the instructions of others (Converse and Pierce 1986; Pitkin 1967). Conceptualising MPs as trustees means in large parts to assume that MPs' intrinsic motivations underlie their legislative behaviour. This is basically what normative arguments in the 'politics of presence' school of thought are based on. In order to represent the interests of disadvantaged groups, representatives need to have a thorough understanding of and similar life experiences to the represented, which can be best achieved by descriptive representation (Mansbridge 1999; Phillips 1995: 159). Previous empirical research in this line of thought (e.g. Aydemir and Vliegthart 2016; Saalfeld 2011; Saalfeld and Bischof 2013; Wüst 2014a) is thus widely based on the assumption that MPs' legislative behaviour hinges on their intrinsic motivations, that is, on the trustee notion of substantive representation.

Conceptualising MPs as delegates, however, makes us aware that substantive representation may also be based on demands external to MPs' conscience and personal experiences. In this view, MPs act as agents of principals who control access to certain goods that MPs value (Carey 2009; Mitchell 2000; Müller 2000). The assumption is that MPs are driven by their ambition to reach certain career-related goals, ordered in the following way. First of all, MPs need to

achieve *reselection* as a necessary precondition for their second goal, *reelection*, which in turn is a necessary condition for the achievement of their third goal, *access to positions of influence within parliament*, such as committee membership and chairs or front-bench membership (Strøm 1997). The achievement of the first goal, reselection, is in most parliamentary democracies controlled by parties' nomination conventions (Müller 2000). The second goal, reelection, can only be achieved, if enough voters support the candidate or the party list bearing him/her (Mitchell 2000). The third goal, positions of legislative influence, is in most cases under control of the leadership of the PPG (Carey 2009). Thus, MPs typically find themselves in the difficult situation of having to please the demands of (at times) three different principals: voters, party selectorates and PPG leaders.

In the view of principal-agent theory, MPs' acting on behalf of disadvantaged immigrant groups can be therefore understood as being part of a strategy supposed to please the demands of one or several principals. The extent to which the demands of one principal outweigh the demands of another one, however, depends on the relative value of the resources controlled by each principal, which is determined by the rules of the game, that is, their institutional environment (cf. Carey 2009: 14). Among the most important institutional variables are electoral rules, candidate selection methods and the internal organisation of parliaments.

Electoral rules, to begin with, are commonly thought to determine the relative weight of local voter groups for MPs' reelection prospects relative to the weight of the party branch controlling the candidate selection process. Under closed-list PR elections, voters have little leverage to change the electoral fate of individual candidates, given they are confronted with fixed and often long lists of candidates, which voters can only take or defect as a whole (Carey and Shugart 1995; Mitchell 2000; Shugart *et al.* 2005). The list position allocated in the selection process will thus determine MPs' future electoral prospects, such that MPs should have strong incentives to follow the demands of a party selectorate (Carey 2009). By contrast, in more

candidate-centred systems, like single-member district elections, voters have more influence over the electoral fate of individual candidates, such that MPs should cultivate a relatively stronger local voter support (Carey and Shugart 1995; Mitchell 2000). Therefore, MPs should see more reasons to provide substantive representation in response to local concentrations of immigrant-origin citizens when elected in single-member districts. Conversely, the demands of national party bodies should weigh stronger on MPs' shoulders with regard to the representation of immigrants' interests under closed-list PR rules.

The candidate selection method is another factor that may affect the relationship between MPs and their principals. As already mentioned, reselection is a necessary precondition for all other career-related goals, such that MPs can be assumed to owe part of their loyalty to the gatekeepers in the candidate-selection process (Müller 2000). In this respect, the degree of territorial decentralisation is an important dimension of candidate selection (Rahat and Hazan 2001). Arguably, local party organisations should attach greater weight to the local visibility of their parliamentary representatives while national party headquarters should value MPs' efforts to cultivate a national party reputation (Gallagher 1988: 15; Karlsen and Narud 2013). Given the reputation and visibility of national MPs, their legislative behaviour should serve local party branches as an important campaigning tool for the purpose of tapping into local voter markets of immigrant-origin citizens in municipality elections. If local party branches have leverage over the reselection of MPs, they possess a means to that end, that is, the means to make their parliamentary agents speak on behalf of disadvantaged immigrant groups. Thus, the link between local concentrations of immigrant-origin citizens and their substantive representation may be the result of a localised candidate selection method. On the other hand, if the national party headquarters maintain control over the reselection of MPs, the demand of this principal should determine immigrants' substantive representation more strongly.

Legislative organisation is a third institutional feature that is particularly consequential for the principal-agent relationship between PPG leaderships and individual MPs. Strøm (1998), distinguishes a vertical and a horizontal dimension of legislative organisation. Vertically, the building blocks of parliaments are hierarchically organised PPGs (Saalfeld and Strøm 2014). At the top of this hierarchy, PPG leaderships seek to further the collective goals of the national party in terms of policy, offices and votes (Strøm and Müller 1999). To achieve these goals, however, PPG leaders depend on the collective effort of the entire party group (Müller 2000), although individual MPs sometimes face deviating cross-pressure from competing principals (Carey 2009). In order to incentivise MPs to work towards the collective goals of the party despite competing demands, PPG leaders often have a number of disciplinary instruments at their disposal: patronage and control of MPs' promotion to influential legislative or executive office, assignment to or withdrawal from certain committees, access to the parliamentary floor/rapporteurship, access to the media, and benefits such as business trips, office space, staff and a variety of other perks (Bailer 2017; Bowler *et al.* 1999; Carey 2009; Sieberer 2006; Strøm 1997). Some of these resources can strongly affect MPs' individual vote-seeking and policy goals. For example, appointment to a leadership position in the PPG can enhance MPs' policy influence, while access to the parliamentary floor in a well-publicised debate provides a public platform to enhance the MP's status among constituents or the local party base.

Disciplinary measures are commonly considered important instruments for the purpose of accomplishing party unity when bills are voted on in the plenary, thus ensuring the collective decision-making ability of the parliament (Bailer 2017; Bowler *et al.* 1999; Sieberer 2006). However, focusing solely on legislative voting in the plenary would neglect the horizontal dimension of legislative organisation, that is, the role of specialised committees. Committees play a crucial role in most parliaments as they constitute the arena in which bills are considered and amended before being mainly 'waved through' in the plenary (Cox and McCubbins 2007: 9–12). Indeed, the scarcity of time and the fact that law-making necessitates a sophisticated

level of policy-specific expertise on the part of MPs makes committee specialisation a necessary and important feature of parliamentary politics (Strøm 1998: 24–27).

Therefore, by necessity, PPG leaders have to consider that policy-making takes place in various policy jurisdictions. Plausibly, the need for an efficient division of labour is intimately connected with the principal-agent relationship between PPG leaderships and their MPs. In that sense, committees can be understood as an extension of legislative party power (Cox and McCubbins 2007; Miller and Stecker 2008; Strøm 1998). On the one hand, the assignment of MPs to the various specialised committees ensures an efficient division of labour within the PPG. On the other, the tight vertical organisation within PPGs provides PPG leaders with a vertical grip over their MPs that often effectively reaches down into MPs' committee-based work. If the PPG leadership possesses effective monitoring devices and has at its disposal the sort of disciplinary measures already discussed, it possesses effective means of incentivising individual MPs to further the collective goals of the party within the confines of the policy jurisdictions of the MP's committee specialisation (Damgaard 1995). Based on these considerations, it is thus plausible to assume that the extent to which MPs' committee assignments shape their legislative behaviour reflects the extent to which they serve their PPG leaderships as policy-specialised agents. Therefore, MPs should have incentives to further the interests of disadvantaged immigrant groups if this is a policy goal of their PPG leaderships in the policy jurisdiction of their committees.

Parliamentary Questions and the Substantive Representation of Disadvantaged Immigrant Groups in the German Bundestag

To examine this theoretical framework, we focus our study on Germany for two main reasons. First, Germany is a very relevant case to the study of immigrants' substantive representation.

Germany accounts for 20% of the entire immigrant population in the European Union (OECD and EU 2015: 40) and the immigrant-origin electorate is sizeable, amounting to 9% in the 2013 Bundestag elections (Bundeswahlleiter 2013). At the same time, however, there are strong structural inequalities separating immigrants' social and economic situations from those of the German majority population (cf. Die Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration 2016).

Second, Germany's institutional environment offers the opportunity to analyse and contrast the effects of institutional variables on the relationship between MPs and their principals. German MPs find themselves in a complex institutional environment combining mixed electoral rules with a localised candidate selection procedure and tightly organised PPGs. This environment provides researchers the opportunity to better disentangle the effects of these factors while holding constant influences of country-specific context (e.g. Moser and Scheiner 2012: 46). Indeed, it remains a matter of controversy whether electoral rules trump the effects of candidate selection methods and legislative organisation, or vice versa. Shugart and coauthors (2005: 441) argue, for example, that parties and MPs alike respond mainly to voters' informational demands encoded in the electoral system, and not, for example, to party-related candidate selection procedures. However, others have argued that centralised candidate selection methods and powerful PPG leaders weaken MPs constituency relations despite strong personal vote-seeking incentives encoded in electoral rules (Martin 2014; Preece 2014). In this article, we take these opposing views as empirical questions, leveraging Germany as an institutional environment in which principals 'pull' their MPs into different directions.

To pursue these empirical questions, we follow previous research and draw on parliamentary questions (PQs) for written answer (*Schriftliche Fragen*) as indicators of substantive representation (Aydemir and Vliegenthart 2016; Saalfeld 2011; Saalfeld and Bischof 2013; Wüst 2014a). PQs are well suited for the purpose of dealing with our research question, because

they indicate MPs' *personal* efforts to represent the interests of disadvantaged immigrant groups in response to external demands. Other legislative activities, for example speeches or roll call votes, are strictly controlled by the PPG leadership, especially in a strongly party-controlled parliament such as the Bundestag (Depauw and Martin 2009; Proksch and Slapin 2015). In comparison, MPs can use PQs relatively freely to raise the attention of the government to certain issues, to acquire information from the bureaucracy or to claim credit for their PQs in their websites, social media or local newspapers (Martin 2011b; Rozenberg and Martin 2011; Russo and Wiberg 2010).

The first major question we seek to answer is whether MPs' election in local constituencies or whether their selection as local candidates determines their responsiveness to local concentrations of immigrant voters. In Germany's electoral system, 299 MPs are elected in single-member plurality districts (SMD tier), and a slightly larger number of MPs is elected in 16 multi-member districts under rules of closed-list proportional representation (PR tier). The system is compensatory in that parties' vote shares in the PR tier determine their overall seat shares, that is, seats won in the SMD tier are used first to fill the allocated seat shares and thereafter remaining seats are drawn from state-based party lists. As MPs are elected under different electoral rules in the same system, it is often assumed that mixed-member systems generate a 'mandate-divide' between the two types of MPs, that is, SMD MPs' representative behaviour focuses more strongly on local constituents, while PR MPs focus mainly on national party bodies (for an excellent literature review see Manow 2013). Scholars of mixed-member systems, however, have expressed scepticism regarding the mandate-divide thesis, arguing that behavioural differences between the two types of MPs blur due to other institutional influences affecting MPs' behaviour in similar ways across electoral tiers (e.g. Crisp 2007; Ferrara *et al.* 2005; Manow 2013).

In the German case, such arguments often highlight how candidates are selected to run for the Bundestag (Manow 2013). Formally, the electoral law stipulates that candidates in the SMD tier must be selected in local constituencies while candidates in the PR tier must be selected at nomination conferences at the upper regional level (Detterbeck 2016). However, the electoral law permits candidates to run as ‘dual candidates’, that is, in both electoral tiers simultaneously¹. In fact, dual candidacy is common, because voters reward parties electorally for the local presence of candidates (Ferrara *et al.* 2005; Hainmueller and Kern 2008). Therefore, parties have vote-seeking incentives to require that their candidates are selected in the SMD tier *before* being allowed access to promising party list positions in the PR tier (Detterbeck 2016; Manow 2013: 289). In other words, as local re-selection is a precondition for realistic list positions in the PR tier, local reselection is de-facto a requirement for MPs’ reelection in both electoral tiers. Therefore, SMD and PR MPs alike may have incentives to represent local constituencies in their PQs.

In line with Crisp, we argue that the finding of behavioural homogeneity across electoral tiers can be taken as evidence that the incentives institutionalized in the electoral tiers are being trumped by the candidate selection process (Crisp 2007: 1462). In other words, if the locus of candidate selection were the driving force behind German MPs’ responsiveness to the demands of local immigrant-origin citizens and not their election in single-member districts, then MPs should respond to the share of immigrant-origin citizens in the constituency where they were locally *selected*. Thus, our first hypothesis reads:

MPs are more responsive to immigrant-origin citizens’ interests the more immigrant-origin voters reside in their local constituencies, regardless of their election mode (H1).

The second major question is whether MPs’ responsiveness to the demands of their PPG leaderships to speak on behalf of disadvantaged immigrant groups is mainly influenced by electoral rules or by the internal organisation of the parliament. In terms of legislative

organisation, the Bundestag could be described as a party-controlled *Arbeitsparlament* ('working parliament') based on the division of labour in policy-specialised committees mirroring the government structure (Ismayr 2012: 162; Miller and Stecker 2008). PPG leaderships maintain strong control over their MPs' committee work, as they have the prerogative of assigning MPs to, *and* withdrawing them from committees (Damgaard 1995; Miller and Stecker 2008). The strong role of parties is also reflected in the fact that committees work behind closed doors, therefore only visible to the PPG leadership as a principal. Party control is further ensured by weekly meetings of the PPGs' working groups, which mirror the committee structure and prepare the parties' positions in the committee (Miller and Stecker 2008). If MPs refuse to work in line with the policy goals of the party, the PPG leadership can apply several sanctions. These range from subtle pressure, to the dissenting MP's withdrawal from the committee, or the ultimate denial of promotion within the hierarchy of the PPG (Damgaard 1995; Ismayr 1992: 169).

While it remains relatively undisputed that PQs can serve MPs for the purpose of cultivating local voter support (Fernandes *et al.* 2018; Martin 2011a; Russo 2011; Saalfeld 2011), it is not as common to use PQs as a measure of MPs' responsiveness to the demands of PPGs (but see Bailer 2011). After all, PQs are widely considered a legislative instrument free of party control. Nevertheless, we argue that PQs matter to the principal-agent relationship between PPG leaderships and MPs, albeit in an indirect way. Our argument is based on the intuition that MPs serve their PPG leaderships as policy experts in specialised committees, as outlined in the previous section of this paper. PQs are informative for this principal-agent relationship, because they afford MPs a low-cost opportunity to gather relevant information from government departments to support their daily committee-based work (Bailer 2011; Russo and Wiberg 2010). To comply with their role as policy-specialised agents, MPs may thus ask PQs on issues in their area of expertise. Consequently, a close relationship between MPs' committee memberships and the type of PQs they ask should reflect their responsiveness to the

expectations of their PPG leaders to further collective party goals within the policy jurisdictions of their committees. Since certain committees are more likely to deal with matters of immigrants' disadvantage, for example the committee for social affairs or education rather than the committee for environment or defence, MPs sitting on these committees should be more likely to ask PQs related to immigrant matters. We thus expect that:

MPs are more responsive to immigrant-origin citizens' interests when they sit on migrant-related committees (H2a).

However, the extent to which the improvement of the living conditions of disadvantaged immigrant groups is defined as a policy goal should vary across PPGs. Plausibly, this variation is reflected in parties' election manifestos, guiding MPs in their pursuit to please the demands of their PPG leadership. For this reason, we expect an interaction between MPs' policy specialisation, reflected in their committee memberships, and the extent to which the integration of immigrants is reflected as a policy goal in the party manifesto (herein called the integration-relatedness of party manifestos).

We thus hypothesise that

the committee effect described in H2a depends on the integration-relatedness of the party manifesto (H2b).

Finally, the question remains to what extent the effects of legislative organisation are countervailed by electoral rules. As Carey (2009: 133) explains 'virtually all legislators are subject to influence by at least one principal – their legislative party leadership', but 'legislators' electoral connection to voters might pull them in directions contrary to the demands of legislative party leaders'. Accordingly, we might expect that the influence of PPG leaders to ask PQs on behalf of disadvantaged immigrant groups measured by an interaction of committee membership and integration-related party ideology will be weaker for MPs elected in SMDs as

compared to MPs elected in the PR tier. Conversely, if legislative organisation can trump effects of electoral rules entirely, we would expect that

the interaction effect of committee membership and integration-related party ideology described in H2b works regardless of electoral rules (H2c).

Data and Methods

Measuring Substantive Representation in Parliamentary Questions

To test the hypotheses laid out in the previous section, we compiled all 20,130 PQs tabled by individual MPs in the 17th German Bundestag. In order to identify PQs tabled on behalf of disadvantaged immigrant groups, we focus on the representation of their *objective interests* rather than on the representation of their *subjective interests* (for a detailed discussion see Swain 1993: 6). That is, PQs are understood to be substantively representative if they raise attention to immigrants' unequal living conditions, for example in terms of level of income, physical well-being or employment status, and/or demand the *integration* of immigrant-origin residents into German society. Integration refers here, according to Alba and Foner (2015: 5), to processes that increase the opportunities of immigrants and their descendants in major institutions such as the educational and political system and the labour and housing market. In order to identify PQs tabled on behalf of disadvantaged immigrant groups, herein called *integration-related PQs*, we combine human and dictionary-based machine coding. A detailed description of the text coding procedure and its validation, the final list of key words, as well as two examples of such questions are provided in the appendices A1 and A2 to this paper.

Based on this coding, the final measure of our dependent variables is the count of integration-related PQs per MP.

Independent and Control Variables

We measure the magnitude of the local demand of immigrant-origin citizens as the *share of foreign nationals in the local district*² and connect this information to all MPs who were running in the election as SMD tier candidates. Thus, all dually nominated MPs are linked to the constituencies in which they were *selected* to run as SMD tier candidates. PR tier legislators who did not run as a candidate in a district race (2.3% of all legislators) were excluded from the analysis. Of course, using the percentage of foreign nationals as a proxy for the immigrant-origin electorate at the constituency-level is not ideal. Nevertheless, it is the only immigrant-related indicator available at the constituency-level, and given it is highly correlated ($r=0.78$) with the group of naturalised residents of immigrant-origin at the level of differently drawn administrative districts (Wüst 2014b) we take this indicator as a reasonable approximation of the immigrant-origin electorate. The difference between SMD and PR tier MPs is captured in a dummy variable which takes values of one for *SMD MPs*.

