Learning to be Unbiased: Evidence from the French Asylum Office*

Mathilde Emeriau†

July 17, 2019

Abstract

What determines whether some asylum seekers are granted refugee status while others are rejected? I draw upon archival records from a representative sample of 4,143 asylum applications filed in France between 1976 and 2016 to provide new evidence on the determinants of asylum decisions. Comparing accepted applicants with those who were rejected, I find that Muslim applicants are 30 percent less likely to be granted asylum than otherwise similar Christian applicants. In addition, linking archival records to detailed administrative data, I show that bureaucrats at the French asylum office initially discriminate against Muslims but stop after about a year on the job. Assessing potential mechanisms of discrimination, I do not find support for the claim that discrimination is driven by bureaucrats' preferences or ideology. Instead, I argue that discrimination occurs because bureaucrats underestimate the probability that Muslims are persecuted. This novel finding has direct implications for strategies to curtail discrimination in courtrooms and administrations.

Word count: 11,910

*The author would like to thank Jens Hainmueller, Dominik Hangartner and David Laitin for invaluable guidance in designing and implementing this research. I am also grateful to Pascal Brice, Aline Angoustures, all members of OFPRA's scientific committee and the French National Archives who granted me access to these archives. Adelaide Choisnet, Marie Gonzalez, Duncan Lawrence, Jean-Paul Levi and Corinne Pallastrelli provided the technical support that made the data collection possible. Catherine Anderson, Astrig Atamian, Guillaume Beaud, Diane Belkheraz, Charles Bena, Lucie Benaiteau, Cécile Bénoliel, Sonia Bilon, Lisa Cavalieri, Emmanuel Charreau, Thomas Emeriau, Elyne Etienne, Jeanne Huybrechts, Anna Kauffmann, Shushan Kerovpyan, Maxime Krumholz, Alice Ledoux, Marion Lotin, Mathilda Loussert, Charles Maillard, Camille Urvoy, and Alice Zivohlava all provided excellent research assistance. This research is based upon work supported by the National Science Foundation Grant No. 1624048 (Principal Investigator: David Laitin), the Stanford Immigration Policy Lab, The Europe Center, and the France Stanford Center for Interdisciplinary Research. Kirk Bansak, Jennifer Cryer, Jane Esberg, Vicky Fouka, Justin Grimmer, Jeremy Weinstein, and Yang-Yang Zhou provided valuable comments.

†Department of Political Science and Immigration Policy Lab, Stanford University, Stanford, CA 94305-6044, USA. memeriau@stanford.edu.

1 Introduction

Over three million refugees have arrived in Europe since 2015, triggering an unprecedented political crisis in the European Union. Anti-immigrant parties are gaining ground across Europe, campaigning on exclusionary policies. Several European countries reintroduced border controls, leading to the de-facto suspension of the open border Schengen area, and in 2016 the United Kingdom voted to exit the European Union. At the root of the crisis is the conflict among countries over the distribution and processing of the massive influx of asylum seekers. All European member states are bound by the 1951 Geneva Convention on Refugees, but they continue to handle asylum applications very differently, as is reflected by the disparity in acceptance rates across Europe – in 2015 they varied from 12 percent in Latvia to 90 percent in Bulgaria.[1]

At the center of the asylum process, decision-makers in asylum courts and offices face the difficult task of determining whether asylum seekers provide truthful and substantiated claims of persecution for reason of "race, religion, nationality, membership of a particular group or political opinion."[2] For asylum seekers this decision is highly consequential: those granted refugee status will be able to stay in Europe while others will be sent back. However, the subjectivity of this definition and the possibility that applicants are falsifying claims give significant discretion to judges and bureaucrats. In addition, the political salience of the issue creates incentives for governments to influence asylum decisions to accommodate domestic and international interests. How then do asylum officers decide? Is asylum granted on the basis of legal merit, or is there discrimination in the attribution of refugee status?

The lack of micro-level data has so far limited researchers' ability to analyze the determinants of asylum decisions. Using country-level data, a number of studies find that both humanitarian and strategic interests explain variation in acceptance rates by country of origin (Rottman et al. 2009; Rosenblum and Salehyan 2004; Keith et al. 2013; Neumayer 2005; Holzer et al. 2000). However, these analyses are unable to account for the possibility that variation in the composition of asylum seekers from the same country of origin could confound these results. In light of this challenge, scholars have turned to countries like the U.S., where cases are randomly assigned to judges,[3] to compare acceptance rates between immigration judges to detect discrimination (Ramji-Nogales et al. 2007; Hausman 2016). However, without micro-level data, they are able neither to identify the types of asylum seekers that are discriminated against nor to infer the most likely mechanism of discrimination.

[1] Eurostat. "Distribution of first instance decisions on (non-EU) asylum applications in 2015."

[2] Article 1(A)2 of the 1951 Geneva Convention Relating to the Status of Refugees.

[3] Hausman (2016) shows that even though the assignment is in fact not random, it is most likely arbitrary with respect to the merits of cases.

In this study, I overcome previous data limitations by taking advantage of an unprecedented effort in transparency from the French asylum office,[4] which opened its archives ten years ago. In total, I digitized 4,143 asylum applications, a representative sample of all applications filed at the French asylum office between 1976 and 2016.[5] I collected all information recorded in the application form (e.g. age, family situation, education, employment, religion) and transcribed the personal narrative in which applicants explain the reasons they need political asylum. Crucially, for each application, I also observe the decision and the anonymized identifier of the bureaucrat who decided the case.

For the first time, these data allow me to rigorously test for deviation from a purely merit-based attribution of refugee status. The outcome of interest is whether an applicant was granted refugee status upon first examination at the French asylum office. I compare accepted and rejected applicants to analyze how individual characteristics affect the probability of getting asylum in France, holding constant the country of origin, the year of arrival and the merit of the case. Doing so, I leverage the fact that I had access to the entire case file, which was all the bureaucrats themselves saw in order to decide whether to grant refugee status, to eliminate concerns over omitted variable bias. To control for differences in the merit of asylum applications, I measure the extent to which the narrative both meets the criteria outlined in the Geneva Convention definition and sounds authentic, using a set of hand-coded narratives and supervised machine learning. The predicted measure correlates as expected with features of the text, like the length of the narrative and its level of detail, as well as with the decision.

This research design improves on previous studies of discrimination. It innovates compared to similar studies (Pope and Sydnor 2011; Laouénan and Rathelot 2017; Hainmueller and Hangartner 2013; List 2004; Levitt 2004) by directly measuring and predicting the legal merit of the case. In addition, studies that leverage random assignment of cases to judges in an attempt to identify cases of comparable quality are limited since they only detect discrimination to the extent that decision-makers discriminate differently. Instead, by directly controlling for the merit of the case, this research design is guaranteed to identify discrimination even if all bureaucrats discriminate against the same type of applicant. Moreover, by using real-world data this design also improves on audit studies (White et al. 2015; Butler and Broockman 2011; McClendon 2016; Hemker and Rink 2017) in which the task is

[4]The full appellation and acronym of this administration is the French Office for Refugee Protection and Stateless Persons (OFPRA).

[5]The French National Archives granted me access on October 9, 2015 (derogation JM/2015/1504) to 10,000 applications filed until December 31, 2015 and on August 30, 2017 to 500 applications filed in 2015 and 2016. The conditions of access are stipulated in a research agreement between the Stanford Immigration Policy lab and the French asylum office. As part of this agreement, the names of applicants and their family members were redacted to ensure the anonymity of asylum seekers and refugees.

purposefully time-insensitive and thus sometimes not very consequential.

The study yields two main findings. First, I demonstrate evidence of discrimination against Muslim applicants in the attribution of refugee status in France since 1976. Muslims are 30 percent less likely to be granted asylum compared to otherwise similar Christian applicants. Moreover, no other individual characteristics explain asylum decisions as consistently as religion does. Having started graduate school, being highly skilled and speaking French are all positively correlated with a positive decision, but these estimates are not robust to alternative specifications. I show that the difference between Christians and Muslims cannot be explained by differences in individual characteristics or in the legal merit of their application. I also show that this gap is not driven by differences in unobservable characteristics revealed during the interview process or by bureaucrats' access to private information regarding the level of persecution applicants face. In magnitude, this effect is comparable to the effect of the merit of the application and to a recent estimate of Muslim bias in European citizens' asylum preferences (Bansak et al. 2016).

Second, I show that French bureaucrats stop discriminating after about a year on the job – Muslims are much more likely to be discriminated against when their case is examined by an inexperienced bureaucrat than by an experienced one. This suggests that bureaucrats learn on the job not to discriminate. How do bureaucrats learn, given that they do not receive feedback on whether their past decisions were accurate? One possibility is that, as bureaucrats read more applications over the course of their tenure, they are exposed to different types of applicants and can more easily single out authentic narratives. Consistent with this learning hypothesis, I find that the predictive power of the legal merit of the case increases as bureaucrats become more experienced. This pattern suggests that the reduction in discrimination that I identify is in part driven by the fact that bureaucrats hone over time their ability to identify authentic narratives.

This finding is consistent with a mechanism of discrimination in which French bureaucrats initially discriminate because they underestimate the probability that Muslims are persecuted but eventually learn on the job to correct this bias. Theoretically, this mechanism of discrimination is closely related to statistical discrimination, the main distinction being that bureaucrats' belief that Muslims are less likely to be persecuted is not grounded in statistics but in personal biases. This intuitive mechanism of discrimination remains to date relatively unexplored. So far, Arnold et al. (2017) are the first and only researchers to make a similar claim in the context of bail court decisions by showing that inexperienced judges are more likely to discriminate against black defendants.

This study offers two primary contributions. First, it provides a different picture than the one usually drawn by commentators of the asylum process. Whether by journalists

(Chesnet 2015; Aho Nienne 2013), scholars (Hausman 2016; Keith and Holmes 2009; Keith et al. 2013; Ramji-Nogales et al. 2007), or even judges (Lepors 2015), the asylum decision system is often described as a form of Russian roulette in which similar asylum seekers face very different outcomes depending on where and when they apply. Instead, I show that even though the discretion built into the Geneva Convention does lead to inconsistencies, the attribution of refugee status in France is not as indiscriminate as these commentators suggest, since Muslim applicants are systematically disadvantaged compared to Christians.

Second, while previous work suggests that bureaucratic discrimination occurs when bureaucrats pursue objectives like personal preferences or political ideology instead of doing their job (Broockman 2013; Marten 2017; Neggers 2017; Shayo and Zussman 2011), this study provides suggestive evidence for an alternative mechanism of discrimination, with vastly different implications for strategies to reduce discrimination in administration and courts. As such it has direct implications for reducing discrimination at the French asylum office, as reducing bureaucratic turnover would automatically reduce overall discrimination. In addition, increasing the length of training of bureaucrats might speed up the learning process. Finally this study opens up a new agenda of research, as there is still little rigorous empirical evidence on the effectiveness of existing anti-bias training (See Paluck (2016) for a review).

In the remainder of the paper, I provide background information on the asylum process in France and review existing literature on the mechanisms of discrimination. I then present the research design, before presenting the results and discussion of the mechanisms. The last section concludes.

2 Background

2.1 Asylum process in France

Upon arrival in France, asylum seekers first need to register at the prefecture of the department in which they settle.[6] There, a bureaucrat registers their fingerprints in the Europe-wide fingerprinting database (EURODAC) and assesses through an informal interview whether France is in fact the country responsible for processing his or her asylum application.[7] If it is, the asylum seeker has twenty-one days to submit his or her application to the French asylum office in Paris. The application includes a standardized application

[6] France counts 101 departments, intermediate administrative units smaller than a region and larger than a commune.

[7] According to the Dublin III Regulation, the responsible Member State will be the state through which the asylum seeker first entered the EU.

form with self-reported demographics and socio-economic information and a personal narrative in which applicants describe, in French, their motives for asking for political asylum in France.[8]

Upon submission to the French asylum office in Paris, the application is dispatched to the relevant division and assigned to a bureaucrat.[9] He first reads the application, decides whether to interview the applicant or not (interviews are mandatory since 2006) and proposes a decision to grant or deny refugee status to his or her supervisor, who makes the final decision.[10] This decision is very high stakes for applicants. Those who are granted asylum receive a ten-year renewable residency permit (or a one-year residency permit in the case of subsidiary protection). Those rejected can appeal at the Cour National du droit d'Asile (CNDA),[11] where asylum cases are reexamined by a three-judge panel during a public audience. In magnitude, the acceptance rate at the appeals court is comparable to the acceptance rate at the French asylum office over that period. Asylum seekers whose appeal is rejected can submit a new application to the French asylum office if they have new elements to provide, but at this stage of the process the acceptance rate is extremely low (around 3 percent).

What determines whether a person will receive political asylum in France? France, like 143 other signatories of the 1951 Geneva Convention, is bound to grant asylum to individuals with a well-founded fear of being persecuted for reasons of "race, religion, nationality, membership in a particular group or political opinion." This definition might surprise some and even seem outdated to others, as it does not encompass what we think of as refugees today, i.e. those who flee civil conflicts. To resolve this tension between the Geneva convention and the needs of asylum seekers, France introduced in 1998 another form of protection called territorial asylum, which was replaced in 2003 by the subsidiary protection when the French asylum office became the unique asylum counter in France. Subsidiary protection is granted in France to those that do not meet the Geneva definition but who face death, torture or indiscriminate violence in the context of an internal or international armed conflict. This status is more precarious since it only gives access to a one-year residency permit and is not used as widely in France as it is in other countries until very recently. Acceptances on the basis of subsidiary protection represent between 20 and 30 percent of all first-time acceptances between 2007 and 2015.[12]

[8]This personal narrative has to be submitted in French, and asylum seekers can get help from family members or friends or even from volunteers at a migrants' association like France Terre d'Asile.

[9]Even though the assignment process to bureaucrats is neither random nor based on a clear decision rule, there are good reasons to believe that this is based only on observable characteristics.

[10]In the applications I only observe the final decision.

[11]Formerly known as the Commission des Recours des Réfugiés (CRR).

[12]Source: OFPRA Activity report 2017, page 53.

2.2 Mechanisms of discrimination

The literature identifies two main reasons to expect deviation from a purely merit-based attribution of refugee status. First, governments might find it profitable to intervene and influence asylum decisions in order to accommodate domestic or international interests, whether they pertain to diplomacy, national security or immigration policy. In the U.S., in addition to the humanitarian situation in the country of origin, whether an applicant comes from a country engaged in important trade relationships with the U.S., or from a country that receives U.S military assistance or faces economic sanctions, helps explain the variation in asylum decisions by country of origin (Rottman et al. 2009; Rosenblum and Salehyan 2004; Keith et al. 2013). In Western Europe, the unemployment rate and the success of right-wing parties are good predictors of the recognition rates (Neumayer 2005). In Switzerland, cantons with a centralized asylum system have lower recognition rates (Holzer et al. 2000). However, these country-level analyses are merely suggestive since they do not take into account possible variation in the composition of the asylum seekers within countries of origin, which could confound these results.

In France, several elements suggest that the French asylum office is similarly subject to attempts at political control by the French government. The French asylum office is a public institution with administrative, financial and functional independence,[13] but its director is nominated every three years by the government. In 2007, after the election of Nicolas Sarkozy as president, the French asylum office was placed under the umbrella of the newly created ministry of Immigration and National Identity, while it had been under the jurisdiction of the Ministry of Foreign Affairs since its creation in 1952. Former directors of the French asylum office themselves have provided anecdotal evidence that politicians did attempt to intervene in the decision process, though often these cases are limited to personal connections or high-profile cases.[14] Moreover, Pereira (2017) shows how the decision to grant refugee status to Portuguese in the 1960s was in part a response to France's diplomatic interests. In addition, a quick look at aggregate statistics suggests that asylum seekers who applied under left-wing presidents are more likely to get asylum than those who applied under right-wing presidents, controlling for country of origin and standard demographic characteristics. Together these elements suggest both a clear attempt by the government to influence asylum decisions and anecdotal evidence that this works in some cases.

Second, absent any attempt at political interference, discrimination might simply result

[13] Article 7 of the 2015 Asylum Act (amending Article L. 721-2 of the Code on the Entry and Stay of Foreigners and the Right of Asylum) stipulates that "the Office shall exercise impartiality in his missions (...) and receives, in their fulfillment, no instructions."

[14] This observation is based on the visioning of filmed interviews of former directors recorded by the French asylum office.

from the fact that decision-makers are prejudiced or biased. In addition to the subjectivity of the asylum criteria, the possibility that applicants are falsifying persecution claims, and are in fact economic migrants, contributes to greater uncertainty. As early as 1970, the annual activity report noted that "The pace of arrivals remains high, although the Office is striving to exclude some elements — Yugoslav in particular — who are in reality, economic refugees, in search of better life and employment conditions." Since the 1980s, after the first massive frauds by applicants from former Zaire and South East Asia, bureaucrats at the French asylum office have been advised to use caution in reading narratives. Bureaucrats at the French asylum office are faced with the difficult task of determining whether applicants provide truthful and substantiated claims of persecution. As a result, deciding whether an applicant meets the definition set by Geneva Convention is a nebulous process. The following quote from one of the former directors illustrates nicely the challenge involved in making asylum decisions: "You are gold diggers. There is huge stack of rocks. In this stack there are a few gold nuggets. You have to find them. But there are a lot of rocks."

How then do bureaucrats decide, and when does bureaucratic discretion result in discrimination? A well-established explanation is that bureaucrats discriminate when they use leeway to maximize their utility. As in taste-based discrimination (Becker 1971), bureaucrats have an intrinsic preference for persons of some type they benefit from favoring. Shayo and Zussman (2011), for example, argue that Israeli judges promote their social identity (Shayo 2009) by favoring members of their own religious group in small-claims courts. Neggers (2017) similarly concludes that own-group favoritism is the mechanism through which the ethnic composition of teams of election officers affect voting outcomes in India. Broockman (2013) shows that black politicians are more intrinsically motivated than their white counterparts in advancing their group's interests. A related explanation is that decision-makers may discriminate when they deviate from their objective to advance their political agenda.[15] Marten (2017) for example finds that liberal laymen sitting in asylum courts in Sweden influenced decisions to be more generous and conservatives did the opposite. Both of these explanations assume that bureaucrats are imperfectly monitored and can shirk at low cost instead of doing their jobs.

