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

Mathilde Emeriau[†]

August 28, 2020

What determines whether some asylum seekers are granted refugee status while others are rejected? I draw upon archival records from a representative sample of 4,142 asylum applications filed in France between 1976 and 2016 to provide new evidence on the determinants of asylum decisions. I find that applicants who are Christian (rather than Muslim), have higher skill levels or have higher educational attainment are more likely to be granted refugee status, controlling for all other individual characteristics available to the asylum officers making the decisions. However, linking archival records to detailed administrative data, I also show that bureaucrats at the French asylum office stop discriminating after about a year on the job. This pattern, I argue, is most consistent with inaccurate statistical discrimination. These findings have implications for strategies to curtail discrimination in courtrooms and administrations.

Word count: 9,798

1 Introduction

Over three million refugees arrived in Europe between 2015 and 2019,[1] triggering an unprecedented political crisis in the European Union. Anti-immigrant parties campaigning on exclusionary policies are gaining ground across Europe. Several European countries reintroduced border controls, leading to the de-facto suspension of the open border Schengen area. At the root of this crisis is the conflict among countries over the distribution and processing of the influx of asylum seekers. All European Union member states are bound by the 1951 Geneva Convention on Refugees, but despite decades of efforts to harmonize the European asylum system, member states continue to handle asylum applications very differently.[2]

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 reasons of "race, religion, nationality, membership of a particular group or political opinion."[3] For asylum seekers, these decisions are highly consequential: Those granted refugee status will be able to stay in Europe, while those not given refugee status will be sent back. However, the subjectivity of the definition and the possibility that applicants are falsifying claims gives significant discretion to judges and bureaucrats. How then do asylum officers decide? Is asylum granted on the basis of the credibility of the claims of persecution, 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. In this study, I overcome previous data limitations by taking advantage of an unprecedented effort in transparency from the French asylum office.[4] By opening its archives in 2009, the French asylum office made close to 1.5 million asylum applications available to researchers. Included was the full extent of the office's administrative archives, as well as hours of filmed interviews with some of the main bureaucrats in the French asylum system. In total, I digitized 4,142 asylum applications, a representative sample of all applications filed at the French asylum office between 1976 and 2016.[5] I collected

---

[1]Eurostat. "Asylum applications (non-EU) in the EU-27 Member States, 2008-2019."

[2]European Commission. (2016). Proposal for a Regulation of the European Parliament and of the Council Establishing a Common Procedure for International Protection in the Union and Repealing Directive 2013/32/EU.

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

[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, the French asylum office and the London School of Economics. As part of this agreement, the names of applicants and their family members have been redacted to ensure the anonymity of asylum seekers and refugees.

all information recorded in the application form (e.g., age, gender, family situation, education, employment and religion) and transcribed applicants' personal narratives, in which they explain the reasons they need political asylum. Crucially, for each application, I also know the decision and the anonymized identifier of the bureaucrat who decided the case.

By comparing accepted and rejected applicants, I analyze how individual characteristics affect the probability of getting asylum in France, holding constant the country of origin and the year of arrival. To control for the effect of the personal narrative, I employ three different strategies. First, I combine hand-coding and supervised learning to measure and predict whether the narrative is establishing credible claims of persecutions. Second, I develop a list of substantive features of the text that *could* be relevant to the decision. Third, I apply the supervised Indian Buffet Process to a random subset of the data to discover the text features that *are* most relevant in explaining the decision, which I then predict in the rest of the sample (Fong and Grimmer 2016). Further, concerns over omitted variable bias are alleviated by the fact that I had access to the entire case file, which was all the bureaucrats themselves saw when deciding an applicant's refugee status.

This study yields two main findings. First, I demonstrate evidence of discrimination in the attribution of refugee status in France from 1976 on. While asylum decisions should only be based on the assessment of the claims of persecutions, I find instead that Christians are substantially more likely to be granted refugee status compared to Muslims. So too are those with graduate education, compared to those with secondary and primary education, and those with high skill levels, compared to those with medium and low skill sets. I show that these differences cannot be explained by differences in other individual characteristics or by differences in their personal narratives. Moreover, these gaps are not driven by differences in unobservable characteristics revealed during the interview process or by a selective assignment of cases to bureaucrats.

Second, I show that French bureaucrats stop discriminating after about a year on the job. Indeed, Muslims, compared to Christians, are much more likely to be discriminated against when their case is examined by an inexperienced bureaucrat than by an experienced one, and the pattern is similar for educational attainment and skill level. This suggests that bureaucrats learn on the job not to discriminate. One explanation for this pattern is that narratives are uninformative at first, but bureaucrats hone their ability to identify credible narratives over time. For instance, as they read narratives over the course of their tenure, they could become better able to identify stereotyped language and cookie-cutter testimonies. I argue that this pattern is most consistent with inaccurate statistical discrimination, in which discrimination results from bureaucrats' inaccurate beliefs about the persecution faced by different groups (Bohren et al. 2019).

This study makes three main contributions. First, it contributes to the "refugee roulette" literature by analyzing the first micro-level dataset on a representative sample of asylum applications.[6] Using country-level data, a number of studies find that both humanitarian and strategic interests explain variation in acceptance rates (Holzer et al. 2000; Rosenblum and Salehyan 2004; Salehyan and Rosenblum 2008; Fariss and Rottman 2009; Rottman et al. 2009; Keith et al. 2013; Neumayer 2005; Schneider and Holzer 2002; Toshkov 2014). However, these country-level analyses are merely suggestive since they do not take into account the possibility that variation in the composition of asylum seekers between countries of origin could confound these results.

Second, this study broadens our understanding of discrimination in bureaucracies. It adds to a growing number of empirical studies showing that bureaucrats discriminate on the basis of ethnicity (White et al. 2015; Butler and Broockman 2011; McClendon 2016; Hemker and Rink 2017; Neggers 2017). Most notably, by linking self-reported religion with administrative decisions, a first in the French context, this dataset provides a unique opportunity to study bureaucratic discrimination on the basis of religion in France using real-world data. Moreover, this study contributes to a recent literature calling for examining a more careful examination of inaccurate statistical discrimination, an often overlooked source of discrimination (Bohren et al. 2019).

Third, this study provides credible evidence that on-the-job experience can mitigate discrimination. Arnold et al. (2017) uncover a similar pattern comparing discrimination by bail judges across different courts in the U.S., with different average levels of experience. By showing that experience mitigates discrimination within a single administration, this study substantially enhances the credibility of this finding.

2 Asylum process in France

To apply for refugee status in France, asylum seekers first need to fill out a standardized application form with self-reported demographics and socio-economic information and provide a personal narrative in which they describe, in French, their motives for seeking political asylum. Upon submission to the French asylum office in Paris, the application, is sent to the relevant division and assigned to a bureaucrat. He first reads the application, decides whether to interview the applicant or not (starting in 2006, interviews became mandatory) and proposes a decision to grant or deny refugee status to his supervisor, who makes the

---

[6]Holzer et al. (2000) is an exception since they control for some individual-level characteristics in their analyses, but these are limited to the age, gender and marital status of applicants. Keith and Holmes (2009) analyze individual-level data but for a non-representative subset of asylum applications.

final decision. Those who are granted asylum receive a 10-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),[7] where asylum cases are reexamined by a three-judge panel with a public audience. Asylum seekers whose appeal is rejected can resubmit an application to the French asylum office if they have new information to provide.

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." In 1998, France introduced another form of protection called territorial asylum, which was replaced in 2003 by the subsidiary protection. Subsidiary protection is granted to those who do not meet the Geneva Convention's definition but who face the death penalty, torture or indiscriminate violence in the context of an internal or international armed conflict.

Bureaucrats working at the French asylum office have substantial discretion when deciding who gets refugee status. In addition to the subjectivity of the definition in the Geneva Convention, 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 Republic of Zaire and countries in South East Asia, bureaucrats at the French asylum office have been advised to use caution in assessing claims of persecution. As a result, deciding whether an applicant meets the definition set by the Geneva Convention is a nebulous process. The following quote from a filmed interview of Jean-François Terral, director of the French asylum office between 1996 and 2000 illustrates the challenge involved in making asylum decisions: "You are gold diggers. There is a 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."

Moreover, asylum officers often lack the time, space, training and documentation needed to make informed decisions. In 1978, the activity report depicts the disastrous working conditions of asylum officers: "Four officers usually worked in a 12 square meter office with two typists typing, while the other two officers try to assist asylum seekers in the remaining space." While working conditions have improved over time, asylum officers report working under very stringent time constraints consistently until today. In 1986, the activity reports lead to a back-of-the-envelope calculation that asylum officers decide on average 3.5 cases per day, twice as many as in Germany at the time. As recently as 2013, an asylum officer

---

[7]CNDA was known as the Commission des Recours des Réfugiés (CRR) until January 2008.

described how the stringent time constraints and the lack of documentation significantly impaired her ability to "discover *the truth*" (Aho Nienne 2013).

In addition, working at the French asylum office has historically been very precarious employment. From the beginning, the French 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 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 office again hired temporary workers to deal with a sudden increase in the number of first-time asylum applications. In many aspects then, bureaucrats at the French asylum office resemble the street-level bureaucrats described by Michael Lispky in his research dating back to 1980. Working under significant time constraints, street-level bureaucrats are often required to develop "routines of practice and psychologically [simplify] their clientele and environment," (Lipsky 2010, xii), a process that, he argues, can give rise to favoritism and stereotyping. While there is evidence from multiple contexts that bureaucrats in various capacities discriminate on the basis of race and ethnicity, these studies overwhelmingly rely on experimental methods (White et al. 2015; Butler and Broockman 2011; McClendon 2016; Hemker and Rink 2017), with very few studies identifying discrimination using real-world data (Neggers 2017). In this study, I leverage unique access to archival records from France to study discrimination (i.e., unequal treatment) on the basis of religion, education, skill level and proficiency in French of otherwise equal applicants.

3   Sources of discrimination

Economists distinguish between two sources of discrimination to explain group-based differential treatment: taste-based discrimination (Becker 1971) and statistical discrimination (Phelps 1972; Arrow 1973). In the case of taste-based discrimination, differential treatment arises because decision-makers derive utility from favoring one group over another. By contrast, statistical discrimination stems from decision-makers' beliefs about the group distributions of the relevant outcome. Bohren et al. (2019) further distinguish between cases where belief-based discrimination is based on accurate beliefs (which they label *accurate statistical discrimination*) or inaccurate beliefs (*inaccurate statistical discrimination*).