To code the integration-relatedness of party manifestos, we utilise data from the Comparative Manifesto Project for the 2009 Bundestag election, following previous work in the field (Alonso and Fonseca 2012; Volkens *et al.* 2015; Wüst 2016). For a detailed description of the coding, please see the online appendix A3. Higher values on this continuous scale indicate more integration-relatedness. While the two right-wing parties (CDU/CSU and FDP) score low on this scale (7.121 and 6.935), the three left-wing parties (SPD, Greens and The Left) score considerably higher (16.894, 16.435 and 24.91). Additionally, party differences are captured in a dummy for the simple *left/right distinction*. For the purpose of identifying *migrant-related*

committees, we rely on a modified categorisation of the dichotomous categorisation scheme proposed by Wüst (2011)³.

In order to test whether the theoretical framework proposed in this article contributes significantly to established explanations, we also add a control variable for the effect of *descriptive representation* as the main focus of previous research. We identified all MPs as being of immigrant-origin (n=24) if they were born with a foreign nationality or if one of the respective person's parents was born with a foreign nationality. In addition to that, we control for the *total number of PQs* asked per MP. Since the extent to which PQs are used overall should depend on MPs' government or opposition status as well as on their seniority and career stages (Bailer and Ohmura 2018), we control for these factors implicitly when including this variable. Table 1 provides a descriptive overview of all variables.

Table 1: Descriptives

| | Min | Max | Mean / Share | SD |
|--|------|-------|-----------------|-------|
| No. of integration-related PQs | 0 | 52 | 0.83 | 3.86 |
| % Foreign Nationals in District | 1% | 28% | 9% | 5% |
| PR (0) vs. SMD tier(1) | 0 | 1 | 0.47 | - |
| Party: Right (0) vs. Left (1) | 0 | 1 | 0.46 | - |
| Integration-relatedness of party manifesto | 6.94 | 24.91 | 12.42 | 6.25 |
| Other (0) vs. immigrant-related committee (1) | 0 | 1 | 0.47 | - |
| Native (0) vs. migratory background (1) | 0 | 1 | 0.04 | - |
| Total no. of PQs | 0 | 196 | 30.68 | 44.58 |
| Observations | 637 | | | |

Statistical Model

The empirical modelling strategy must take into account two related methodological aspects. First, as our unit of analysis is the MP and the dependent variable captures counts of integration-related questions asked per MP, negative-binomial regression models are an appropriate choice.⁴ Second, the share of zeros in our dependent variable amounts to 82.7%. Zeros may be generated according to two different mechanisms. First, an MP decides not to ask a single integration-related PQ. Second, an MP decides not to ask any PQs at all. The latter mechanism is strongly related to the tendency of MPs representing government parties to ask no or only few PQs, while MPs of opposition parties typically ask a lot more PQs. Obviously, a major precondition to the tabling of integration-related questions is that an MP asks PQs at all. In our dataset 399 out of 637 MPs asked at least one PQ, and 110 MPs asked at least one integration-related question.

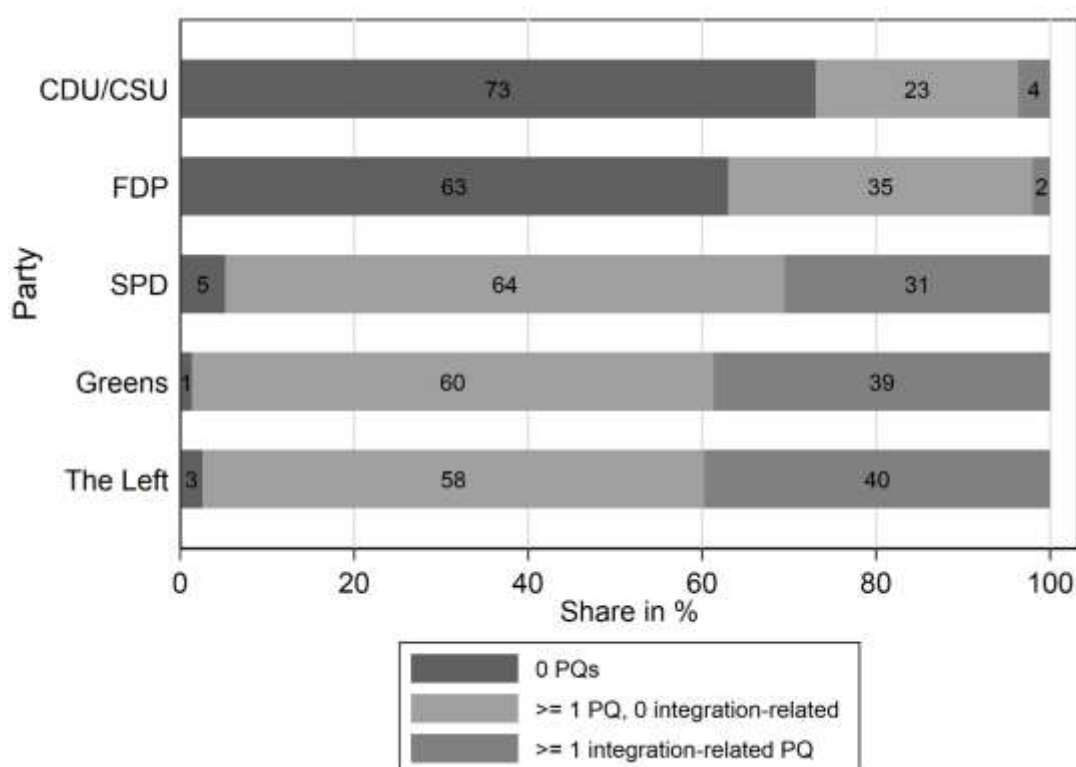


Figure 1. Percentages of MPs by party tabling no PQ, at least one PQ but no integration-related PQs, or at least one integration-related PQ.

Figure 1⁵ confirms this suspicion by showing the distribution of MPs who either tabled no PQs, at least one PQ but no integration-related PQs, or at least one integration-related PQ. In comparison to government MPs (CDU/CSU, FDP), members of opposition parties (The Left, SPD, Greens) are more likely to table more PQs overall. This is especially problematic since in the legislative term under study the division between opposition and government is clear-cut with regard to the left/ right divide. In order to better disentangle party and opposition effects and better handle the zero-inflation in our dependant variable we fit zero-inflated negative binomial regression models. These models are mixture models that combine two regression equations: a logit model to explain the zero inflation, and a negative binomial model to explain non-zero counts in the dependant variable (Cameron and Trivedi 2013: 111–76). In order to explain MPs' propensity of asking at least one integration-related PQ, we include the total number of PQs asked by each MP in the inflation equation⁶. Doing so allows us to control for factors that make MPs ask few or no PQs at all, as outlined in the previous section, in the explanations of zero-observations.

Results

In Table 2 we present the results of four estimated zero-inflated regression models. The models estimate the effects of the independent and control variables on the number of integration-related PQs in the count regression equation. Model 1 includes all independent variables without interactions, indicating that MPs ask more integration-related PQs the more immigrants reside in their constituencies, the higher the integration-relatedness of their parties' manifestos and if they sit on migration-related committees.

Model 2 extends the regression model by the interaction between the local share of foreign nationals and the distinction between MPs' election modes to test whether the constituency

effect works regardless of the electoral tier (H1). Here, the coefficient for the share of foreign nationals, which stands for the constituency effect of list MPs, is positive and statistically significant at $p < 0.1$. By contrast, neither the coefficient of the SMD tier, which stands for the average difference between list and SMD MPs, nor the coefficient of the interaction term, which stands for the difference of the constituency effect for SMD MPs relative to list MPs, reaches conventional levels of statistical significance.

Table 2: Determinants of the number of integration-related PQs

| | Model 1 | Model 2 | Model 3 | Model 4 |
|--|--------------------|--------------------|--------------------|---------------------|
| | b/se | b/se | b/se | b/se |
| Negative binomial count model: | | | | |
| % Foreign Nationals ^a | 0.06** (-0.03) | 0.06* (-0.03) | 0.06** (-0.02) | 0.06** (-0.03) |
| SMD MP | -0.14 (-0.25) | -0.15 (-0.25) | -0.13 (-0.25) | |
| % Foreign Nationals ^a * SMD MP | | 0.02 (-0.04) | | |
| Integration-relatedness of manifesto content ^a | 0.07** (-0.03) | 0.07** (-0.03) | 0.04 (-0.04) | |
| Migrant-related committee | 0.92*** (-0.25) | 0.93*** (-0.26) | 0.75** (-0.29) | |
| Manifesto ^a * committee | | | 0.04 (-0.04) | |
| Migratory background | 1.35*** (-0.42) | 1.37*** (-0.43) | 1.34*** (-0.42) | 1.28*** (-0.45) |
| <i>Reference category: SMD/ left-wing/ migrant-related committee</i> | | | | |
| PR/ left-wing/ migrant-related committee | | | | 0.29 (-0.36) |
| SMD/ left-wing/ other committee | | | | -0.79* (-0.46) |
| PR/ left-wing/ other committee | | | | -0.77* (-0.4) |
| PR/ right-wing/ other committee | | | | -3.00*** (-1.16) |
| SMD/ right-wing/ other committee | | | | -1.47** (-0.71) |
| PR/ right-wing/ migrant | | | | -0.94 |

| | | | | |
|------------------------------------|----------|----------|----------|----------|
| -related committee | | | | (-0.57) |
| SMD/ right-wing/ migrant | | | | -1.26** |
| -related committee | | | | (-0.64) |
| Intercept | -0.50* | -0.52* | -0.41 | 0.65** |
| | (-0.3) | (-0.3) | (-0.31) | (-0.33) |
| Zero-inflation logit model: | | | | |
| Total no. of PQs | -0.07*** | -0.07*** | -0.08*** | -0.07*** |
| | (-0.02) | (-0.02) | (-0.02) | (-0.02) |
| Intercept | 2.41*** | 2.40*** | 2.44*** | 2.24*** |
| | (-0.38) | (-0.38) | (-0.38) | (-0.4) |
| Intercept alpha | 0.81*** | 0.81*** | 0.81*** | 0.80*** |
| | (-0.21) | (-0.21) | (-0.2) | (-0.2) |
| N | 637 | 637 | 637 | 637 |
| Nonzero N | 110 | 110 | 110 | 110 |
| BIC | 928.47 | 934.78 | 934.4 | 951.91 |

Note: Zero-inflated negative binomial regression models; Table entries show unstandardised coefficients with robust standard errors in parentheses; ^a variable centered at global mean;

* p<0.10, ** p<0.05, *** p<0.01

Figure 2 visualises these effects. Based on model 1, the left-hand panel shows how the predicted number of integration-related PQs changes when the foreign national share increases from roughly two standard deviations below the mean up to two standard deviations above the mean. The predicted change is roughly one integration-related PQ. While this effect may seem substantially negligible, it is important to note that the mean number of integration-related PQs for our analysis is only at 0.83. Based on model 2, the right-hand panel of figure 2 shows the marginal effect of being an SMD MP conditional on the local share of foreign nationals. As can be seen, the election mode does not interact with the size of the immigrant electorate in the constituency. Overall, these findings support the contention that MPs increase their number of integration-related PQs as the share of foreign nationals rises in their local constituencies where they were *selected* rather than *elected* (H1).

Turning to the analysis of the party focus in MPs' integration-related PQs (H2a-c), Model 1 already provides evidence that the manifesto's integration-relatedness and the policy

specialisation in migration-related committees shape MPs' parliamentary questioning behaviour considerably (in line with H2a). Model 3 examines the extent to which the committee effect is contingent on the integration-related content of the party manifesto by extending Model 1 by the interaction of both variables.

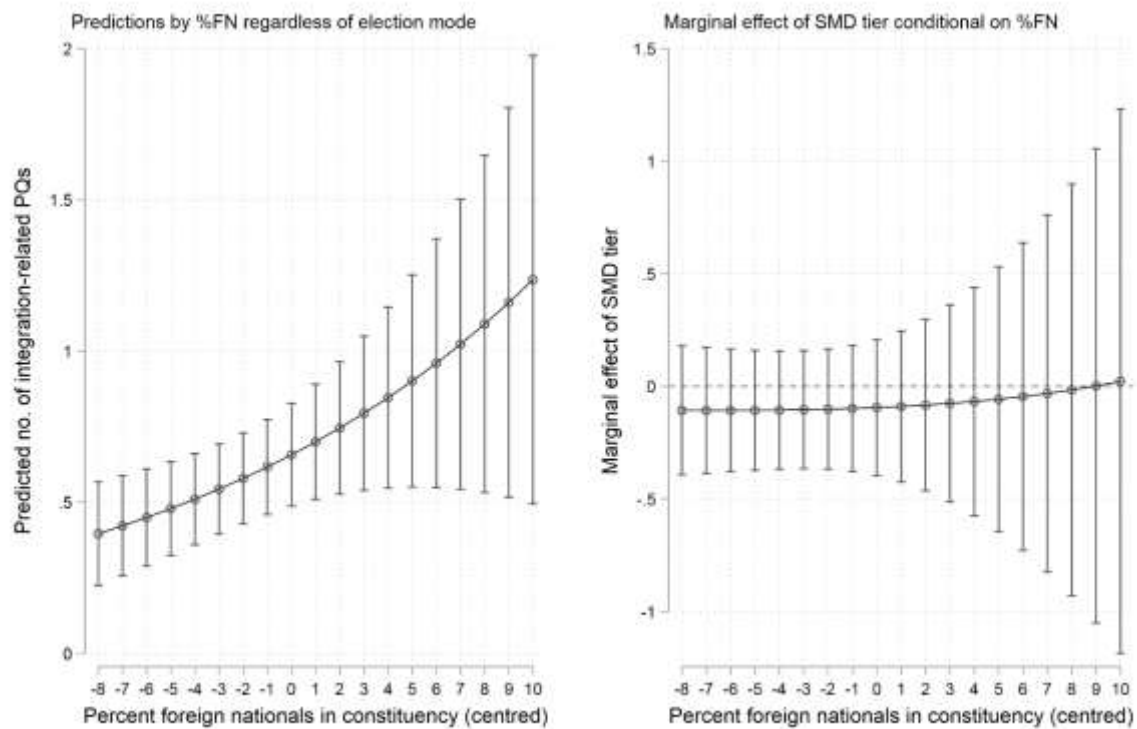


Figure 2. MPs' integration-related PQs in response to local demands with 95% confidence intervals

The coefficient for migration-related committee indicates that the effect of committee membership remains positive and statistically significant. Moreover, the interaction term indicates that as the integration-relatedness of the manifesto rises, so does the effect of migration-related committee. The calculated joint significance of the interaction term and migration-related committee is at $p < 0.01$. Estimating the marginal effects of the committee membership conditional on the manifesto's integration-relatedness (see the left-hand panel of Figure 3) indicates further that the committee effect is only noticeable if the centred manifesto scale takes values higher than -2 (10 on the non-centred scale). While right-wing MPs

(CDU/CSU and FDP) fall below, left-wing MPs are all above this threshold. Taken together, this suggests that the effect of migration-related committee membership depends on a higher degree of the manifesto's integration-relatedness (H2b).

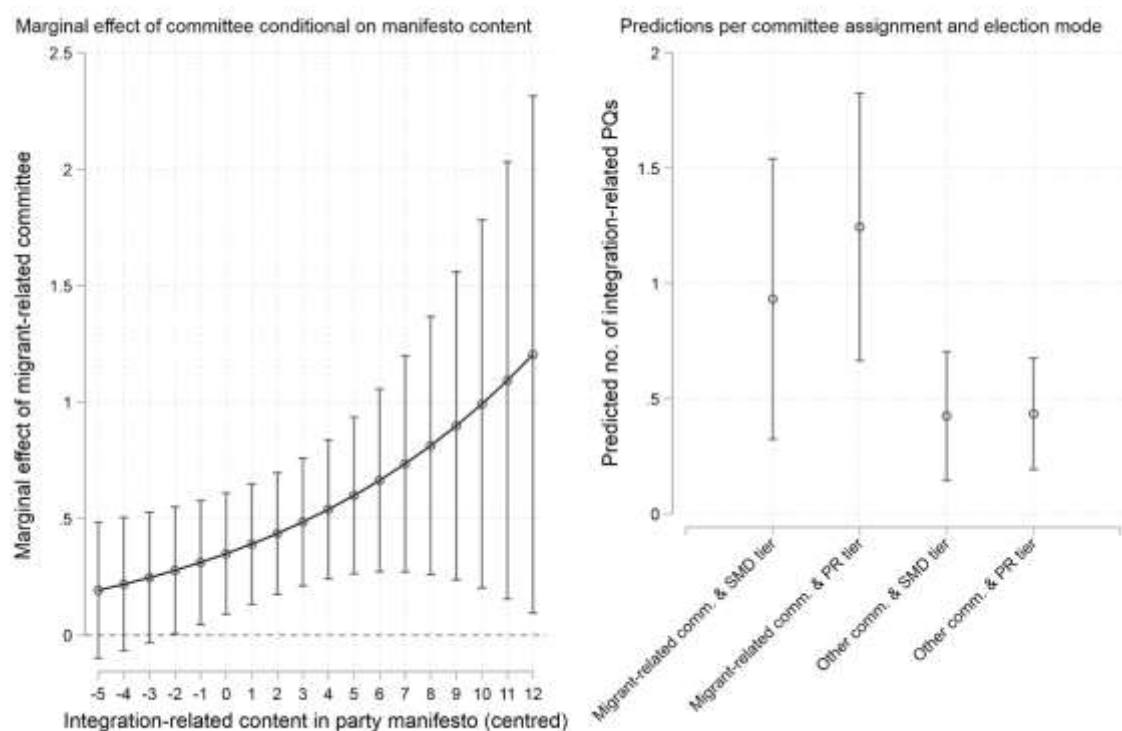


Figure 3. MPs' integration-related PQs in response to PPGs' demands with 95% confidence intervals

In model 4 we intend to test whether the party focus in MPs' integration-related PQs further depends on their election in the PR tier (H2c). For this purpose, we create a three-way interaction between SMD MP, migration-related committee and their affiliation with a left-wing as opposed to a right-wing party⁷. Since this regression table is an unwieldy format to assess the model coefficients, we direct the reader to the visualisation of the predicted counts of integration-related PQs, shown in the right panel of Figure 3. In this visualisation, the left/right PPG distinction is held at 'left-wing', while MPs committee assignments and election modes vary. As can be seen, the effect of the committee assignment does not vary greatly

between MPs' elected in the PR and SMD tier and the confidence intervals are widely overlapping. Therefore, H2c cannot be falsified based on this empirical evidence.