A second set of explanations does not assume that bureaucrats deviate from their objective. Statistical discrimination stipulates that bureaucrats discriminate when they rely on accurate beliefs to infer the most likely outcome under imperfect information (Phelps 1972; Arrow 1973; Aigner and Cain 1977). In the context of asylum decisions, this would be the

[15]The terminology for this is subject to discussion as White et al. (2015) refer to the mechanism in which bureaucrats discriminate to maximize political return as statistical discrimination. Distelhorst and Hou (2014) talk about strategic discrimination instead.

case if bureaucrats were more likely to accept Christian applicants because they knew for a fact that they are more likely than Muslims to tell the truth, for example. However, statistical discrimination, originally developed on the labor market, applies only to situations in which it is possible to gather data on the distribution of the unobservable characteristics (e.g. productivity, guilt) by group. This is a strong assumption that does not hold in the context of asylum decisions since bureaucrats do not have access to such data. An alternative explanation, which remains vastly under-examined despite being very intuitive, is the possibility that bureaucrats discriminate when they have *inaccurate* beliefs (Bordalo et al. 2016; Fryer and Jackson 2008). Arnold et al. (2017) recently showed suggestive empirical evidence that discrimination results from incorrect stereotypes in bail courts.

As Phelps (1972, p. 661) himself noted, "discrimination is no less damaging to its victim for being statistical." However, the relevant distinction for reducing discrimination is not whether bureaucrats have preferences for or biases against certain groups, as both could be referred to as prejudice, but rather, whether bureaucrats are playing by the rules or whether they are furthering their agenda. Indeed if bureaucrats discriminate even though they are trying to do the right thing, then changing their incentives will not reduce discrimination. To date, and despite the different implications of these two theories for developing interventions to reduce discrimination, the lack of data has limited researchers' ability to properly test for mechanisms of discrimination in bureaucracies. In this paper, I leverage a wealth of data from the French asylum office that allows me to assess these mechanisms with more precision.

3 Research design

3.1 Data

I digitized 4,143 asylum applications filed at the French asylum office between 1976 and 2016. These applications of both accepted and rejected applicants were randomly selected from a representative sample of 100,000 asylum applications filed at the French asylum office. The study starts in 1976, even though the French asylum office opened its doors in 1952. After an initial peak, the number of first-time applicants slowly trickled down to a couple thousand until France ratified the Bellagio protocol in 1976 – an amendment that lifted the geographic and time limitations included in the definition of the Geneva Convention and opened the asylum process to non-European nationalities for events that happened also after 1951 (Figure C.1). Moreover between 1952 and 1976, the acceptance rate at the French asylum office is close to 100 percent, and there are no Muslim applicants. I further restrict

the sample to decisions that were taken solely based on the merit of the application, thus excluding cases of family unification and special procedures that South East Asian applicants benefited. In all specifications, I use weights to adjust the sample to the population margins (See Appendix A.4 for more information about the weights).

For each application, I coded the complete form including the following topics: Condition of arrival in France (itinerary, former countries of residence, modality of entry on French territory, whether legal or illegal), origin of the applicant (country of origin, nationality, ethnicity added in the 1990s), family situation (date and place of wedding, divorce, information about family members), military service (date, role and country), skills (education and qualification, last job in country of origin), languages (mother tongues and other languages), and self-reported mentions of documents (passport, diplomatic laissez-passer).[16] To ensure the quality of the data collection, I checked each coded application for inconsistencies and missing values and resolved discrepancies manually on a daily basis. The error rate among 210 randomly selected folders for double data entry was less than 5 percent.

The main independent variables used in the analysis fall under two broad categories, and all are listed in Table 1. The first category includes all demographic characteristics: gender, age and family situation at the time of the application. These characteristics are used as controls in all specifications and do not constitute the primary focus of this study. The second set of variables includes proxies for the applicant's potential for integration. These were chosen in order to detect the most likely deviation from a merit-based attribution of refugee status. Since most of the questions were open-ended, I recoded the responses into categorical variables. When the applicant did not answer a question, the variable is coded as *Missing*. I recoded the religion of the applicant into the following categories: *Christian*, *Muslim*, *Other*, *None* or *Missing*. Close to 80 percent of applicants are either Christian or Muslim, but there is significant variation over time: Muslims represent 60 percent of applicants in 2016, up from 5 percent in 1976 (Figure C.2). I also recoded the highest level of education into whether the applicant had started any of the following education levels: *primary education*, *secondary education* and *graduate education*. Applicants who specified that they had not started any education were coded as *No education* (6 percent) and those who did not answer this question were coded as *Missing* (20 percent). Half of all applicants had started secondary education and 14 percent had started university. Using information on applicants' current occupation and last job in their country of origin, I recoded their skill level into *High*, *Middle* and *Low/None*. Highly skilled jobs include academics, doctors,

[16]The application form has changed six times since 1952, but the questions asked remain stable. Few questions did not remain and among them: "Are you planning on staying permanently in France?" and "Are you registered in the consulate of their country of origin?"

Table 1: Summary statistics on dependent and independent variables

	N	Mean	Std. Dev.	Min	Max
Got refugee status	4,142	0.15	0.36	0	1
Narrative					
Provided a narrative	4,142	0.93	0.25	0	1
Arrival in France					
>1 year in France	4,142	0.15	0.36	0	1
Gender					
Female	4,142	0.31	0.46	0	1
Age					
Less than 20	4,142	0.09	0.29	0	1
Between 20 and 30	4,142	0.50	0.50	0	1
Between 30 and 40	4,142	0.28	0.45	0	1
Between 40 and 50	4,142	0.09	0.28	0	1
Between 50 and 60	4,142	0.03	0.17	0	1
More than 60	4,142	0.02	0.13	0	1
Family Situation					
Single	4,141	0.54	0.50	0	1
Married/Partnership	4,141	0.42	0.49	0	1
Divorce	4,141	0.03	0.16	0	1
Widowed	4,141	0.02	0.14	0	1
Religion					
Christian	4,142	0.38	0.48	0	1
Muslim	4,142	0.40	0.49	0	1
Other	4,142	0.12	0.33	0	1
None/Missing	4,142	0.11	0.31	0	1
Education					
None	4,142	0.06	0.24	0	1
Primary	4,142	0.11	0.32	0	1
Secondary	4,142	0.49	0.50	0	1
Graduate	4,142	0.14	0.35	0	1
Missing	4,142	0.19	0.39	0	1
Occupation					
High	4,142	0.15	0.35	0	1
Middle	4,142	0.34	0.47	0	1
Low/None	4,142	0.31	0.46	0	1
Missing	4,142	0.20	0.40	0	1
Speaks French					
No	4,142	0.68	0.47	0	1
Yes	4,142	0.27	0.44	0	1
Missing	4,142	0.05	0.22	0	1

Note: This table presents weighted summary statistics on the main variables used in the analysis. All variables except the first two are self-reported.

engineers, high level executives, lawyers, journalists and students. 15 percent of applicants fall into that category. Blue collar workers, civil servants, mechanics, farmers, guards, and other such employees are coded as *Middle* and represent 34 percent of applications in the sample; and drivers, hair dressers, sale clerks, none, homemaker as *Low/None* (31 percent). 20 percent did not respond. Using self-reported year of arrival, I also calculate the number of years spent in France prior to submitting the application. On average 15 percent of applicants in the sample had been in France for more than one year by the time they applied. Even though asylum seekers are required to apply within 21 days of registration at the prefecture, it is possible for applicants to enter France legally and ask for asylum after their visa or residency permit expires. An applicant was coded as French-speaking if he listed French as a native or spoken language. Under this coding rule, 27 percent of applicants speak French. To ensure that results are not dependent on the decision to include these variables, I show in the next section that results are robust to the inclusion of additional variables including the number of children, whether the application was expedited or not, whether the applicant declared a passport or a diplomatic laissez-passer, whether the applicant has a refugee family member in France, whether the applicant completed military service, and whether his entry on the territory was legal or illegal (See Table C.1 for summary statistics on these additional variables).

For each application, I also transcribed the personal narrative, a majority of which were handwritten. Of the 4,143 applications included in the final sample, 93 percent submitted a narrative in French. For each, I extract features that relate substantively to the quality of the text (See Appendix Table B.5 for summary statistics). The number of words (including stop words) and the number of dates and location mentioned (extracted using Stanford CoreNLP) are proxies for the level of detail of the narrative. Narratives vary greatly in length, but on average they are 777 words long and mention 7 dates and 6 locations. To get at the extent to which the narrative is personal and individualized, I count the number of first person pronouns.[17] On average, narratives use first person pronouns 34 times. In order to control for the originality of the narrative, I also compute for each narrative its average euclidean distance to other narratives of asylum seekers from the same country of origin in the corpus. Finally, I estimate a structural topic model (STM) with 20 topics using country of origin as a covariate (Roberts et al. 2014).

For confidentiality reasons, I was not authorized to collect any individual-level information about the bureaucrats working at the French asylum office. However, I am able to infer basic information using the administrative database of 261,357 asylum applications filed be-

[17]Specifically, I count the number of times any of the following terms appear in the narrative "je," "j'ai," "me," "mon," "mes," "moi."

tween 2001 and 2015 (about half of all applications filed in that period).[18] In this database, 688 unique bureaucrats starting working at the French asylum office between January 1, 2001 and December 31, 2015. For each of them, I can infer their start date using the date of their first decision and then count the number of decisions taken and the order in which they were taken. On average, bureaucrats decide 389 cases over the course of their tenure, work a little over 3 years (though 41 percent are still working in 2016), and handle cases from an average of 12 different countries. Summary statistics for the administrative data on the bureaucrats are displayed in Table 2.

Table 2: Summary Statistics on the bureaucrats (Administrative data)

	N	Mean	Std. Dev.	Min	Max
# of years worked	688	3.39	3.96	0	16
# number of decisions	688	389.16	477.41	1	3445
# of countries	688	11.53	8.36	1	44
Start Date					
Started in 2001	688	0.08	0.27	0	1
Started in 2002	688	0.04	0.19	0	1
Started in 2003	688	0.10	0.30	0	1
Started in 2004	688	0.04	0.20	0	1
Started in 2005	688	0.01	0.09	0	1
Started in 2006	688	0.02	0.13	0	1
Started in 2007	688	0.04	0.19	0	1
Started in 2008	688	0.06	0.24	0	1
Started in 2009	688	0.06	0.24	0	1
Started in 2010	688	0.04	0.19	0	1
Started in 2011	688	0.10	0.29	0	1
Started in 2012	688	0.06	0.23	0	1
Started in 2013	688	0.08	0.26	0	1
Started in 2014	688	0.05	0.21	0	1
Started in 2015	688	0.12	0.33	0	1
Still working	688	0.41	0.49	0	1

Note: This table presents summary statistics on bureaucrats' employment at the French asylum office.

[18] Before 2000, the identity of the bureaucrats was not systematically recorded. I also drop cases without a decision, and bureaucrats that took at least one decision in 2000 such that I do not know whether they started in 2000 or before, since the data is left censored.

3.2 Estimating discrimination

Identifying discrimination using observational data remains an empirical challenge. Women, for example, tend to get refugee status more often than men (Angoustures et al. 1995; Keith and Holmes 2009), but this does not necessarily imply that women are favored over men in the attribution of refugee status. To conclude to group-based discrimination, one also needs to ensure that there are no unobserved characteristics that differ systematically across gender that could explain the decision as well (Aigner and Cain 1977). In an attempt to test for discrimination, studies have focused on countries, like the U.S., Sweden or Switzerland, where cases are randomly assigned to judges in order to detect unusual inter-adjudicator variation (Marten 2017; Price and Wolfers 2010; Shayo and Zussman 2011). These studies are limited, however, since they only detect discrimination to the extent that decision-makers discriminate differently. For example, if all judges discriminate against Muslims in a similar fashion, Muslims would have the same expected acceptance rate across judges, and the study would incorrectly conclude to an absence of discrimination. Alternatively, scholars have used audit surveys to compare response rates across comparable groups of applicants (Butler and Broockman 2011; McClendon 2016; White et al. 2015), but as it is often the case in audit studies, the task is time-insensitive and thus sometimes less relevant.

The unique, individual level dataset I collected allows me to compare successful applicants to unsuccessful ones to see if they differ systematically on characteristics that are not related to the merit of the case. This methodology tests for discrimination by measuring the difference in expected outcomes between two groups that is unaccounted for by observable characteristics. This methodology has been used in a variety of settings, such as a peer-to-peer lending market (Pope and Sydnor 2011), an online vacation rentals platform (Laouénan and Rathelot 2017), naturalization decisions (Hainmueller and Hangartner 2013), games fairs (List 2004) and TV shows (Levitt 2004). One necessary condition for the estimation is overlap, which is often difficult with highly collinear characteristics like religion and country of origin. I check for overlap in the data by plotting countries of origin in the sample as a function of the share of applicants from that country who are Christian on the x-axis, and the proportion who are Muslims on the y-axis (Figure C.5). While a number of countries cluster on the bottom left and the upper right corner of the graph, a non-negligible proportion is spread out on the diagonal, suggesting substantial overlap.

I use ordinary least squares to estimate the gap. In practice, I regress an indicator variable for whether the applicant was granted asylum on the religion of the applicant, a measure of the quality of the application, which I describe below, and a set of individual characteristics (education, profession, knowledge of French, family situation, age and gender), controlling for the country of origin and year of application. The coefficient on the indicator variable for

whether the applicant reported to be Muslim is an estimate of the difference in percentage point in the chance of getting refugee status between Muslims and Christians.

To conclude that a negative coefficient is evidence of discrimination against Muslims, I rely on the selection on observables assumption. This assumption is well supported because I have access to the same information than the bureaucrats making the decisions. As a result, I can fully account for the extent to which Christian and Muslim applicants differ. In other words, I assume that there are no unobserved applicant characteristics that are both correlated with the religion and the decision. This works because the application of the Geneva Convention is based on the examination of individual claims to persecution rather than on group belonging. This practice, referred to as "individual screening," is introduced by the International Refugee Organization (IRO) at the end of the 1940s, and specifies that asylum seekers have to demonstrate that they are individually persecuted (Cohen 2000). This practice rules out the possibility that the observed Muslim penalty is driven by the fact that bureaucrats know Muslims to be less persecuted than Christians. In contrast, the same assumption does not hold for the attribution of the subsidiary protection since it is based, among other things, on the security situation in the region of origin of the applicant. In this case, the origin of the applicant carries additional information as to whether he or she needs protection. This, however, is not a concern for the analysis since, in France, asylum applications are examined sequentially, such that only the ones that did not qualify for protection under the Geneva Convention are examined for subsidiary protection. I can thus recode applicants who received refugee status under subsidiary protection as having been rejected for the Geneva Convention protection. Since France does not award often subsidiary protection, results are not sensitive to this coding rule.

3.3 Measuring the merit of the case

The efficacy of this research design hinges on the possibility to rule out that differences in the probability of being granted refugee status between Christians and Muslims is driven by differences in applicants' narratives.[19]

I use a combination of hand-coding and supervised learning to measure the quality (or credibility or merits) of the personal narrative, that is the extent to which the narrative, first, meets the criteria outlined in the definition of refugee in the Geneva Convention and, second, sounds authentic and convincing. The authenticity of a text is not easily captured by

[19]Specifically, it is essential for the validity of the research design to control for differences in the narratives between Christians and Muslims that also explain the decision. In Appendix Table B.1, I show that the narratives provided by Muslims do differ from those provided by Christians, but mostly on the content as the differences in topic proportions suggest, rather than strength of the argument.

a single dimension, as several features contribute to it. The level of detail provided and the extent to which the narrative is personal and original are two examples of several dimensions that may contribute to it (Greslier 2007). A common strategy to measure the effect of a text is to identify the relevant features of the text, hand-code these for a subset of the corpus and use supervised learning to predict them for the rest of the corpus. Three coders read eight percent of the narratives in the sample and,[20] for each narrative, coded whether the narrative mentioned any form of persecution, rated the text along different dimensions (e.g. detailed, personal, original), and finally coded (on a scale from 0 to 3) whether the applicant made a reasonable claim to political asylum as defined by the Geneva Convention.[21] This last score, converted to a binary indicator, is what I refer to as the merit of the application, which I use as a control in all specifications. On average, they answered yes 50 percent of the time (Table B.2), a proportion that is significantly higher than the average acceptance rate over the period (15 percent).

Three elements raise confidence that this variable measures the merits of the case. First, the coders were selected from the pool of research assistants who helped with the digitization and who had Master's degrees in law and political theory. This selection process ensured the confidentiality of the narratives, a first-order parameter in this study. It also ensured that coders were knowledgeable about the Geneva convention. By the time they read and coded these narratives, they had each digitized several hundred applications, an important requirement for reliable hand-coding (Kripendorff 1980). Second, among the 59 narratives that were triple-coded, the agreement rate between these three coders is about three times higher (they agree 61 percent of the time) than the agreement one would anticipate from coders deciding randomly (using simulation I find that random coders agree 25 percent of the time). The Krippendorff α is .44 (See Table B.4).[22] Finally, several elements point to strong internal validity of this measure. Questions about the content and objective features of the text, like the level of detail and the cohesiveness of the text, which were designed to prime coders to think fully about the different aspects of the text before rating narratives as credible or not, correlate strongly with the credibility measure (Table B.3). In addition, the principal component analysis of all of these features reveals a single component with eigenvalue above one (eigenvalue of 4.39). Moreover, when predicting a credibility score

[20] In total 450 ratings and 350 unique narratives were coded. Though relatively small a total of 341 hand coded documents is relatively satisfying as compared to the benchmark provided by Hopkins and King (2010) (between 100 and 500).

[21] The question wording is *"In your opinion, is this person entitled to claim the right to asylum according to the Geneva Convention"* {*No/Rather no/Rather yes/Yes*}.

[22] Krippendorff α is computed as (1 - (number of disagreements observed)/(number of disagreements observed by chance). This ratio is used to determined inter-coder reliability, where 0 means perfect disagreement and 1 perfect agreement.

using this principal component analysis, the prediction correlates strongly with the measure of credibility I obtain from hand-coding (.79).