I start by outlining how these different sources of discrimination might be at play at the French asylum office. The set up is the following: A bureaucrat decides whether to grant refugee status (decision v) to an asylum seeker who has observable characteristics $g =\{$A,

B} and faces persecution ω. In this study, I examine how different asylum seekers' characteristics g (religion, skill level, educational attainment and proficiency in French) affect their probability of getting asylum in France. Since the attribution of refugee status should only be based on the assessment of the credibility of the claims of persecution, these characteristics should not be relevant to the decision.

The bureaucrat does not observe the level of persecution ω that the asylum seeker faces at home. Instead, he observes the narrative written by the asylum seeker. The credibility of the narrative, $q = \omega + \epsilon$, is related to the level of persecution ω, but with some noise ϵ introduced by the fact that the applicant needs to write about her persecution in French and that the asylum seeker can get help from family, friends and volunteers from migrants' associations. For simplicity, I assume $\epsilon \sim N(0, \sigma_\epsilon^2)$ is an independent random shock. Upon reading the narrative, the bureaucrat extracts a noisy signal of the credibility of the narrative, $s = q + \eta$. The noise η comes from the fact that the bureaucrat may only imperfectly be able to assess the credibility of the narrative. I model the bureaucrat's ability as a random noise drawn from a normal distribution with mean 0 and variance σ_η^2.

The bureaucrat receives $-(v - (\omega - c_g))^2$ as payoff from decision v for asylum seeker with characteristic g and persecution ω. The introduction of a taste parameter, c_g, in the bureaucrat's payoff function allows for the possibility that the bureaucrat derives utility from favoring one group over another. Normalizing $c_A = 0$, a bureaucrat has a taste-based partiality for group B if $c_B \geq 0$. In that case, the bureaucrat has higher standards for asylum seekers from group B than for those from group A, even when they face similar levels of persecution. This opens up the possibility of taste-based discrimination.

The second source of discrimination, belief-based discrimination, arises instead from the way bureaucrats reach decision v. I assume that bureaucrats have prior beliefs about the level of persecution faced by group g, which are distributed according to the normal distribution $N(\hat{\mu}_g, \sigma_\omega^2)$. A bureaucrat updates his beliefs about the level of persecution faced by the asylum seeker upon reading the narrative. The bureaucrat has belief-based partiality toward group A if he believes that asylum seekers from group A are more persecuted than asylum seekers from group B ($\hat{\mu}_A \geq \hat{\mu}_B$). These beliefs can be accurate, in which case they are equal to the true mean μ_g, but they can also be inaccurate ($\hat{\mu}_g \neq \mu_g$).

The decision $v(s, g)$ that maximizes the bureaucrat's payoff is

$$v(s, g) = \frac{\sigma_\omega^2}{\sigma_\omega^2 + \sigma_\epsilon^2 + \sigma_\eta^2} s + \frac{\sigma_\epsilon^2 + \sigma_\eta^2}{\sigma_\omega^2 + \sigma_\epsilon^2 + \sigma_\eta^2} \hat{\mu}_g - c_g \tag{1}$$

Equation (1) provides an important insight for the empirical strategy. To estimate discrimination in the attribution of refugee status, it is important to control for the signal s

that the bureaucrat receives to distinguish between the signal effect and the discrimination effect. In practice, however, I can only estimate q, the credibility of the narrative, since I do not observe η. Moreover, the estimation inevitably introduces some noise λ. As a result, I control for $s' = q + \lambda$. However, as long as λ and η are uncorrelated with group characteristics, I can recover the effect of individual characteristics on the decision. I return to this in Section 5.

Discrimination occurs when two asylum seekers with the same signal, one from group A and one from group B, receive different decisions. Assuming $c_A = 0$,

$$D(s) = v(s, B) - v(s, A) = \frac{\sigma_\epsilon^2 + \sigma_\eta^2}{\sigma_\omega^2 + \sigma_\epsilon^2 + \sigma_\eta^2}(\hat{\mu}_B - \hat{\mu}_A) - c_B \tag{2}$$

Equation (2) illustrates that disparate decisions can arise from two sources: taste-based discrimination, which results from bureaucrats' preference for group A ($c_B \geq 0$), and belief-based discrimination, which results from bureaucrat's beliefs that group A is more persecuted on average than group B ($(\hat{\mu}_B - \hat{\mu}_A) \leq 0$). As such, we cannot distinguish between these two sources in a static setting since both types of partiality could be present (Bohren et al. 2019). Linking archival records to detailed administrative data on the history of decisions made by bureaucrats allows me to adjudicate between competing mechanisms of discrimination, by taking advantage of the fact that both experienced and inexperienced bureaucrats are in my sample. Indeed, many bureaucrats indicate they learn on the job how to identify credible narratives. Reflecting on her time as an asylum officer, Aho Nienne (2013) reports: "My colleagues assured me that, with experience, I would acquire a gift that is crucial to our profession: intimate conviction. That indescribable feeling when an asylum seeker lies." 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, she continues, "At first you do not know what to look for, what to base [decisions] on" (Greslier 2007, p.8). The intuition is simple: It takes time to be able to single out truly authentic narratives, that is, narratives that distinguish themselves through original storytelling and language. As they read more and more applications, bureaucrats discover the universe of narratives, which allows them to identify cookie-cutter narratives. Testifying anonymously, a bureaucrat, working at the French asylum office, powerfully illustrates this intuition: "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). In short, with experience, bureaucrats become better able to distinguish fake narratives from

true narratives, increasing the precision — decreasing the variance σ_η^2 — of the signal they receive when reading narratives. From equation (1), we see that as σ_η^2 decreases, bureaucrats place more weight on the signal $\left(\frac{\sigma_\omega^2}{\sigma_\omega^2+\sigma_\epsilon^2+\sigma_\eta^2}\right)$, than on their prior beliefs $\left(\frac{\sigma_\epsilon^2+\sigma_\eta^2}{\sigma_\omega^2+\sigma_\epsilon^2+\sigma_\eta^2}\right)$.

With this assumption in mind, we can see from equation (2) that absent belief-based partiality ($\hat{\mu}_B = \hat{\mu}_A$), discrimination will not vary with bureaucrats' experience, since $D(s) = -c_B$. However, if bureaucrats have belief-based partiality ($\hat{\mu}_B \neq \hat{\mu}_A$), discrimination will reduce with bureaucrats' experience: $\frac{\sigma_\epsilon^2+\sigma_\eta^2}{\sigma_\omega^2+\sigma_\epsilon^2+\sigma_\eta^2}$ decreases when σ_η^2 decreases. As a result, I expect that experienced bureaucrats will discriminate less than inexperienced bureaucrats in the case of belief-based discrimination, but not in the case of taste-based discrimination.

4   Data

I digitized 4,142 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 (See Appendix A for more information on the sampling strategy and the construction of the weights.) The outcome of interest is whether an applicant was granted refugee status upon first examination at the French asylum office. On average, 15 percent of applicants in the sample were granted political asylum in France based on the Geneva Convention between 1976 and 2016 (Figure 1, grey bars). By construction, this rate is lower than the overall first-time acceptance rate at the French asylum office (Figure 1, dashed line bars) for two reasons. First, I exclude cases of family unification and resettled refugees from the sample because these cases are not evaluated exclusively on the basis of their application. Second, I focus on the decision to grant refugee status based on the Geneva Convention, which presents a unique feature crucial to the research design. Indeed, the application of the Geneva Convention in France is based solely on the examination of individual claims to persecution rather than on group belonging (Cohen 2000). This practice rules out the possibility that the observed group differences can be driven by the fact that bureaucrats know some groups to be less persecuted since this information should in theory not matter. By 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 also carries additional information as to whether she needs protection. In France, asylum cases are examined sequentially, such that only the ones that did not qualify for protection under the Geneva Convention are examined for subsidiary protection. I thus recode applicants who received refugee status under subsidiary protection as having been rejected under the Geneva Convention protection.

Figure 1: Acceptance rate at the French asylum office (1976-2016)
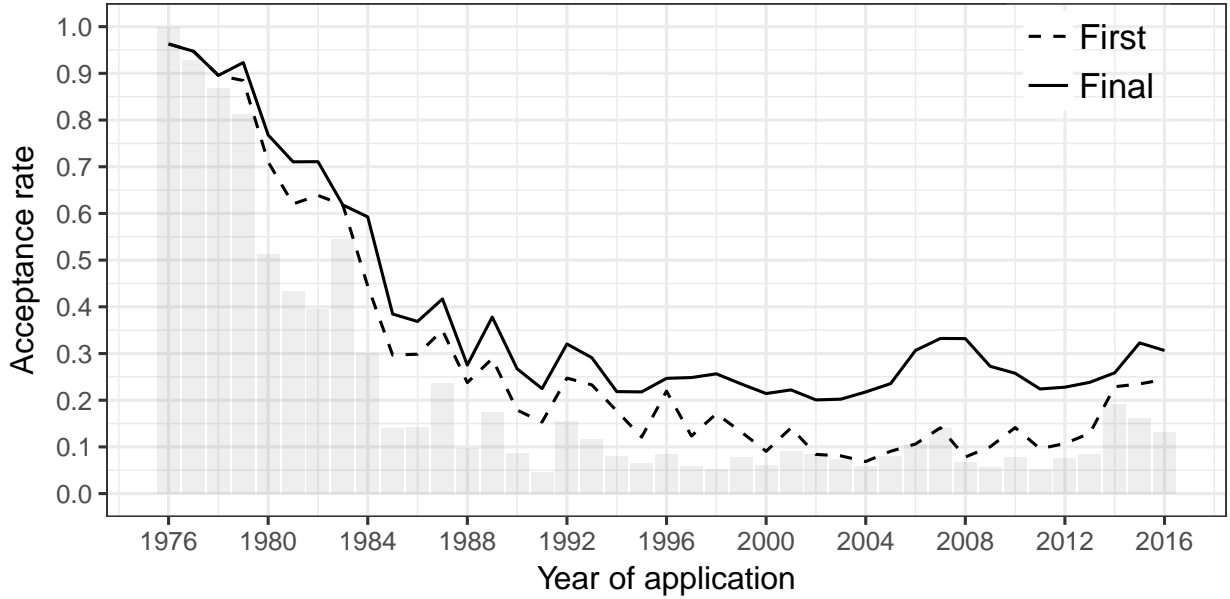
Notes: This figure displays the acceptance rate for applications filed at the French asylum office between 1976 and 2016. The dashed and solid line plots the acceptance rate on the basis of the Geneva Convention and the subsidiary protection combined after first examination at the French asylum office (dashed), and after examination by the appeals court and subsequent reexaminations by the French asylum office (solid). The gray bars plot the acceptance rate based on the Geneva Convention only for applicants in the sample. Source: Archival data.