Furthermore, in all four models it is found that MPs of immigrant-origin are more likely to ask integration-related questions, corroborating findings from previous studies. However, the empirical evidence indicates that descriptive representation is only part of the story of immigrants' substantive representation.

Three major findings can be summarised from this analysis. First, MPs tend to ask more integration-related PQs the more foreign nationals reside in local constituencies where they were *selected* to run as district candidates. Second, they ask more of such questions when they sit on committees more likely to deal with matters of immigrants' integration as representatives of parties that make a commitment to improving the living conditions of disadvantaged immigrant groups in their manifestos. Third, these relationships seem to exist irrespective of MPs' own national backgrounds and regardless of whether they were elected under SMD or PR electoral rules in Germany's mixed-member system.

Our results are robust to different modelling strategies, which are presented in the online appendix A4.

Concluding remarks

Western representative democracies face new challenges due to the pressures of large-scale immigration creating multi-ethnic societies (Bird *et al.* 2011). Drawing on principal-agent models of democratic representation, this paper examines how institutional factors shape MPs' responsiveness to the disadvantages that immigrants and their descendants face in German society. Arguing that the role of native MPs has been underappreciated in previous research,

we conceptualise MPs irrespective of their own national backgrounds as delegates who act based on the instructions of their most important principals: local constituents, party selectorates and PPGs. This analytical perspective constitutes a contrast to the ‘politics of presence’ approach, which sees MPs rather as trustees whose conscience and personal experiences determine their legislative behaviour (Mansbridge 1999; Phillips 1995). However, we do not seek to contradict previous work based on the trustee conception. Rather, we argue that in order to improve our knowledge of the political representation of disadvantaged immigrant groups, it is fruitful to investigate relevant phenomena through a variety of analytical perspectives.

Drawing on a new dataset which includes a corpus of all 20,130 parliamentary questions (PQs) tabled by individual MPs in the 17th Bundestag, we find that the demands of MPs’ principals shape profoundly the substantive representation of disadvantaged immigrant groups in PQs. Moreover, the German institutional context, which confronts MPs with mixed electoral rules, a localised candidate selection process and tightly organised PPGs, allows us further to derive hypotheses about the behavioural consequences of these institutional features and to what extent they outperform each other. Putting these hypotheses to the test, our empirical results provide little support for the idea that differences in electoral rules shape immigrants’ substantive representation in MPs’ parliamentary questions. However, our findings do suggest, first, that MPs’ responsiveness to local concentrations of immigrant-origin citizens hinges on a localised candidate selection method. Second, their responsiveness to the demands of national party bodies to speak on behalf of disadvantaged immigrant groups is a consequence of tightly organised PPGs in the Bundestag.

Taken together, our study makes therefore two important contributions to the literature. First, it shows that our understanding of the substantive representation of immigrant-origin minorities can be advanced by conceptualising MPs irrespective of their national backgrounds as delegates

of principals inside and outside parliament. Second, this paper outlines also the limits of institutional explanations given the finding that candidate selection rules and legislative organisation are found to outperform electoral rules in their effects on immigrants' substantive representation in MPs' parliamentary questions.

Thus, future research should recognise more strongly the role of native MPs and the factors that affect their legislative behaviour. As long as different immigrant groups remain descriptively underrepresented in national legislatures, native MPs remain the most important vessel for this group's substantive representation. In other words, more research is needed to better understand MPs' legislative behaviour irrespective of their national backgrounds.

In this regard, our study of the German case is a first step. Comparative research would be a valuable extension to the present study in order to deepen our understanding of the consequences of candidate selection and legislative organisation for substantive representation across different electoral system regimes. Moreover, future research may also include other characteristics of MPs' institutional environments. For example, in many party-centred electoral systems national MPs pursue local political careers simultaneously (Fernandes *et al.* 2018; Russo 2011) or are subject to powerful local party branches in other ways (Tavits 2011). These factors can have the effect that MPs remain responsive to the demands of local concentrations of immigrant-origin citizens despite party-centred electoral rules. Given its parsimony, principal-agent theory should be a useful tool to explore the consequences of these factors in other parliamentary democracies, as well. In this light, the present contribution should be understood as a point of departure for future studies interested in the political representation of immigrants and their descendants in Western democracies.

References

- Alba, Richard, and Nancy Foner (2015). *Strangers No More: Immigration and the Challenges of Integration in North America and Western Europe*. Princeton: Princeton University Press.
- Alonso, Sonia, and Saro Claro da Fonseca (2012). 'Immigration, left and right', *Party Politics*, 18:6, 865–884.
- Aydemir, Nermin, and Rens Vliegenthart (2016). "“Minority Representatives” in the Netherlands: Supporting, Silencing or Suppressing?", *Parliamentary Affairs* , 69:1, 73–92.
- Bailer, Stefanie (2011). 'People's voice or information pool? The role of, and reasons for, parliamentary questions in the Swiss parliament', *Journal of Legislative Studies*, 17:3, 302–314.
- Bailer, Stefanie (2017). 'To use the whip or not: Whether and when party group leaders use disciplinary measures to achieve voting unity', *International Political Science Review*, 39:2, 163–177.
- Bailer, Stefanie, and Tamaki Ohmura (2018). 'Exploring, Maintaining, and Disengaging-The Three Phases of a Legislator's Life', *Legislative Studies Quarterly*, 43:3, 493–520.
- Bird, Karen, Thomas Saalfeld, and Andreas M Wüst (2011). 'Ethnic Diversity, Political Participation and Representation: a Theoretical Framework', in Karen Bird, Thomas Saalfeld, and Andreas M Wüst (eds.), *The Political Representation of Immigrants and Minorities. Voters, Parties and Parliaments in Liberal Democracies*, vol. London and New York: Routledge, 1–22.
- Bischof, Daniel (2017). 'New graphic schemes for Stata: plotplain and plottig', *Stata Journal*,

17:3, 748–759.

Bloemraad, Irene, and Karen Schönwälder (2013). ‘Immigrant and Ethnic Minority Representation in Europe: Conceptual Challenges and Theoretical Approaches’, *West European Politics*, 36:3, 564–579.

Bowler, Shaun, David M Farrell, and Richard S Katz (1999). ‘Party Cohesion, Party Discipline, and Parliaments’, in Shaun Bowler, David M Farrell, and Richard S Katz (eds.), *Party Discipline and Parliamentary Government*, vol. Columbus: Ohio State University Press, 3–22.

Bundeswahlleiter (2013). ‘5,8 Millionen Deutsche mit Migrationshintergrund sind wahlberechtigt’,
https://www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_13/presse/W13013_Wahlberechtigte_Migrationshintergrund.html (Accessed March 9, 2016).

Cameron, Adrian Colin, and Pravin K Trivedi (2013). *Regression analysis of count data*. Cambridge: Cambridge University Press.

Carey, John M (2009). *Legislative Voting and Accountability*. Cambridge: Cambridge University Press.

Carey, John M, and Matthew Soberg Shugart (1995). ‘Incentives to cultivate a personal vote: A rank ordering of electoral formulas’, *Electoral Studies*, 14:4, 417–439.

Converse, Philip E, and Roy Pierce (1986). *Political Representation in France*. Cambridge: MA: Belknap Press.

Cox, Gary W, and Matthew D McCubbins (2007). *Legislative Leviathan. Party Government in the House*. 2nd ed. Cambridge: Cambridge University Press.

Crisp, Brian F (2007). ‘Incentives in Mixed-Member Electoral Systems: General Election

- Laws, Candidate Selection Procedures, and Cameral Rules', *Comparative Political Studies*, 40:12, 1460–1485.
- Dahl, Robert A (1971). *Polyarchy: Participation and Opposition*. New Haven: Yale University Press.
- Damgaard, Erik (1995). 'How Parties control Committee Members', in Herbert Döring (ed.), *Parliaments and Majority Rule in Western Europe*, vol. Frankfurt am Main: Campus Verlag, 308–324.
- Depauw, Sam, and Shane Martin (2009). 'Legislative Party Discipline and Cohesion in Comparative Perspective', in Daniela Giannetti and Kenneth R Benoit (eds.), *Intra-Party Politics and Coalition Governments in Parliamentary Democracies*, vol. London: Routledge, 103–120.
- Detterbeck, Klaus (2016). 'Candidate Selection in Germany: Local and Regional Party Elites Still in Control?', *American Behavioral Scientist*, 60:7, 837–52.
- Die Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration (2016). *11. Bericht der Beauftragten der Bundesregierung für Migration, Flüchtlinge und Integration – Teilhabe, Chancengleichheit und Rechtsentwicklung in der Einwanderungsgesellschaft Deutschland*.
- Fernandes, Jorge M, Cristina Leston-Bandeira, and Carsten Schwemmer (2018). 'Election proximity and representation focus in party-constrained environments', *Party Politics*, 24:6, 674–685.
- Ferrara, Federico, Erik S Herron, and Misa Nishikawa (2005). *Mixed Electoral Systems. Contamination and its Consequences*. New York: Palgrave Macmillan.
- Gallagher, Michael (1988). 'Introduction', in Michael Gallagher and Michael Marsh (eds.),

- Candidate Selection in Comparative Perspective. The Secret Garden of Politics*, vol. London: Sage, 1–19.
- Hainmueller, Jens, and Holger Lutz Kern (2008). ‘Incumbency as a source of spillover effects in mixed electoral systems: Evidence from a regression-discontinuity design’, *Electoral Studies*, 27:2, 213–227.
- Ismayr, Wolfgang (1992). *Der Deutsche Bundestag. Funktionen - Willensbildung - Reformansätze*. Opladen: Leske and Budrich.
- Ismayr, Wolfgang (2012). *Der Deutsche Bundestag*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Karlsen, Rune, and Hanne M Narud (2013). ‘Nominations, Campaigning and Representation: How the Secret Garden of Politics Determines the Style of Campaigning and Roles of Representation’, in Peter Essaiasson and Hanne M Narud (eds.), *Between-Election Democracy. The Representative Relationship after Election Day*, vol. Colchester: ECPR Press, 77–101.
- Manow, Philip (2013). ‘Mixed Rules, Different Roles? An Analysis of the Typical Pathways into the Bundestag and of MPs’ Parliamentary Behaviour’, *The Journal of Legislative Studies*, 19:3, 287–308.
- Mansbridge, Jane (1999). ‘Should Blacks Represent Blacks and Women Represent Women? A Contingent “Yes”’, *The Journal of Politics*, 61:03, 628–657.
- Martin, Shane (2011a). ‘Using Parliamentary Questions to Measure Constituency Focus: An Application to the Irish Case’, *Political Studies*, 59:2, 472–488.
- Martin, Shane (2011b). ‘Parliamentary Questions, the Behaviour of Legislators, and the Function of Legislatures: An Introduction’, *The Journal of Legislative Studies*, 17:3,

259–270.

Martin, Shane (2014). ‘Why electoral systems don’t always matter: The impact of “mega-seats” on legislative behaviour in Ireland’, *Party Politics*, 20:3, 467–79.

Miller, Bernhard, and Christian Stecker (2008). ‘Consensus by Default? Interaction of Government and Opposition Parties in the Committees of the German Bundestag.’, *German Politics*, 17:3, 305–322.

Mitchell, Paul (2000). ‘Voters and their representatives: electoral institutions and delegation in parliamentary democracies’, *European Journal of Political Research*, 37, 335–351.

Moser, Robert G, and Ethan Scheiner (2012). *Electoral Systems and Political Context. How the Effects of Rules Vary Across New and Established Democracies*. Cambridge: Cambridge University Press.

Müller, Wolfgang C (2000). ‘Political parties in parliamentary democracies: making delegation and accountability work’, *European Journal of Political Research*, 37:3, 309–33.

OECD, and EU (2015). *Indicators of Immigrant Integration 2015: Settling In*. Paris: OECD Publishing.

Phillips, Anne (1995). *The Politics of Presence*. Oxford: Clarendon Press.

Pitkin, Hanna Fenichel (1967). *The Concept of Representation*. Berkeley: University of California Press.

Preece, Jessica Robinson (2014). ‘How The Party Can Win in Personal Vote Systems: The “Selector Connection” and Legislative Voting in Lithuania’, *Legislative Studies Quarterly*, 39:2, 147–167.

- Proksch, Sven-Oliver, and Jonathan Slapin (2015). *The Politics of Parliamentary Debate*. Cambridge: Cambridge University Press.
- Rahat, Gideon, and Reuven Y Hazan (2001). 'Candidate selection methods - An analytical framework', *Party Politics*, 7:3, 297–322.
- Rozenberg, Olivier, and Shane Martin (2011). 'Questioning Parliamentary Questions', *The Journal of Legislative Studies*, 17:3, 394–404.
- Russo, Federico (2011). 'The Constituency as a Focus of Representation: Studying the Italian Case through the Analysis of Parliamentary Questions', *The Journal of Legislative Studies*, 17:3, 290–301.
- Russo, Federico, and Matti Wiberg (2010). 'Parliamentary Questioning in 17 European Parliaments: Some Steps towards Comparison', *The Journal of Legislative Studies*, 16:2, 215–232.
- Saalfeld, Thomas (2011). 'Parliamentary Questions as Instruments of Substantive Representation: Visible Minorities in the UK House of Commons, 2005–10', *The Journal of Legislative Studies*, 17:3, 271–289.
- Saalfeld, Thomas, and Daniel Bischof (2013). 'Minority-ethnic MPs and the substantive representation of minority interests in the house of commons, 2005-2011', *Parliamentary Affairs*, 66:2, 305–328.
- Saalfeld, Thomas, and Kaare W Strøm (2014). 'Political Parties and Legislators', in Shane Martin, Thomas Saalfeld, and Kaare Strøm (eds.), *Oxford Handbook of Legislative Studies*, vol. Oxford: Oxford University Press, 371–398.
- Shugart, Matthew, Melody Ellis Valdini, and Kati Suominen (2005). 'Looking for locals: Voter information demands and personal vote-earning attributes of legislators under

- proportional representation', *American Journal of Political Science*, 49:2, 437–449.
- Sieberer, Ulrich (2006). 'Party unity in parliamentary democracies: A comparative analysis', *The Journal of Legislative Studies*, 12:2, 150–178.
- Strøm, Kaare (1997). 'Rules, Reasons and Routines: Legislative Roles in Parliamentary Democracies', in Wolfgang C Müller and Thomas Saalfeld (eds.), *Members of Parliament in Western Europe: Roles and Behaviour*, vol. London: Frank Cass, 155–174.
- Strøm, Kaare (1998). 'Parliamentary Committees in European Democracies', *Journal of Legislative Studies*, 4:1, 21–59.
- Strøm, Kaare, and Wolfgang C Müller (1999). 'Political Parties and Hard Choices', in Wolfgang C Müller and Kaare Strøm (eds.), *Policy, Office or Votes? How Political Parties in Western Europe Make Hard Decisions*, vol. Cambridge: Cambridge University Press, 1–35.
- Swain, Carol (1993). *Black Faces, Black Interests - The Representation of African Americans in Congress*. London: Harvard University Press.
- Tavits, Margit (2011). 'Power within Parties: The Strength of the Local Party and MP Independence in Postcommunist Europe', *American Journal of Political Science*, 55:4, 923–936.
- Volgens, Andrea et al. (2015). *The Manifesto Data Collection. Manifesto Project (MRG / CMP / MARPOR). Version 2015a*. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB).
- Wüst, Andreas M (2011). 'Migrants as parliamentary actors in Germany', in Karen Bird, Thomas Saalfeld, and Andreas M Wüst (eds.), *The Political Representation of*

Immigrants and Minorities. Voters, Parties and Parliaments in Liberal Democracies, vol. London and New York: Routledge, 250–65.

Wüst, Andreas M (2014a). ‘A Lasting Impact? On the Legislative Activities of Immigrant-origin Parliamentarians in Germany’, *Journal of Legislative Studies*, May, 37–41.

Wüst, Andreas M (2014b). ‘Immigration into Politics: Immigrant-origin Candidates and Their Success in the 2013 Bundestag Election’, *German Politics & Society*, 32:3, 1–15.

Wüst, Andreas M (2016). ‘Incorporation beyond Cleavages? Parties, Candidates and Germany’s Immigrant-Origin Electorate’, *German Politics*, 25:2, 414–432.

Notes

¹ If dual candidates are entitled to seats in both electoral tiers, they are automatically considered elected in the SMD tier and the PR tier seat will be allocated to the next candidate on the list. Due to the seat compensation mechanism between electoral tiers, parties' seat shares are not affected by these rules.

² All data used in this article, including the raw text of parliamentary questions, have been collected in the context of the PATHWAYS project (www.pathways.eu).

³ Immigrant-related committees are labour and social affairs; education and research; family, elderly and women; domestic affairs; culture and media; human rights; economic development; petitions; and the investigation committee on the fascist terror of the 'Nationalsozialistischer Untergrund' (NSU).

⁴ We chose a negative binomial model as diagnostics for a poisson model indicated overdispersion. Vuong tests further provided strong support for the use of zero-inflated mixture models.

⁵ All figures shown in this paper were generated using the Stata scheme plotplain (Bischof 2017).

⁶ We tested other specifications of the zero -inflation equation, including other variables used in the count equation. However, since these variables did not turn out to be significant and further increased the complexity of the models without improving their explanatory power indicated by a growing BIC value (Bayesian Information Criterion), we decided against their inclusion.

⁷ Because MPs' election modes are strongly dependent on their party affiliation (almost all SMD MPs are either CDU/CSU or SPD), we would run into collinearity issues if we used the manifesto variable or party dummies in the interaction. Thus, we rely here on the rather simplistic left-right distinction. However, we would argue that it is reasonable to use this rather crude measure in interaction with the committee variable to capture PPGs' demands, because as Model 3 and the left-hand plot of Figure 3 have already shown, the committee effect is only significant for the three left-wing parties, such that it should make sense to compare the questioning behaviour of left-wing SMD and PR MPs who sit on migrant-related committees.