Using this set of hand-coded narratives, I compare the performance of three classification algorithms to predict the merits of the application for the rest of the corpus using weighted word frequencies and additional text features. To convert the corpus of narratives into a document-term matrix, I first remove capitalization, punctuation, word order and stop words. I then stem words and drop unigrams that are too uncommon (those that occur in less than 1 percent of the narratives) or too common (those that appear in more than 99 percent of the narratives) (Hopkins and King 2010). For each of the 2,842 remaining unigrams, I compute term frequency weighted by the inverse document frequency. Using the training set, I tune and evaluate the performance of algorithms using the following procedure: I first randomly partition the labeled set into the training set (75 percent) and the test set (25 percent). I tune the parameters using 10-fold cross-validation of three algorithms (Random Forests, Gradient Boosted Trees and the Lasso) and select the model that performed best on the test set using a combination of four different metrics, the area under the ROC curve, the percent correctly predicted, the Brier score and the calibration of predicted probabilities. See the Appendix B for more information on the tuning of parameters and performance metrics. All three algorithms perform relatively well on standard metrics (Table B.7), but predicted probabilities from Random Forest are better calibrated (Figure B.1). Using Random Forest, I am able to accurately predict the quality measure 76 percent of the time in the left-out sample of 81 narratives, a substantial reduction in error compared to a baseline of 48 percent.

In addition to a reasonable inter-coder agreement, a reasonable accuracy performance and a very good calibration, the predicted probability of being a credible narrative correlates as expected with features of the text like length, the number of dates and location (Table B.8). Moreover, as I will show in the next section, applicants whose narrative was predicted to be credible were 6.2 percentage points more likely to get refugee status upon first examination, a 41 percent increase compared to the baseline acceptance rate.

4 Results

4.1 Discrimination against Muslim asylum seekers

The outcome of interest is whether an applicant was granted refugee status upon first examination at the French asylum office. On average, between 1976 and 2016, 15 percent of applicants in the sample, i.e. after excluding cases of family unification and resettled refugees, got political asylum in France as the basis of the Geneva Convention, but the acceptance

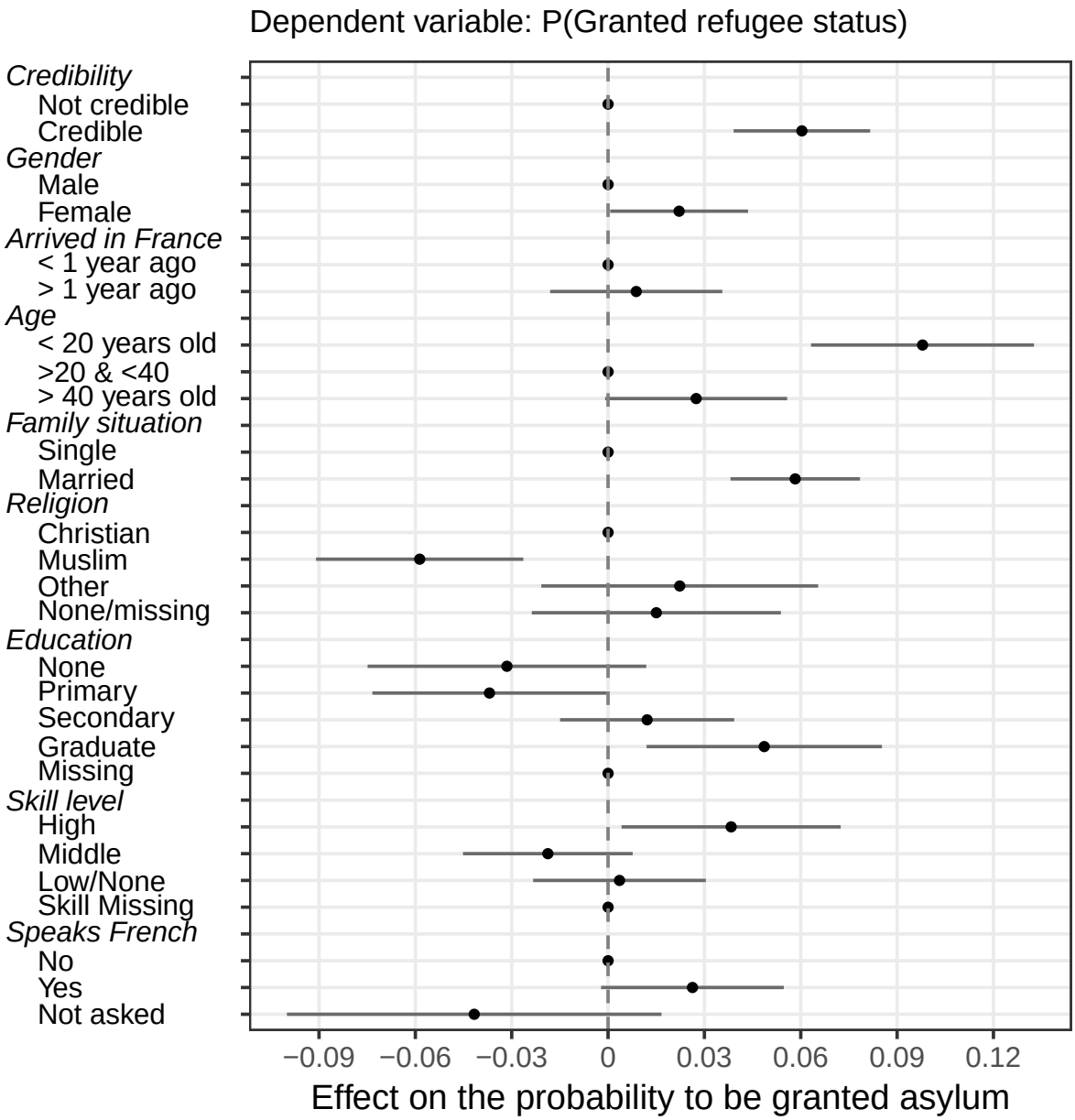
rate varies greatly over the period.[23] From close to 95 percent in 1976, the acceptance rate was down to 20 percent twenty years later, reaching its lowest point in 2004 (8 percent) (See Appendix Figure C.3). It has increased slightly since the beginning of the refugee crisis, and 25 percent of first-time applicants who applied in 2016 received political asylum. This variation over time is consistent with a gradual trend toward stricter immigration policies observed across Europe but does not in itself constitute evidence that equal applicants are treated unequally. Similarly, the acceptance rate also varies greatly by country. 75 percent of asylum seekers from Iraq received political asylum on the basis of the Geneva Convention in 2015, while only 4 percent of applicants from Kosovo did that same year. This variation, however, could reflect differences in the situation of countries of origin, which are difficult to account for empirically.

Instead, to analyze discrimination in the attribution of refugee status, I turn to regression results. Figure 1 shows the marginal effect along 95 percent confidence intervals of the main specification, and regression results are presented in Table 3. The effect of the religion of the applicant on the probability of getting asylum is large. Muslim applicants are 5.6 percentage points less likely to get refugee status than Christians who are similar on all other characteristics. This represents a substantial difference (30 percent) given that the average acceptance rate is 15 percent over the period. Moreover, this effect compares in magnitude to the effect of the merit of the application on the probability of getting asylum (6 percentage points). In Figure C.8, I show that this result is robust to two alternative strategies for controlling for the effect of the personal narrative. First, I control for the text features that *should* be relevant in assessing asylum applications (They are listed in Table B.5). I developed this list through informal discussions with bureaucrats working at the French asylum office. Second, I use the supervised Indian Buffet process developed by Fong and Grimmer (2016) to discover the features of the text that *are* relevant in explaining the decision in a training set and estimate the causal effect of the religion controlling for these features in the rest of the sample. Regardless of how I control for the effect of the personal narrative, I find that Muslims are much less likely to get refugee status compared to Christians who are otherwise similar.

To ensure that this result is not driven by differences between Christians and Muslims between countries and not within countries, I estimate the Muslim gap within country by controlling for interaction effects between an indicator for whether the applicant is Muslim and each of the country of origin fixed effects. For each country, I add the estimated Muslim penalty for the omitted country and the coefficient on the interaction to recover an estimate

[23]Note that this is an underestimate of the overall first-time acceptance rate because cases of asylum granted for subsidiary protection are coded as rejection.

Figure 1: Determinants in the attribution of refugee status.

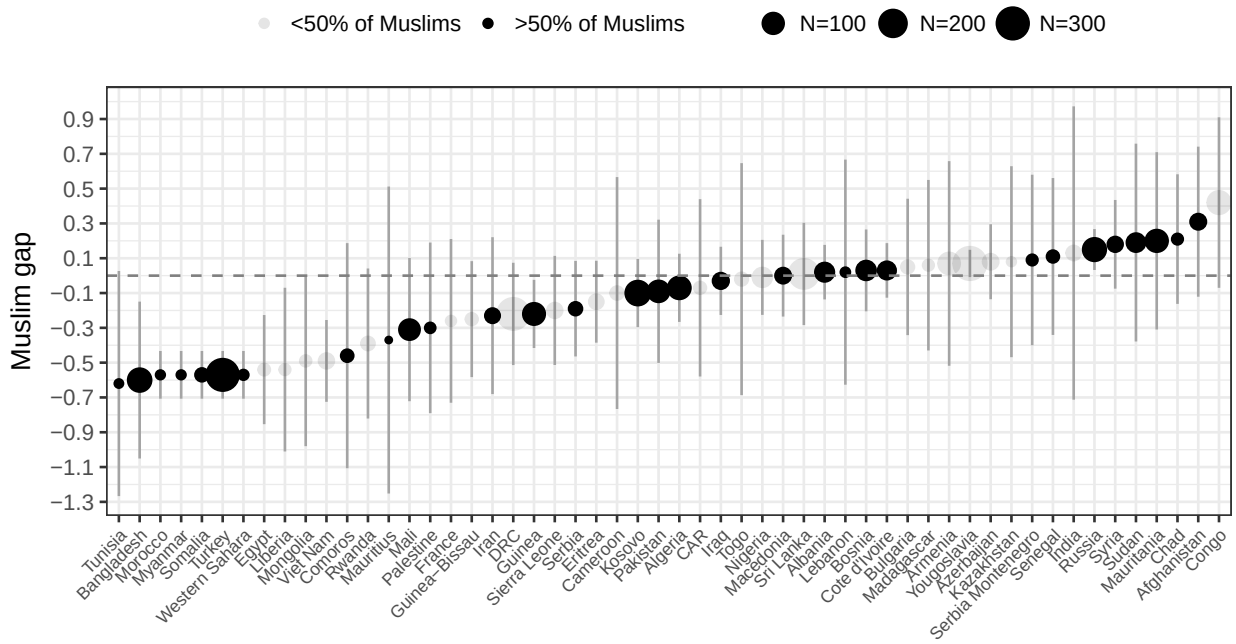


Note: This figure displays the results of a weighted OLS regression. The dependent variable is a dummy variable indicating whether the applicant received status upon first examination at the French asylum office. In addition to the variables reported in the figure, the regression includes country of origin and year of application fixed effect.

of the Muslim penalty within country. These estimates are plotted in Figure 2, along with the 95 percent confidence intervals. This figure reveals that even though there is variation in the intensity of the Muslim penalty between countries, Muslims incur a penalty in the

majority of the countries in the sample. The pattern is the same when considering only countries in the sample with at least 4 Christians and 4 Muslims (Figure C.9). In addition, I find that the estimated penalty for Muslim applicants is robust to the inclusion of interaction effects between the year of application and the country of origin (Table 3, column 2) and to the addition of additional applicants' covariates (Table 3, column 3).

Figure 2: Muslim gap by country of origin



Note: This figure displays estimates the difference in acceptance rate between Christians and Muslims within country along with 95 percent confidence intervals from a weighted OLS regression including fixed effects for year of application, country of origin, and interaction between country of origin and the religion of the applicant. The size of the dots is proportional to the size of the population from this particular country in the sample, and the shade of gray indicates whether this is a majority Muslim (Dark) or Christian (Light) in the sample.

Is this effect driven by differences between applicants that became apparent during the interview process? I show that results are robust to restricting the analysis to the sample of applicants who were not called in for an interview, suggesting that the interview is not an important confounder (Table 3, column 4). The estimated Muslim penalty is almost twice as large in this subsample, as restricting the sample in this way shifts the composition of applicants toward the beginning of the period. For this reason, I also measure the Muslim gap

among those who were not interviewed since 1986 only (Table 3, column 5). Alternatively, could this difference in the probability of getting asylum between Christians and Muslims merely result from a process through which asylum applications submitted by Muslims are systematically assigned to stricter bureaucrats, who grant asylum on more restrictive conditions? I reject this hypothesis by showing that the Muslim gap holds after controlling for bureaucrat fixed effects. This analysis is complicated by the fact that the information about bureaucrats is incomplete and unreliable before 2000 both in the administrative database and the paper applications I digitized. For the purpose of this analysis, I binned all applications for which the identifier of the bureaucrats was missing into one additional "missing" category (Table 3, column 6).

To further alleviate concerns that Muslims are less likely to be granted asylum because they are less likely to be persecuted, and not, as I argue, because they are discriminated against, I perform two additional tests. I first show that results hold when I control for whether the applicant belongs to a minority religious group in his home country (Table 3, column 7). To do so, I use data from the World Religion dataset from the Association of Religion Data Archives, which provides estimates on the percentage of the population that identifies with Christianity or Islam since 1945 for most countries in the world. Using these estimates, I generate a binary variable which indicates, for each applicant in the sample, whether he or she belongs to a religious group that comprises less than 20 percent of the population in his home country.[24] Second, I show that results also hold in the subset of applicants who did not claim persecution on religious grounds. The rationale for this test is as follows: If it is the case that Christian applicants are more likely to get asylum because they are more likely to be persecuted, then the Muslim gap should not sustain in the subsample of applicants who did not claim persecution on religious grounds. Identifying applicants that claim persecution for religious reasons is not straightforward. About 1,021 of the narratives (22 percent) mention at least one religious keyword from a relatively short list, whereas only 6 percent of narratives claim persecution for religious reasons according to the coding realized by research assistants (Table B.2). To identify narratives that claim persecution for religious reasons (one of the five motives of the Geneva Convention), I read 292 narratives, a random sample of the 1,021 narratives that contained at least one religious keyword, and for each coded whether the applicant claims persecution on religious grounds. Restricting the sample to narratives that either did not contain a religious keyword or contained one but was not coded as persecuted due to religion, I find that the Muslim penalty holds (Table 3, column 8).

[24] 165 observations did not merge due to unequal overlap in coverage by country. For example, Armenia, Bosnia and Croatia appear in the World Religion data only after 1995.

Table 3: Robustness Table

	1(Granted refugee Status)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Full Sample	Country-year interaction	Additional covariates	Not Interviewed	Not Interviewed since 1986	Bureaucrats fixed effects	Controls for minority	No mention of religious persecution
Credible narrative	0.060***	0.055***	0.030*	0.049**	0.028*	0.065***	0.057***	0.040***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Religion (Ref: Christian)								
Muslim	-0.059***	-0.065***	-0.047**	-0.083***	-0.049*	-0.059**	-0.058***	-0.045*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Other	0.022	0.032	0.016	-0.039	-0.007	0.011	0.057*	0.021
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)
None/Missing	0.015	0.010	0.010	-0.023	-0.019	0.012	0.036	0.040
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Education (Ref: Missing)								
None	-0.031	-0.032	-0.024	-0.052	-0.015	-0.043	-0.024	-0.029
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Primary	-0.037*	-0.059**	-0.045*	-0.061*	-0.043	-0.048*	-0.040*	-0.025
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Secondary	0.012	0.012	-0.000	-0.025	-0.018	0.013	0.016	0.028
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
Graduate	0.049**	0.040*	0.028	0.039	0.024	0.048*	0.053**	0.063**
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Skill level (Ref: Missing)								
High	0.038*	0.020	0.040*	0.020	-0.004	0.030	0.042*	0.027
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Middle	-0.019	-0.023	-0.009	-0.001	0.008	-0.017	-0.019	-0.025
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
Low/None	0.004	0.006	0.007	0.018	0.027	0.010	0.006	-0.003
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Speaks French (Ref: No)								
Yes	0.026	0.007	0.018	0.016	0.005	0.015	0.020	0.018
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Missing	-0.042	-0.032	-0.036	-0.087	-0.052	-0.053	-0.041	-0.019
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
$>$ 1 year in France	0.009	0.020	0.017	0.020	0.017	0.004	0.005	0.006
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Observations	3,930	3,930	3,754	1,481	1,337	3,930	3,773	3,103
R^2	0.377	0.633	0.407	0.652	0.354	0.499	0.374	0.413

Note: This table displays the results of eight weighted OLS regressions. The dependent variable is a dummy variable indicating whether the applicant received refugee status upon first examination at the French asylum office. In addition to the variables reported in the table, all regressions include demographic characteristics, country of origin and year of application fixed effects.

In contrast, no other individual characteristic, whether higher education or skill level, is as consistently associated with a higher probability of being granted asylum. In the main specification (Table 3, column 1), applicants who reported having started primary (graduate) school are 3.6 (5) percentage points less (more) likely to be granted asylum than those who did not report their education level. The sign of these effects is consistent but not significant in all specifications. Similarly, highly skilled applicants were 3.9 percentage points more likely to receive refugee status, but as before, even if the sign of the coefficient is consistent,

the coefficient is not significant in all specifications. Speaking French has a small and positive effect (2.9 percentage points), but the coefficient is significant only in the first specification. Having spent more than one year in France prior to application has systematically no effect on the decision.

Overall, these results suggest that religion is the most robust characteristic that influences the decision both in terms of magnitude and robustness to alternative specifications. This is consistent with recent survey evidence that citizens from 15 European countries systematically prefer applicants who are more educated and more skilled (Bansak et al. 2016). This result complements a literature that demonstrates discrimination against Muslims in the labor market and beyond (Adida et al. 2016).