The main independent variables used in the analysis are listed in Table 1. They fall under two broad categories. 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 are 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 recode the responses into categorical variables. I recode the religion of the applicant into the following categories: *Christian, Muslim, Other* or *None/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. I also coded 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 (6 percent) and those who

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 40	4,142	0.77	0.42	0	1
More than 40	4,142	0.14	0.34	0	1
Family Situation					
Single	4,141	0.58	0.49	0	1
Married	4,141	0.42	0.49	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					
Graduate	4,142	0.14	0.35	0	1
Secondary	4,142	0.49	0.50	0	1
Primary	4,142	0.11	0.32	0	1
None/Missing	4,142	0.25	0.43	0	1
Skill level					
High	4,142	0.17	0.37	0	1
Middle	4,142	0.40	0.49	0	1
Low	4,142	0.19	0.39	0	1
None/Missing	4,142	0.25	0.43	0	1
Speaks French					
Yes	4,142	0.27	0.44	0	1
No/Missing	4,142	0.73	0.44	0	1

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

did not answer the question (20 percent) were coded as *None/Missing*. Half of all applicants had started secondary education, and 14 percent reported starting a graduate degree. Using information on applicants' current occupations and last jobs in their country of origin, I recoded each question into *High*, *Middle* and *Low* levels of skills and used the highest of the two as a measure of skill level. Highly skilled jobs include academics, doctors, engineers, high-level executives, lawyers, journalists and students. 17 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 40 percent of applicants; drivers, hair dressers, sale clerks and homemakers are coded as *Low* (19 percent). 20 percent did not respond and are coded as *None/Missing*. 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 proficient in French if she listed French as a native or spoken language (27 percent of applicants did). For each application, I also transcribed the personal narrative, a bulk of which were handwritten. Of the 4,142 applications included in the final sample, 93 percent submitted a narrative in French.

Table 2: Summary statistics on French asylum office bureaucrats

	N	Median	Mean	Std. Dev.	Min	Max
# of years worked	601	3.00	3.80	4.08	0	16
Total # of decisions	601	296.00	440.52	489.90	1	3445
# of countries	601	12.00	12.96	8.42	1	44
Start Date						
Started in 2001	601	0.00	0.09	0.29	0	1
Started in 2002	601	0.00	0.04	0.20	0	1
Started in 2003	601	0.00	0.11	0.31	0	1
Started in 2004	601	0.00	0.05	0.22	0	1
Started in 2005	601	0.00	0.01	0.09	0	1
Started in 2006	601	0.00	0.02	0.13	0	1
Started in 2007	601	0.00	0.04	0.20	0	1
Started in 2008	601	0.00	0.07	0.26	0	1
Started in 2009	601	0.00	0.07	0.25	0	1
Started in 2010	601	0.00	0.04	0.20	0	1
Started in 2011	601	0.00	0.11	0.31	0	1
Started in 2012	601	0.00	0.06	0.24	0	1
Started in 2013	601	0.00	0.09	0.28	0	1
Started in 2014	601	0.00	0.05	0.22	0	1
Started in 2015	601	0.00	0.14	0.35	0	1

Notes: This table presents summary statistics on bureaucrats' employment at the French asylum office. Source: Administrative data.

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 extract some information about bureaucrats' case-loads using administrative data from

the French asylum office on about half of all asylum applications filed between 2001 and 2015. The French asylum office only records limited information about the case (the asylum seeker's nationality, demographic characteristics and the decision, and as well as the identity of the bureaucrat who decided the case). Thus, I restrict this dataset to the 232,679 asylum applications decided by bureaucrats who took their first decision between January 1, 2001 and December 31, 2015. For bureaucrats who made at least one decision in 2000, I cannot determine whether they started in 2000 or before, and I also excluded those who started after 2015, since I only have applications filed until the end of 2015. In this subset of the data, 601 unique bureaucrats starting working at the French asylum office between January 1, 2001 and December 31, 2015 (Table 2). For each of them, I can infer their start date using the date of their first decision and compute the total number of decisions made, as well as the order in which they were made. On average, bureaucrats decide 441 cases over the course of their tenure, work almost 4 years and handle cases from an average of 13 different countries.

5 Research design

This unique individual-level dataset allows me to estimate discrimination on the basis of religion, education, skill level and proficiency in French when bureaucrats determine refugee status in France. In practice, I regress an indicator variable for whether the applicant was granted asylum on these characteristics, while holding constant the applicant's age, gender, family situation, and time spent in France; the credibility of her narrative; the country of origin; and the year of application. One necessary condition for the estimation is overlap, which is often difficult with highly collinear variables like religion and country of origin. I check for overlap between religion and country of origin 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.2). 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.

The selection on observables assumption is well supported because I have access to the same information as the bureaucrats making the decisions. However, the fact that interviews are mandatory starting in 2006 raises the concern that the decision could be determined by unobservable characteristics revealed during the interview. Fortunately, close to 40 percent of applicants in the sample did not go through an interview, allowing me to check that results hold in this subsample. Moreover, the exact assignment mechanism of cases to bureaucrats is not known, but we have good reasons to believe that it is based on observable

characteristics. Cases are first dispatched into different divisions, each handling a different set of nationalities. The division head then assigns cases to bureaucrats on a monthly basis. Since this assignment happens before the interview takes place, we can assume that this assignment is based exclusively on observable characteristics. To completely alleviate this concern, I show that results hold when controlling for bureaucrat fixed effects.

As outlined in Section 3, the efficacy of this research design hinges on the possibility to control for the signal s that the bureaucrat receives from the narrative. Indeed, highly educated and highly skilled applicants and those who are proficient in French might provide better structured narratives that more easily convince bureaucrats that their claims of persecution are credible. As a first step, I explore how the narratives of Christian applicants, those who are highly educated, those who are highly skilled and those proficient in French differ, respectively, from the narratives of Muslim applicants, those who are less educated, those who are less skilled, and those not proficient in French. To do so, I extract the features of the narratives that, based on informal discussions I had with bureaucrats at the French asylum office, could relate substantively to the credibility of the narrative. Appendix Table B.1 provides summary statistics on these features. The number of words (including stop words) and the number of dates and locations mentioned (extracted using Stanford CoreNLP) are proxies for the level of detail of the narrative. Narratives vary greatly in length, but on average are 777 words long and mention 7 dates and 6 locations. To understand the extent to which the narrative is personal and individualized, I count the number of first person pronouns.[8] On average, narratives use first person pronouns 34 times. To control for the originality of the narrative, I also compute each narrative's average euclidean distance to other narratives of asylum seekers from the same country of origin. Moreover, to gain insight into the topics covered in the narratives, I estimate a structural topic model (STM) with 20 topics using country of origin as a covariate (Roberts et al. 2014). Descriptive patterns reveal the narratives differ significantly by religion, education, skill level and proficiency in French, emphasizing the necessity to control for the effect of the narrative on the decision (Appendix Table B.2, B.3, B.4 and B.5).

I use a combination of hand-coding and supervised learning to measure the credibility 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 in its claims of persecution. Three coders read eight percent of the narratives in the sample.[9] Importantly, they were not informed of the decision that was reached by

---

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

[9]Research assistants coded 350 unique narratives. Though relatively small, a total of 350 hand coded documents satisfies the criteria provided by the benchmark in Hopkins and King (2010) (between 100 and

the bureaucrat. For each narrative, they first coded whether the narrative mentioned any form of persecution, and, if so, for which of the five motives of the Geneva Convention (race, political opinion, religion, nationality or social group). Second, they assessed whether the claims were believable, convincing, detailed, individualized and coherent on a 4 point-scale. Third, they coded whether the narrative made reference to a historical event and whether the applicant mentioned members of her family who were already living in France. Finally, research assistants were asked to determine whether the applicant made a reasonable claim to political asylum, as defined by the Geneva Convention, by answering this question, *"In your opinion, is this person entitled to claim the right to asylum according to the Geneva Convention?"* {*No/Somewhat no/Somewhat yes/Yes*}. This last score, converted to a binary indicator, is what I refer to as the credibility of the narrative, which I use as a control in all specifications. On average, they answered yes 50 percent of the time, a proportion that is significantly higher than the average acceptance rate over the period (15 percent). Appendix Table B.6 provides summary statistics on all of the variables collected from this hand-coding.

Two elements inspire confidence that this variable measures the credibility of the narrative. 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, any two coders agree between 70 and 80 percent of the time, while coders coding at random would only agree 50 percent of the time. 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 a simulation, I find that random coders agree 25 percent of the time). Using this set of hand-coded narratives, I compare the performance of three classification algorithms to predict the credibility of the narrative on a left-out sample. 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 (see Appendix B for more information). Far from claiming that research assistants are better than civil servants at their jobs, or even that we should replace bureaucrats with computers, this procedure is an attempt at summarizing, in a single variable, important variation among applicants' narratives that could explain the decision to grant or withhold refugee status.

I use this measure to control for the credibility of the narrative, q. It is important to

---

500).

note here that I do not control for the level of persecution ω. This is not a concern because bureaucrats use a signal of persecution, not observation of persecution itself, to make the decision. However, the estimation of the credibility of the narrative inevitably introduces some noise, λ. As a result, the first key identification assumption is that the noise introduced by the estimation of the credibility of the narrative (λ) is uncorrelated with the characteristics of asylum seekers I examine (religion, education, skill level and proficiency in French). This would be problematic if some research assistants were more likely to find the same narrative more credible than others — for example, more likely to find the narrative credible if the asylum seeker is Christian than if she is Muslim. To mitigate this concern, I check the robustness of the results by controlling for the effect of the narratives in two additional ways: First, by directly controlling for features of the narratives that *should* be relevant in explaining the decision (the features described above). Second, I use the supervised Indian Buffet Process developed in Fong and Grimmer (2016) to discover the features of the text that *are* relevant in explaining the decision. The model takes a document term matrix as its input and learns from a training set (30 percent of the narratives chosen at random (N=1,180)), a set of latent binary features that are predictive of both the text and the outcome. I set the number of features to 8 and search over a range of parameters to select the model that ranked the highest on a quantitative measure of model fit (Fong and Grimmer 2016). Using this model, I infer the latent treatments for the test set (the remaining 70 percent of narratives), which I use as controls when predicting the decision. By using a split-sample design, this procedure solves the identification and estimation problems that arise from using the same documents for discovering treatments and estimating causal effects (Egami et al. 2018).