Appendix to paper

“MPs’ Principals and the Substantive Representation of Disadvantaged Immigrant Groups”

A1: Measuring “integration-related” PQs

The texts of German PQs were extracted from official online archives of the Bundestag using Python programming scripts. All files were available as PDF documents and were converted to raw text. Subsequently, several string matching procedures were used to isolate questions and subsequently match them with MP-level information.

The measurement goal is to identify PQs that raise attention to and demand the improvement of the living conditions of immigrants and their descendants. It is important to note that we do not intend to capture sceptical positions on the integration of immigrant-origin residents, i.e. content that relates to the protection of German national identity or expresses reservations against the integration of immigrants and multicultural society. In other words, our measure should not be mistaken as a measure of saliency or positioning on a pro- vs. anti-immigrant continuum.

The following two translated examples illustrate how parliamentary written questions are used by MPs in order to raise attention to and demand the improvement of the living conditions of immigrants and their descendants.

„How does the government justify the Federal Office for Migration and Refugees recent announcement to cut the budget for integration courses in the light of the CDU, CSU and FPDs‘ coalition agreements‘ plan to qualitatively and quantitatively upgrade those courses?“ (PQ tabled by Aydan Özoğuz, SPD, May 7th 2010)

„... how does the government want to ensure that the Federal Employment Office will bring residents with a migratory background into vocational training in similar proportions in their respective age groups as compared to Germans.“ (PQ tabled by Mechthild Rawert, SPD, March 18th 2011)

Ideally, in order to identify integration-related PQs, every single question in our corpus would be inspected qualitatively to determine whether it addresses immigrants’ disadvantages in German society or not. As this is not feasible for over 20,000 questions we combine human and machine coding to identify integration-related PQs. The procedure involved four steps.

In a first step we pre-defined a list of terms which have been manually extracted from the minutes of a parliamentary debate in which integration-related issues were discussed⁷. We also added other terms that we gathered from comprehensive qualitative inspections of the PQs. We then used this list of terms to filter the corpus. If, for example, a PQ includes the term “Migrationshintergrund” (German for “migratory background”) or any other term in the list, this PQ would remain in our filtered corpus. A PQ without any terms on the list would be excluded from the corpus.

In a second step, we combined this filtered corpus with a random sample of remaining, non-filtered PQs. Two hand coders were familiarised with our definition of substantive representation and then were asked to classify each question as either integration-related or

not⁷. The intercoder-reliability in form of Cohen's Kappa (Cohen 1960) between human coders was 0.79. All coding disagreements were discussed and recoded after consensus accordingly. Additionally, from each validated question, our hand-coders again collected specific key terms which indicate that the question is integration-related. We updated the key term list accordingly.

In a third step we used the hand-coded corpus to test our updated key term list for the identification of integration-related PQs. By using the updated list of key terms⁷, 82% of all questions in our validated corpus were classified correctly⁷. In a fourth step we applied our updated key term list to all 20,130 PQs, identifying a total of 869 potentially integration-related questions in the whole corpus.

One concern with key term-based textual analyses is its susceptibility to falsely capturing irrelevant documents (false positives), while at the same time failing to capture relevant documents (false negatives). In order to keep such bias at a minimum, we inspected in a final step all 869 positives qualitatively in order to discard false positives, which left us with a total of 544 PQs as a final measure of integration-related PQs. This amounts to 2.7% of all PQs in our corpus.

Concerns regarding false negatives cannot be quantified to the same extent, but we are confident that this does not pose too great a problem to our measurement, given that we have included a random subset of the unfiltered corpus in our validation approach in step 2.

Nevertheless, to be fair, we cannot completely rule out that the captured number of integration-related PQs constitutes an underestimation of the real number of integration-related PQs in the analysed text corpus.

A2: Final term dictionary to identify questions

abgeschoben, abschiebehaftbedingungen, abschiebestopps, abschiebung, abschiebungen, altübersiedler, aufenthaltstitel, antidiskriminierungsrichtlinie, antidiskriminierungsstelle, arbeitserlaubnis, asylbewerberleistungsbezug, assoziationsrecht, asyl, asylantrag, asylantragstellern, asylanträge, asylbewerber, asylbewerberinnen, asylbewerberleistungsbezug, asylbewerberleistungsgesetz, asylbewerberleistungsgesetzes, asylbewerberleistungsgesetz, asylbewerberleistungsgesetz, asylbewerbern, asylbewerbers, asylblg, asylsuchende, asylsuchenden, asylsuchendenzahlen, asylsuchender, asylsystem, asylsystems, asylverfahren, asylverfahrenrichtlinie, asylverfahrensgesetz, asylverfahrensgesetzes, asylverfahrensrecht, asylverfahrensrichtlinie, asylverfahrensgesetz, aufenthaltsgesetz, aufenthaltsstatus, aufenthaltserlaubnis, aufenthaltserlaubnisse, aufenthaltserlaubnis, aufenthaltsgesetz, aufenthaltsgesetze, aufenthaltsgesetzes, aufenthaltsgestaltung, aufenthaltsgewährung, aufenthaltspapiere, aufenthaltsrecht, aufenthaltstitel, ausländer, ausländerbeschäftigungsrecht, ausländerförderung, ausländerjagdschein, ausländerzentralregister, ausländischer, aussiedler, balkanflüchtlinge, bleiberechtsregelung, bleibeberechtigung, bürgerkriegsflüchtlinge, bürgerkriegsflüchtlingen, diskriminierung, doppelstaatlers, drittstaatsangehörige, drittstaatsangehörige, drittstaatsangehörigen, dublin-ii, dublinüberstellungsverfahren, ehgattennachzug, einbürgerung, einbürgerungstest, einbürgerungstests, einbürgerungsverhalten, eingebürgert, einreiseerlaubnis, einreisevisum, einwanderern, einwanderungsgruppen, eu-aufnahmerichtlinie, eu-aufnahmerichtlinien, fachkräfteanwerbung, familiennachzug, familienzusammenführung, familienzusammenführungsrichtlinien, familienzusammenführungsrichtlinie, flüchtlinge, flüchtlingen, flüchtlingselend, flüchtlingskonvention, flüchtlingslager, frontex, grenzsicherug, grenzübergangsstellen, herkunftsfamilie, herkunftsland, herkunftsstaaten, integration, integrationsansprüche, integrationsarbeit, integrationscoaching, integrationsfördernd, integrationsförderung, integrationsgipfel, integrationsherausforderungen, integrationskurs, integrationskursbeteiligung, integrationskurse, integrationskursen, integrationsleistung, integrationsleistungen, integrationsministerkonferenz, integrationspolitik, integrationspolitisch, integrationsprogramm, integrationsprogramms, integrationsprojekte, integrationsgesprachkursleiter, integrationstest,

integrationsunwillig, integrationsverordnung, integriert, interkulturelle bildung, integrationsprojekte, islam, jugendintegrationskurse, jugendmigrationsdienst, jugendmigrationsdienstes, migranten, migrantinnen, migration, migrationsabkommen, migrationsbiographie, migrationshintergrund, migrationshintergrund, migrationshintergrundes, minderheitsangehoerige, minderheitsangehörige, immigranten, optionskind, optionskinder, optionspflicht, optionspflichtige, rassismus, resettlement-programms, roma-minderheit, rückführungsabkommen, rückführungsentscheidungen, rücknahmeabkommen, rückübernahmeabkommen, rückübernahmeabkommens, rücküberstellung, sammelunterkünfte, sammelunterkünften, scheineheverdachts, scheineheverdachtsfälle, sprachförderung, sprachkurs, sprachkurse, sprachkursen, sprachtest, spätaussiedler, staatenlose, staatsangehörigkeit, staatsangehörigkeitsgesetz, staatsangehörigkeitsrecht, staatsbürgerschaft, visa, visagebühren, visapflicht, visavergabe, visum, visumantrags, visumanträge, visumbefreiung, visumfreiheit, visumgebühren, visums, visumsanträge, visumsbefreiung, visumsfreiheit, visumsgebühren, visumspflicht, visumverfahren, zugewandert, zuwanderer, zuwanderern, zuwanderung

A3: Coding of party manifestos' integration relatedness

Following previous work in the field, we measure the degree to which party manifestos contain claims of integrating immigrant-minorities into society (integration-relatedness) based on data from the comparative manifesto project (Alonso and Fonseca 2012; Wüst 2016; Volkens *et al.* 2015). Similarly to Alonso and Fonseca (2012) as well as Wüst (2016), we build an additive index based on the following items: positive values for per602 (national way of life: negative), per607 (multiculturalism: positive), per705 (favourable references to underprivileged minorities); and negative values for the items per601 (national way of life: positive) and per608 (multiculturalism: negative). However, in addition to these items and in difference to the cited literature, we also add positive values for the item per503 (Equality: positive). Including the equality item per503 takes into account that policy agendas with a

focus on redistribution, equal opportunities and racial equality, tend to intersect “with the material and subjective aspirations of immigrant voters who generally find themselves socioeconomically disadvantaged or the objects of racial prejudice or social exclusion” (Messina 2007: 208). Thus, by including this item in the index, our measure comes closer to the running definition of immigrant-origin citizens’ integration (see page 16 in the main article). Nevertheless, as a robustness check, we re-estimated the first three models of Table 2 shown in the main article using a more parsimonious index that excludes per503. As can be seen in the section on the robustness checks (robustness check 3 in this appendix file), results do not change considerably when per503 is considered or not. Based on our operationalisation, the five parties achieve the following scores in 2009:

| | |
|----------|--------|
| CDU/CSU | 7.121 |
| FDP | 6.935 |
| SPD | 16.894 |
| Greens | 16.435 |
| The Left | 24.910 |

A4: Robustness Checks

As robustness checks, we refitted the models as standard negative binomial regression models on the whole sample of MPs (Robustness check 1) and on a reduced sample of MPs who have asked at least one PQ (Robustness check 2). Robustness check 3 replicates Models 1-3 from the main article using the same party manifesto coding as Wüst (2016) does.

Robustness Check 1 – Negative binomial regression models

| | Model 1 b/se | Model 2 b/se | Model 3 b/se | Model 4 b/se |
|--|--------------------|--------------------|--------------------|--------------------|
| % Foreign Nationals ^a | 0.07*** (0.02) | 0.06** (0.03) | 0.07*** (0.02) | 0.06*** (0.02) |
| SMD MP | -0.20 (0.26) | -0.22 (0.26) | -0.20 (0.26) | |
| % Foreign Nationals ^a * SMD MP: | | 0.02 (0.04) | | |
| Integration-related manifesto content ^a | 0.09*** (0.02) | 0.09*** (0.02) | 0.07** (0.03) | |
| Migrant-related committee | 1.08*** (0.26) | 1.08*** (0.26) | 0.99*** (0.29) | |
| Manifesto ^a * committee | | | 0.03 (0.04) | |
| Migratory background | 1.36*** (0.42) | 1.38*** (0.43) | 1.35*** (0.42) | 1.27*** (0.39) |
| <i>Reference category: SMD/ left-wing/ migrant-related committee</i> | | | | |
| PR/ left-wing/ migrant -related committee | | | | -0.08 (0.34) |
| SMD/ left-wing/ other committee | | | | -1.35*** (0.46) |
| PR/ left-wing/ other committee | | | | -1.12*** (0.38) |
| PR/ right-wing/ other committee | | | | -3.68*** (1.09) |
| SMD/ right-wing/ other committee | | | | -2.29*** (0.57) |
| PR/ right-wing/ migrant -related committee | | | | -1.36** (0.63) |
| SMD/ right-wing/ migrant -related committee | | | | -2.11*** (0.71) |
| Total no. of PQs | 0.02*** (0.00) | 0.02*** (0.00) | 0.02*** (0.00) | 0.02*** (0.00) |
| Intercept | -2.74*** (0.29) | -2.73*** (0.29) | -2.67*** (0.28) | -0.87*** (0.31) |
| Intercept alpha | 1.28*** (0.16) | 1.28*** (0.16) | 1.27*** (0.16) | 1.17*** (0.17) |
| N | 637.00 | 637.00 | 637.00 | 637.00 |
| BIC | 943.77 | 950.09 | 949.63 | 956.05 |

Note: Negative binomial regression models; Table entries show unstandardised coefficients with robust standard errors in parentheses; ^a variable centred at global mean; * p<0.10, ** p<0.05, *** p<0.01

Robustness Check 2 – Negative binomial regression models only for MPs who asked at least one question

| | Model 1 b/se | Model 2 b/se | Model 3 b/se | Model 4 b/se |
|--|--------------------|--------------------|--------------------|--------------------|
| % Foreign Nationals ^a | 0.05*** (0.02) | 0.04* (0.02) | 0.05*** (0.02) | 0.05*** (0.02) |
| SMD MP | -0.04 (0.23) | -0.05 (0.24) | -0.04 (0.23) | |
| % Foreign Nationals ^a * SMD MP: | | 0.02 (0.04) | | |
| Integration-related manifesto content ^a | 0.02 (0.02) | 0.02 (0.02) | 0.03 (0.03) | |
| Migrant-related committee | 0.73*** (0.22) | 0.73*** (0.23) | 0.75*** (0.27) | |
| Manifesto ^a * committee | | | -0.00 (0.04) | |
| Migratory background | 1.11*** (0.28) | 1.15*** (0.29) | 1.12*** (0.29) | 1.13*** (0.28) |
| <i>Reference category: SMD/ left-wing/ migrant-related committee</i> | | | | |
| PR/ left-wing/ migrant -related committee | | | | -0.08 (0.33) |
| SMD/ left-wing/ other committee | | | | -0.73* (0.43) |
| PR/ left-wing/ other committee | | | | -0.77** (0.34) |
| PR/ right-wing/ other committee | | | | -2.58** (1.06) |
| SMD/ right-wing/ other committee | | | | -1.01* (0.57) |
| PR/ right-wing/ migrant -related committee | | | | 0.02 (0.57) |
| SMD/ right-wing/ migrant -related committee | | | | -0.98 (0.66) |
| Total no. of PQs | 0.08*** (0.01) | 0.08*** (0.01) | 0.08*** (0.01) | 0.07*** (0.01) |
| Intercept | -1.91*** (0.21) | -1.91*** (0.21) | -1.92*** (0.23) | -0.91*** (0.30) |
| Intercept alpha | 0.80*** (0.17) | 0.80*** (0.17) | 0.80*** (0.17) | 0.74*** (0.18) |
| N | 387.00 | 387.00 | 387.00 | 387.00 |
| BIC | 861.30 | 866.84 | 867.25 | 878.21 |

Note: Negative binomial regression models; Table entries show unstandardised coefficients with robust standard errors in parentheses; ^a variable centred at global mean; * p<0.10, ** p<0.05, *** p<0.01

Robustness Check 3 – Zero-inflated negative binomial regression models with alternative manifesto coding

| | Model 1 b/se | Model 2 b/se | Model 3 b/se |
|---|--------------------|--------------------|--------------------|
| Negative binomial count model: | | | |
| % Foreign Nationals ^a | 0.06** (0.02) | 0.06 (0.03) | 0.06** (0.02) |
| SMD MP | -0.09 (0.26) | -0.10 (0.26) | -0.08 (0.26) |
| % Foreign Nationals ^a * SMD MP: | | 0.01 (0.04) | |
| Integration-related manifesto content ^a | 0.27* (0.14) | 0.28* (0.14) | 0.14 (0.19) |
| Migrant-related committee | 0.95*** (0.26) | 0.96*** (0.27) | 0.78*** (0.29) |
| Manifesto ^a * committee | | | 0.21 (0.22) |
| Migratory background | 1.31*** (0.42) | 1.32*** (0.43) | 1.29*** (0.42) |
| Intercept | -0.41 (0.30) | -0.42 (0.31) | -0.31 (0.31) |
| Zero-inflation logit model: | | | |
| Total no. of PQs | -0.08*** (0.02) | -0.08*** (0.02) | -0.08*** (0.02) |
| Intercept | 2.54*** (0.35) | 2.53*** (0.35) | 2.57*** (0.35) |
| Intercept alpha | 0.82*** (0.20) | 0.82*** (0.20) | 0.81*** (0.20) |
| N | 637 | 637 | 637 |
| Nonzero N | 110 | 110 | 110 |
| BIC | 929.98 | 936.36 | 935.71 |

Note: Zero-inflated negative binomial regression models; Table entries show unstandardised coefficients with robust standard errors in parentheses; ^a variable centred at global mean; * p<0.10, ** p<0.05, *** p<0.01

Appendix sources

Alonso, Sonia, and Saro Claro da Fonseca (2012). 'Immigration, left and right', *Party Politics*, 18:6, 865–884.

Cohen, Jacob (1960). 'A coefficient of agreement for nominal scales', *Educational and Psychological Measurement*, 20:1, 37–46.

Deutscher Bundestag (2010). 'Stenografischer Bericht. 68. Sitzung (17/68). 28 October 2010.', <http://dip21.bundestag.de/dip21/btp/17/17068.pdf>.

Grimmer, Justin, and Brandon M Stewart (2013). 'Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts', *Political Analysis*, 1:1–31.

Messina, Anthony M (2007). *The Logics and Politics of Post-WWII Migration to Western Europe*. Cambridge: Cambridge University Press.

Volken, Andrea et al. (2015). *The Manifesto Data Collection. Manifesto Project (MRG / CMP / MARPOR). Version 2015a*. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB).

Wüst, Andreas M (2016). 'Incorporation beyond Cleavages? Parties, Candidates and Germany's Immigrant-Origin Electorate', *German Politics*, 25:2, 414–432.

4 Third Article: Social Media Strategies of Right-Wing Movements - The Radicalization of Pegida

This single-authored manuscript has been submitted to the international peer-reviewed Sociology journal *Acta Sociologica*. At the time of writing, the manuscript is still under review. A preprint is available online.

Carsten Schwemmer (2019b). “Social Media Strategies of Right-Wing Movements - The Radicalization of Pegida”. URL: <https://osf.io/preprints/socarxiv/js73z/>

Social Media Strategies of Right-Wing Movements - The Radicalization of Pegida

[This paper is currently under review]

Carsten Schwemmer¹

This paper investigates how right-wing movements strategically utilize social media for communication with supporters. I argue that movements seek to maximize user activity on social media platforms for increasing on-site mobilization. To examine what factors affect social media activity and how right-wing movements strategically adjust their content, I analyze the German right-wing movement Pegida, which uses Facebook for spreading its anti-Islam agenda and promoting events in the Internet. Data from Pegida's Facebook page are combined with news reports over a period of 18 months to measure activity on Facebook and in the public sphere simultaneously. Results of quantitative text and time series analysis show that the quantity of posts by Pegida does not increase user activity, but it is the content of posts that matters. Moreover, findings highlight a strong connection between Facebook activities and the public sphere. In times of decreasing public attention, the movement changes its social media strategy in response to exogenous shocks and resorts increasingly to radical mobilization methods.