4.2 Bureaucrats stop discriminating after a year on the job

I then show that bureaucrats stop discriminating against Muslims after about a year on the job. For this analysis, I computed the order in which each of the 2,132 applications filed after 2001 was taken. Using this variable, I analyze changes in the predictive power of the applicant's religion on the decision over time *within* the tenure of a bureaucrat. In practice, I compare discrimination patterns between two groups of applications: those decided by inexperienced bureaucrats (who had decided fewer than 200 decisions, which corresponds to approximately a year of tenure) and decisions taken by experienced bureaucrats (who had already decided at least 200 decisions). In Table 4, I first compare the Muslim gap in these two groups of decisions using all applications submitted in or after 2001 (Column 1). I find that Muslims are 7.6 percentage points less likely to get refugee status than Christians when their case is decided by an inexperienced bureaucrat. But while switching from an inexperienced to an experienced bureaucrat does not improve chances of getting asylum for Christians (coefficient: -0.013, standard error: 0.02), it improves Muslims' chances by 6.1 percentage points (standard error: 0.03), such that the difference between Christians and Muslims becomes small and insignificant.

To ensure that this reduction in discrimination is not driven by the changing composition of bureaucrats over time, I show that this result holds when I restrict the sample to the first 500 decisions of all bureaucrats (Column 2) who had decided at least 500 decisions (Column 3). Adding bureaucrat fixed effects (Column 4) does not change the result. In all specifications the Muslim penalty among inexperienced bureaucrats is negative and sizable, even though the coefficient does not reach conventional levels of significance in the last two specifications because of the small sample size. But being examined by an experienced bureaucrat always significantly improves Muslims' prospects compared to Christians. Could

this result reflect the fact that bureaucrats are assigned different types of decisions over the course of their tenure? In Appendix Figure C.7, I show that these two groups of applications do not differ systematically.

This pattern suggests that bureaucrats learn over time not to discriminate. How do bureaucrats learn even though there is no way to know whether an applicant they approved had falsified the claim, or whether an applicant they rejected was in fact telling the truth and ought to have been approved? One possibility is that bureaucrats at the French asylum office learn on the job how to distinguish authentic narratives from fake ones. The intuition is simple: it takes time to be able to single out truly authentic narratives, that is, the ones that distinguish themselves through original storytelling and language. Bureaucrats working at the appeals court confirm that they learn how to infer a sharp signal over time. "Being experienced means to be able to recognize when an applicant is lying or when he is a 'fake refugee"'; testifies one bureaucrat working at the appeals court. As a result, he continues, "At first you do not know what to look for, what to base [decisions] on" (Greslier 2007). As they read applications, bureaucrats discover the universe of narratives, which allows them to identify cookie-cutter narratives. In a recent anonymous testimony, a bureaucrat working at the French asylum office describes a feeling of deja vu: "When you see for the 80th time the same story written with the same font, the same line spacing, in which just a few details change... Sometimes, there is even the name of another asylum seeker in the story at one place." (Aubry and Le Loët 2019).

Consistent with the hypothesis that the reduction in discrimination over time is the result of bureaucrats learning how to infer a sharper signal from the personal narratives, I show, in Figure 3, that the explanatory power of the credibility is small and indistinguishable from zero in the first 200 decisions, but that the coefficient of the case merit is positive and significant afterward, with the exception of decisions between 400 and 500, which is the result of the case merit being a noisy measure. These results provide suggestive evidence for a mechanism of discrimination in which bureaucrats discriminate against Muslims because they underestimate the probability that they are persecuted. This mechanism is closely related to statistical discrimination, but here instead the beliefs are *inaccurate* since they are not based on any underlying data. This intuitive mechanism of discrimination remains to date relatively unexplored. Arnold et al. (2017) are the first to provide suggestive evidence that more experienced judges are less likely to discriminate against blacks in bail courts than unexperienced judges and to conclude that discrimination results from incorrect stereotypes.

Table 4: Learning effect: Bureaucrats stop discriminating after 200 decisions

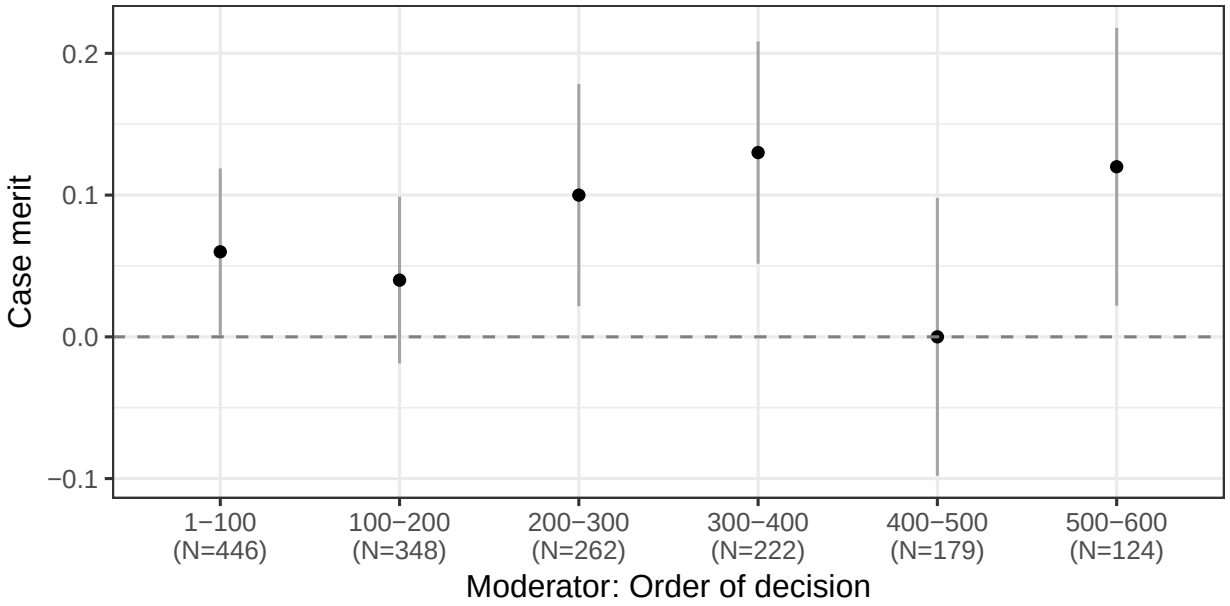
	1(Granted refugee Status)			
	(1) All decisions	(2) First 500 decisions of each bureaucrats ...	(3) ... who decided at least 500 decisions	(4) ... who decided at least 500 decisions
Muslim	-0.078*	-0.093*	-0.088	-0.091
	(0.03)	(0.04)	(0.05)	(0.07)
Other	0.039	0.048	0.027	0.005
	(0.04)	(0.04)	(0.06)	(0.08)
After 200	-0.014	-0.026	-0.009	-0.038
	(0.02)	(0.02)	(0.03)	(0.05)
Muslim × After 200	0.062*	0.072*	0.124**	0.149**
	(0.03)	(0.04)	(0.04)	(0.06)
Other × After 200	0.010	0.042	0.033	0.045
	(0.04)	(0.04)	(0.05)	(0.07)
Bureaucrats Fixed Effects	No	No	No	Yes
R^2	0.237	0.256	0.314	0.486
Observations	2087	1403	762	762

Note: This table shows point estimates and standard errors clustered by bureaucrats in parenthesis from OLS regressions with individual covariates and fixed effects for year of application and country of origin. *Other* is an indicator variable for whether the applicant identifies with a religion other than Islam or Christianity, or with no religion. Model 1 is based on the full sample of decisions decided by bureaucrats who started working at the French asylum office after January 1, 2001. For model 2, the sample is restricted to the first 500 decisions of each bureaucrat. Model 3 further restricts the sample used for model 2 to bureaucrats who decided at least 400 decisions. Model 4 uses the same sample as model 3 but includes bureaucrats fixed effects.

5 Discussion

In this section, I assess the explanatory power of two alternative mechanisms of discrimination: taste-based discrimination and political control. Previous work points to prejudices as the cause of discrimination, showing that adjudicators are more lenient toward members of their own race or ethnic group. In this study, this hypothesis is hard to test directly without any information regarding the bureaucrats themselves. Moreover, the fact that the decision-making process is centralized in Paris does not allow me to leverage regional variation in order to infer bureaucrats' preferences (Hainmueller and Hangartner 2013). However, one pattern already established in the previous section invalidates the hypothesis that discrim-

Figure 3: Learning effect: Testing for the increasing predictive power of the credibility measure



Note: This figure shows the estimated conditional marginal effect of the credibility of the narrative, along with the 95 percent confidence interval based on standard errors clustered by bureaucrats, from the binning estimator (Hainmueller et al. 2016). The moderator is the order in which the decision was taken by each bureaucrat. The evaluation point is the average decision order within in each bin. The specification includes covariates, and fixed effects for year of application and country of origin.

ination results from an attempt by bureaucrats to select applicants who are most likely to integrate in France (instead of picking those who are persecuted). Indeed, if this was the mechanism at play, we would expect that other dimensions associated with better integration prospects would also increase the chances of getting asylum, such as speaking French or having been in France for more than a year. The fact that this is not the case is weak evidence for this mechanism.

In addition, I argue that the assumption, central to taste-based discrimination, that bureaucrats can easily shirk instead of doing their job does not seem to hold at the French asylum office. Instead, I argue that bureaucrats are incentivized to grant refugee status to applicants who provided the most convincing narratives. Surprisingly, working at the French asylum office has historically been very precarious. From the beginning, the asylum office has relied mostly on temporary workers to deal with frequent fluctuations in the number of applications (Figure C.1). In fact, there was not a single bureaucrat on a permanent contract at the French asylum office until 1993, when the unions eventually succeeded in

negotiating that temporary contracts be converted into permanent positions. The share of temporary workers dropped radically afterward, but in the early 2000s the French asylum offices again hired temporary workers (Figure C.6) to deal with a sudden increase in the number of first-time asylum applications. The turnover rate among bureaucrats is also relatively high throughout the period. Using administrative data, I estimate that two-thirds of the bureaucrats hired between 2005 and 2010 stayed only two years. Recently, this proportion has dropped significantly as the result of a strategy from the administration to stabilize the workforce.

To keep their jobs, bureaucrats need to perform well in terms of aggregate statistics. First, they have an interest in granting refugee status at a rate that is comparable to the average acceptance rate so that they do not stand out as too strict or too lenient. Second, they should also aim to limit the number of their decisions that are overturned at the appeals court (note that only rejected cases are reexamined at the appeals court). The 1985 activity report illustrates that the percent of overturned applications is a sign of quality: "This quality [of the decisions] has, so far, been preserved or even improved, as evidenced by the fact that the percentage of cancellations pronounced by the [appeals court], which remained around 12 percent at the beginning of the 1980s, has fallen to 7.8 percent in 1984 and 5.7 percent in 1985." Consistent with these two predictions, I find, using administrative data, that the probability of a bureaucrat remaining employed in a given year is negatively correlated with the acceptance rate and the proportion of overturned decisions in the previous year (Table 5). I argue that the combination of these two objectives (low acceptance rate and small proportion of overturned decisions) create a strong incentive for bureaucrats to grant asylum to those who have a credible narrative rather than to those they prefer.[25]

[25]This reasoning assumes that bureaucrats anticipate that judges at the appeals court are not biased against Muslims, otherwise bureaucrats could discriminate against Muslims without seeing their decisions being overturned more often. This assumption is not testable as such but I believe that bureaucrats are unlikely to hold such belief. Indeed, because of the large number of judges working at the appeals court, bureaucrats do not know which judge will be in charge of a particular application.

Table 5: Determinant of employment continuation

	Employed in year t			
	(1)	(2)	(3)	(4)
Acceptance rate (t-1)	-0.334***		-0.374***	-0.293***
	(0.08)		(0.08)	(0.09)
Overturned rate (t-1)		-0.188	-0.302*	-0.302*
		(0.12)	(0.12)	(0.12)
Constant	0.515***	0.505***	0.525***	0.518***
	(0.00)	(0.00)	(0.01)	(0.04)
Bureaucrat Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	N	N	N	Y
Tenure Length	N	N	N	Y
R^2	0.436	0.420	0.443	0.490
Observations	2072	2072	2072	2072

Note: This table shows point estimates and standard errors clustered by bureaucrats in parenthesis from OLS regressions with fixed effects for bureaucrats. Model 4 also includes indicator variable for the starting year and controls for tenure length. The unit of observation is the bureaucrat year.

Finally, could discrimination against Muslims instead result from a deliberate attempt by the French government to interfere with asylum decisions? This would be consistent with the observed pattern if bureaucrats became more independent after a year (if they change from temporary to permanent contracts) and became less responsive to supervisors' instructions. Does the French government issue direct instructions to favor Christians over Muslims, a practice that would place France in direct violation of international agreements? I argue that no such instructions were formulated, at least formally in the administrative archives.[26] I did find anecdotal evidence suggesting a double standard for Christians and Muslims for two countries, Iraq and Turkey. To illustrate, in 1979, the French Minister of Foreign Affairs noted in a letter to the director of the French asylum office that "[Christians in Turkey] seem to be more and more systematically the target of bullying" and concluded that he had "no objection to the granting of refugee status to Christians of Turkish nationality."[27] In

[26]They contain board of director's minutes, internal memos and correspondence with the ministry of jurisdiction since 1952.

[27]Source: Ofpra Administrative Archives (DIR3/39).

addition, in 2008, a circular addressed among others to the director of the French asylum office announced that "France decided to host on its soil a number of Iraqi nationals belonging to minorities, in a situation of great vulnerability.", implicitly referring to Iraqi Christians.[28] In 1991, a former director of the French asylum office, writing under a pseudonym, stated himself that "Assyro-Chaldeans [Christians from the Middle East] (...) obtain refugee status with a certain benevolence, even if the criteria for persecution are not really met." (François 1991). However, as I have shown in the previous section, the Muslim gap largely extends beyond these two countries which these examples, rather limited in scope, do not seem to explain fully.

6 Conclusion

Are European countries abiding by the Geneva Convention and granting asylum to those who fear persecution for reason of "race, religion, nationality, membership of a particular group or political opinion"? Up to now, the lack of micro-level data on asylum decisions has limited researchers' ability to answer this question. In this study, I take advantage of an unprecedented effort in transparency from the French asylum office, which opened its archives ten years ago. I digitized a representative sample of more than 4,000 asylum applications filed in the last 40 years, putting together the first in-depth, micro-level dataset on asylum decisions. For the first time, these data allow me to examine whether France is indeed granting asylum to those in need of protection. To do so, I compare accepted and rejected applicants in order to causally identify the effect of applicants' characteristics on the probability of being granted asylum in France.

This study provides the first empirical evidence of discrimination against Muslims in the attribution of refugee status in France using real-world data. I find that Muslims are systematically less likely to be granted asylum than Christians who are similar in every respect. Importantly, I show that this effect is comparable in magnitude to the effect of providing a credible application and is consistent with prior experimental survey evidence. Moreover, I show that this effect is driven neither by unobserved differences revealed during the interview process nor by the fact that Muslims are less likely to be persecuted than Christians. These findings unambiguously call for an increased scrutiny in the attribution of refugee status within the European Union more broadly and encourage similar data collection in other countries, either by researchers themselves or even by asylum offices directly, in order to determine whether the discrimination patterns identified in the French case apply more broadly.

[28] Source: Ofpra Administrative Archives (DIR3/10).

In addition, this work is the first to show empirical evidence that bureaucrats self-correct their, by now, well established discriminatory behaviors. In particular, I find that the discrimination against Muslims reduces over time while the predictive power of the merits of the case increases. I argue that these patterns are more consistent with a mechanism of discrimination in which bureaucrats discriminate because they underestimate the probability that Muslims are persecuted, an intuitive but under-examined mechanism in the literature of bureaucratic discrimination. Alternative explanations like taste-based discrimination or political control find only weak support in the data. As such, this paper has direct policy implications for strategies to reduce discrimination at the French asylum office – increasing the tenure of bureaucrats would mechanically reduce the overall level of discrimination – and beyond. This research also opens up a new avenue for research on the effectiveness of possible interventions to reduce the influence of bureaucrats' beliefs on decisions in order to curtail discrimination in courtrooms and administrations. Finally, this work contributes to the literature on the effectiveness of bureaucratic decision-making (Brehm and Gates 1997; Gulzar and Pasquale 2017; Huber 2000) by showing that French bureaucrats are correctly incentivized to grant asylum to those with the best application, both through the internal monitoring of their performance and through appellate review.

References

Adida, C. L., D. D. Laitin, and M.-A. Valfort (2016). *Why Muslim Integration Fails in Christian-Heritage Societies*. Harvard University Press.

Aho Nienne, C. (2013). Pourquoi il ne faut pas demander l'asile politique en France en 2013.

Aigner, D. and G. Cain (1977). Statistical Theories of Discrimination in Labor Markets. *ILR Review 30*(2), 175–187.

Angoustures, A., L. Legoux, and C. Jaggers (1995). Que sont devenus les demandeurs d'asile de 1990? *Revue européenne des migrations internationales 11*(3), 105–121.

Arnold, D., W. Dobbie, and C. S. Yang (2017). Racial Bias in bail decisions. *The Quarterly Journal of Economics*.

Arrow, K. J. (1973). The Theory of Discrimination. In *Discrimination in Labor Markets*, pp. 3–33.

Aubry, A. and K. Le Loët (2019). Il y a des gens qui craquent. Et ils ne font pas semblant de craquer.

Bansak, K., J. Hainmueller, and D. Hangartner (2016). How economic, humanitarian, and religious concerns shape European atttitudes toward asylum seekers. *Science 354*(6309), 217–222.

Becker, G. (1971). *The Economics of Discrimination.*

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2016). Stereotypes. *Quartely Journal of Economics* (February), 1753–1794.

Brehm, J. and S. Gates (1997). *Working, Shirking, and Sabotage Bureaucratic Response to a Democratic Public.*

Broockman, D. E. (2013). Black politicians are more intrinsically motivated to advance blacks' interests: A field experiment manipulating political incentives. *American Journal of Political Science 57*(3), 521–536.

Butler, D. M. and D. E. Broockman (2011). Do Politicians Racially Discriminate Against Constituents? A Field Experiment on State Legislators. *American Journal of Political Science 55*(3), 463–477.