The second identifying assumption is that the bureaucrat's specific noise that proxies for his ability (η) — which I do not observe — is uncorrelated with the characteristics of asylum seekers I examine. This would be a concern if bureaucrats were better able to assess the credibility of a given narrative when the asylum seeker belongs to one group over another. The fact that bureaucrats specialize by nationality, and not by any of the characteristics that I examine, helps mitigate this concern for education, skill level and proficiency in French. This, however, is not implausible for religion, since this is also one of the motives of the Geneva Convention. To address this concern, I show that results on religion are robust to restricting the analysis to the sample to asylum seekers who did not claim persecution on the basis of religion.

6   Results

What are the determinants of asylum decisions? In Table 3, I analyze how the applicants' religion, educational attainment, skill level and proficiency in French affect asylum decisions. In addition to country of origin and year of application fixed effects, the basic specification in column 1 includes a limited set of controls: age, gender, family situation, time spent in France prior to applying for refugee status and the credibility of the narrative. This analysis reveals that religion, educational attainment and skill level are all important predictors of the decision. Muslim applicants are 5.9 percentage points less likely to get refugee status compared to Christians who are similar across all other characteristics. This represents a substantial difference (30 percent) given that the average acceptance rate is 15 percent over the period. The effect of educational attainment and skill level is large as well. Compared to those who reported starting graduate education, those who reported secondary or primary level education are, respectively, 3 and 7 percentage points less likely to be granted refugee status, though only the difference between primary and graduate education is significant. Those with middle, or low, skill level are both significantly much less likely (6 and 8 percentage points) to get refugee status. Compared to those who speak French, not speaking French has a small negative effect (3.5 percentage points) across all specifications, but the coefficient is significant only in the first specification.

To ensure that these results are not driven by omitted variables, I perform four robustness checks. First, I show that these results are robust to including additional variables and a more complex set of interactions. In column 2, I include the additional covariates reported in the application form: the number of children, whether the application was expedited, 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 her entry to the territory was legal (see Table C.1 for summary statistics on these additional variables). In column 3, I add country of origin-year of application interactions. Results for religion, education and skill level are robust to these additional controls, but the result for proficiency in French loses statistical significance.

Second, I show that these results are robust to alternative ways of controlling for the effect of the personal narrative. In columns 4 and 5, in addition to the credibility of the narrative, I control, respectively, for substantive text features (described in Table B.1) and for text features discovered by the supervised Indian Buffet Process in one-third of the sample. In this last specification, the coefficients are less precise, which could be a function of the smaller sample size. Overall, results are not sensitive to these alternative ways of controlling for the effect of the narrative.

Table 3: Determinants of the attribution of refugee status

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				1(Granted refugee Status)			
	Full Sample	Additional covariates	Country-year interaction	Substantive text features	Discovered text features	Not Interviewed	Bureaucrats fixed effects
Credible narrative	0.058***	0.058***	0.049***	0.025	0.063***	0.047**	0.063***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
> 1 year in France	0.003	0.009	0.011	0.011	0.006	0.016	-0.001
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Speaks French (Ref: Yes)							
No/Missing	-0.032*	-0.028	-0.014	-0.025	-0.037*	-0.025	-0.019
	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
Religion (Ref: Christian)							
Muslim	-0.059***	-0.047**	-0.058***	-0.053**	-0.051**	-0.077***	-0.055**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Other	0.061**	0.060**	0.080***	0.044	0.075**	-0.016	0.033
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)
None/Missing	-0.017	-0.022	-0.017	-0.007	-0.022	-0.047	-0.010
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
Education (Ref: High)							
Secondary	-0.028	-0.027	-0.024	-0.020	-0.017	-0.060*	-0.026
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Primary	-0.070***	-0.063**	-0.084***	-0.070**	-0.070**	-0.080**	-0.081***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)
None/Missing	-0.038*	-0.027	-0.029	-0.030	-0.031	-0.044	-0.048*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
Skill level (Ref: High)							
Middle	-0.063***	-0.063***	-0.041*	-0.051**	-0.059**	-0.045*	-0.053**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Low	-0.084***	-0.082***	-0.066***	-0.072***	-0.077***	-0.046	-0.069***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
None/Missing	-0.025	-0.024	-0.001	-0.019	-0.019	0.001	-0.011
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	3,930	3,930	3,930	3,754	2,772	1,481	3,930
R^2	0.359	0.377	0.610	0.384	0.334	0.640	0.491
Mean acceptance rate	0.152	0.152	0.152	0.152	0.152	0.176	0.152

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. This table displays the results of seven 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. All regressions include demographic characteristics (age, gender and family situation), an indicator variable for whether they spent more than one year in France before applying, country of origin and year of application fixed effects.

Third, 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 results are not driven by unobservable characteristics that became apparent during the interview process (Table 3, column 6). Finally, by showing that results hold after controlling for bureaucrat fixed effects, I reject the possibility that these differences result from a process through which

asylum applications submitted by Muslims, less educated, and less skilled applicants, are systematically assigned to stricter bureaucrats. This analysis is complicated by the fact that information about bureaucrats is incomplete for applications filed before 2000, both in the administrative database and in the paper applications I digitize. Overall, the identifier of the bureaucrat who decided the case is missing for 17 percent of the total sample and for 53 percent of applications filed in 1999 or earlier. To deal with these missing values, I binned all applications for which the identifier of the bureaucrats was missing into one additional "missing" category (Table 3, column 7).

I perform a last set of robustness checks to ensure that differences are not driven by the fact that asylum officers have access to information regarding the persecution faced by these different groups. Because this is highly unlikely in the case of education and skill level, I focus on the difference between Christians and Muslims. To rule out the possibility that Muslims are less likely to be granted refugee status because they are, in fact, less persecuted than Christians, I first show that results hold when I control for whether the applicant belongs to a minority religious group in her home country (Figure 2). The World Religion dataset from the Association of Religion Data Archives 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 she belongs to a religious group that comprises less than 20 percent of the population in her home country.[10]

Second, I show that differences between Christians and Muslims 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 than Muslims because they are more likely to be persecuted than Muslims, then the Muslim gap should not sustain in the subsample of applicants who did not claim persecution on religious grounds. Only 6 percent of narratives claim persecution on religious reasons according to the coding completed by research assistants on a representative sample of applicants (Table B.6), and the proportion of the text dedicated to the topic *Religion* discovered by the structural topic model exceed 20 percent in only 2 percent of the narratives (Table C.1). This suggests that this is unlikely to explain the full difference between Christians and Muslims. Identifying applicants who 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. To identify narratives that claim persecution on religious grounds (one of the five motives of the Geneva Convention), I read 292 narratives, a random sample of the 1,021 narratives that

---

[10]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.

Figure 2: Testing for the robustness of the difference between Christians and Muslims
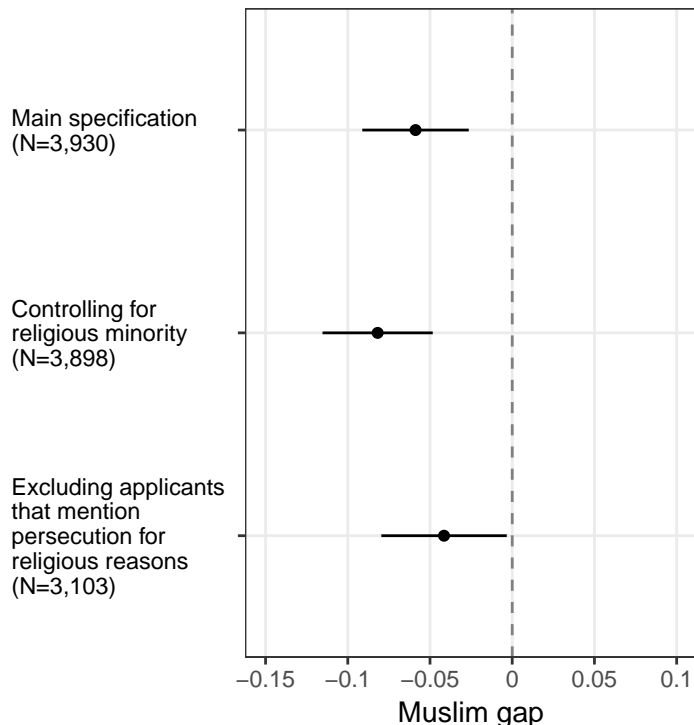
Notes: This figure displays estimates of the difference in the probability of getting asylum between Christian and Muslim applicants. At the top, this difference is estimated using specification 1 of Table 3. In the middle, the difference is estimated from the same specification to which I add, as a control, whether the applicant belongs to a religious minority in her country of origin. At the bottom, the difference is estimated among the subset of applicants who did not claim asylum on religious grounds.

contained at least one religious keyword, and for each I code whether the applicant is claiming 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 (Figure 2). Overall, these additional tests suggest that the difference between Christians and Muslims is unlikely to be explained entirely by the fact that Christians are more persecuted than Muslims.

In short, these results show that applicants who are Christian, highly educated and highly skilled are much more likely to be granted asylum than their respective Muslim, less educated and less skilled counterparts. Through a series of robustness tests, I rule out that these differences are explained by omitted variables, both observable and unobservable, by the selective assignment of cases to bureaucrats and by the fact that Muslims are less

persecuted than Christians.

7   Mechanisms

I now investigate whether these results are driven by the fact that bureaucrats derive utility from favoring Christians, highly educated and highly skilled applicants (taste-based discrimination) or by the fact that bureaucrats have priors beliefs about the persecution levels faced by these different groups (belief-based discrimination). As outlined in section 3, I expect bureaucrats' experience to mitigate discrimination if it is belief-based, but not if it is taste-based. For this analysis, I use the number of past decisions, which I extract from a representative sample of about half of all asylum applicants filed between 2001 and 2015, as a proxy for bureaucrats' experience. I restrict the sample to the 2,398 applications filed between 2000 and 2015 and for which a decision was reached between January 2001 and December 2015. In this sample, the number of past decisions ranges from 1 to 3,224, with a mean of 524 and a median of 355.

To examine the influence of bureaucrats' level of experience on discrimination, I follow Hainmueller et al. (2016)'s guidelines to estimate conditional marginal effects. I first divide the number of past decisions into three equal sized groups: a low (from 1 to 213), middle (from 214 to 582) and high number of past decisions (from 583 to 3224). The average bureaucrats' tenure length in the first bin is about a year (10 months), a little over 2 years (28 months) in the second bin and about 6 years in the third bin. I pick the median number of past decisions as the evaluation point within each bin and estimate a model that includes interactions between the indicator variables for each of the three bins, the individual characteristics and the number of past decision minus the evaluation point, as well as the triple interactions of these. I conduct this analysis separately for religion (comparing Muslims to Christians), education (comparing secondary and primary to graduate) and skill level (comparing middle and low to high).