Radical right, social media, social movement, pegida, automated text analysis, computational social science

¹University of Bamberg

Corresponding author:

Carsten Schwemmer, University of Bamberg, Chair of Political Sociology, Feldkirchenstr. 21, 96052 Bamberg.
Email: c.schwem2er@gmail.com

Introduction

The use of social media to mobilize participants has become more common for political movements and protest groups in the past few years. Similarly, radical right-wing and populist forces have increasingly gained influence in many Western-European countries (Arzheimer 2015). Previous research showed that social media played an important role for this development and results of several studies indicate that online representations of social movements are no isolated bubbles of interaction, but instead do affect on-site user mobilization (Budak and Watts 2015, Harlow 2012, Poell et al. 2016, Suh et al. 2017). However, few studies investigate exactly how right-wing movements use social media sites to reach their desired audience and which strategies are used to increase user participation for spreading xenophobic propaganda in the Internet. Moreover, little empirical evidence is available about how social media strategies of such movements are affected by their received public attention and exogenous shocks. This work aims to fill this gap in the literature by posing two research questions:

RQ1: What factors influence the activity of right-wing movement supporters on social media platforms?

RQ2: How do right-wing movements adjust their social media content over time to mobilize supporters?

These questions will be examined empirically by analyzing the social media activities of the right-wing populist movement Pegida. Starting as a Facebook group in 2014, the political movement “Patriotic Europeans Against the Islamization of the West” - in short Pegida organized weekly demonstrations in the German town Dresden to protest against the allegedly ongoing Islamization of Germany and policy decisions related to the refugee crisis (Dostal 2015). To organize street-rallies, communicate with sympathizers and distribute its anti-Islam agenda on the Internet, Pegida predominantly uses the social networking site Facebook, where the movement reached over 100.000 likes within a few months (Patzelt 2016b, 140). As Pegida provides an ideal example for an emerging right-wing movement that has been using social media since its creation, a case study of its Facebook page provides valuable insights into the online mobilization strategies of such movements. To analyze activity on Facebook and in the public sphere simultaneously, I combine data from Pegida’s Facebook page with news reports from digital archives over a period of 18 months from December 2014 until May 2016. This research design allows to understand how social media activities relate to the public attention of right-wing movements, which is not possible with approaches of other studies on the subject that solely focus on social media data. By applying an automated method for measuring salience with news reports extracted from digital archives, this study therefore goes beyond analyzing social media content in isolation and examines the interplay with the public activities. Furthermore, the research design proposed in this work is applicable to many other use cases, thus allowing further research to build upon this work. Results suggest that user activities are strongly connected to public attention and that Pegida’s leadership cannot influence user activity by simply creating more content. Instead, it is the content of posts that matters. Posts with xenophobic

material generate more user activity than others. Moreover, it will be shown that the movement leadership strategically changes topics in response to decreasing salience in the public sphere and exogenous shocks. As public attention for Pegida decreases, its leadership resorts to more and more radical mobilization methods on Facebook.

Related work

Social movements, mobilization and social media

A large body of literature highlights that in order to reach their goal of achieving some kind of social change, social movements are constantly examining ways to mobilize motivated people for their cause and to gain access to more resources (Opp, 2009). In order to be successful at mobilization, factors like a group's resources, its common interests and shared identity, as well as its political power and available resources play a vital role (Harlow 2012; Tilly 1978). In the digital era, online platforms are promising tools for social movements in this regard. Social media pages enable transnational communication to reach potential supporters and only require very limited resource investment. Moreover, communication between movement organizers and followers enhances the shaping of a collective identity and connects users according to their ideological beliefs (Van de Donk et al. 2004, 5ff.).

Scholars have recently examined the use of social media platforms by social movements, showing that networking services like Twitter and Facebook are commonly used to organize social, political and protest movements. Harlow studied a Guatemalan justice movement by interviewing the leadership and analyzing the content of Facebook comments. He concluded from the interviews that organizers of the movement were “never imagining the site would prove to be such a powerful force for uniting tens of thousands of Guatemalans in protest” (2012, 238). Kavada (2015, 872) found that “social media helps to blur the boundaries between the inside and the outside of the [occupy] movement” and that social media platforms are important in the process of creating collective identity. Budak and Watts (2015) used Twitter data to analyze party support of users in times of the Gezi uprising. Their results suggest that incorporating time in research design is a crucial factor for understanding dynamics of social movements. So far, little research has been done on right-wing movements. Stier et al. (2017) compared Facebook use by Pegida and German political parties, showing that both right-wing populist actors, Pegida and the AfD party, appeal to similar target groups. The current work aims at narrowing this research gap, while at the same time providing new insights about the dynamic interplay between substantive content generated by right-wing movements, their public attention and exogenous shocks.

In order to derive expectations for the strategic social media use of right-wing movements, it is important to highlight that the activity of users is a crucial factor because it influences how fast and to whom right-wing propaganda can spread on the platform. What are the motives for right-wing movements to strategically maximize user activity on social media platforms like Facebook? The mechanisms behind the diffusion of information on Facebook, e.g. the diffusion of right-wing propaganda, are of great importance to protest groups. Facebook has a complex, algorithmically driven method of organizing news feeds of users. Information in the network spreads as users interact with content, which can be seen by befriended users. This possibly leads to friends also interacting with that content, promoting its diffusion across the network and influencing how fast and to whom information spreads in the network. In general, Facebook users seem to underestimate their potential reach (Bernstein et al. 2013). Diffusion mechanisms also apply when users interact with pages like the official Pegida page. As users subscribe to a Facebook page, its content will appear in their news feed. A high number of subscribers would benefit Pegida or any other political actor on Facebook because it simplifies distributing content to an already established user base. More importantly however, when users engage with a post of a Facebook page through commenting, liking, or sharing, this also increases the probability that this post will appear in other people's news feed (Rieder et al. 2015, 4). Moreover, if friends of a user interact with a Facebook page, this content can also appear on the users' news feed although the user her or himself is not a subscriber.

Salience in the public sphere

It is reasonable to assume that mobilization potentials of emerging social movements are related to the amount of public attention they receive and to occurring exogenous shocks that can be exploited to push a movement's agenda. However, over the course of their existence, many social movements - including Pegida - suffered from decreasing attention. As I argue that social media platforms of right-wing movements are no isolated environments, a dynamic salience measure is therefore required to understand the connection between social media activities and the public sphere. In the context of this work, salience is understood as a measure for tracking received public attention over time. Several methods for measuring issue salience have already been developed over the last decades. The majority of traditional approaches rely on survey data to measure public opinion (Weaver 1991). It is certainly feasible to analyze the salience of important issues for movements. e.g. for Pegida immigration in Germany (Czymara and Dochow 2018) with survey data. However, capturing salience of social movements themselves by using survey data is not a viable alternative for analyzing emerging movements. Corresponding survey items first would have to be developed and integrated in surveys. As social movements can appear out of nowhere and disappear just as fast, surveys are usually not sufficiently responsive. Other methods used in the literature rely on textual data. Epstein and Sengal (2000) compared several salience measures employed in studies of the US Supreme Court. Helbling and Tresch (2011) used a qualitative coding scheme for newspapers and compared the results with other

approaches to measuring positions and issue salience of parties. They found that coding media coverage data is resource draining, but nevertheless allows for establishing long-time data series and retrospective data collection. A more recent alternative for capturing issue salience relies on aggregated data from the Google search engine Google Trends allows researchers to analyze time series of search term's popularity at no cost and thus provides "an attractive data source for social scientists" (Mellon 2013, 1). However, Mellon also notes that the potential of Google Trends depends on a lot of factors, including the specificity of the search terms used and the method only captures relative trends. In summary, survey data are not an adequate method for measuring issue salience or public attention of social movements. The use of media data, e.g. newspaper articles, allows the use of textual content to not only aggregate counts but also to provide additional context information. Both media data and Google Trends enable time series analysis and retrospective data collection, but qualitative coding of media data is very time-intensive and Google Trends only provides relative indicators. Because of these shortcomings, I apply a scalable approach based upon digital news archives, which will be described further in the data and methods section.

Expectations

As outlined above, movements in general have an incentive to maximize user activity because they want to reach new sympathizers. Therefore, it can be expected that Pegida tries to generate a high quantity of posts, as more content can potentially stimulate more user interaction. In addition, it is reasonable to assume that movements like Pegida will qualitatively choose topics that either directly affect mobilization on the streets, or stimulate a lot of user interaction, which in turn indirectly increases mobilization potential. Furthermore, it can be expected that, besides Facebook content, salience in the public sphere is an important factor for user activity on the platform, as issues covered in traditional media are also likely to increase public interest in corresponding social media channels (RQ1). Regarding expectations for the adjustment of social media content over time, right-wing movements seek to exploit external events if they can be used to warrant their agenda, which can eventually result in more user activity and therefore increase potential for on-site mobilization. For the case of the right-wing movement Pegida, its leadership is expected to dynamically adjust its Facebook content in response to important exogenous shocks that are salient in the public sphere so long as such events can be utilized to spread xenophobic and islamophobic content (RQ2).

The right-wing populist movement Pegida

Since its creation, Pegida very effectively utilized Facebook as a platform for propaganda and mobilization. Within a few months, the movement reached over 100.000 likes on Facebook (Patzelt 2016b) and received international media attention (e.g. Connolly 2014). In addition,

Pegida played an important role for the establishment of the right-wing populist party AfD in Germany. For these reasons and due to the fact that only few research has been conducted on social media usage of right-wing populist movements, much can be learned from analyzing Pegida in a case study. To briefly describe its historical development and political agenda at this point, the movement emerged in 2014 and was founded by a small number of citizens living in Saxony, which has been dominated by conservative politics in the last years (Dostal 2015, 523). Primarily driven by political motives, Pegida supporters fear an allegedly Islamization of Western culture. According to Pegida, this Islamization would lead to an increasing “alienation” of German culture and language and would increase the danger of religious wars on German territory. For this reason, the movement’s supporters claim that the German asylum policy should be more restrictive and delinquent immigrants be deported immediately. Since October 2014, Pegida has been organizing weekly demonstrations in the German city of Dresden to protest against the Islamization of the West. While first events only attracted a low number of participants, the movement quickly experienced an enormous upswing. According to police reports, 25,000 protesters attended an event in January 2015, resulting in even more public interest and media coverage since Pegida’s creation. However, following a number of crises the organization has gone through, including a rift between its leaders, Pegida’s public attention declined steadily soon after its peak (Dostal 2015, 525f). During the summer of 2015, counts of protesters rarely reached over 3000 which also resulted in decreasing media interest. In October 2015 demonstrations reached a second upswing when Pegida celebrated its anniversary, reaching over 15,000 supporters. With a few exceptions, events then attracted a somewhat stable number between 2,000 and 3,000, all the way until May 2016.

Two phrases in particular, often shouted by protesters during demonstrations, were seen as trademarks of Pegida: “We are the People!” (in German “Wir sind das Volk!”), illustrated a strong group identity representing ordinary citizens and “Lying press!” (in German “Lügenpresse!”), referred to news coverage which, according to the Pegida, misrepresented their actions. Both phrases also have important historical meanings. “Lying press!” was used by the Nazis to agitate against Jewish and leftist newspapers, whereas “We are the People” was shouted during demonstrations in Eastern Germany in 1989 and 1990. These trademark phrases - and Pegida’s content in general - is in line with populist views of representing “the people”, opposing to “the corrupt elite” (Mudde and Kaltwasser 2017, 4).

Regarding Pegida’s supporters, several research teams conducted field studies to survey the protesters and understand why Pegida reached its unexpected public attention (Vorländer et al. 2015, Daphi et al. 2015, Patzelt 2016a). While these studies were not able to analyze representative samples of movement supporters, they nevertheless report a rather consistent image of participants being predominantly male, working- and middle-class members with an average age over 50. In comparison to other Germans, Pegida survey participants were ranked as considerably farther right on the political spectrum, ranging from centrist up to extreme right (Patzelt 2016b, 160ff). Regarding the content of speeches held by members of the organizers and invited guests at demonstrations, speakers were ranked as clearly

islamophobic and xenophobic. Overall, speeches were classified as populist and for most topics radical. Signs of cultural racism were apparent wherever Muslims were mentioned.

Data and Methods

Data for this work was collected over a period of 18 months between December 2014 and May 2016. The netvizz application (Rieder 2013) was used to connect to the Facebook Programming Interface and extract texts and summary statistics from Pegida's page. In total, 3,765 posts and 1,312,397 user comments were retrieved. The posts represent all the content generated by the movement's leadership and user activity was measured with all comments on Pegida posts. I focus on comments instead of likes and shares as commenting can be done repeatedly. Every comment in turn raises the chance of visibility and therefore affects mechanisms for information diffusion (Bene 2017, 6). To capture Pegida's salience over time, I propose an automated procedure by extracting and processing data from the LexisNexis archive. LexisNexis is a digital news archive which stores news reports in several languages for local newspapers and magazines, but also major world publications like the Guardian. Articles can be retrieved as plain text and include meta data like time stamps and subject terms for each document. I extracted articles of 116 available German news sources, which included Pegida as a subject term to create a dataset of 24,279 news articles. Subject terms were chosen as filter criteria in comparison to filtering by any instance of Pegida in the text to ensure a minimum of false positives. An overview of the top 50 sources and the corresponding number of articles is available in Supplementary Appendix A. Out of this dataset, aggregated daily counts for the number of news articles related to Pegida are used to analyze salience over time in comparison to activities on Facebook.

In addition, all news report texts are analyzed to understand time-dependent context of Pegida articles. For this purpose, the time period of the dataset is split into intervals of three months length. Afterwards a support vector machine (Crammer and Singer 2001, Pedregosa et al. 2011) is trained on the time interval categories. Support vector machines are supervised models that learn features with the most predictive power for some values of interest. This allows to discover the most important terms for correctly classifying an article as being published in the corresponding interval. The model further allows to examine issue-related content over time without the need of hand-coding or similar resource intensive procedures. In conjunction with aggregated daily counts for the number of news articles related to Pegida, this provides measure of context-enriched issue salience.

For automated text analysis, corresponding texts first are processed into a corpus with common methods of text preprocessing (Grimmer and Stewart 2012): terms within documents are treated as bags of words, where each term represents a single feature and information on word order is discarded. Terms were also reduced to their stem form, such that for example "family" and "families" become a common feature "famili". In addition, stop words with no semantic meaning, like German equivalents for "the" or "a", were removed from the corpus.

As for the interplay between Pegida's Facebook content, user activity and salience over time, correlations and granger causality tests were applied (Granger 1969). Granger tests are useful to examine whether values of a time series X provide more predictive power to forecast the development of another time series Y than by only using lagged values of Y . In the context of this paper granger tests are applied to analyze whether either the amount of Facebook content generated by Pegida or its salience in the public sphere substantively influence the activity of Facebook users.

Moreover, to shed light on the determinants of user activity, it is important to not only compare how many posts are generated, or how salient Pegida is over time, but also which topics are discussed on Facebook and whether specific subjects generate more user activity than others. To categorize posts into different topics, a structural topic model was fitted to the corpus (Roberts et al. 2014). Topic models are unsupervised models – like dimensionality reduction techniques such as cluster analysis - and help to automatically discover latent topics from text documents. In these models, a topic can be understood as a set of words representing interpretable themes and documents are represented as a mixture of these topics. For each document, proportions across all topics sum up 100%. As an example, after fitting a topic model, a post could for instance mostly be capturing a topic “Islamization” with a proportion of 60%, “foreign policy” with 30% and other topics with 10%. In addition to representing documents as a distribution of topics, structural topic models further allow the inclusion of document-specific covariates that are meaningful to affect both document-topic proportions and word distributions over topics. Drawing on this feature, I incorporated dates of posts as an explanatory covariate to analyze how topic proportions vary over time. While topic models are very useful for reducing the dimensionality of textual data, one disadvantage is that the number of topics must be chosen in advance by the analyst. As the corpus with 3.765 posts is rather small and a classification into broader themes is more useful for this work than high levels of granularity, a model for ten topics was fitted to the corpus of Pegida posts.¹ Afterwards, topics were examined qualitatively to assign labels by finding representative posts with high proportions for a given topic. Additionally, the FREX metric was utilized, which indicates terms that are both frequent and exclusive for each topic (Lucas et al. 2015, 19). Finally, the most prevalent topics were determined for each post and used in combination with time stamps to model effects of topical content and time on the number of comments each post received. Using a negative-binomial model for comment counts, this allows to analyze whether topic and/or time effects are more meaningful for explaining user activity on Facebook. Goodness of Fit tests indicated that a negative-binomial distribution is more appropriate than a poisson distribution to model comment counts due to overdispersion.

Results

An overview of Pegida content

Before answering the research questions for this paper, it is worth to first provide descriptive information about the content Pegida disseminates on Facebook. Inspecting the 100 most common (translated) terms

in the posts reveals that Pegida very often refers to itself within posts (“#Pegida”). This can be interpreted as an attempt to manifest a collective identity and is in line with a general populist view of “we down here against the upper class”. Many common terms are used in the context of protest mobilization, where Pegida prompts the users to take to the streets (“#OnTheStreet”) for weekly demonstrations on Monday (“#MondayIsPegidaDay”) at the usual times of (“18,30”) in the German city (“Dresden”). Several terms also illustrate Pegida’s xenophobic core issues, frequently using terms for Islamization (“#Islamization”), closing borders (“#CloseBorders, #SuspendSchengen”) and demands for deportations (“#GetOutAsylumBetrayers”). References and criticism against politicians, especially the German Chancellor Angela Merkel (“#MerkelNeedsToDisappear”), are another common theme.