Chesnet, P. (2015). CNDA. La roulette russe des immigrés.

Cohen, D. G. (2000). Naissance d'une nation: les personnes déplacées de l'après-guerre, 1945-1951. pp. 56–78.

Distelhorst, G. and Y. Hou (2014). Ingroup Bias in Official Behavior: A National Field Experiment in China. *Quarterly Journal of Political Science 9*(2), 203–230.

Fong, C. and J. Grimmer (2016). Discovery of Treatments from Text Corpora. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1600–1609.

François, J. (1991). De l'exil à l'asile. *Approches polémologiques, Paris*, 416–428.

Fryer, R. G. and M. O. Jackson (2008). A Categorical Model of Cognition and Biased. *The B.E. Journal of Theoretical Economics*.

Greslier, F. (2007). La commission des Recours des Réfugiés ou "l'intime conviction" face au recul du droit d'asile en France. *Revue européenne de migrations internationales 23*(2).

Gulzar, S. and B. J. Pasquale (2017). Politicians, Bureaucrats, and Development: Evidence from India. *American Political Science Review 111*(1), 162–183.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis 20*(1), 25–46.

Hainmueller, J. and D. Hangartner (2013). Who Gets a Swiss Passport? A Natural Experiment in Immigrant Discrimination. *American Political Science Review 107*(01), 159–187.

Hainmueller, J., J. Mummolo, and Y. Xu (2016). How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice.

Hausman, D. (2016). The Failure Failure of Immigration Appeals. *University of Pennsylvania Law Review 164*(5), 1177–1238.

Hemker, J. and A. Rink (2017). Evidence from German Welfare Offices. *American Journal*

of Political Science 61(4), 786–803.

Holzer, T., G. Schneider, and T. Widmer (2000). Discriminating Decentralization Federalism and the Handling of asylum application in Switzerland, 1988-1996. *Journal of Conflict Resolution 44*(2), 250–276.

Hopkins, D. J. and G. King (2010). A Method of Automated Nonparametric Content Analysis for Social Science. *American Journal of Political Science 54*(1), 229–247.

Huber, J. D. (2000). Delegation to civil servants in parliamentary democracies. *European Journal of Political Research 37*(3), 397–413.

Keith, L. C. and J. S. Holmes (2009). A rare examination of typically unobservable factors in US asylum decisions. *Journal of Refugee Studies 22*(2), 224–241.

Keith, L. C., J. S. Holmes, and B. P. Miller (2013). Explaining the divergence in asylum grant rates among immigration judges: An attitudinal and cognitive approach. *Law and Policy 35*(4), 261–289.

Kripendorff, K. (1980). *Content Analysis: An Introduction to its Methodology*, Volume 5.

Laouénan, M. and R. Rathelot (2017). Ethnic Discrimination on an Online Marketplace of Vacation Rental. (318).

Lepors, A. (2015). La grande Loterie du droit d'asile.

Levitt, S. (2004). Testing Theories of Discrimination: Evidence from Weakest Link. *The Journal of Law and Economics 47*(2), 431–453.

List, J. A. (2004). The Nature and Extent of Discrimination in the Marketplace : Evidence from the Field. *The Quarterly Journal of Economics 119*(1), 49–89.

Marten, L. (2017). Political Bias in Court? Lay Judges and Asylum Appeals.

McClendon, G. H. (2016). Race and Responsiveness: An Experiment with South African Politicians. *Journal of Experimental Political Science 3*(01), 60–74.

Neggers, Y. (2017). Enfranchising Your Own? Experimental Evidence on Bureaucrat Diversity and Election Bias in India. *American Economic Review*.

Neumayer, E. (2005). Asylum Recognition Rates in Western Europe: Their Determinants, Variation, and Lack of Convergence. *Journal of Conflict Resolution 49*(1), 43–66.

Paluck, E. L. (2016). How to overcome prejudice. *Science 352*(6282), 147.

Pereira, V. (2017). De l'exclusion à la protection : l'Ofpra et les réfugiés portugais entre 1952 et 1974. In *Réfugiés et apatrides. Administrer l'asile en France (1920-1960)*, pp. 243–266.

Phelps, E. (1972). The Statistical Theory of Racism and Sexism. *The American Economic Review 62*(4), 659–661.

Pope, D. G. and J. R. Sydnor (2011). What's in a Picture ? Evidence of Discrimination from Prosper.com. *The Journal of Human Resources 46*(1), 53–92.

Price, J. and J. Wolfers (2010). Racial Discrimination Among NBA Referees. *The Quarterly*

Journal of Economics 125(4), 1859–1887.

Ramji-Nogales, J., A. Schoenholtz, and S. Philipp (2007). Refugee Roulette: Disparities in Asylum. *Stanford Law Review 62*(2), 125.

Roberts, M. E., B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. K. Gadarian, B. Albertson, and D. G. Rand (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science 58*(4), 1064–1082.

Rosenblum, M. R. and I. Salehyan (2004). Norms and Interests in US Asylum Enforcement. *Journal of Peace Research 41*(6), 677–697.

Rottman, A. J., C. J. Fariss, and S. C. Poe (2009). The path to asylum in the US and the determinants for who gets in and why. *International Migration Review 43*(1), 3–34.

Shayo, M. (2009). A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution. *American Political Science Review 103*(02), 147–174.

Shayo, M. and A. Zussman (2011). Judicial ingroup bias in the shadow of terrorism. *Quarterly Journal of Economics 126*(3), 1447–1484.

White, A., N. L. Nathan, and J. Faller (2015). What Do I Need to Vote? Bureaucratic Discretion and Discrimination by Local Election Officials. *American Political Science Review 109*(1), 129–142.

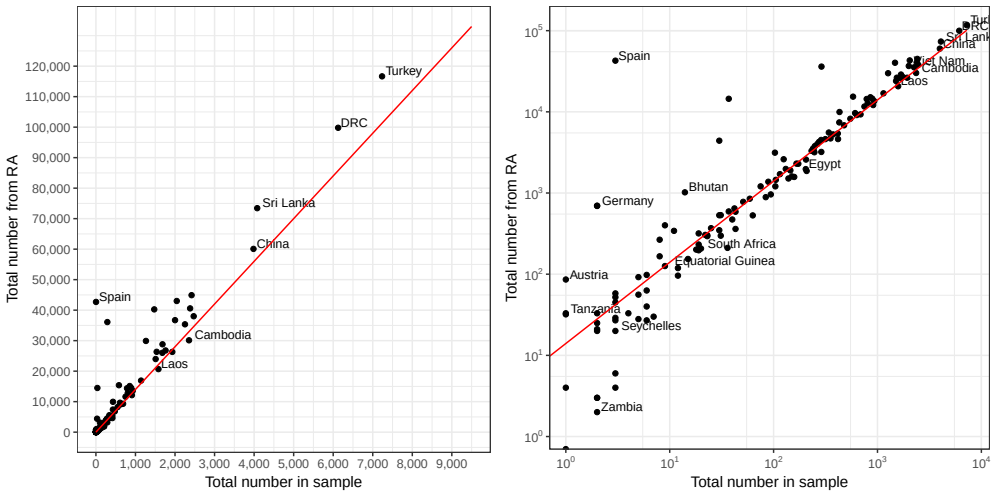
Online Appendices

A Sampling

A.1 Representativity of the sampling frame

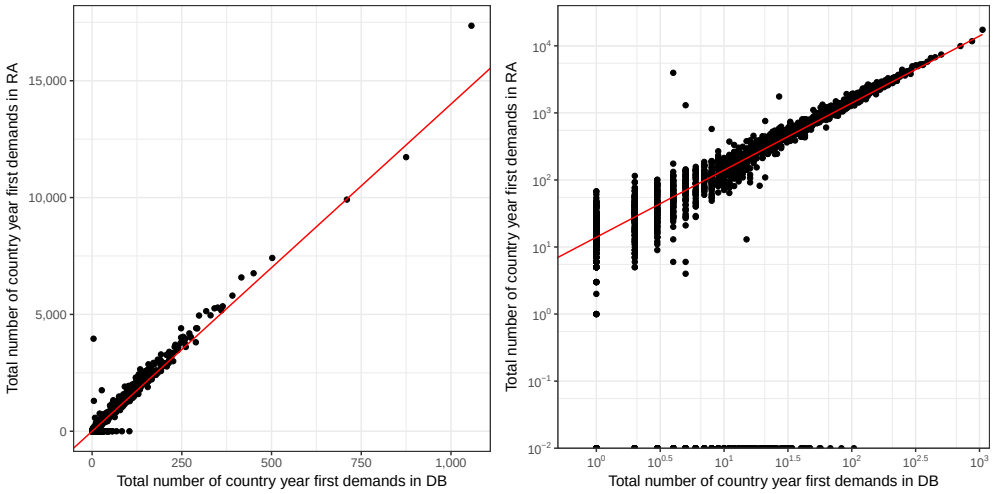
In 2015, the French asylum office provided me with a list of 100,000 asylum applications filed between 1952 and 2014, randomly drawn from their administrative databases. I show, in this section, that this list is representative of the universe of asylum applications filed during this period. To do so, I use official statistics published every year by the French asylum office that are accessible online (since 2001) or in the archives of the French asylum office (before 2000). I compiled a dataset of the total number of first-time applications by country of origin and year of application. The type of statistics published every year by the French asylum office vary slightly over time. Between 1952 and 1970, they only reported the number of refugees accepted. But as I mentioned, the acceptance rate was almost 100 percent until 1976. Between 1970 and 1980, they reported the number of accepted and the number of rejected applications. I add these two numbers to get at the number of first-time applications. Since 1981, they reported directly the number of first-time applications. In Figure A.1, I compare, on the left panel, the total number of first-time asylum applications by country (1952-2014) in the administrative sample of 100,000 (x-axis) to the number published in the activity reports (RA) (y-axis). The right panel displays the same data using a log scale. For asylum applications filed after 1989, I know the year of application, so for the period between 1989 and 2014, I can compare the number of asylum applications by country of origin and year of application (Figure A.2). Overall, these figures show that the sample of 100,000 observations that I use to sample applications for in-depth data collection is representative of the population of asylum seekers who applied for asylum between 1976 and 2014.

Figure A.1: Total number of country-level first-time applications (1952-2014)



Note: Each dot represents a country of origin in the sample of 100,000 observations, the x axis represents the number of applications from that country of origin in the sample, and the y axis the number of applications in the official statistics. The same figure is displayed twice, on a regular scale on the left and using a log scale on the right.

Figure A.2: Total number of year-country level first-time applications (1989-2014)



Note: Each dot represents a country of origin/year of application in the sample of 100,000 observations, the x axis represents the number of applications from that country of origin in the sample, and the y axis the number of applications in the official statistics. The same figure is displayed twice, on a regular scale on the left and using a log scale on the right.

A.2 Destructions

In 2009, the archivists at the French asylum office started to destroy folders in order to make space in the archives. The protocol for the destruction is the following: Every six months, all applications rejected for more than ten years are selected. Qualitatively interesting folders are put aside and 90 percent of remaining folders are destroyed. As applications rejected for more than 10 years qualify every round for destruction, they are very few rejected folders left.

A.3 Sampling Process

In this section, I describe in details the sampling process for the selection of the applications for in-depth data collection. The final sample was selected in three stages. First, in December 2015, I sampled 10,000 folders for digitization accounting for the non-random destruction of folders and oversampling applications posterior to 1989, after exclusion of the small nationalities (I limited the sample to the top 74 countries of origin (out of 155) for which I had at least 100 applications). Second, in May 2016, I requested an additional sample of 50,000 administrative records filed between 1990 and 1992 to sample an additional 500 folders from these three years. I drew an additional 500 folders in order to fill in the margins. Third, in June 2016, I sampled an additional 500 application filed between 2015 and 2016. In what follows, I describe in details the procedure for the first stage.

I first dropped 3 perfect duplicates and 1,572 applications from small nationalities. I then split the data into 4 different subsets depending on the date of application (no destruction after 2005) and the database.

- 26,433 applications filed prior to 1989 (Garonor). For these, I only know the nationality of the applicant accurately. None of these folders were destroyed. From these I draw 1,292 folders.

- 14,259 applications filed prior to January 1, 1990 (Inerec). For these and all other subset I know the nationality, the date of application, and whether the applicant received refugees status. Some of these had been destroyed. From these I draw 699 folders.

- 35,398 applications filed between January 1, 1990 and December 31, 2004 from which I draw 4,868 folders.

- 22,335 applications filed between January 1, 2005 and December 31, 2014 from which I draw 3,054 folders.

Within each of these subsets of the data, I generated a new nationality variable that is equal to the nationality for the top ten nationalities within each sample, and generated a code for all others. I then predicted the probability of being observed (which is one minus the probability of being destroyed) using different linear models and a combination of different variables of refugee status, year of application, country of origin and all possible interactions. For each model, I drew several hundreds of samples. I choose the model that minimizes the average distance according to the average margin difference for the following: frequencies for nationality, year of application, the overall rejection rate, frequencies restricted to the top 10 nationalities and the yearly rejection rate. Finally for each sample, model, I selected the sample that minimized deviation.

In total, with the help of a team of research assistants, I searched 5,421 applications and found 5,220.[29] Of these, 447 applications were excluded after a rapid examination of the content of the folder on the basis that they did not meet the criteria of being a first-time application with a decision and an application form in French.[30]

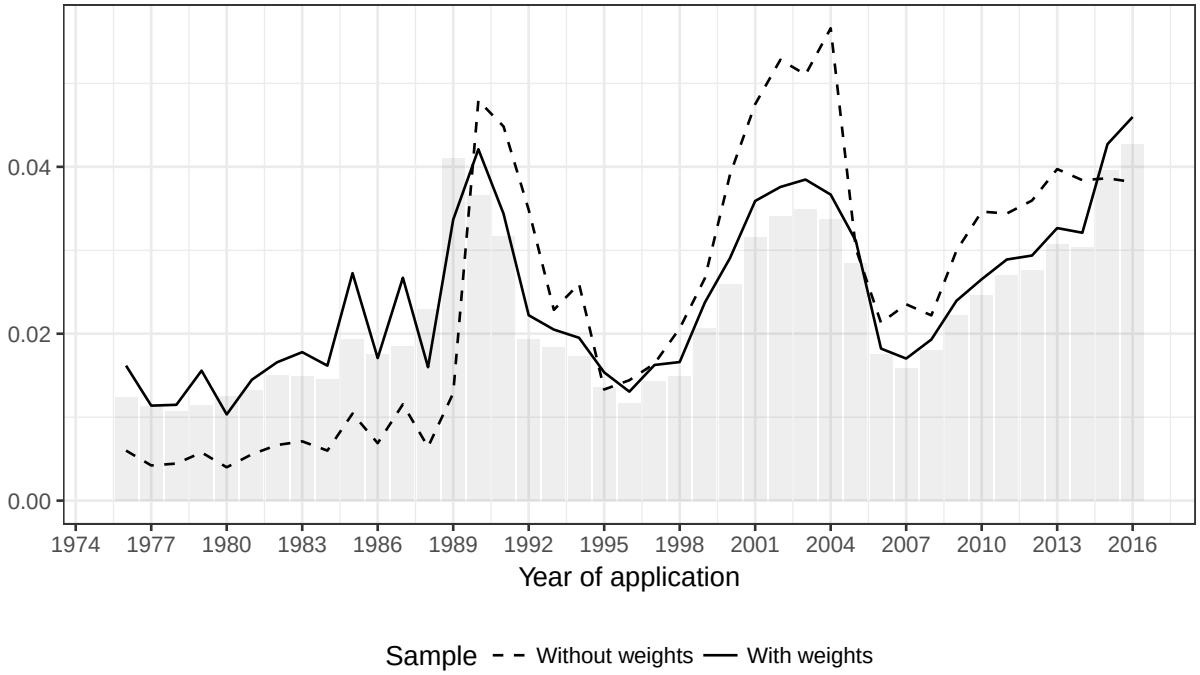
A.4 Weights

I generate weights to first adjust for the fact that the sampling correction was not perfect and second correct for the oversampling. I first use entropy balancing (Hainmueller 2012) to match the proportion of accepted applications by year in the data collected to the proportion of accepted applications in the administrative sample. I then multiply these weights by 6.5 for observations prior to 1990, 2 for observations between 1990 and 2015 included, and finally by 14.8 for observations for 2016 in order to match the number of first-time applications. The following figures compare the density in the administrative sample (gray bar) to the density in the unweighted data (dashed or round dot) and in the weighted dataset (solid or triangle) of the year of application (Figure A.3), the acceptance rate among first-time application (Figure A.4), the final acceptance rate (Figure A.5) and the acceptance rate by country of origin (Figure A.6).

[29] Applications were randomly ordered and I searched for them in that order. 142 had been destroyed between the sampling and the data collection and 59 were lost.