In Figure 3, I plot the marginal effect of these three characteristics within each bin with the relevant confidence intervals. These analyses reveal a consistent pattern for all three characteristics: The marginal effect is the most negative in the lowest bin, and it is reduced in subsequent bins (becoming positive in the highest bin for education), suggesting that discrimination is most pronounced at lower levels of experience.

To ensure that this reduction in discrimination is not driven by the changing composition of bureaucrats over time, I further restrict the sample to the first 582 applications (second tercile of the moderator) of the 185 bureaucrats who decided at least 582 applications. Restricting the sample in this way ensures that the composition of the bureaucrats is similar

Figure 3: Learning effect: Comparing discrimination across different levels of bureaucrats' experience
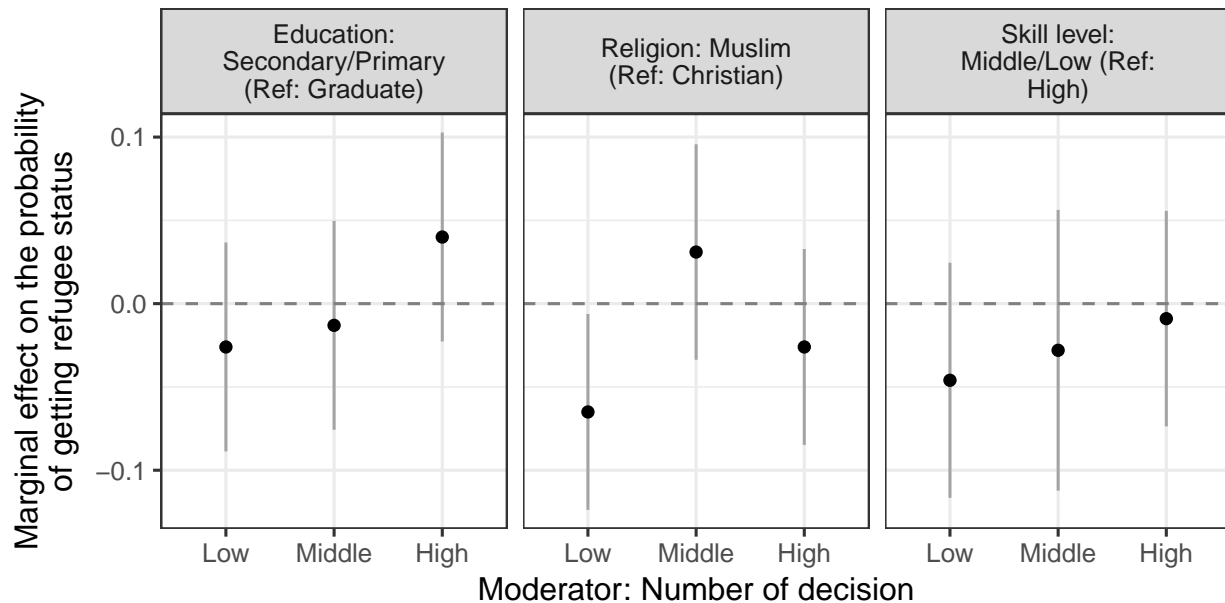
Notes: This figure shows the estimated conditional marginal effect, along with the 95 percent confidence intervals based on standard errors clustered by bureaucrats, from the binning estimator (Hainmueller et al. 2016). The moderator is the order in which the decision was made by each bureaucrat. The evaluation point is the median decision order within in each bin. The specification includes covariates and fixed effects for year of application and country of origin.

across the first two bins. I estimate the same specification as before, separately for each characteristic, both with and without bureaucrat fixed effects. Focusing on the second specification in Table 4, I find that Muslims are 12 percentage points less likely to receive 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: -0.042, standard error: 0.05), it improves Muslims' chances by 16.2 percentage points (standard error: 0.06), such that the difference between Christians and Muslims is small and insignificant among experienced bureaucrats. The pattern is similar for educational attainment and skill level (Table 5). Compared to highly educated and highly skilled, those with lower levels of education and skills are less likely to get refugee status when their case is examined by an inexperienced bureaucrat (the relevant coefficients are all negative but do not all reach statistical significance), but consistent with a reduction in discrimination among more experienced bureaucrats, the coefficients of the interactions terms are all positive (though only one reaches statistical significance). In

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

	1(Granted refugee Status)	
	(1)	(2)
More than 214	0.000	-0.042
	(0.03)	(0.05)
Religion (Ref: Christian)		
Muslim	-0.062	-0.120*
	(0.04)	(0.06)
Other	0.062	0.074
	(0.07)	(0.10)
None/Missing	0.032	0.009
	(0.09)	(0.11)
Interactions		
More than 214 × Muslim	0.109*	0.165**
	(0.04)	(0.06)
More than 214 × Other	-0.021	-0.019
	(0.08)	(0.11)
More than 214 × None/Missing	-0.053	-0.002
	(0.08)	(0.11)
Bureaucrats Fixed Effects	No	Yes
R^2	0.273	0.461
Observations	844	844

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$, $^*p < 0.1$. This table shows point estimates and standard errors clustered by bureaucrats in parentheses 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. The sample is restricted to the decisions to the first 582 decisions of bureaucrats who made at least 582 decisions. In Model 2, I add bureaucrat fixed effects.

Appendix Table C.2, I show that these two groups of applications do not differ systematically, allowing me to rule out that this result reflects the fact that bureaucrats are assigned different types of decisions over the course of their tenure.

These patterns suggest, consistent with the theoretical expectations of belief-based discrimination, that bureaucrats stop discriminating after about a year on the job. But results could also be driven by a reduction in bureaucrats' prejudice over time. Indeed, contact hypothesis proposes that personal contact across social lines can reduce prejudice (Allport 1974). However two elements suggest that this is unlikely to be the case here. First, the interview, which is the only point of contact between the bureaucrat and the asylum seeker,

Table 5: Learning effect: Bureaucrats stop discriminating after 200 decisions

	Education (Ref: Graduate)		Skill level (Ref: High)	
	(1)	(2)	(3)	(4)
More than 214	-0.014	-0.102	-0.025	-0.067
	(0.07)	(0.09)	(0.09)	(0.11)
Characteristics (Ref: High)				
Medium	-0.043	-0.053	-0.116	-0.112
	(0.06)	(0.07)	(0.07)	(0.09)
Low	-0.029	-0.069	-0.128	-0.180*
	(0.08)	(0.10)	(0.07)	(0.09)
None	-0.134*	-0.185*	-0.050	-0.065
	(0.06)	(0.09)	(0.07)	(0.09)
Interactions				
More than 214 × Medium	0.044	0.106	0.075	0.070
	(0.07)	(0.09)	(0.09)	(0.11)
More than 214 × Low	0.002	0.131	0.069	0.164
	(0.09)	(0.11)	(0.09)	(0.11)
More than 214 × None	0.132	0.233*	0.035	0.071
	(0.08)	(0.11)	(0.09)	(0.12)
Bureaucrats Fixed Effects	No	Yes	No	Yes
R^2	0.255	0.441	0.258	0.442
Observations	844	844	844	844

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. This table shows point estimates and standard errors clustered by bureaucrats in parentheses from OLS regressions with individual covariates and fixed effects for year of application and country of origin. The sample is restricted to the first 582 decisions of bureaucrats who made at least 582 decisions. In Models 2 and 4, I add bureaucrat fixed effects.

does not fit the traditional assumptions of contact theory: Contact should be cooperative, have a shared goal and be endorsed by communal authority, and participants should have equal power status. The interview quite strongly violates all of these, with the possible exception of communal authority. For instance, in his study (Lowe 2020) shows that while collaborative contact reduces own-caste favoritism in the context of cricket leagues, adversarial contact does not lead to a similar reduction in favoritism. Additionally, all bureaucrats in this study, both inexperienced and experienced, are in contact with applicants during the interview. Absent any variation in exposure to asylum seekers, contact theory is a very unlikely candidate to explain the observed reduction in discrimination.

8 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 level of transparency by the French asylum office, which opened its archives 10 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 of asylum decisions. For the first time, this dataset allows an in-depth examination of whether France is indeed granting asylum to those in need of protection. To do so, I compare accepted and rejected applicants in order to identify the effect of applicants' characteristics on the probability of being granted asylum in France.

This study provides empirical evidence that Christian (over Muslim), higher educated and higher skill level applicants are much more likely to be granted refugee status. Moreover, I show that these effects are driven neither by unobserved differences revealed during the interview process nor by a selective assignment of cases to bureaucrats. Importantly, this study also reveals that bureaucrats can self-correct their established discriminatory behaviors. Indeed, as bureaucrats get more on-the-job experience, they discriminate less on the basis of religion, education and skill level. These patterns, I argue, are most consistent with belief-based discrimination: The narrative provided by applicants is relatively uninformative at first, but bureaucrats learn on the job to assess credibility, allowing them to place more weight on the narrative than on their prior beliefs. Alternative explanations like taste-based discrimination find only weak support in the data. As such, this study has direct policy implications for strategies to reduce discrimination at the French asylum office. Its findings suggest that exposure to past narratives during bureaucrats' training could help bureaucrats better assess the credibility of persecution claims once they are faced with actual applicants. Moreover, increasing the tenure of bureaucrats would mechanically reduce the overall level of discrimination. This research also opens up a new avenue of study on the effectiveness of possible interventions to reduce the influence of bureaucrats' beliefs on decisions in order to curtail discrimination in administrations. These findings unambiguously call for increased scrutiny in the attribution of refugee status in France, and within the European Union more broadly. The Geneva Convention being the common framework for granting asylum across Europe, asylum officers elsewhere in Europe are afforded similar levels of discretion as bureaucrats at the French asylum office. However, we currently lack the data needed to compare working and training conditions across European asylum offices. Similar data collection in

other countries, either by researchers themselves or even by asylum offices directly, should be undertaken in order to determine whether the discrimination patterns identified in France apply more broadly.

References

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

Allport, G. (1974). The Nature of Prejudice.

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.

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

Bohren, J. A., K. Haggag, A. Imas, and D. G. Pope (2019). Inaccurate Statistical Discrimination. *SSRN Electronic Journal*.

Bohren, J. A., A. Imas, and M. Rosenberg (2019). The dynamics of discrimination: Theory and evidence. *American Economic Review 109*(10), 3395–3436.

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.

Cohen, D. G. (2000). Naissance d'une nation: les personnes déplacées de l'après-guerre, 1945-1951. *Genèses 38*, 56–78.

Egami, N., C. J. Fong, J. Grimmer, M. E. Roberts, and B. M. Stewart (2018). How to Make Causal Inferences Using Texts.