It is also striking that Pegida very frequently uses hashtags within Facebook posts, which mainly serves two purposes: labeling the content with expressions known by its supporters and indexing posts to enlarge their reach on the platform. One might question whether Pegida administrators are fully aware of the possibility to search for hashtags on the Facebook platform, which is another important aspect of information diffusion. An exemplary translated post from January 19th of May 2015 shows that they are:

“Thanks Kathrin! You took our view very well and held your ground against the constantly interrupting, aggressive and arrogant CDU politician Spahn. Next time together with Rene or Lutz! This was only the first round which was clearly won by you! #DresdenShowsHowToDoIt PS: All the stupid comments on some watch-site - for which we do not want to provide reach with links or hashtags - obviously show how they boil with rage because of Kathrin’s confident performance. Beforehand, they predicted a big disaster. Well, once again a prove that do-gooders just don’t have a clue about anything.”²

The post relates to the appearance of a Pegida member, Kathrin Oertel, as a discussant in a German TV show. The text clearly indicates that Pegida knows about the effects of links and hashtags on information diffusion as they explicitly caution against the use of such features to refer to another anti-Pegida Facebook page.

Dynamics of user activity on the platform

Facebook pages benefit from high rates of user participation because more activity increases the probability of reaching new supporters on the platform. Therefore, organizations like Pegida are encouraged to positively affect user participation, where a straightforward way of doing so is to create more content which users can interact with. This raises the important question of whether Pegida is able to influence user activity by simply posting more often. Another important aspect to consider is Pegida’s salience as a general public issue. In times were Pegida receives more attention, one can expect that this also leads to more people participating in related online activities. On average Pegida created seven posts per day, on which the users commented 2,524 times, and 43 news reports about Pegida were

published per day. However, magnitudes for these measures vary substantially over time. To allow for comparisons, time series for posts, comments, and news reports were first smoothed by rolling means over 15 days to remove seasonality noise. Second, time series were normalized, such that value 0 (1) indicates minimum (maximum) activity. Figure 1 illustrates these normalized series in combination with annotations for important external events.

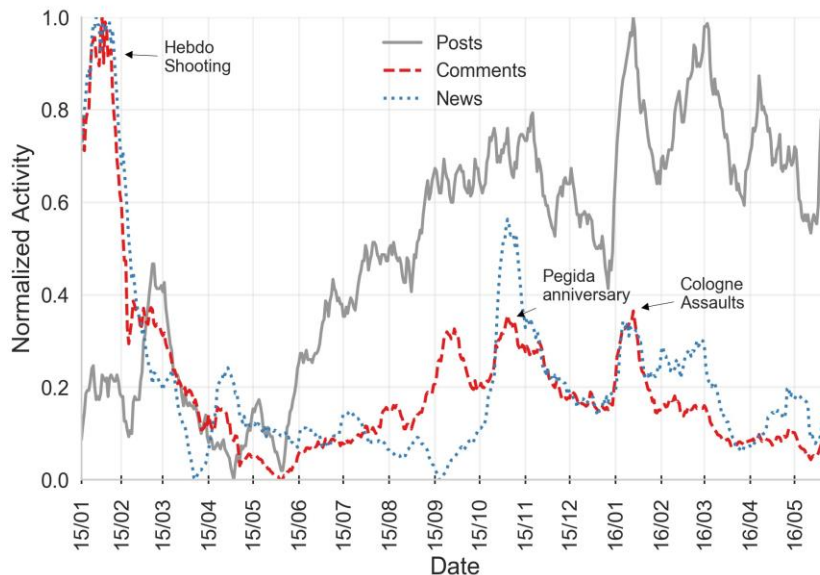


Figure 1. Normalized time series for posts, comments and news reports

As for the quantity of Pegida posts, the figure shows that Pegida continuously increased its content output over time. In contrast, the number of comments declined, reaching a peak in mid-January after the Charlie Hebdo shooting (10.000 daily comments), followed by very little activity in consecutive months (2000 daily comments). Overall peaks in user activity are in line with external events. For instance, between October and November 2015, Pegida celebrated its extensively advertised anniversary.

In January 2016, the activity increased again after New Year's Eve sexual assaults on women in the German town Cologne, for which mostly Northern-African and Arabic men were held responsible in the media. Past research indicates that the topic of immigration and sexual violence became more salient after this event in German media outlets (Czymara and Schmidt-Catran 2017). As will be shown below, the Cologne Assaults were also heavily exploited by Pegida to legitimate its xenophobic agenda. Overall, data does not support the assumption that Pegida can increase user participation by simply creating more content, with a correlation of -0.10 (0.09 unsmoothed) between post and comment counts. With regards to the salience of Pegida as indicated by the relevant media coverage, similar to user

activity, salience decreased over time. Major salience spikes correspond with those for user activity and are also related to external events.

Most importantly, there is a strong connection between salience and user activity, with a correlation of 0.88 (0.59 unsmoothed) between news and comment counts. As a robustness check, I also used Facebook likes instead of comments as an alternative measure for user activity. The relation to salience is somewhat weaker but nevertheless substantial, with a correlation of 0.56 (0.38 unsmoothed). As for the relation to the number of on-site protesters at Pegida events in Dresden, Supplementary Appendix C includes an additional comparison of normalized activities. The graph shows that attendance at protest events is also strongly connected to both Facebook activity as well as the salience of Pegida. To provide further evidence that user activities are more strongly related to external events than the content generation by Pegida organizers, granger hypothesis tests were applied. If user participation is caused by either an increase of Pegida posts or media coverage, these measures should have significantly more predictive power for participation than just using lagged values of participation in isolation. Table 1 shows test results in form of F statistics and p values for effects of Facebook posts and news articles on the number of Facebook comments. Results are displayed for included time lags between one and seven days, meaning that earlier values of posts or news between one day up to one week are tested as predictors for user activity.

Table 1. Granger test results for predictors of user activity

| Lags | <i>Fposts</i> | <i>Pposts</i> | <i>Fnews</i> | <i>Pnews</i> |
|------|---------------|---------------|--------------|--------------|
| 1 | 0.961 | 0.327 | 1.329 | 0.250 |
| 2 | 0.753 | 0.471 | 2.869 | 0.058 |
| 3 | 1.861 | 0.135 | 14.155 | 0.000 |
| 4 | 1.744 | 0.139 | 6.162 | 0.000 |
| 5 | 1.573 | 0.166 | 7.786 | 0.000 |
| 6 | 1.632 | 0.136 | 6.104 | 0.000 |
| 7 | 1.511 | 0.161 | 11.025 | 0.000 |

The table provides further evidence that the quantity of Pegida posts is not an important factor for explaining user activity. In comparison, for including time lags between three and seven days, Pegida's salience provides significant predictive power for forecasts of user activity. However, these results need to be interpreted with caution and do not clearly indicate a causal relation, as this procedure does not control for other potential causes of user activity. In addition, the effect might also be reversed in a small number of cases.³ Nevertheless, based upon the results of time series comparisons and granger tests, it is reasonable to assume that media coverage is an adequate measure of Pegida's salience at a given point in time and that public attention is most important for increasing user activity on Facebook.

What is in the news?

As described in the data and methods section, the extraction of news reports not only allows to create a times series for analyzing Pegida's salience, but also the use of report texts to show in which context Pegida was a common subject within several time intervals. In doing so, I can reveal which events were important for the movement and its received attention at a given point in time. For this purpose, a support vector machine was used to find terms with the highest probability for correctly classifying news articles into corresponding time intervals. Supplementary Appendix A includes the top ten most distinguishing terms for each interval. In the first three months Pegida's name and stemmed terms for Islam criticism and Islamization were used more frequently in comparison to other intervals. This is not surprising as the movement was a rather new phenomenon and journalists used these terms to introduce Pegida to readership. Other terms relate to important national and international events which were also utilized by the movement to warrant its position. For example, "charlie" in the first period was mentioned in articles about Pegida's reactions to the Charlie Hebdo shooting. Terme for New Year's Eve in the fifth period were used in context of the Cologne Assaults. Shortly after, Pegida protesters used signs referring to refugees responsible for sexual assaults as "rapefugees". In later intervals, references to the refugee crisis, to protests against refugee accommodations and between March/May 2016 also to the right-wing party AfD, were dominant in news reports about Pegida.

Topics and their variation over time

After showing that Pegida is not able to influence user activity on Facebook by simply increasing its post output, an important question remains to be answered: Does it at least matter what kind of material is distributed? As described above, scholars observed that while Pegida's salience and the number of on-site protesters decreased over time, contents of speeches during Pegida demonstrations shifted to more extreme positions on Islam, the refugee crisis and other related topics (Patzelt 2016a). To analyze whether similar changes can also be observed for online content, a structural topic model was fitted on all available posts. This approach makes it possible to uncover latent themes and topical variation over time. For each topic one example of a highly representative post by Pegida is available in Supplementary Appendix D for this paper. An overview of topic proportions and labels is given in Figure 2.

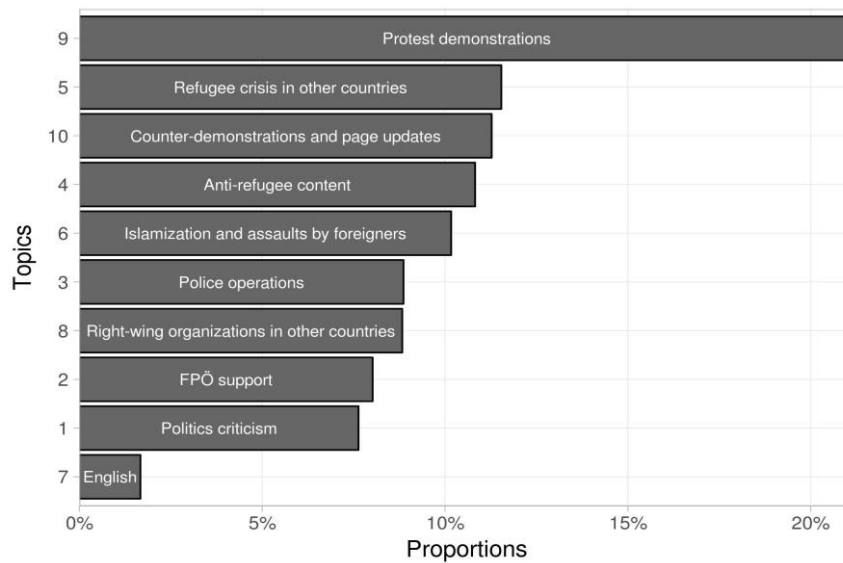


Figure 2. Topic proportions in Facebook posts by Pegida

Overall, more than 20% of Pegida's online content deals with demonstrations, which is not surprising, as this is the primary topic for increasing on-site mobilization. Also, references to how the refugee crisis is handled in countries other than Germany are common. Moreover, Pegida often distributes content about police operations in which predominantly foreigners were involved. Criticism of politicians, the government and elites in general is also apparent, as is Pegida's public support of the Austrian right-wing party FPÖ. Furthermore, Pegida generated a small number of English posts. A lot of content falls into a broader anti-refugee category, where foreigners and especially refugees are discriminated against. More than 10 percent issue Islamization and foreigner assaults. Two of these topics are of special interest: firstly, the topic about demonstrations, as related posts are most important for Pegida to potentially mobilize online users and convince them to join protest demonstrations on the street. Secondly, the topic about Islamization and assaults by foreigners, as the analysis of news report texts above demonstrated that foreigner assaults were an important issue for Pegida after the exogenous shock of the Cologne Assaults.

Table 2 lists translated terms for both topics that are frequently used and exclusive in both topics, determined by the FREX metric (Lucas et al. 2015, 19). Terms for the remaining topics are available in Supplementary Appendix D.

Table 2. Terms associated with topics about demonstrations and Islamization / assaults

| Demonstrations | Islamization/Assaults |
|---|--|
| watch, www.youtube.com, join, #Legida, livestream, thank, #DresdenShowsHowToDoIt, #JoinUs, theater_place, clock, #dresden, monday, face, patriot, tomorrow, youtube | #islamization, cologne, religion, christ, muslim, mosque, school, islam, woman, sexual, religious, book, paris, islamist, arab, church |

Pegida uses hashtags, asking supporters to join them in common places for demonstrations usually performed on Monday. In posts about demos, the movement also distributes links to corresponding live streams on YouTube. In comparison, the topic about Islamization and assaults by foreigners is strongly related to many religious terms and includes references to New Year's Eve sexual assaults in Cologne. After the Cologne incidence, sexual assaults and other types of attacks by foreigners became a dominant theme in the German media and, as described above, were also utilized by the movement to warrant its xenophobic position. To answer the second research question, how content of Pegida's posts changed over time, variations in topic proportions for demonstrations and Islamization / foreigner assaults are illustrated in Figure 3.

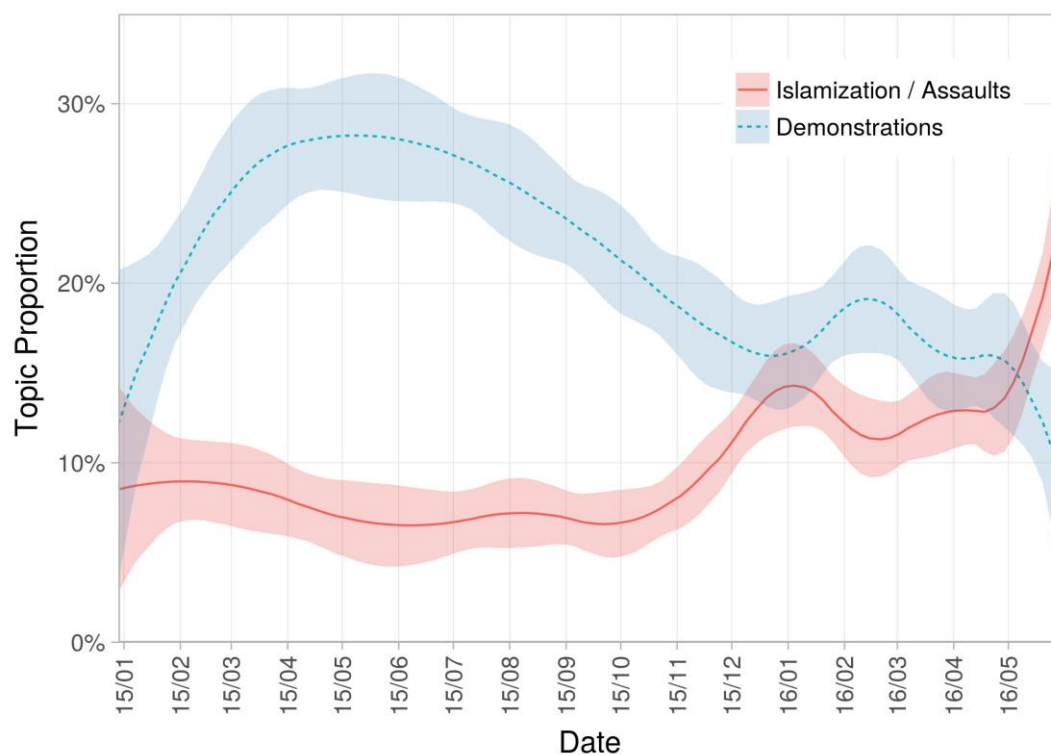


Figure 3. Estimates of topic proportions over time with 95% confidence intervals

The figure shows that over 25% of Pegida's Facebook content issues demonstrations in the first months, but proportions are decreasing and reach a share of less than 10% in May 2016. In contrast, Islamization and foreigner assaults were not a major topic in the early stages but increased over time, reaching a midway spike after the Cologne Assaults and are used more often than demonstrations in the last months of observation. These results show that Pegida's leadership dynamically adjusts online content, posting less about demonstrations when its salience decreases and exploiting external events by distributing more xenophobic material. What is more, this finding is also consistent with qualitative

analysis of speeches held during Pegida demonstrations, where negative and extreme attitudes towards immigration and Islam were increasingly used over time (Patzelt 2016b).

The relation between topics and user activity

Is this strategic change of topical content on Facebook working for the radical movement? If so, posts that predominantly contain content about Islamization and foreigner assaults should result in more user activity than others, ultimately leading to an effective dissemination of rightwing propaganda. If content of topics matters, it should be relevant even when accounting for time, which, as shown above, is strongly related to the salience of the movement. To analyze effects of the interaction between time and topic categories on user participation, topics with the highest corresponding proportions were identified for each post. Afterwards, negative binomial regressions were fitted, with the number of comments per post as a dependent variable and topic category in addition to time as explanatory variables. Regression tables for all models are given in Supplementary Appendix E. The full model with the best fit includes a quadratic term for non-linear time trends and an interaction effect between topic and time. Results suggest that, in comparison to the reference topic on Islamization and assaults, most other topics generate less user activity. Even when controlling for time, posts about Islamization and assaults result in more users participating than for demonstration related posts. Figure 4 illustrates topic effects over time with estimates from the regression model.

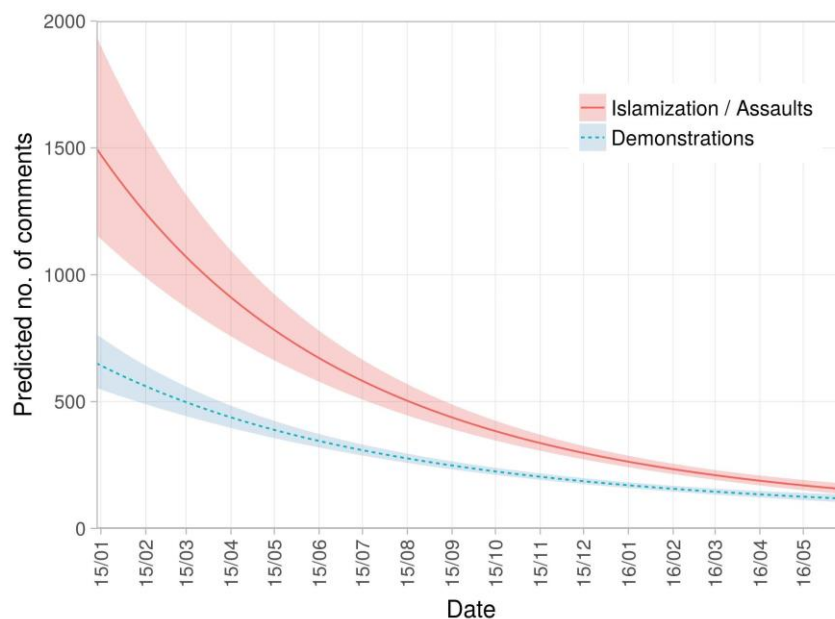


Figure 4. Predicted comment counts with 95% confidence intervals

The figure displays estimated comment counts for posts about demonstrations in comparison to posts about Islamization and assault. Across the whole observation period, posts about Islamization and assaults lead to more people commenting on Facebook than mobilization posts about demonstrations. Furthermore, user activity per post strongly decreases over time - similar to Pegida's received public

attention and the number of protesters on the street - regardless of a post's topic. This result shows that, besides time as the most important factor, topical content of posts nevertheless does affect the user activity on the platform and that xenophobic material leads to more activity than content for protest mobilization. One possible explanation for the stronger effect of xenophobic content on user activity is that related posts are more likely to affect mood or emotion of Facebook users, which eventually also increases their need to express opinions by leaving a comment on the public page (Jasper 2011). An important implication of this finding is that in general, radical and xenophobic content of right-wing movements potentially diffuses faster and reaches a wider audience than other content.