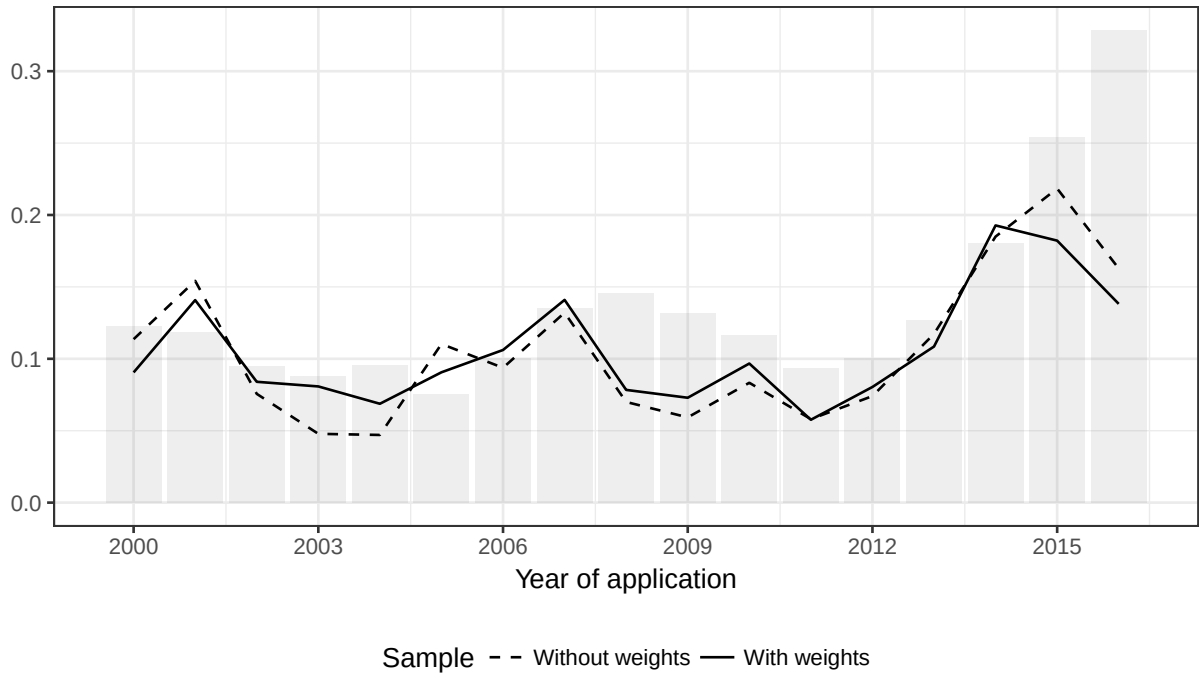
[30] These include applications for stateless status, those that were submitted prior to 1952, applications that were not submitted in French, and the 96 applications that did not have a decision.

Figure A.3: Number of first-time applications (1976-2016)



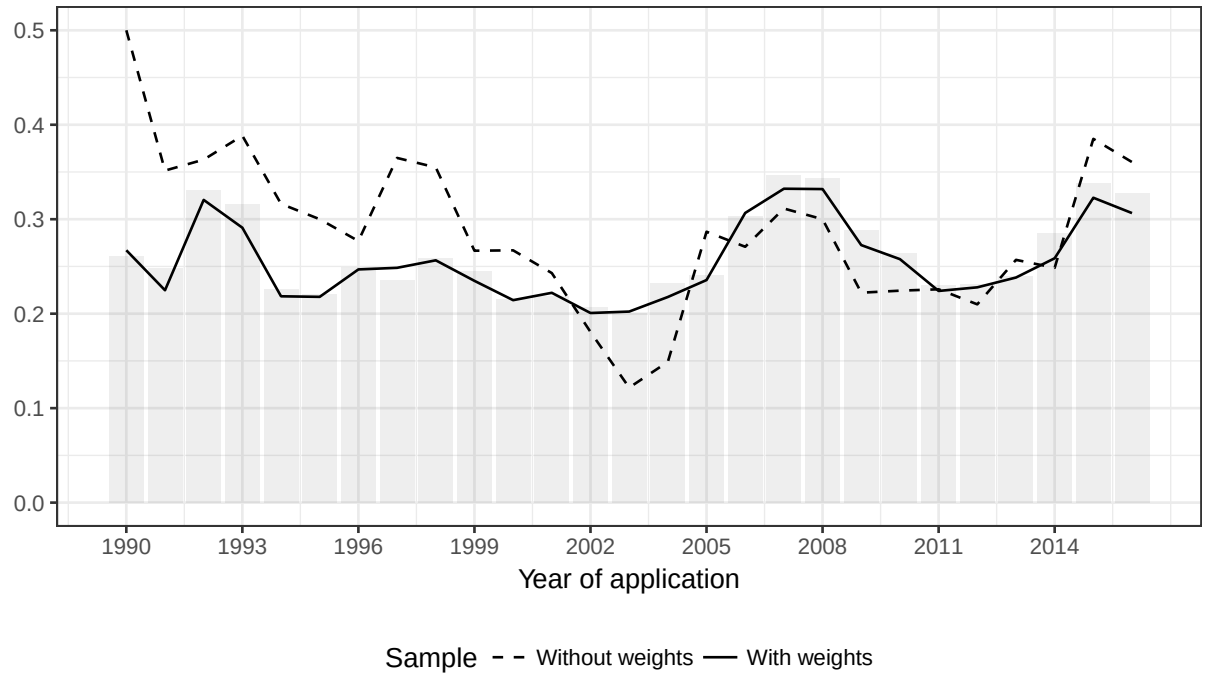
Note: This figure displays the density of first-time applications in the administrative sample (gray bar), in the unweighted sample of extracted folders (dashed line) and in the weighted sample of extracted folders (solid line).

Figure A.4: First-time acceptance rate by year of application (2000-2016)



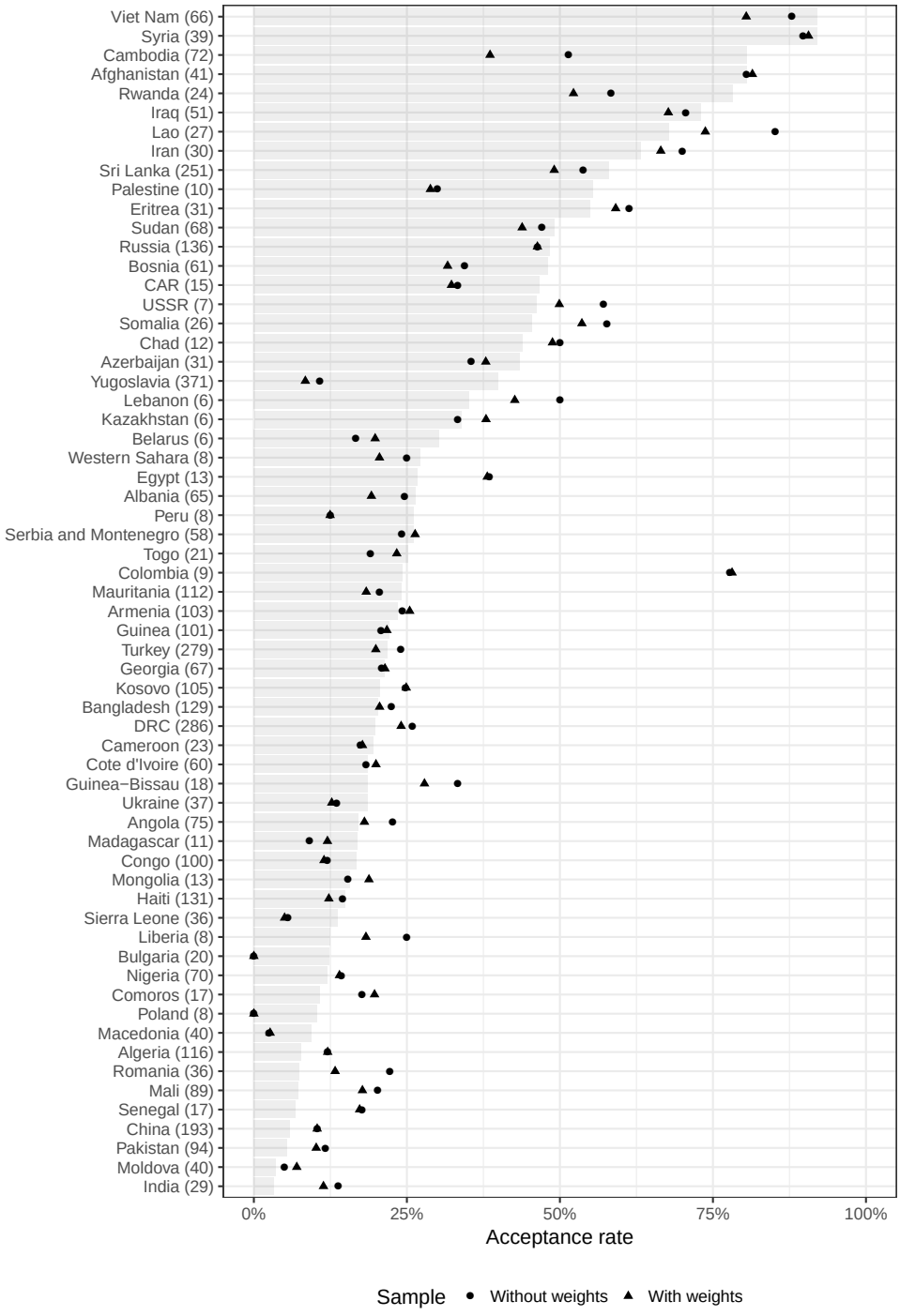
Note: This figure displays the density of first-time acceptance rate in the administrative sample (gray bar), in the unweighted sample of extracted folders (dashed line) and in the weighted sample of extracted folders (solid line).

Figure A.5: Final acceptance rate by country of origin by year of application (1990-2016)



Note: This figure displays the density of final acceptance rate in the administrative sample (gray bar), in the unweighted sample of extracted folders (dashed line) and in the weighted sample of extracted folders (solid line).

Figure A.6: Final acceptance rate by country of origin



Note: This figure displays the density of final acceptance rate by country of origin in the administrative sample (gray bar), in the unweighted data (round) and in the weighted dataset (triangle).

B Measuring and predicting the merit of the case

B.1 Comparing the narratives of Muslim and Christian applicants

I find that narratives of Christians and Muslims differ substantially in content (Table B.1). Muslims are substantially more likely to apply for asylum for reasons related to race (25 percent do) compared to Christians (only 13 percent). They are thus slightly less likely to ask for asylum for each the four other reasons in the Geneva Convention though none shows up significantly. The narratives provided by Muslims also differ substantially on most of the twenty topics discovered by the structural topic model. However, reassuringly, Muslims are no less likely to claim persecution (84 percent of Christians and Muslims do) and they are no more likely to not mention a motive of persecution (30 percent of narratives do not explicitly mention one of the five motives in the Geneva Convention). Moreover, Muslims do not provide narratives that substantially differ in quality. Even though they are slightly longer (46 additional words on average), they are no less likely to be coded as probable, convincing, detailed, or coherent and do not differ in the number of dates or location nor in the number of personal pronouns. Together these patterns suggest that even though narratives provided by Muslims and Christians do differ substantially in content, they do not seem to differ in the quality of the argument.

Table B.1: Differences between the narratives of Christian and Muslim applicants

	Christians		Muslims		t-test	
	Mean	N	Mean	N	Diff.	p
Panel A: Narrative Coding						
Persecution						
Is persecuted	0.84	159	0.84	192	-0.00	0.98
For reasons related to						
Race	0.13	159	0.25	192	-0.12	0.01
Political opinion	0.47	159	0.39	192	0.08	0.13
Religion	0.09	159	0.06	192	0.03	0.36
Nationality	0.08	159	0.05	192	0.03	0.18
Social group	0.08	159	0.08	192	0.00	0.90
None	0.30	159	0.32	192	-0.02	0.75
Narrative was coded as						
Probable	2.65	159	2.66	192	-0.00	0.98
Convincing	2.39	159	2.34	192	0.05	0.60
Detailed	2.43	159	2.31	192	0.12	0.26
Singular	2.46	159	2.44	192	0.02	0.80
Coherent	2.55	159	2.67	192	-0.12	0.15
The narrative mentions						
A historical event	0.46	159	0.46	192	-0.00	0.93
Family in France	0.13	159	0.07	192	0.06	0.07
Credibly claims the						
Geneva Convention	0.53	159	0.50	192	0.03	0.52
Panel B: Narrative Features						
Level of detail						
Number of words	828.93	1547	780.13	1719	48.80	0.07
# of dates mentioned	7.47	1557	7.30	1721	0.17	0.55
# of places mentioned	6.90	1557	6.61	1721	0.29	0.29
Personal						
Number of personal pronoums	35.52	1474	36.44	1648	-0.92	0.45
Originality						
Distance	49.83	1543	48.39	1718	1.44	0.09
Topics						
Family membesr	0.10	1547	0.11	1717	-0.00	0.67
Bangladesh	0.00	1547	0.08	1717	-0.07	0.00
Student protest	0.08	1547	0.05	1717	0.02	0.01
Escape	0.08	1547	0.13	1717	-0.05	0.00
Sri lanka	0.02	1547	0.01	1717	0.02	0.00
RDC	0.12	1547	0.00	1717	0.12	0.00
Live in France	0.21	1547	0.17	1717	0.04	0.00
Former Yugoslavia	0.05	1547	0.13	1717	-0.08	0.00
China	0.02	1547	0.01	1717	0.01	0.00
Ethnic minorities	0.09	1547	0.03	1717	0.06	0.00
Kurds in Turkey	0.00	1547	0.13	1717	-0.13	0.00
Political instability	0.10	1547	0.02	1717	0.07	0.00
Political opposition	0.03	1547	0.12	1717	-0.09	0.00
Court hearing	0.02	1547	0.02	1717	-0.00	0.79
Narrative	0.18	1547	0.15	1717	0.02	0.06
Angola	0.07	1547	0.03	1717	0.04	0.00
Salutation	0.06	1547	0.07	1717	-0.00	0.89
Family at risk	0.04	1547	0.06	1717	-0.02	0.03
Zaire	0.05	1547	0.01	1717	0.05	0.00
Religion	0.02	1547	0.02	1717	0.00	0.66

B.2 Hand-coding the credibility of the narrative

In order the measure the credibility of the narrative, three students read 459 randomly sampled narratives and coded them on several dimensions. This section provides additional elements on the data collected and the reliability of the main measure of the credibility of the narrative. Table B.2 provides summary statistics on the main variables asked in the narrative survey. In Table B.4, I report the proportion of narratives for which every combinations of coders agreed. Next to it, I report the frequency with which coders coding randomly (with mean 50 percent) would agree. Coders agree with another coder 70 percent to 80 percent of the time whereas random coders agree only 50 percent of the time. provides an assessment of inter-coder agreement on this main measure and this table shows that agreement is substantially higher than random.

Table B.2: Summary statistics on the narratives

	N	Mean	Std. Dev.	Min	Max
Persecution					
Is persecuted	459	0.80	0.40	0	1
For reasons related to					
Race	459	0.17	0.38	0	1
Political opinion	459	0.42	0.49	0	1
Religion	459	0.06	0.24	0	1
Nationality	459	0.05	0.21	0	1
Social group	459	0.07	0.25	0	1
None	459	0.35	0.48	0	1
Probable	459	2.60	0.83	1	4
Convincing	459	2.31	0.82	1	4
Detailed	459	2.24	0.99	1	4
Singular	459	2.40	0.81	1	4
Coherent	459	2.59	0.75	1	4
The narrative mentions					
A historical event	459	0.46	0.50	0	1
Family in France	459	0.11	0.31	0	1
Credibly claims protection under					
Geneva Convention	459	0.50	0.50	0	1

Note: This table reports summary statistics on the main variables collected by coders for a representative sample of 459 narratives. The unit of observation is narrative/coder.

Table B.3: Correlates of credibility

	(1) Credible Narrative
Is persecuted	0.113*
	(0.05)
Probable	0.120***
	(0.04)
Convincing	0.200***
	(0.04)
Detailed	0.053
	(0.03)
Singular	-0.047
	(0.03)
Coherent	0.083*
	(0.03)
Constant	-0.603***
	(0.07)
Observations	459
R^2	0.450

Note: This table shows point estimates and standard errors in parenthesis from a OLS regression of the main credibility variable collected in the narrative hand-coding on other leading questions about features of the text like the level of detail. The unit of observation is narrative/coder.

Table B.4: Assessment of inter-coder agreement for the measure of the quality of the narrative

	Observed	Chance
Agreement		
Coder 1 and 2	0.69	0.49
Coder 2 and 3	0.81	0.53
Coder 1 and 3	0.71	0.49
All three coders	0.61	0.25

Note: This table reports the observed inter-coder agreement for all combination of coders and the agreement generated by random coding.

B.3 Summary statistics on the personal narratives

Table B.5: Summary statistics on narrative features

	N	Mean	Std. Dev.	Min	Max
Detail					
Number of words	3,932	777.98	765.20	1	7169
# of dates mentioned	3,932	7.29	8.32	0	95
# of places mentioned	3,932	6.54	7.51	0	84
Personal					
Number of personal pronoums	3,760	34.58	33.07	0	445
Originality					
Distance	3,927	47.20	23.49	6	347
Topic proportion >.2					
Family membesr	3,930	0.09	0.28	0	1
Bangladesh	3,930	0.04	0.19	0	1
Student protest	3,930	0.07	0.26	0	1
Escape	3,930	0.09	0.29	0	1
Sri lanka	3,930	0.04	0.20	0	1
RDC	3,930	0.04	0.20	0	1
Live in France	3,930	0.22	0.42	0	1
Former Yugoslavia	3,930	0.08	0.27	0	1
China	3,930	0.04	0.20	0	1
Ethnic minorities	3,930	0.05	0.22	0	1
Kurds in Turkey	3,930	0.07	0.25	0	1
Political instability	3,930	0.05	0.22	0	1
Political opposition	3,930	0.05	0.23	0	1
Court hearing	3,930	0.02	0.14	0	1
Narrative	3,930	0.15	0.36	0	1
Angola	3,930	0.06	0.23	0	1
Salutation	3,930	0.06	0.23	0	1
Family at risk	3,930	0.04	0.20	0	1
Zaire	3,930	0.03	0.18	0	1
Religion	3,930	0.02	0.14	0	1

Note: This table presents summary statistics on the features of the narratives used for predicting the credibility for the rest of the corpus.