Fariss, C. J. and A. J. Rottman (2009). The Political Economy of Decision Making Bias: An Analysis of Asylum Merit Hearings in U.S. Immigration Court.

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.

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).

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

Hemker, J. and A. Rink (2017). Multiple Dimensions of Bureaucratic Discrimination: 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.

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.

Lipsky, M. (2010). *Street-Level Bureaucracy*.

Lowe, M. (2020). Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration.

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.

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

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.

Salehyan, I. and M. R. Rosenblum (2008). International Relations, Domestic Politics, and Asylum Admissions in the United States. *Political Research Quarterly 61*(1), 104–121.

Schneider, G. and T. Holzer (2002). *Asylpolitik auf Abwegen: nationalstaatliche und eu-*

ropäische Reaktionen auf die Globalisierung der Flüchtlingsströme.

Toshkov, D. D. (2014). The dynamic relationship between asylum applications and recognition rates in Europe (1987-2010). *European Union Politics 15*(2), 192–214.

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   Representativeness 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. In this section, I show that this list is indeed representative of the universe of asylum applications filed during this period. To do so, I use the 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 compile 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, the office only reported the number of refugees accepted, but the acceptance rate was almost 100 percent until 1976. Between 1970 and 1980, the French Asylum Office 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 (x-axis) to the number published in the activity reports (y-axis). The right panel displays the same data using a log scale.

Figure A.1: Total number of first-time applications by country of origin (1952-2014)
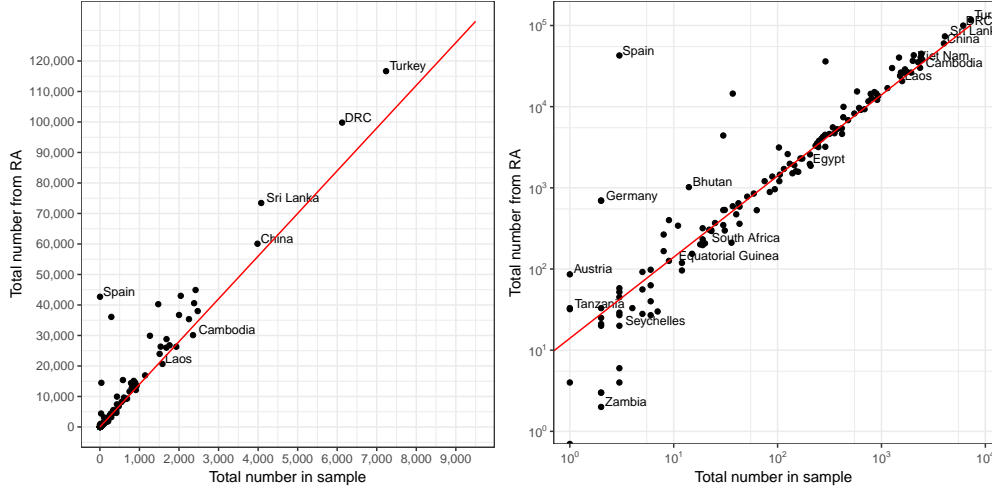
Notes: Each dot represents a country of origin, the x axis represents the number of applications from that country of origin in the administrative sample of 100,000 observations, and the y axis represents the number of applications in the official statistics. The same figure is displayed twice, using a linear scale on the left and using a log scale on the right.

2

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 used to select the folders for destruction is the following: Every six months, among the applications filed after 1989, all applications rejected more than 10 years earlier are selected. Qualitatively interesting folders are put aside, and 90 percent of remaining folders, selected at random, are destroyed.

A.3   Sampling Process

In this section, I describe the sampling of applications for in-depth data collection. The final sample was selected in three stages.

In December 2015, I sampled 10,000 folders for digitization, accounting for the non-random destruction of folders and the oversampling of applications filed after 1989. After excluding 3 perfect duplicates and 1,572 applications from small nationalities (I limited the sample to the top 74 countries of origin, out of 155, for which I had at least 100 applications in the administrative list of 100,000 observations), I split the data into 4 different subsets depending on the date of application (because no applications filed after 2005 have been destroyed) and the database. The French asylum office has used two non-overlapping databases to manage asylum applications since 1952: All applications filed since 1990 are stored in *Inerec*, the most recent database, and *Inerec* also stores some applications prior to 1990, for non-obvious reasons. *Garonor*, the oldest database, has all the remaining applications. This is relevant because for applications stored in *Garonor* we only know the nationality of the applicant with accuracy. Splitting the sample this way yields the following subsamples:

- 26,433 applications filed prior to 1989 from *Garonor*. From this subsample, I draw 1,292 folders.

- 14,259 applications filed prior to January 1, 1990 from *Inerec*. For these and all other subsets, I know the nationality, the date of application and whether the applicant received refugee status. Some of these have been destroyed. From this subsample, I draw 699 folders.

- 35,398 applications filed between January 1, 1990 and December 31, 2004 from *Inerec*. From this subsample, I draw 4,868 folders.

- 22,335 applications filed between January 1, 2005 and December 31, 2014 from *Inerec*. From this subsample, I draw 3,054 folders.

Within each of these subsamples, I first generate a new nationality variable that is equal to the nationality for the top 10 nationalities within each subsample, and I put all other nationalities into one bin. I then predict for each application the probability of being in the archives (which is one minus the probability of having been destroyed) using different linear models with a combination of the following variables: whether the applicant was granted refugee status, the year of application and the country of origin, as well as all possible interactions. For each model, I then draw several hundred samples weighted by the inverse of the probability of still being in the archives. I use these samples to choose the model that minimizes the average distance between the sample and the population on the proportion of nationality, year of application, the overall rejection rate, the yearly rejection rate, and the proportion of the top 10 nationalities. Finally, for each model, I select the sample that minimizes overall deviation.

For the reasons described in section A.2, the first iteration did not yield enough rejected applications for the period 1990-1992. 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. Finally, in June 2016, I requested an additional sample of 8,000 asylum applications filed in 2015 and 2016, and I sampled an additional 500 applications.

A.4 Data Collection

In total, with the help of a team of research assistants, I searched 5,421 applications and found 5,220. 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. Of these, 447 applications were excluded because they did not meet the criteria of being a first-time application with a decision and an application form in French. 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.

The application form changed a few time over the period but the questions remain similar. Only a few questions were removed. Among them were: "Are you planning on staying permanently in France?" and "Are you registered in the consulate of their country of origin?" The application form contains questions about the applicant's origin (country of origin, nationality and ethnicity, which was added in the 1990s), family situation and family members, conditions of arrival in France (itinerary, past countries of residence, modality of entry on French territory and whether entry was legal or illegal), religion, education and profession (in France and the last job in country of origin), languages spoken, military service (date, role and country), and official documents (passport or diplomatic laissez-passer).

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. After these initial checks, the error rate among 210 randomly selected folders for double data entry was less than 5 percent.

For this study, I restrict the sample to applications filed after 1976 for two reasons. 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 after 1951 (Figure C.1). Moreover, between 1952 and 1976, the acceptance rate at the French asylum office was close to 100 percent.

A.5 Weights

I generate weights to, first, adjust for the fact that the sampling correction was not perfect and, second, to correct for the oversampling of applications filed after 1989. 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. In order to match the total number of first-time applications, I then multiply these weights by 6.5 for observations prior to 1990, by 2 for observations between 1990 and 2015 included, and by 14.8 for observations for 2016.

B Measuring the credibility of the narrative

B.1 Features of the narratives

Table B.1: 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 pronouns	3,760	34.58	33.07	0	445
Originality					
Distance	3,927	47.20	23.49	6	347
Topic proportion >.2					
Family members	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

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

B.2 Comparing the narratives by religion, education, skill level and proficiency in French

In this section, I compare the narratives of Christians to the narratives of Muslims (Table B.2), the narratives of those with graduate education to those with secondary and primary education (Table B.3), the narratives of those who are highly skilled to those who have middle or low skill levels (Table B.4), and finally the narratives of those who are proficient in French to those who are not (Table B.5)

Table B.2: Differences between the narratives of Christian and Muslim applicants

	Christians			Muslims			t-test	
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Diff.	p
Level of detail								
Number of words	827.23	810.64	1,547	780.79	738.23	1,719	-46.45	0.09
# of dates mentioned	7.47	8.37	1,557	7.37	7.87	1,721	-0.10	0.72
# of places mentioned	6.85	8.47	1,557	6.68	7.16	1,721	-0.17	0.54
Personal								
Number of personal pronouns	35.25	34.86	1,474	36.20	32.82	1,648	0.95	0.43
Originality								
Distance	49.60	25.38	1,543	48.03	22.33	1,718	-1.57	0.06
Topics								
Family members	0.10	0.30	1,547	0.11	0.31	1,717	0.01	0.40
Bangladesh	0.00	0.05	1,547	0.08	0.27	1,717	0.08	0.00
Student protest	0.08	0.27	1,547	0.05	0.23	1,717	-0.03	0.00
Escape	0.08	0.26	1,547	0.12	0.33	1,717	0.05	0.00
Sri Lanka	0.02	0.14	1,547	0.01	0.07	1,717	-0.02	0.00
RDC	0.12	0.32	1,547	0.00	0.02	1,717	-0.11	0.00
Live in France	0.22	0.41	1,547	0.17	0.38	1,717	-0.05	0.00
Former Yugoslavia	0.05	0.23	1,547	0.14	0.35	1,717	0.09	0.00
China	0.02	0.13	1,547	0.01	0.08	1,717	-0.01	0.00
Ethnic minorities	0.10	0.30	1,547	0.03	0.17	1,717	-0.07	0.00
Kurds in Turkey	0.00	0.06	1,547	0.13	0.34	1,717	0.13	0.00
Political instability	0.10	0.30	1,547	0.02	0.15	1,717	-0.07	0.00
Political opposition	0.03	0.16	1,547	0.12	0.32	1,717	0.09	0.00
Court hearing	0.02	0.14	1,547	0.02	0.15	1,717	0.00	0.59
Narrative	0.17	0.38	1,547	0.15	0.36	1,717	-0.02	0.09
Angola	0.07	0.26	1,547	0.03	0.18	1,717	-0.04	0.00
Salutation	0.06	0.23	1,547	0.06	0.24	1,717	0.00	0.57
Family at risk	0.04	0.20	1,547	0.06	0.24	1,717	0.02	0.03
Zaire	0.06	0.23	1,547	0.01	0.07	1,717	-0.05	0.00
Religion	0.02	0.15	1,547	0.02	0.14	1,717	-0.00	0.48