Conclusion

Results of this work about the Facebook use of Pegida have shown that in order to explain what factors influence participation on social media sites of right-wing movements, it is important to not only consider activities taking place on the platform itself, but also how the public attention towards movements change over time. Although Pegida tried to create increasingly more content, high quantities alone do not lead to more user interactions, which are mostly determined by changes in the public attention that Pegida receives. Most importantly, it is the content of posts that matters. Over time, the movement created more xenophobic material about topics like Islamization and foreigner assaults, which attracted more users than other themes. This suggests that right-wing movements resort to more and more radical mobilization methods over time, which underlines the responsibility of social media platforms to successfully detect and remove obnoxious content. Although disentangling the causal relation between online activities and the public sphere is notoriously difficult, findings of this work further suggest a possible reinforcement process between the strategies of right-wing movements and the reactions of the audience: more radical posts lead to more user reactions and more reaction will eventually lead to more radicalized posts by Pegida. However, the radicalized mobilization of the movement also leads to less mobilization from the public, since more radical methods do not appeal to an audience with moderate ideology. Over the long run, a lack of exogenous shocks that can be utilized to push xenophobic agenda as well as the radicalization of Pegida are possible reasons for a declining supporter base.

Are these changes in content distribution strategic in nature, and were thus planned by Pegida's leadership? Observations from this work strongly support the assumption that these changes were indeed not undertaken without reason. Firstly, as shown by a post in the results section, Pegida administrators are well aware of how social media features like hashtags and links can be used to increase the reach of organizations on Facebook. Therefore, one can also expect that administrators analyze user activities on posts. Secondly, when comparing context terms contained in Pegida-related news articles and topic changes within Facebook posts over time, results suggest that Pegida adapts Facebook content based upon which issues - if they can be exploited to warrant its position - are salient in the public sphere. This case study also showed that research on social media usage of political groups can greatly benefit from

incorporating media coverage. The dynamic measure of Pegida's salience used in this paper not only unfolded a strong connection with changes in user activity over time, but also shows why the rightwing movement adjusted its social media strategy. Future research should therefore consider the interdependency between social media activities and the public sphere and be cautious with analyzing platforms like Facebook in isolation.

Despite novel insights into social media strategies of right-wing movements, conclusions from this study are also constrained by limitations. While user activity is an important factor for shaping social media pages, it is not only relevant how often users interact, but in addition who participates in such debates. Due to Facebook's data policies it is difficult to provide valid estimates of the sociodemographic attributes of its users, although some results suggest that in general, Pegida's Facebook users are younger and more conservative than the average demonstration participant (Patzelt 2016b, 323ff). In addition, the majority of comments on Pegida's page are by its supporters, but there is also a small number of people on the page who dislike Pegida and disagree with the position it takes. However, when it comes to reaching a maximum number of people for mobilization purposes, negative comments are still more useful than no activity at all, because mechanisms of information diffusion on Facebook apply regardless of user positions. With regards to the increasing use of radical mobilization methods by Pegida, this paper focused on the prevalence of related textual content. Qualitative coding could be used in future research to examine whether the toning of posts became more negative or included even more extreme arguments over time.

At last, the question remains to what extent findings of this study can be generalized to movements with ideologies other than (populist) radical right. While different political movements have different agendas, they can all be expected to share the interest in maximizing social media user engagements for increasing on-site mobilization. At the very least, it is therefore reasonable to assume that for instance also left-wing social movements adjust their social media content strategically. However, it is still unclear whether the strategic use of platforms like Facebook in times of decreasing public attention generally leads to increasingly radical mobilization methods, regardless of ideology. To answer this question, further research about the connection of social media activity and the public sphere of other political movements is necessary. Unfortunately, in 2018 Facebook closed its data interface for Facebook pages, limiting the potential for future studies on that matter. It is to hope that future scholars find other ways of studying the social media use of (right-wing) movements in an era where companies are increasingly restrictive about providing their data for scientific research.

Endnotes

1. I used the R package *stminsights* to further inspect models with 20 and 30 topics (Schwemmer 2018). The model with ten 10 topics provided the best substantive fit.
2. The German original version of the post is available in Supplementary Appendix B.
3. In one case, a member of the Pegida leadership, Lutz Bachmann, posted a picture of him styled as Adolf Hitler, which went viral and also received a lot of news coverage.

References

- Arzheimer, K. (2015). The AfD: Finally a Successful Right-Wing Populist Eurosceptic Party for Germany? *West European Politics* 38(3): 535–556.
- Bene, M. (2017). Go viral on the Facebook! Interactions between candidates and followers on Facebook during the Hungarian general election campaign of 2014. *Information, Communication & Society* 20(4): 513–529.
- Bernstein, M.S., Bakshy, E., Burke, M. and Karrer, B. (2013). Quantifying the invisible audience in social networks. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*. Association for Computing Machinery (ACM). ISBN 9781450318990, p. 21.
- Budak, C. and Watts, D. (2015). Dissecting the Spirit of Gezi: Influence vs. Selection in the Occupy Gezi Movement. *Sociological Science* 2: 370–397.
- Connolly, K. (2014). Estimated 15,000 people join ‘pinstriped Nazis’ on march in Dresden. The Guardian. URL <http://www.theguardian.com/world/2014/dec/15/dresden-police-pegida-germany-far-right>.
- Crammer, K. and Singer, Y. (2001). On The Algorithmic Implementation of Multiclass Kernel-based Vector Machines. *Journal of Machine Learning Research (JMLR)* 2(Dec): 265–292
- Czymara, C.S. and Dochow, S. (2018). Mass Media and Concerns about Immigration in Germany in the 21st Century: Individual-Level Evidence over 15 Years. *European Sociological Review* 34(4): 381–401.
- Czymara, C.S. and Schmidt-Catran, A.W. (2017). Refugees Unwelcome? Changes in the Public Acceptance of Immigrants and Refugees in Germany in the Course of Europe’s ‘Immigration Crisis’. *European Sociological Review* 33(6): 735–751.
- Daphi, P., Rucht, D., Kocyba, P., Neuber, M., Roose, J., Scholl, F., Sommer, M., Stuppert, W. and Zajak, S. (2015). *Protestforschung am Limit: Eine soziologische Annäherung an Pegida*. WZB Berlin.
- Dostal, J.M. (2015). The Pegida Movement and German Political Culture: Is Right-Wing Populism Here to Stay? *Political Quarterly* 86(4): 523–531.

- Epstein, L. and Segal, J.A. (2000). Measuring Issue Salience. *American Journal of Political Science* 44(1): 66.
- Granger, C.W.J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* 37(3): 424.
- Grimmer, J. and Stewart, B.M.(2012). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis* 21(617): 267–297.
- Harlow, S. (2012). Social media and social movements: Facebook and an online Guatemalan justice movement that moved offline. *New Media & Society* 14(2): 225–243.
- Helbling, M. and Tresch, A. (2011). Measuring party positions and issue salience from media coverage: Discussing and cross-validating new indicators. *Electoral Studies* 30(1): 174–183.
- Jasper, J. M. (2011). Emotions and social movements: Twenty years of theory and research. *Annual Review of Sociology*, 37, 285-303.
- Kavada, A. (2015). Creating the collective: social media, the Occupy Movement and its constitution as a collective actor. *Information Communication and Society* 18(8): 872–886.
- Lucas, C., Nielsen, R.A., Roberts, M.E., Stewart, B.M., Storer, A. and Tingley, D. (2015). Computer-Assisted Text Analysis for Comparative Politics. *Political Analysis* 23(2): 254–277.
- Mellon, J. (2013). Where and When Can We Use Google Trends to Measure Issue Salience? *PS: Political Science & Politics* 46(02): 280–290.
- Mudde, C. and Kaltwasser, C.R. (2017). *Populism: a Very Short Introduction*. Oxford University Press. ISBN 9780190234874 0190234873.
- Opp, K.D. (2009). *Theories of Political Protest and Social Movements: A Multidisciplinary Introduction, Critique, and Synthesis*. London: Routledge.
- Patzelt, W. and Klose, J. (2016a). *PEGIDA. Warnsignale aus Dresden*. Number 3 in Social coherence studies. Dresden: Thelem. ISBN 9783945363447.
- Patzelt, W. (2016b). "Rassisten, Extremisten, Vulgärdemokraten!" Hat sich PEGIDA radikalisiert? Dresden.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12(Oct): 2825–2830.
- Poell, T., Abdulla, R., Rieder, B., Woltering, R. and Zack, L. (2016). Protest leadership in the age of social media. *Information Communication and Society* 19(7): 994–1014.
- Rieder, B. (2013). Studying Facebook via data extraction. In: *Proceedings of the 5th Annual ACM Web Science Conference on - WebSci '13*. Association for Computing Machinery (ACM). ISBN 9781450318891, pp. 346– 355.
- Rieder, B., Abdulla, R., Poell, T., Woltering, R. and Zack, L. (2015). Data critique and analytical opportunities for very large Facebook Pages: Lessons learned from exploring "We are all Khaled Said". *Big Data & Society* 2(2): 205395171561498.

- Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder-luis, J., Gadarian, S.K., Albertson, B. Rand, D.G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science* 58(4): 1064–1082.
- Schwemmer, C. (2018). stminsights. A 'Shiny' Application for Inspecting Structural Topic Models. URL <https://cran.r-project.org/web/packages/stminsights/index.html>.
- Stier, S., Posch, L., Bleier, A. and Strohmaier, M. (2017). When populists become popular: comparing Facebook use by the right-wing movement Pegida and German political parties. *Information Communication and Society* 20(9):1365–1388.
- Suh, C.S., Vasi, I.B. and Chang, P.Y. (2017). How social media matter: Repression and the diffusion of the Occupy Wall Street movement. *Social Science Research* 65: 282–293.
- Tilly, C. (1978). From Mobilization to Revolution. Reading, MA: Addison-Wesley Publishing Company.
- Van de Donk, W., Loader, B.D., Nixon, P.G. and Rucht, D. (2004). *Cyberprotest: New media, citizens and social movements*. Routledge.
- Vorländer, H., Herold, M. and Schäller, S. (2015). Wer geht zu PEGIDA und warum? Eine empirische Untersuchung von PEGIDA-Demonstranten in Dresden.
- Weaver, D. (1991). Issue salience and public opinion: Are there consequences of agenda-setting? *International Journal of Public Opinion Research* 3(1): 53–68.

Supplementary Appendix: Social Media Strategies of Right-Wing Movements - The Radicalization of Pegida

A. Newspapers

Newspaper Sources

The following table shows the top 50 out of 114 newspapers used to measure issue salience, ordered by the number of articles.

| Newspaper | No. of articles | | |
|---|-----------------|-----------------------------------|-----|
| Sächsische Zeitung Regionalausgaben | 2243 | Hamburger Abendblatt | 373 |
| Sächsische Zeitung Stammausgabe Dresden | 1330 | Kölner Express | 365 |
| Frankfurter Rundschau | 1069 | Berliner Morgenpost | 361 |
| taz, die tageszeitung | 967 | SDA - Basisdienst Deutsch | 324 |
| dpa-AFX ProFeed | 905 | SPIEGEL ONLINE | 293 |
| Agence France Presse – German | 896 | Berliner Kurier | 263 |
| Rheinische Post Duesseldorf | 85 | Allgemeine Zeitung | 247 |
| Berliner Zeitung | 773 | ZEIT-online | 230 |
| Der Tagesspiegel | 717 | Welt kompakt | 222 |
| Frankfurter Neue Presse (Regionalausgaben) | 684 | B.Z. | 198 |
| abendblatt.de - Hamburger Abendblatt Online | 622 | Die ZEIT (inklusive ZEIT Magazin) | 158 |
| Kölner Stadt-Anzeiger | 606 | Wiesbadener Tagblatt | 156 |
| Nürnberger Nachrichten | 582 | Wiesbadener Kurier | 156 |
| Mitteldeutsche Zeitung | 533 | Main-Taunus-Kurier | 146 |
| Stuttgarter Zeitung | 483 | Aar-Bote | 146 |
| WELT ONLINE | 455 | Idsteiner Zeitung | 144 |
| Aachener Nachrichten | 434 | Main-Spitze | 144 |
| Berliner Morgenpost Online | 434 | Wormser Zeitung | 143 |
| Aachener Zeitung | 426 | Neuss Grevenbroicher Zeitung | 137 |
| Stuttgarter Nachrichten | 425 | Giessener Anzeiger | 132 |
| Nürnberger Zeitung | 422 | Bergische Morgenpost | 125 |
| Kölnische Rundschau | 417 | Der Spiegel | 121 |
| General-Anzeiger (Bonn) | 415 | Lampertheimer Zeitung | 121 |
| Die Welt | 389 | Burstädter Zeitung | 121 |
| Südwest Presse | 374 | Solinger Morgenpost | 119 |

Pegida related reports - most distinguishing words

The following table includes ten most important terms for correctly assigning a news report to the corresponding time interval with a support vector machine.

| rank | 2014-12/2015-02 | 2015-03/2015-05 | 2015-06/2015-08 | 2015-09/2015-11 | 2015-12/2016-02 | 2016-03/2016-05 |
|------|-----------------|-----------------|--------------------------|-----------------|-----------------|-----------------|
| 1 | neujahrsempfang | trogritz | verfassungsschutzbericht | galg | silvesternacht | afd |
| 2 | Pegida | geert | flüchtlingsheim | jahrestag | flüchtlingskris | katholikentag |
| 3 | islamkrit | wuppertal | heidenau | flüchtlingskris | silv | kinderschokolad |
| 4 | charli | wild | flüchtlingsunterkunft | 1938 | russlanddeutsch | clausnitz |
| 5 | islamisier | islamkrit | freital | transitzon | clausnitz | jena |
| 6 | kathrin | gey | austritt | rek | connewitz | böhmermann |
| 7 | demonstration | befreiung | ramadan | einjahr | europaweit | flüchtlingskris |
| 8 | kogida | luck | alfa | gift | 59 | hof |
| 9 | abendland | blockupy | jag | asylchaos | obergrenz | geldstraf |
| 10 | ukrain | henkel | zeltstadt | schaff | warschau | hattk |

Columns show the top 10 most distinguishing stemmed terms for news articles in corresponding time periods.

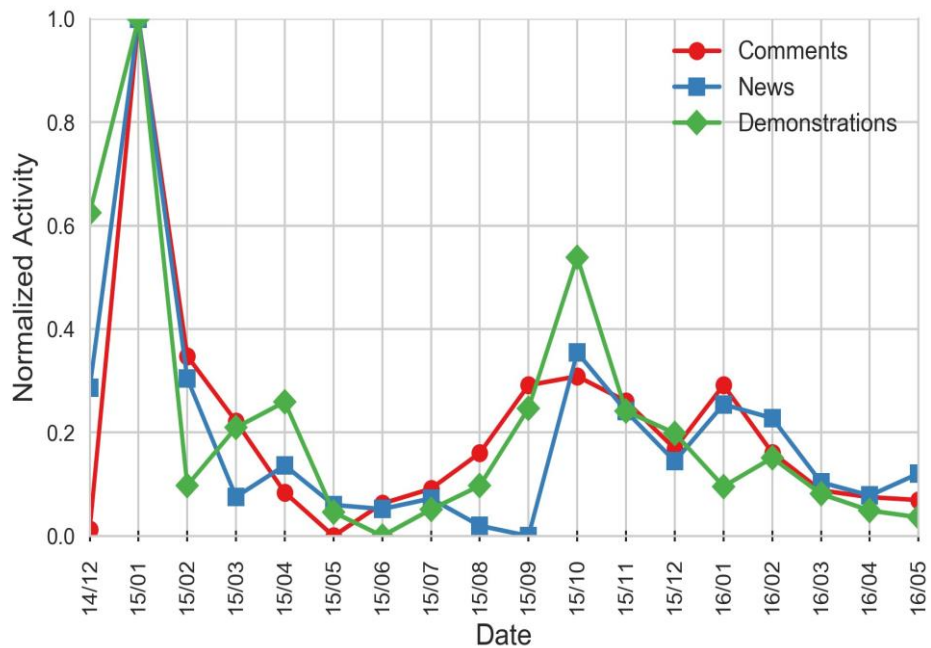
B. Pegida's awareness of social media functions

The following text shows the German original version of the translated post in the results section which illustrates that Pegida knows about information diffusion functions of links and hashtags.

“Danke Kathrin! Du hast unsere Standpunkte sehr gut vertreten und Dich super gegen einen ständig ins Wort fallenden CDU-Spahn behauptet welcher vor Aggression und Arroganz nur so strotzte. Beim nächsten Mal gemeinsam mit Rene oder Lutz! Das war nur die erste Runde aber die ging klar an Dich! #DresdenZeigtWiesGeht PS: An den dümmlichen Kommentaren auf irgendeiner Watch-Seite - der wir jetzt nicht durch links oder hashtags Reichweite verschaffen wollen - sieht man deutlich wie sie innerlich vor Wut über Kathrins souveränen Auftritt kochen. Man hatte dort vorher ein großes Desaster prophezeit. Tja abermals der Beweis das Gutmenschen einfach von nichts Ahnung haben.”

C. Pegida demonstrations in Germany

The following graph includes normalized time series for Facebook Comments, news articles and the number of on-site protesters of Pegida. Data for aggregated counts of protesters is not available on a daily basis, but instead aggregated per month and was retrieved from the website durchgezahlt.org.