Table B.6: Topic description (FREX)

Topic								
Family members	mar	fill	mer	enfant	fil	mari	soeur	frer
Bangladesh	awam	bangladesh	ligu	bnp	malfaiteur	terror	inde	commissariat
Student protest	régim	opposit	étudi	social	populair	démocrat	manifest	professeur
Escape	voitur	argent	environ	passeur	soudan	pai	emmen	puis
Sri Lanka	ltte	sri	tamoul	tigr	lank	colombo	jaffn	camp
RDC	kabil	kinshas	congo	congol	brazzavill	udp	congolais	cellul
Live in France	vivr	pay	franc	veux	peux	vi	peut	espoir
Former Yugoslavia	serb	kosovo	alban	yougoslav	albanais	rom	bosn	guerr
China	chin	chinois	ouvri	usin	fonctionnair	entrepris	licenci	démocrat
Ethnic minority	russ	armen	appart	arménien	tchétchen	moscou	géorg	azerbaïdjan
Kurds in Turkey	kurd	turqu	turc	kurdistan	istanbul	pkk	tortur	villag
Political instability	etait	president	apre	sassou	arret	decid	mem	haït
Political opposition	guin	mauritan	sénégal	mauritanien	conakry	rebel	gendarm	élect
Court hearing	joint	alger	lettr	journal	convoc	tribunal	univers	écrit
Narrative	dit	dis	c'et	qu'il	quand	rien	dir	ça
Angola	militair	angol	prison	titr	unit	fronti	camp	soldat
Salutation	agré	statut	ofpra	madam	monsieur	salut	express	distingu
Family at risk	épous	époux	agress	menac	domicil	syr	physiqu	2012
Zaire	rwand	mobutu	zaïr	président	ministr	zair	zaïrois	annex
Religion	conseil	religion	chrétien	islam	musulman	religi	églis	communaut

Note: This table presents, for each of the 20 topics in the structural topic model, the eight words that scored the highest on the FREX metric.

B.4 Predicting the credibility

In this section, I provide additional details regarding the choice of Random Forest to predict the credibility of the narrative. To choose the best predictive model, I proceed as follow:

- I partitioned the labeled set of 341 of hand-coded documents into the labeled training set (75 percent) and the labeled test set (25 percent).
- I train three different algorithms using cross validation to tune the model using as predictors the tf-idf of unigrams and additional narratives covariates (length, number of pronouns, number of dates and locations mentioned, and indicator for whether one of the twenty topics covers more than 20 percent of a narrative).
- For each of these models, I get the predicted probabilities and compute AUC, Brier Score, and accuracy (Table B.7), and the calibration plots (Figure B.1).

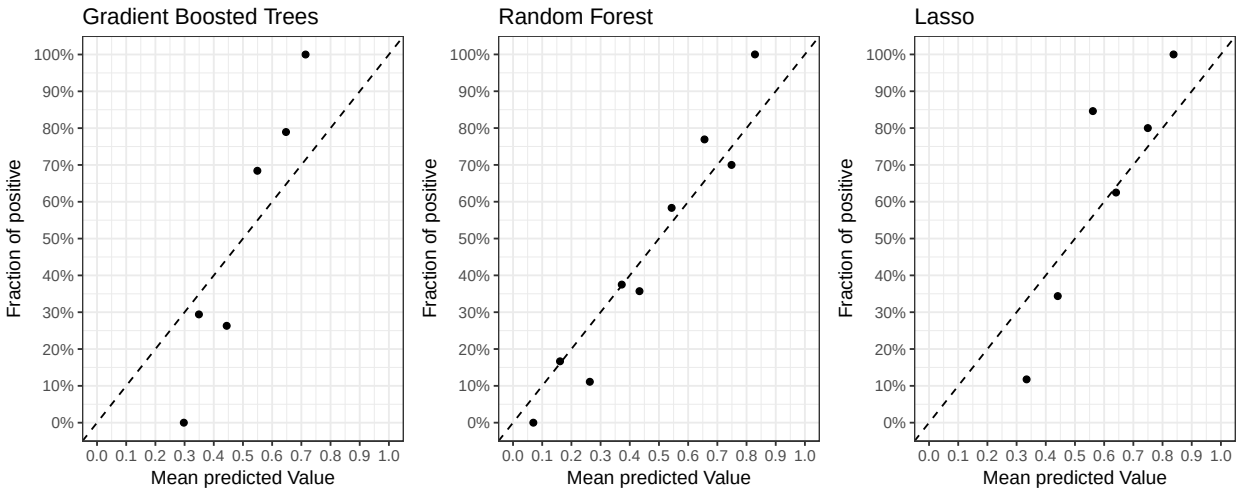
In addition, Table B.8 shows the predicted measure of merit correlates as expected with extracted features of the text.

Table B.7: Model performance

Model	AUC	Accuracy	Brier Score
Random Forest	0.82	0.76	0.20
Gradient Boosted trees	0.82	0.75	0.18
Lasso	0.83	0.76	0.19

Note: This table presents three performance statistics for three different algorithms computed on the left-out sample of 85 narratives.

Figure B.1: Calibration plots



Note: This figure plots the calibration of predicted probabilities from three different algorithms. In a left-out sample of 85 narratives, I predict the credibility for each narrative from each of the three algorithms after the tuning of the parameters. I then divide the predicted probabilities in ten equally sized bins. For each bin, I compute the mean predicted value (x-axis) and the fraction of observations in that bin with a true positive (y-axis).

B.5 Assessing the quality of the credibility measure

Table B.8: Correlates of credibility

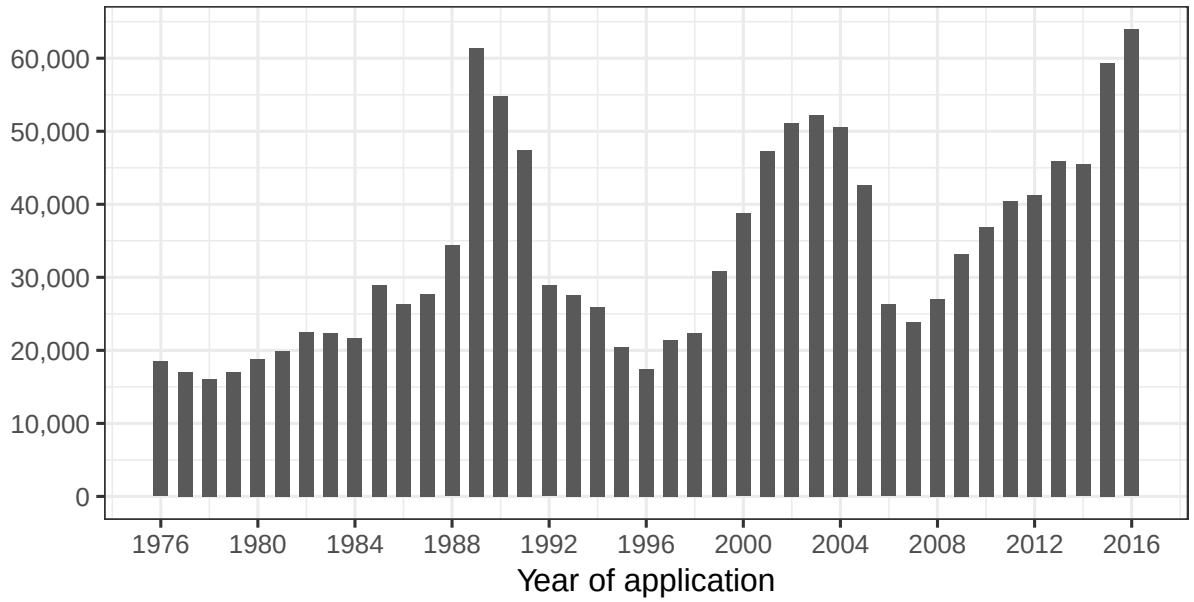
	Credibility of the narrative	
	(1)	(2)
	Hand Coded	Predicted
Number of words ('000)	0.211*	0.255***
	(0.09)	(0.02)
Distance ('000)	0.259	1.359***
	(1.06)	(0.30)
# of dates mentioned ('000)	13.996*	9.679***
	(6.21)	(1.33)
# of places mentioned ('000)	-1.544	2.510*
	(3.97)	(1.22)
Number of personal pronoums ('000)	-1.405	0.254
	(1.45)	(0.42)
Constant	0.274***	0.186***
	(0.05)	(0.02)
Observations	320	3472

Note: This table displays the results of an OLS regression in the subsample of hand-coded narratives in column 1 and the whole sample of narratives in column 2 (excluding hand coded narratives). The dependent variable is a binary variable for whether the predicted credibility was above 50 percent. No additional controls.

C Additional tables and figures

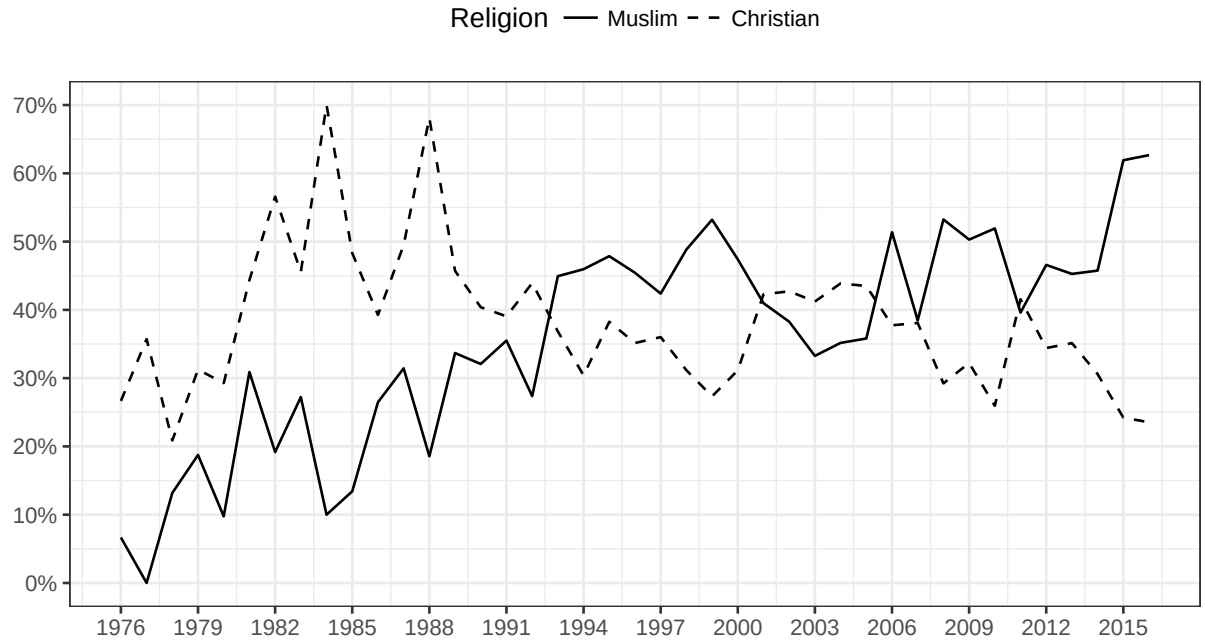
C.1 Descriptive statistics on asylum applications

Figure C.1: Number of first-time applications filed at the French asylum office (1976-2016)



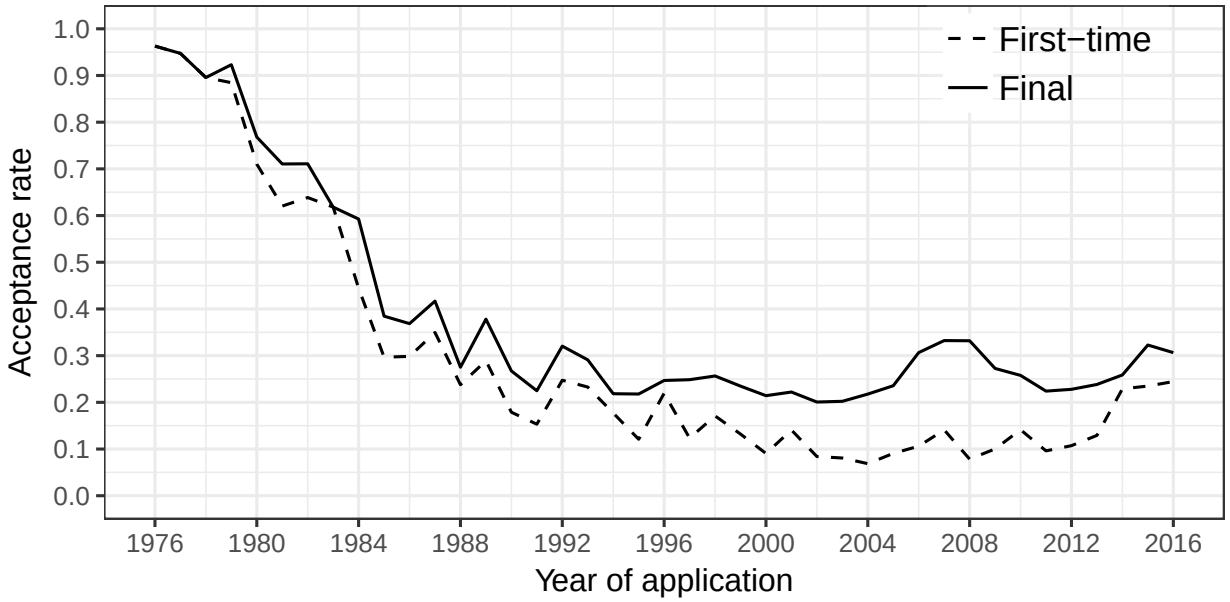
Note: This figure displays the total number of first-time asylum applications filed at the French asylum office between 1976 and 2016, excluding unaccompanied minors. Source: Official statistics published yearly by the French Asylum Office.

Figure C.2: Proportion of Muslims and Christians among first-time applicants (1976-2016)



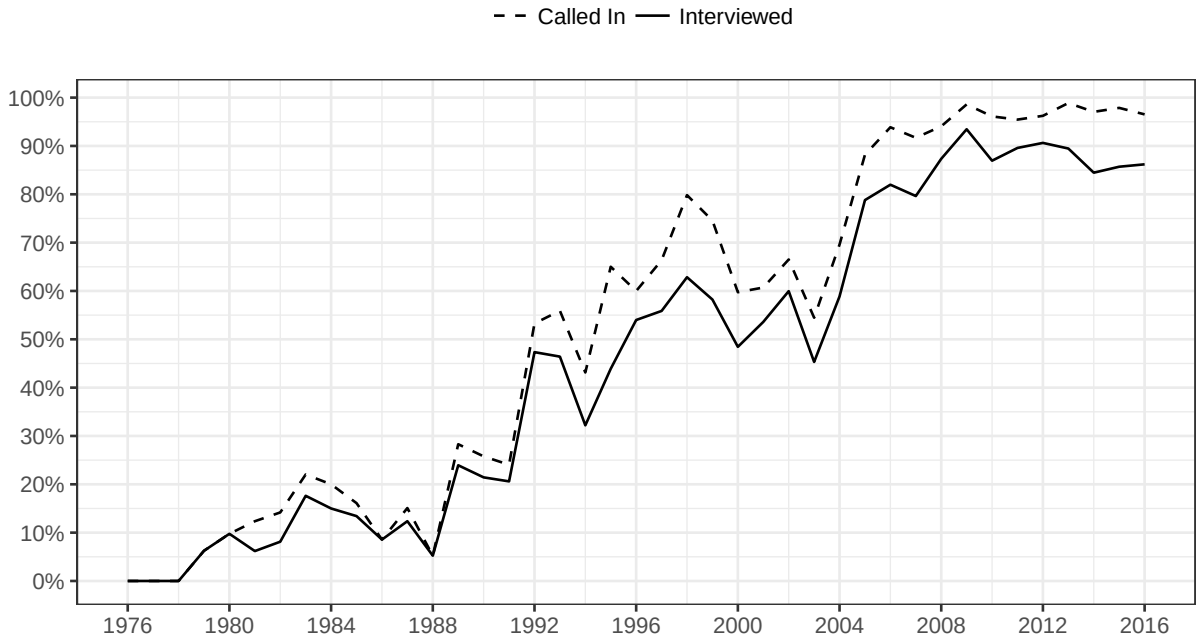
Note: This figure plots the weighted proportion of Muslim (solid line) and Christian (dashed) applicants in the sample.

Figure C.3: Acceptance rate at the French asylum office (1976-2016)



Note: This figure displays the acceptance rate after the first decision (dashed line) and the final decision (solid line) at the French asylum office between 1976 and 2016 (Geneva Convention only). Source: Archival data.

Figure C.4: Proportion of applicants called in for an interview (1976-2016)



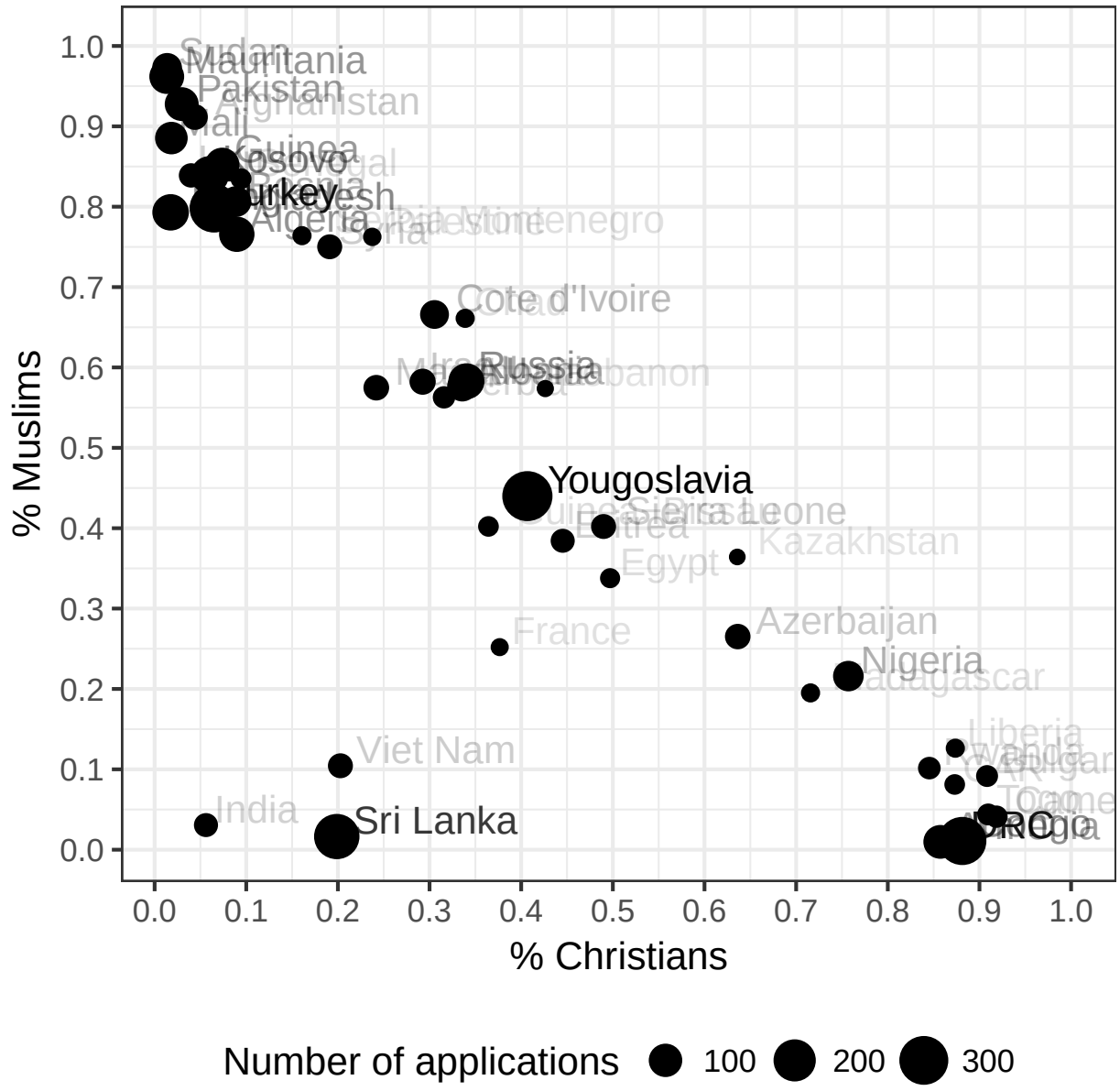
Note: This figures displays the weighted proportion of applicants that are called in for an interview (dashed line), and the proportion of applicants that actually went through an interview (solid line).