Table B.3: Differences between the narratives of highly educated applicants and the narratives of applicants with secondary and primary education

	Graduate			Secondary and Primary			t-test	
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Diff.	p
Level of detail								
Number of words	1090.08	1007.51	594	808.46	743.66	2,579	-281.62	0.00
# of dates mentioned	10.35	10.68	596	7.70	8.10	2,602	-2.65	0.00
# of places mentioned	8.29	9.31	596	6.94	7.83	2,602	-1.35	0.00
Personal								
Number of personal pronouns	42.85	38.37	563	36.23	34.17	2,477	-6.63	0.00
Originality								
Distance	51.51	24.93	593	48.32	24.67	2,574	-3.19	0.00
Topics								
Family members	0.08	0.27	593	0.10	0.30	2,577	0.02	0.11
Bangladesh	0.05	0.23	593	0.04	0.19	2,577	-0.02	0.06
Student protest	0.17	0.37	593	0.05	0.22	2,577	-0.11	0.00
Escape	0.06	0.25	593	0.09	0.29	2,577	0.03	0.02
Sri Lanka	0.01	0.11	593	0.05	0.22	2,577	0.04	0.00
RDC	0.07	0.25	593	0.05	0.22	2,577	-0.02	0.06
Live in France	0.15	0.36	593	0.20	0.40	2,577	0.05	0.01
Former Yugoslavia	0.04	0.19	593	0.08	0.27	2,577	0.04	0.00
China	0.05	0.22	593	0.05	0.21	2,577	-0.00	0.78
Ethnic minorities	0.11	0.31	593	0.06	0.23	2,577	-0.05	0.00
Kurds in Turkey	0.02	0.14	593	0.08	0.27	2,577	0.06	0.00
Political instability	0.05	0.22	593	0.05	0.22	2,577	-0.00	0.79
Political opposition	0.04	0.19	593	0.05	0.22	2,577	0.01	0.13
Court hearing	0.04	0.19	593	0.02	0.14	2,577	-0.02	0.03
Narrative	0.15	0.36	593	0.16	0.36	2,577	0.00	0.77
Angola	0.03	0.18	593	0.04	0.21	2,577	0.01	0.23
Salutation	0.03	0.17	593	0.06	0.24	2,577	0.03	0.00
Family at risk	0.06	0.24	593	0.05	0.21	2,577	-0.01	0.22
Zaire	0.05	0.22	593	0.03	0.16	2,577	-0.02	0.00
Religion	0.02	0.14	593	0.02	0.14	2,577	0.00	0.98

Table B.4: Differences between the narratives of applicants who are highly skilled and the narratives of applicants who have middle and low levels of skills

	Highly skilled			Middle and Low			t-test	
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Diff.	p
Level of detail								
Number of words	894.88	909.00	655	778.73	765.78	2,508	-116.15	0.00
# of dates mentioned	8.25	9.45	666	7.33	8.02	2,528	-0.92	0.01
# of places mentioned	7.11	8.34	666	6.51	7.79	2,528	-0.60	0.08
Personal								
Number of personal pronouns	35.61	34.21	629	35.03	34.35	2,399	-0.58	0.70
Originality								
Distance	45.69	23.94	655	46.75	24.50	2,500	1.06	0.32
Topics								
Family members	0.06	0.23	654	0.08	0.27	2,506	0.02	0.06
Bangladesh	0.05	0.22	654	0.04	0.19	2,506	-0.01	0.12
Student protest	0.15	0.36	654	0.06	0.23	2,506	-0.10	0.00
Escape	0.05	0.21	654	0.10	0.30	2,506	0.05	0.00
Sri Lanka	0.05	0.21	654	0.03	0.18	2,506	-0.01	0.21
RDC	0.04	0.20	654	0.05	0.22	2,506	0.01	0.53
Live in France	0.25	0.43	654	0.22	0.42	2,506	-0.03	0.13
Former Yugoslavia	0.04	0.20	654	0.09	0.29	2,506	0.05	0.00
China	0.03	0.16	654	0.04	0.20	2,506	0.01	0.09
Ethnic minorities	0.07	0.25	654	0.06	0.23	2,506	-0.01	0.29
Kurds in Turkey	0.02	0.14	654	0.07	0.25	2,506	0.05	0.00
Political instability	0.06	0.24	654	0.05	0.22	2,506	-0.01	0.37
Political opposition	0.04	0.20	654	0.07	0.25	2,506	0.02	0.02
Court hearing	0.04	0.19	654	0.02	0.14	2,506	-0.02	0.01
Narrative	0.14	0.35	654	0.13	0.34	2,506	-0.01	0.53
Angola	0.04	0.20	654	0.06	0.24	2,506	0.02	0.07
Salutation	0.06	0.25	654	0.05	0.23	2,506	-0.01	0.27
Family at risk	0.04	0.19	654	0.04	0.20	2,506	0.01	0.50
Zaire	0.04	0.21	654	0.03	0.17	2,506	-0.01	0.08
Religion	0.02	0.12	654	0.02	0.14	2,506	0.00	0.48

Table B.5: Differences between the narratives of applicants who are proficient in French and those who are not

	Proficient in French			Not proficient in French			t-test	
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Diff.	p
Level of detail								
Number of words	719.44	841.77	1,225	777.86	721.22	3,127	58.42	0.02
# of dates mentioned	6.06	7.93	1,225	7.66	8.32	3,165	1.60	0.00
# of places mentioned	6.14	8.25	1,225	6.59	7.39	3,165	0.46	0.08
Personal								
Number of personal pronouns	30.86	34.79	1,169	34.63	31.98	2,988	3.77	0.00
Originality								
Distance	44.39	26.48	1,222	47.46	23.37	3,122	3.07	0.00
Topics								
Family members	0.06	0.23	1,224	0.11	0.31	3,124	0.05	0.00
Bangladesh	0.00	0.03	1,224	0.05	0.22	3,124	0.05	0.00
Student protest	0.10	0.29	1,224	0.06	0.23	3,124	-0.04	0.00
Escape	0.04	0.19	1,224	0.11	0.31	3,124	0.07	0.00
Sri Lanka	0.00	0.00	1,224	0.06	0.24	3,124	0.06	0.00
RDC	0.11	0.32	1,224	0.02	0.12	3,124	-0.10	0.00
Live in France	0.29	0.45	1,224	0.20	0.40	3,124	-0.09	0.00
Former Yugoslavia	0.03	0.17	1,224	0.11	0.31	3,124	0.08	0.00
China	0.00	0.04	1,224	0.06	0.23	3,124	0.05	0.00
Ethnic minorities	0.01	0.07	1,224	0.08	0.27	3,124	0.07	0.00
Kurds in Turkey	0.01	0.09	1,224	0.09	0.28	3,124	0.08	0.00
Political instability	0.10	0.30	1,224	0.03	0.18	3,124	-0.06	0.00
Political opposition	0.11	0.31	1,224	0.04	0.19	3,124	-0.07	0.00
Court hearing	0.02	0.15	1,224	0.02	0.14	3,124	-0.00	0.57
Narrative	0.16	0.36	1,224	0.14	0.35	3,124	-0.02	0.13
Angola	0.07	0.25	1,224	0.04	0.20	3,124	-0.02	0.00
Salutation	0.09	0.28	1,224	0.06	0.24	3,124	-0.03	0.00
Family at risk	0.03	0.18	1,224	0.05	0.22	3,124	0.02	0.01
Zaire	0.08	0.26	1,224	0.01	0.08	3,124	-0.07	0.00
Religion	0.01	0.08	1,224	0.02	0.16	3,124	0.02	0.00

B.3 Hand-coding the credibility of the narrative

Table B.6 provides summary statistics on the main variables collected during the hand coding of the narratives.

In Table B.7, I assess the degree of inter-coder agreement. In the first column, I report the proportion of narratives for which every combination 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 to 80 percent of the time, whereas random coders agree only 50 percent of the time. The Krippendorff α is .44.[11]

In Table B.8, I assess the internal validity of the hand-coding. Questions about the content and objective features of the text, like the level of detail and the cohesiveness of the text, were designed to prime coders to think fully about the different aspects of the text before rating a narrative's credibility and correlate strongly with the credibility measure. 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 using this principal component analysis, the prediction correlates strongly with the measure of credibility I obtain from hand-coding (.79).

---

[11]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.

Table B.6: 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
Narrative was coded as					
Believable	459	2.60	0.83	1	4
Convincing	459	2.31	0.82	1	4
Detailed	459	2.24	0.99	1	4
Individualized	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

Notes: 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.7: 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

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

Table B.8: 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

Notes: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. This table shows point estimates and standard errors in parentheses from a OLS regression of the main credibility variable collected in the narrative hand-coding of other leading questions focused on features of the text such as the level of detail. The unit of observation is narrative/coder.

B.4 Predicting the credibility of the narrative

Using the set of hand-coded narratives, I compare the performance of three classification algorithms to predict the credibility of the narrative 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 three classification algorithms (Random Forests, Gradient Boosted Trees and the Lasso) using the following procedure:

- 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 10-fold cross validation to tune the model and using the tf-idf of unigrams and additional narratives covariates as predictors (e.g., 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.9), as well as the calibration plots (Figure B.1).

All three algorithms perform relatively well on standard metrics (Table B.9), 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, Table B.10 shows that the predicted measure of credibility correlates as expected with extracted features of the text.

Table B.9: 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

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

Figure B.1: Calibration plots
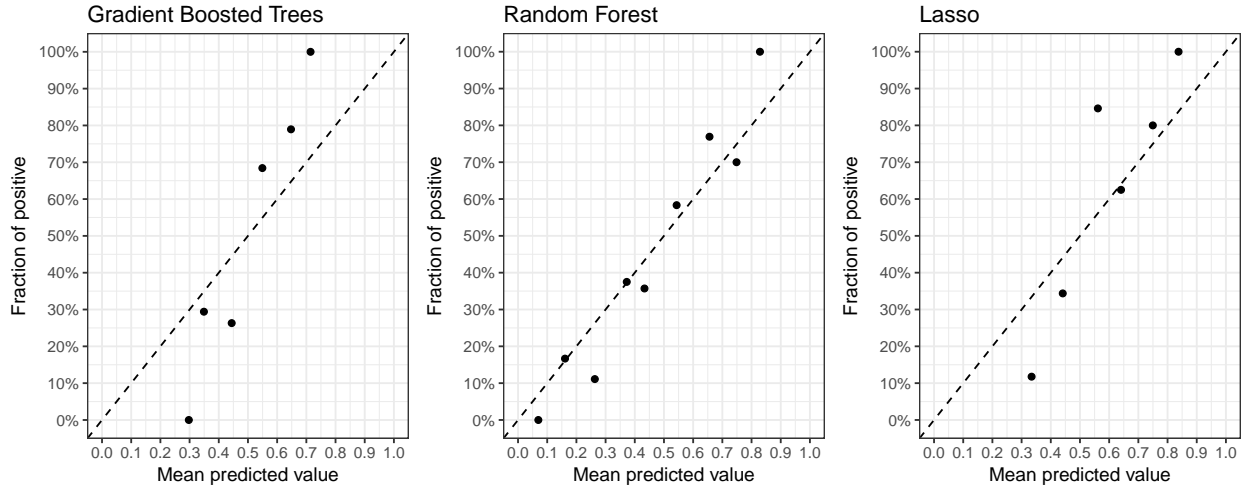
Notes: This figure plots the calibration of predicted probabilities from three different algorithms. In a left-out sample of 85 narratives, I predict the credibility for each narrative from each of the three algorithms after the tuning of the parameters. I then divide the predicted probabilities intp 10 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).