D. Topic models

Frex Terms

The following table shows the top 20 FREX terms for all topics not analyzed in the main article.

| Topic | FREX terms |
|--|---|
| 1: Politics criticism | merkel, cdu, angela, #fastsonderschulersigmar, kanzlerin, #mischpokec, #bundesinnenminis, #claudiafatimaroth, #bundesgauckl, bundeskanzlerin, seehof, #gohringeckardt, maas, #imerika, #volksverrat, www.i-finger.d, spd, fluchtlingsspolit, #ausdenpalastenjag, merkel |
| 2: FPO support | #fpo, #aufdiestrasseuberall, www.wiedenroth-karikatur.d, #tatjana4dresd, strach, #kannstedirnichtausdenk, #widerstand, osterreich, fpo, hc, wahl, lauft, wild, #dresden4tatjana, geert, freital, erreicht, #aufdenpunkt, hof, wien |
| 3: Police operations | schwed, polizist, polizei, tat, strafat, word, schwedisch, beamt, verletzt, einzelfall, kam, flughaf, mehr, unterkunft, anwoh, thuring, mess, berlin, stadt, wohnung |
| 4: Anti-refugee content | #asylbetrugerraus, #schengenaussetz, #glucksritterzuruckverschiff, arzt, eigent, wohl, haus, grun, gar, gerade, ingenieur, ide, leut, gern, bunt, wirklich, gluck, tut, rein, lach |
| 5: Refugee crisis in other countries | turkei, ungarn, griechenland, #verabschiedungskultur, griechisch, migration, ungar, pass, syri, abschieb, russisch, migrant, viktor, russland, milliard, orban, slowakei, fluchtling, idomeni, #remigration |
| 7: English posts / radio fees | of, to, we, for, are, volksbegehren-sachsen.d, bord, frist, volksantrag, #ausgezahlt, and, amtsblatt, jorg, rundfunkstaatsvertrag, abgeb, on, german, zustimm, sachsische, ausfullen |
| 8: Right-wing organizations in other countries | niederland, frankreich, html, www.focus.d, franzos, prozent, bla, wenig, europa, blick, bevolker, putin, erstmal, allerding, eben, sitz, #absurdistan, front, national, liegt |
| 10: Counter-demos / page updates | photo, updated, Pegida, friedlich, bitt, seit, cov, demo, fref, beweg, link, lutz, dah, aktion, demonstration, their, post, zitat, spaziergang, lug |

Highly representative topic documents

This appendix contains one Pegida post in original German spelling for each of the ten topics analyzed in the main article. Examples are chosen to be highly representative for the corresponding topic, which is determined by MAP estimates of topic proportions from the structural topic model.

Topic 1 - Politics criticism

“Unglaublich... statt Schuldenabbau Rentenerhöhungen eine nachhaltige Familienpolitik oder Steuer-oder Lohnnebenkostensenkungen zur Entlastung mittelständischer Unternehmen geht alles in die Vollversorgung illegaler Einwanderer die durch einen nie dagewesenen Rechtsbruch der #Merkeldiktatur in s Land geschleust werden!

#IMerika und Ihre #Volksverräterbände bekämpfen also Probleme die ohne sie nie dagewesen wären mit Geld welches das Volk fleißig erwirtschaftet hat #MerktEuchDieNamen

#MerkelMussWeg #FastSonderschülerSigmar

#ClaudiaFatimaRoth #MischpokeCem #GohringEckardt #Bundesgauckler #BundesinnenMisere

#AusDenPalästen-Jagen #Volksverräter #JudgementDay

http://m.focus.de/magazin/kurzfassungen/focus-09-2016-milliardenuberschuss-des-bundessoll-vollstaendig-in-fluechtlingskrise-fliesen-finanz-staatssekretaer-spahn-wir-wollen-die-schwarze-null-halten_id_5318636.html”

Topic 2 - FPÖ support

“#PEGIDA #FPÖ #Sensationell BRAVO“ osterreich die ersten Hochrechnung sind da #Hofer fuhr mit großem Abstand! Auf in die Stichwahl da geht noch was!“ #ÖsterreichZeigtWiesGeht”

Topic 3 - Police operations

“#PEGIDA #DiskoTrauma #Absurdistan #Bereicherung Tja wenn man als #Flüchtiger #traumatisiert feiern geht kann man schon mal #kulturell bedingt aufgrund der Trauer ob der zurückgelassenen Frauen und Kinder überreagieren ein sogenannter Einzelfall“TM..... (...)Massenschlägerei in Diskothek In der Nacht von Donnerstag zu Freitag hat es in einer Diskothek eine Schlägerei zwischen deutschen und syrischen Männern gegeben. Nach Informationen der Volksstimme handelte es sich bei den Deutschen um eine Gruppe von neun Polizeibeamten die in der Diskothek gefeiert hatten. Die Polizisten waren mit den Syrern in Streit geraten. Der Hintergrund der Auseinandersetzung ist noch unklar. Auf Nachfrage bestätigte eine Polizeisprecherin der Volksstimme dass es sich um Polizisten handelte. Die Syrer haben nach dem Streit das Lokal verlassen sind aber nach Zeugenaussagen mit Tischbeinen und Flaschen bewaffnet wiedergekommen und attackierten die Polizisten. Bei der Schlägerei wurden mehrere Personen verletzt.(...) <http://www.volksstimme.de/lokal/magdeburg/20151211/kriminalitaet-massenschlaegerei-in-diskothek>”

Topic 4 - Anti-refugee-content

“Das Gute an der ganzen gigantischen lächerlichen - Mainstream Medienoffensive zur Schmachhaftmachung angeblich traumatisierter Refutschiies ist dass immer mehr Leute die grottenschlechten Inszenierungen von ZDF und Co durchschauen. Da wird immernoch die arme Flüchtlingsfamilie mit Kind gezeigt obwohl jeder halbwegs denkende und mit gesundem Augenlicht gesegnete Bürger mittlerweile selbst die unendlichen Afrikaner-Horden - bestehend aus kräftigen jungen Männern von 15-35 Jahren (natürlich alle aus Syrien zumindest mit syrischem Pass) ausgestattet mit modernster Technik und bestens gekleidet - in der eigenen Stadt gesehen hat und weiß wie sie sich verhalten. Da werden Unmengen an Geld verpulvert damit B/C/D/E-Prominente sich für die gescheiterte Asylpolitik von #IMerika und #FastSonderschülerSigmar aussprechen und

Betroffenheit vorgaukeln. Aus dem Artikel: überflüssigster Teil der prominenten Selbstbeweihräucherung war dabei wohl die Live-Schalte zu Til Schweiger nach Moskau der fleißig über die Erfolge der Til Schweiger Foundation berichten durfte – als ob die im Studio anwesenden Prominenten nicht ausreichend waren und die Medien nicht schon für genug Aufmerksamkeit für Schweiger gesorgt hätten.“ #MerktEuchDieNamen #AsylbetrugerRaus #GlücksritterZurückverschiffen #SchengenAussetzen #GrenzenDicht #PEGIDA #TilSchweiger #TilUndSigmar #TilDo #AusGEZahlt #Lügenpresse #LügenZDF #Kerner <http://m.welt.de/vermisches/article146250293/ZDF-Fluechtlingsgala-wird-zum-totalen-Reinfall.html>”

Topic 5 - Refugee crisis in other countries

“#PEGIDA #OrbanViktor #Fidesz #GrenzenDicht #Ungarn #Hungary (...) Ungarn Orban will keine Flüchtlinge mehr durchs Land lassen Ungarn will jetzt auch an der Grenze zu Rumänien einen Zaun errichten. Es sei bereits alles vorbereitet sagte Ministerpräsident Orban in einem Rundfunk-Interview. Grundsätzlich sollten gar keine Flüchtlinge mehr durch sein Land kommen. Ungarn hatte bereits im Herbst die Übergänge zu Serbien und Kroatien abgeschottet.(...) Viktor Orban weiß was auf Europa zukommt wenn der Frühling einkehrt und das Mittelmeer ruhiger wird. (y) Und er handelt für sein Land und sein Volk Bravo Herr Orban Köszönöm szépen!

<http://www.deutschlandfunk.de/ungarn-orban-will-keine-fluechtlinge-mehrdurchs-land-lassen.1947.de.html?drn:news id=572509>”

Topic 6 - Islamization and assaults by foreigners

“#PEGIDA #InformiertEuch #SchautHin Text von Sabatina James: Das Attentat auf Christen in Pakistan und die Terrorattacken von Brüssel und Paris sind keine Ausnahmen mehr. Sie sind Teil einer grausamen Kette von Massenmorden die immer länger und blutiger wird. Der westlichen christlichen Welt ist einseitig Krieg erklärt worden vor Jahren schon. Und mit jedem Jahr nimmt dieser Krieg an Intensität zu. Europa will es nicht wahrhaben dass Samuel Huntingtons exakt vor 20 Jahren veröffentlichtes Buch vom Kampf der Kulturen und seinen Bruchlinienkonflikten“ grausame Realität geworden ist. Und doch wirkt die Verdrängung des islamischen Großangriffs zusehends naiv. Denn die Schlachtfelder dieses Krieges sind blutiger als es unsere Abendnachrichten erahnen lassen. Alle Ränder der islamischen Welt sind blutig geworden. Von Indonesien und den Philippinen ganz im Osten bis zur Elfenbeinküste ganz im Westen wo vor wenigen Tagen zwei Dutzend Tote bei Angriffen auf westliche Hotels gemeldet wurden.”

Topic 7 - English posts and protest against radio fees

“PEGIDA – 10 demands to the German asylum politics 1.) We call for an immediate stop for asylum seekers and we call for a German asylum-emergency law - now! Our asylum laws were conceived after the war for manageable quantities of approximately 2 000 refugees per year and not for 1 5

millions we expected to reach already in 2015! 2.) We call for strict border controls! We demand to suspend the Schengen Agreement IMMEDIATELY - for all the borders of Germany! Other EU countries control their national borders - and that although the completely failed Dublin procedure goes almost entirely at the expense of Germany. The temporary reintroduction of border controls during the G7 summit has proved that border controls are an appropriate mean to prevent illegal border crossings the flourishing business of smuggling mafia and the entry of criminals. 3.) We demand that the group of safe countries of origin will be expanded on ALL Council of Europe member countries! This European Council has 47 member countries with 830 million citizens and over 1 800 European officials. All Member States have committed themselves to the preservation of democracy and rule of law as well as the recognition of the fundamental and human rights. That should be enough to count these countries to safe countries! 4.) We call for a TEMPORARY right of asylum for refugees of war! Of course real war refugees and accepted asylum seekers is to grant temporary protection and full coverage in the modest scale. But once the situation in the country improves the refugees have to leave our country again. 5.) We call for a binding limit for the annual reception of asylum seekers namely defined by ourselves the host country Germany! This vital question about the future of our country must be carried out by means of direct democracy through a referendum! 6.) We finally demand honesty in the integration debate and the end of the red-green social-romantic tale of wanting to integrate masses of male African asylum seekers here! No one wants that. The green socialists use the refugees to create a red-green job wonder for bachelor graduates of chatter Sciences here. The pathological altruism and feigned empathy gooders are moral invisibility cloaks which should cover the mega-lucrative migrant market. 7) We demand that immediately all rejected asylum seekers and hundreds of thousands of illegal immigrants to be banished at once! Again: We call MASS deportations - and do it now! 8) We demand that the refugee problem has to be resolved in locally in their own cultures! Our so-called representatives of the people should finally show backbone and take Saudi Arabia Qatar and the United Arab Emirates in charge. These wealthy huge Sharia-paradises are much better suited to accommodate the crowds of Muslim asylum seekers as an Europe of unbelievers! And we finally need asylum procedure-spot audits in the countries of origin. Even in North Africa has to be decided by fast-track procedure on applications for asylum in Germany! 9.) We demand that foreign criminals which are connected with Islamic terrorist organizations are banished immediately! This naturally also includes the adopted sons and daughters of German Minister of Internal Affairs de Maiziere all these jihad returnees and all known and violent Salafists - these people are to be deported outside Europe immediately! 10.) There will be expected resistance from Brussels about any changes in our German asylum policies – so then we all have to leave this bullying dump EU! The future French President Marine Le Pen has summarized it in the destruction of these EU - quote. It's only this radical way which works! These EU will never be to reform - who should himself rationalize his highly-paid job? Asylum seekers driven by nothing than economical reasons - are NOT welcome! Christian refugees specially those who are suppressed by slaughtering Islamists are absolutely welcome in Germany and we provide every shelter food and life-support they need because this belongs to the German helping nature. To all others: STAY OUT! We the people of European nations need to unite to conserve and to defend our values our culture our freedom. We need to unite against the self-declared kings and queens in Brussels. We the German people need international support against our own politicians in our German parliaments. Our

politicians want to change the form of government of the Federal Republic of Germany they want to abolish the German state people in Germany to replace us by a multicultural society they want to establish a multiethnic state on German soil - this is a behavior like high traitors! #PEGIDA"

Topic 8 - Right-wing organizations in other countries

"(...)Europas Rechte schließen sich zusammen Le Pen Vilimsky und Wilders formieren sich zu einer EU-Fraktion. Was das finanziell und rechtlich bringt.(...) Na endlich für die FREIHEIT!! (y) Deutschland wird folgen! (...) Lange wurde daran gebastelt nun haben sich die Rechtspopulisten Europas im EU-Parlament tatsächlich zu einer Fraktion zusammengeschlossen. Heute wurde sie in Brüssel von Harald Vilimsky (FPÄ) Marine Le Pen und Geert Wilders präsentiert. Es ist historisch verkündete Wilders. Bereits gestern wurde auf Twitter die Gründung der Fraktion angekündigt. Der Name ist Europa der Nationen und der Freiheit . Im vergangenen Jahr war der Plan noch gescheitert weil es Marine Le Pen nicht gelang Parlamentarier aus genügend EU-Ländern zu gewinnen. Zur Bildung einer Fraktion im Europaparlament sind 25 Abgeordnete nötig die in mindestens sieben Mitgliedstaaten gewählt sind. Die Front National war bei der Europawahl in Frankreich stärkste Partei geworden und stellt derzeit 23 Abgeordnete. Neben Frankreich Österreich und den Niederlanden sind Parteien aus Italien Großbritannien Belgien und Polen dabei.(...) #PEGIDA #FürDieFreiheit
<http://kurier.at/politik/eu/rechte-fraktion-im-eu-parlament -europas-rechte-schliessen-sich-zusammen/136.376.617>"

Topic 9 - Demonstrations

"+++ Montag ist es wieder soweit. +++ GESICHT ZEIGEN! 07.03.2016 - 19 UHR - Richard-Wagner-Platz Leipzig. Treffpunkt zur sicheren Anreise wie immer: Ab 18:30 Hauptbahnhof Leipzig vor McDonalds. Gemeinsam SICHER zum Richard-Wagner-Platz! Redner am Montag: LUTZ BACHMANN SIEGFRIED DAEBRITZ FRIEDRICH FRÖBEL einige weitere... Teilt die Veranstaltung! Bringt eure Freunde / Nachbarn / Arbeitskollegen mit! Es geht um unser Land / unsere Zukunft und die Zukunft unserer Kinder! #LEGIDA #PEGIDA
#AufDieStraße
<https://www.facebook.com/events/1080481985319052/D10>"

Topic 10 - Counter-demonstrations and page updates

"Nochmal der Aufruf! Am Sonnabend den 14.3.2015 wird PEGIDA die friedliche GEGENDEMO gegen Pierre Vogel und seine Salafisten sein. (Y) Also jeder der kann ab noch Wuppertal! Den Link zur Veranstaltung findet ihr im ersten Kommentar. like"-Emoticon " TEILEN und EINLADEN! :-) Es werden noch ORDNER gesucht bitte melden und keine Angst das ist keine Zauberei. (Y) Hier ein Link zu dieser Gruppe:
<https://www.facebook.com/groups/377245535788683/?fref=ts> Schreibt dem Administrator Chris Ko:
<https://www.facebook.com/christian.konig.330>
Hier eine Möglichkeit Fahrgemeinschaften zu bilden! (Y)
<https://www.facebook.com/groups/1804474533110935/>"

E. Predicting user engagement for Pegida posts

The following table shows nested negative-binomial regression models for the number of user comments each Pegida received.

| | Only Topics | | Topics and Date | | Full Model | |
|------------------------|--------------|-----------------|-----------------|------|-------------|------|
| Variables | Coefficients | Standard Errors | | | | |
| Topic 1 | -0.134 | 0.08 | -0.015 | 0.07 | 0.013 | 0.07 |
| Topic 2 | -0.550*** | 0.08 | -0.346*** | 0.07 | -0.368*** | 0.07 |
| Topic 3 | -0.177 | 0.07 | -0.174** | 0.07 | -0.162* | 0.07 |
| Topic 4 | 0.104 | 0.07 | -0.035 | 0.06 | 0.015 | 0.06 |
| Topic 5 | -0.378*** | 0.06 | -0.164** | 0.06 | -0.149** | 0.06 |
| Topic 6 | Ref. | - | Ref. | - | Ref. | - |
| Topic 7 | -0.135 | 0.17 | -0.137 | 0.16 | -0.073 | 0.15 |
| Topic 8 | 0.150 | 0.08 | -0.172* | 0.07 | -0.144* | 0.07 |
| Topic 9 | -0.324*** | 0.06 | -0.510*** | 0.05 | -0.478*** | 0.05 |
| Topic 10 | 0.860*** | 0.07 | 0.012 | 0.07 | -0.062 | 0.07 |
| Date | | | -0.004*** | 0.00 | -0.006*** | 0.00 |
| Date \times Date | | | | | 0.000004*** | 0.00 |
| Topic 1 \times Date | | | | | 0.00164** | 0.00 |
| Topic 2 \times Date | | | | | 0.00130* | 0.00 |
| Topic 3 \times Date | | | | | 0.00181*** | 0.00 |
| Topic 4 \times Date | | | | | 0.00196*** | 0.00 |
| Topic 5 \times Date | | | | | 0.000139 | 0.00 |
| Topic 6 \times Date | | | | | Ref. | - |
| Topic 7 \times Date | | | | | -0.00388* | 0.00 |
| Topic 8 \times Date | | | | | -0.000288 | 0.00 |
| Topic 9 \times Date | | | | | 0.00108** | 0.00 |
| Topic 10 \times Date | | | | | -0.00144*** | 0.00 |
| Constant | 5.847*** | 0.05 | 7.051*** | 0.06 | 7.255*** | 0.07 |
| Observations | 3738 | | 3738 | | 3738 | |
| Nagelkerke R^2 | 0.159 | | 0.366 | | 0.388 | |
| BIC | 19948,65 | | 18901,83 | | 18862,76 | |