Table C.1: Summary statistics on additional independent variables

	N	Mean	Std. Dev.	Min	Max
Accelerated procedure	4,142	0.07	0.26	0	1
Passport (reported)	4,142	0.25	0.44	0	1
Laissez-Passez (reported)	4,142	0.03	0.18	0	1
Number of children	4,142	1.08	1.57	0	12
Has family refugee in France	4,142	0.09	0.28	0	1
Military service					
No	4,142	0.69	0.46	0	1
Yes	4,142	0.16	0.37	0	1
Missing	4,142	0.15	0.36	0	1
Arrival in France (reported)					
Irregular	4,142	0.54	0.50	0	1
Regular	4,142	0.16	0.36	0	1
Missing	4,142	0.30	0.46	0	1

Note: This table presents weighted summary statistics on the additional variables used in the analysis reported in Table 3 column 3. The variable "Accelerated procedure" indicates whether the application was expedited or not. Whether an application follows the normal or the accelerated procedure is determined by the prefecture which can refuse entry on the territory to those that the country is deemed safe, or when the asylum application is deemed fraudulent. In this case, the applicant has to be notified of the decision within 15 days after the interview. The variables "Passport (reported)" and "Laissez-Passez (reported) " indicate whether the applicant reported providing these documents in the application. The "Number of children" corresponds to the total number of children that were listed by the applicant on the form. The variable "Has family refugee in France" indicates whether the applicant listed a family member currently residing in France. Finally, the variables "Military service" and "Arrival in France" code self-reported information by the applicant.

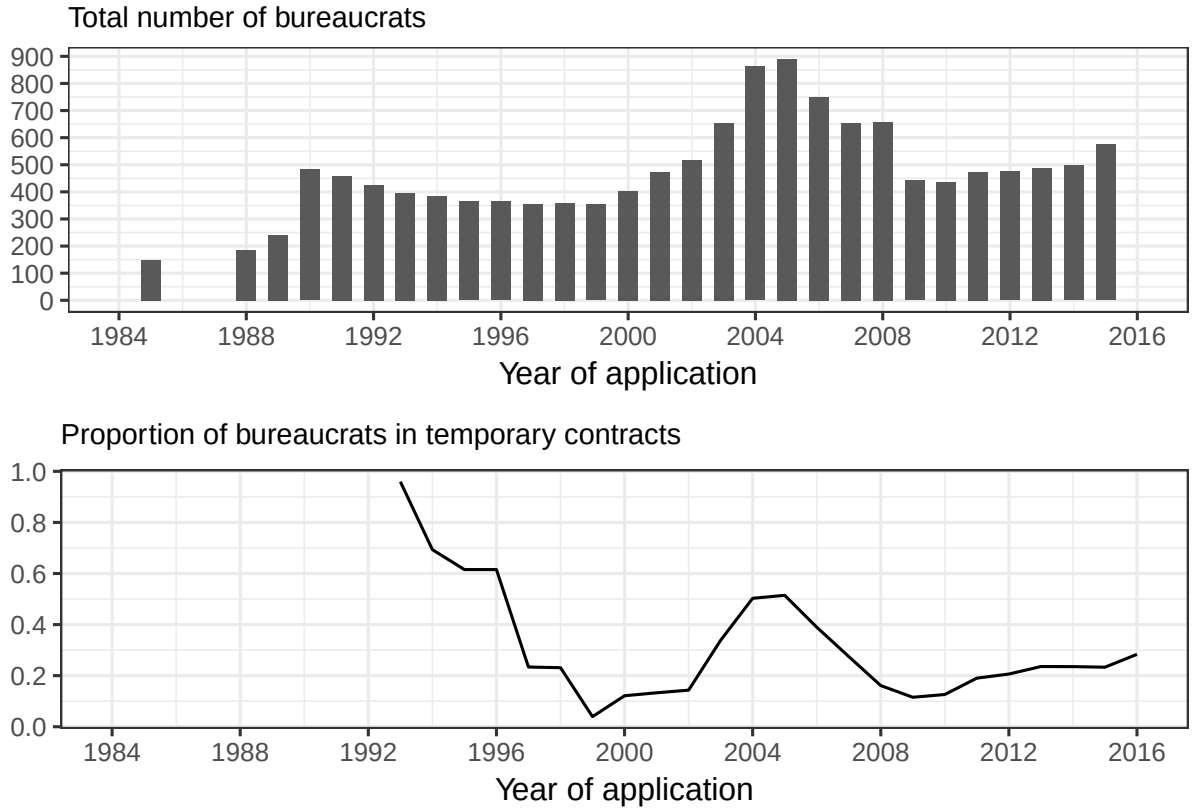
Figure C.5: Proportion of Muslim and Christian applicants within country (1976-2016)



Note: This figure plots, for each country included in the sample, the proportion of the applicants from this country that are Christians (x-axis) and the proportion that are Muslims (y-axis). This figure shows for example that Turkish applicants comprise of 75 percent Muslims, and 10 percent Christians over the period under study (1976-2016).

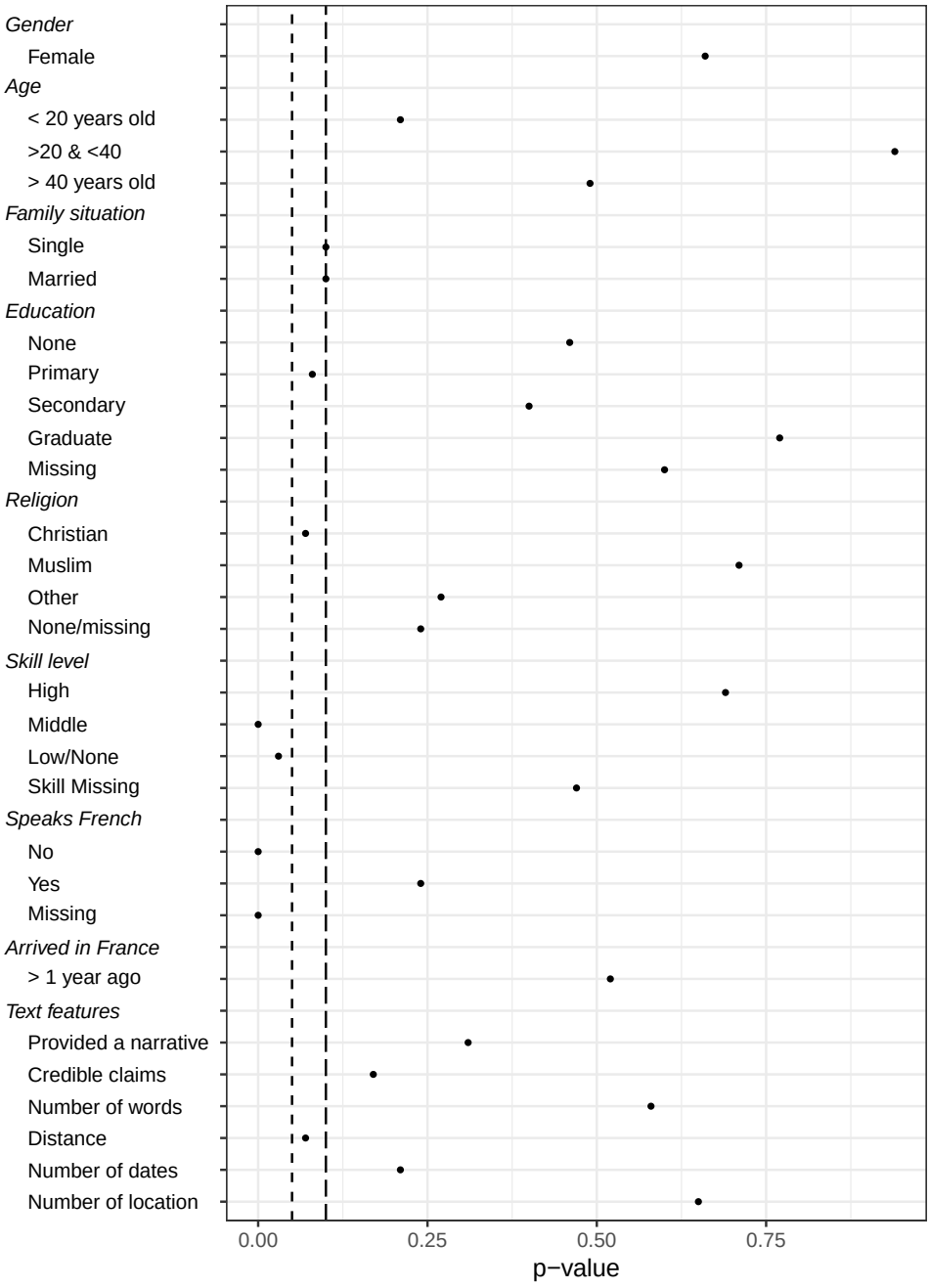
C.2 Descriptive statistics on bureaucrats

Figure C.6: Number and employment status of bureaucrats working at the French asylum office (1984-2016)



Note: This figure displays the number of bureaucrats working at the French asylum office (top panel) and the proportion of bureaucrats on temporary contract since 1993 (bottom panel). This data was collected from the activity reports published yearly by the French asylum office. Unfortunately the total number of bureaucrats working at the French asylum office was not reported 1986 and 1987.

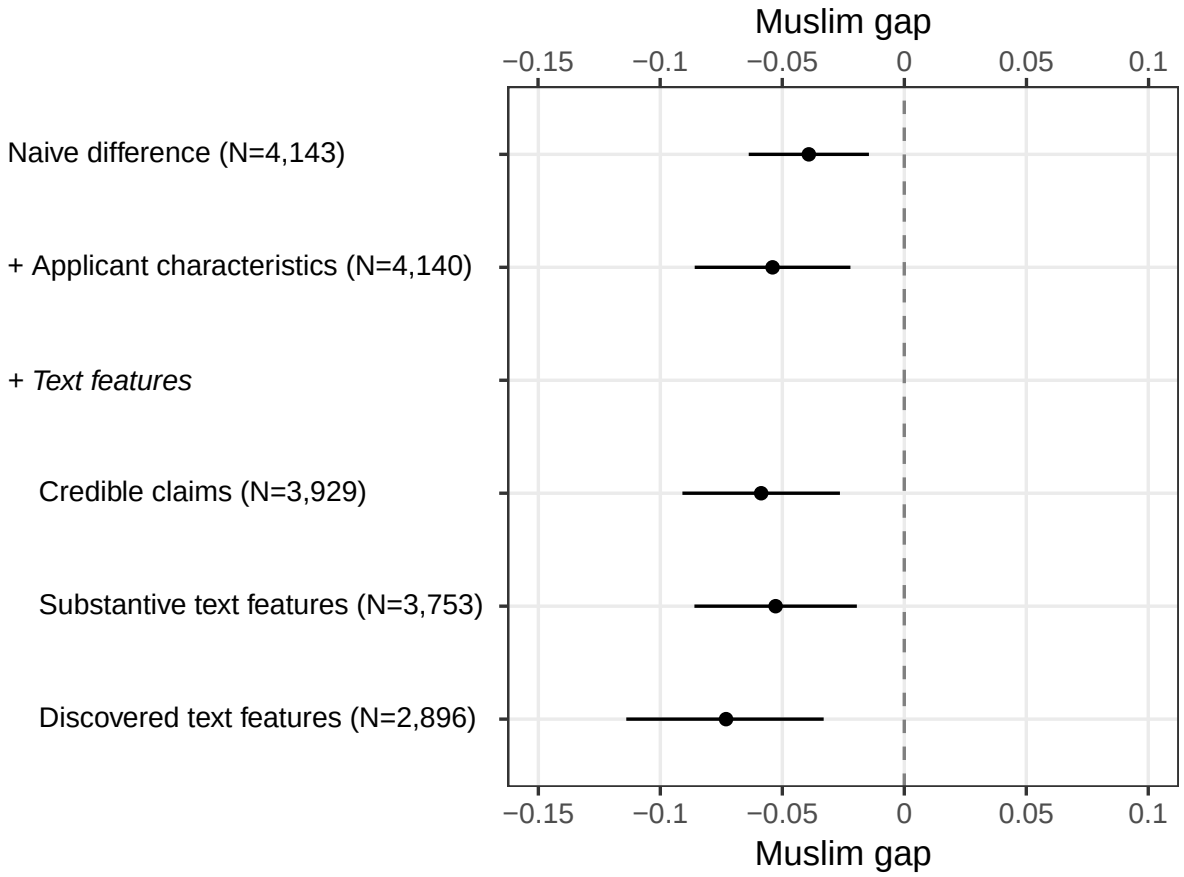
Figure C.7: Compare applications examined by experienced and inexperienced bureaucrats



Note: This figure plots, for a selection of individual characteristics, the standardized difference (difference in means divided by the standard error of the difference) between two groups of applications: Applications examined by inexperienced bureaucrats (those who decided less than two hundred applications) and applications examined by experienced bureaucrats (those who already decided between 200 and 500 applications.

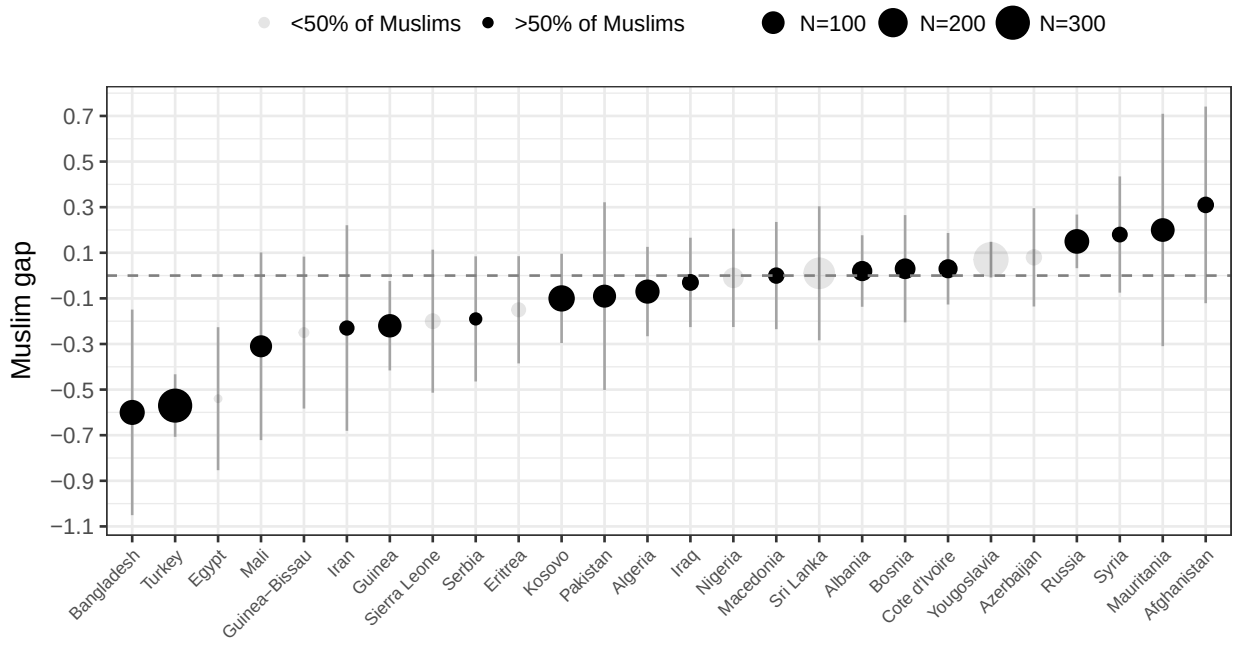
C.3 Additional analysis

Figure C.8: Robustness: Alternative strategies for controlling for the effect of the personal narrative



Note: This figure displays five different estimates of the Muslim penalty, along with their 95 percent confidence intervals. The first one is the naive difference, that is the simple difference in means between the acceptance rate among Christians and Muslims. The second is the estimate of the Muslim penalty only controlling for applicant characteristics. The next three are estimates of the Muslim penalty controlling for the personal narrative in three different ways: using the credibility measure, using the substantive text features listed in Table B.5, and controlling for the discovered text features using the supervised Indian Buffet process with 8 treatments (Fong and Grimmer 2016).

Figure C.9: Robustness: Muslim gap by country of origin



Note: This figure displays estimates of the marginal effect of being a Muslim within country, along with 95 percent confidence intervals from a weighted OLS regression including fixed effects for year of application, country of origin, and all interactions between country of origin and the religion of the applicant. The size of the dots is proportional to the size of the population from this particular country in the sample, and the shade of gray indicates whether this is a majority Muslim (Dark) or Christian (Light) in the sample. Only countries with at least 4 Christians and 4 Muslims are displayed.