Table B.10: Correlates of credibility

	Credibility of the narrative	
	(1) Hand Coded	(2) 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 pronouns ('000)	-1.405	0.254
	(1.45)	(0.42)
Constant	0.274***	0.186***
	(0.05)	(0.02)
Observations	320	3472

Notes: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. This table displays the results of an OLS regression in the subsample of hand-coded narratives in column 1 and the full sample of narratives in column 2 (excluding hand coded narratives). The dependent variable is coded 1 if the predicted credibility was above 50 percent. No additional controls were included in this regression.

C Additional tables and figures

Figure C.1: Number of first-time applications filed at the French asylum office (1952-2016)
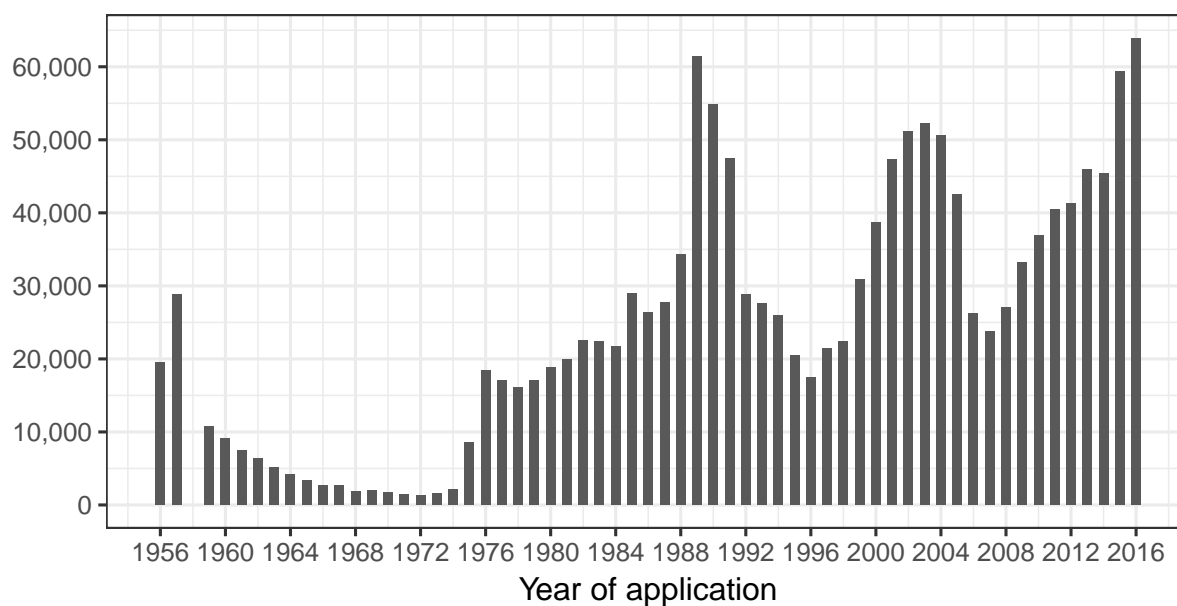
Notes: This figure displays the total number of first-time asylum applications filed at the French asylum office between 1956 and 2016, excluding unaccompanied minors. Source: Annual official statistics, French Asylum Office.

Figure C.2: Proportion of Muslim and Christian applicants by country of origin (1976-2016)
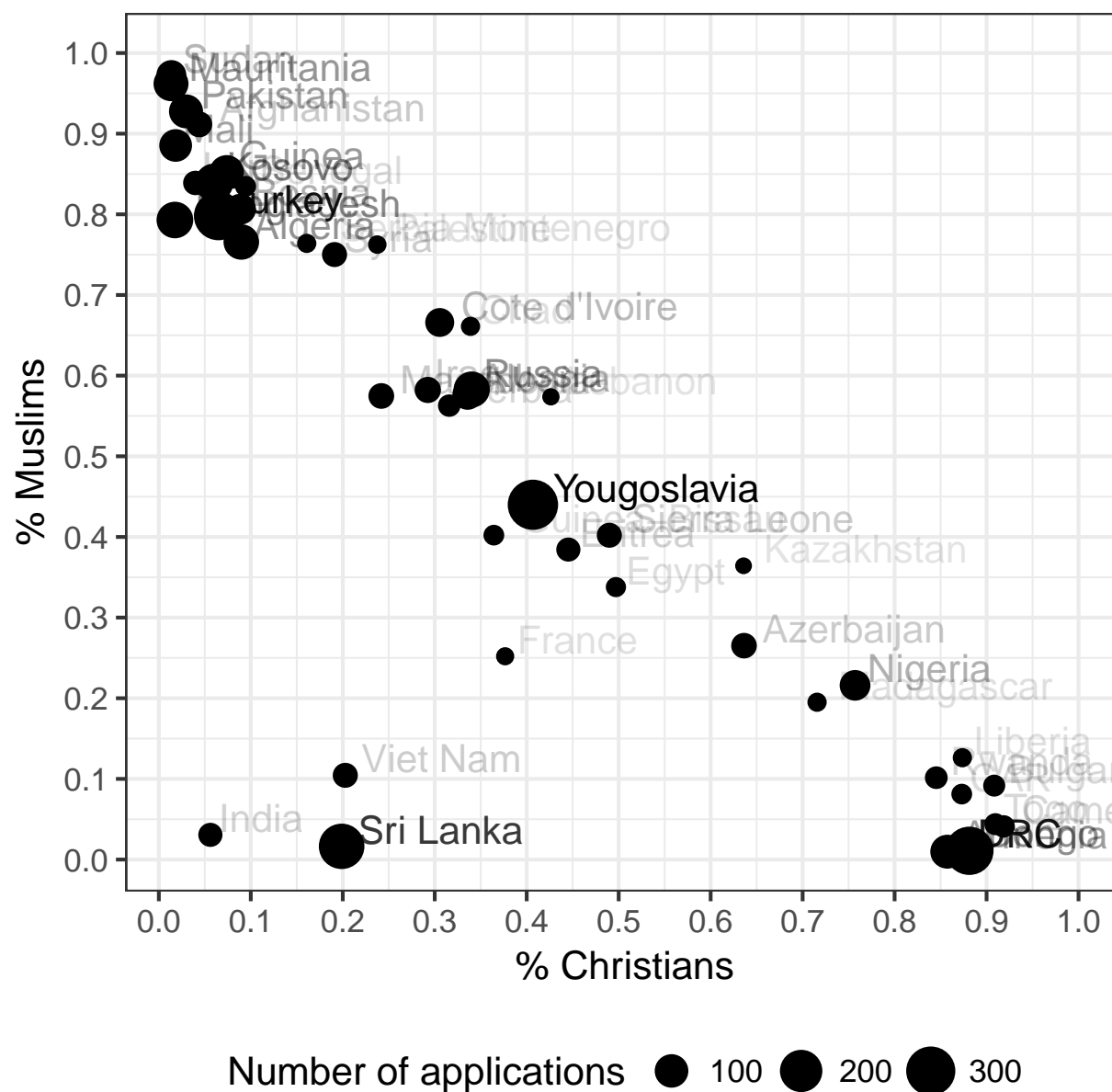
Notes: This figure plots, for each country of origin included in the sample, the proportion of the applicants from this country that are Christian (*x*-axis) and the proportion that are Muslim (*y*-axis). This figure shows, for example, that 75 percent of Turkish applicants are Muslims and 10 percent are Christian over the period under study (1976-2016).

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 refugee family member 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

Notes: 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 grounds that the country is deemed safe or that the 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 refugee family member 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.

Table C.2: Compare applications examined by experienced and inexperienced bureaucrats

	Inexperienced Bureaucrats		Experienced Bureaucrats		t-test	
	Mean	N	Mean	N	Diff.	p
Demographics						
Less than 20	0.08	292	0.05	583	0.02	0.24
Between 20 and 40	0.78	292	0.80	583	-0.01	0.65
More than 40	0.14	292	0.15	583	-0.01	0.78
Single	0.56	292	0.52	583	0.04	0.33
Married	0.44	292	0.48	583	-0.04	0.33
Education						
Graduate	0.14	292	0.14	583	0.00	0.84
Secondary	0.54	292	0.57	583	-0.04	0.30
Primary	0.10	292	0.13	583	-0.02	0.30
None/Missing	0.22	292	0.16	583	0.06	0.04
Religion						
Christian	0.40	292	0.36	583	0.04	0.31
Muslim	0.40	292	0.40	583	0.01	0.86
Other	0.08	292	0.12	583	-0.04	0.08
None/Missing	0.12	292	0.12	583	-0.00	0.93
Skill level						
High	0.13	292	0.10	583	0.02	0.29
Middle	0.34	292	0.39	583	-0.05	0.15
Low	0.22	292	0.20	583	0.02	0.52
None/Missing	0.31	292	0.30	583	0.01	0.81
Proficient in French						
Yes	0.24	292	0.22	583	0.02	0.54
No/Missing	0.76	292	0.78	583	-0.02	0.54
Narrative						
Provided a narrative	0.97	292	0.96	583	0.01	0.43
Credibility of the narrative	0.66	284	0.62	561	0.04	0.24
Number of words	850.24	284	888.01	561	-37.76	0.49
Distance	49.91	284	50.31	561	-0.40	0.80
Number of dates	7.99	285	8.75	563	-0.76	0.22
Number of location	7.93	285	7.81	563	0.12	0.85

Notes: This table presents, for a selection of individual characteristics, the difference in means between applications decided by inexperienced bureaucrats (less than 214 decisions), and experienced bureaucrats (more than 214 decisions) in the subsample of the first 582 applications decided by bureaucrats who decided at least 582 applications.