

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

Mathilde Emeriau<sup>†</sup>

July 31, 2021

## Abstract

What determines whether some asylum seekers are granted refugee status while others are rejected? I draw upon archival records from a representative sample of 4,142 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) 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. These findings have implications for strategies to curtail discrimination in courtrooms and administrations.

Word count: 9,825

---

\*I would like to thank Jens Hainmueller, Dominik Hangartner, David Laitin and Stephane Wolton for their invaluable guidance in designing and implementing this study. I am also grateful to Pascal Brice, Aline Angoustures, members of OFPRA's scientific committee and the French National Archives for granting me access to the archives of the French Asylum Office. Adelaide Choisset, Marie Gonzalez, Duncan Lawrence, Jean-Paul Levi and Corinne Pallastrelli provided the technical support that made the data collection possible. Catherine Anderson, Astrig Atamian, Guillaume Beaud, Diane Belkheraz, Charles Bena, Lucie Benaiteau, Cécile Bénoliel, Sonia Bilon, Lisa Cavalieri, Emmanuel Charreau, Thomas Emeriau, Elyne Etienne, Jeanne Huybrechts, Anna Kauffmann, Shushan Kerovpyan, Maxime Krumholz, Alice Ledoux, Marion Lotin, Mathilda Loussert, Charles Maillard, Camille Urvoy and Alice Zivohlava all provided excellent research assistance. This research is based upon work supported by National Science Foundation Grants No. 1624048 and 1627339 (Principal Investigator: David D. Laitin), the Stanford Immigration Policy Lab, The Europe Center and the France Stanford Center for Interdisciplinary Research. Kirk Bansak, Sarah Brierley, Darin Christensen, Jennifer Cryer, Jane Esberg, Vicky Fouka, Saad Gulzar, Justin Grimmer, Jeremy Weinstein and Yang-Yang Zhou all provided valuable comments.

<sup>†</sup>Department of Government, London School of Economics and Political Science, and Immigration Policy Lab, Stanford University, Stanford, CA 94305-6044, USA. m.emeriau@lse.ac.uk

# 1 Introduction

The arrival of over three million refugees in Europe between 2015 and 2019<sup>1</sup> triggered an unprecedented political crisis in the European Union (EU). 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. The conflict among countries over the distribution and processing of the influx of asylum seekers is at the root of this crisis. All EU member states are bound by the Convention Relating to the Status of Refugees (hereafter the Geneva Convention), but despite decades of efforts to harmonize the European asylum system, member states continue to handle asylum applications very differently.<sup>2</sup>

At the center of the asylum process, decision-makers in asylum courts and offices face the difficult task of determining whether asylum seekers are providing truthful and substantiated claims of persecution for reasons of “race, religion, nationality, membership of a particular group or political opinion.”<sup>3</sup> For asylum seekers, these decisions are very consequential: Those granted refugee status will be able to stay in Europe, while those who are not have to return to their country of origin. However, the subjectivity of the definition of persecution and the possibility that applicants are falsifying claims gives judges and bureaucrats significant discretion to determine the veracity of their claims. On what basis do asylum officers decide? Is asylum granted based on the credibility of the claims of persecution, or do bureaucrats discriminate in the attribution of refugee status?

The lack of micro-level data has thus 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 by the French Office for Refugee Protection and Stateless Persons (hereafter the French asylum office) to increase transparency. In 2009 the office

---

<sup>1</sup>Eurostat. “Asylum applications (non-EU) in the EU-27 Member States, 2008–2019.”

<sup>2</sup>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.

<sup>3</sup>Article 1(A)2 of the Geneva Convention.

opened its administrative archives and made nearly 1.5 million asylum applications available to researchers, which included hours of filmed interviews with some of the main actors in the French asylum system. I digitized a total of 4,142 asylum applications, a representative sample of all applications filed at the French asylum office between 1976 and 2016.<sup>4</sup> I collected all the information recorded on the application form (e.g., age, gender, marital status, education, employment and religion) and transcribed applicants’ personal narratives, in which they explain why 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 obtaining asylum in France, holding the country of origin and year of arrival constant. Since I had access to the same information as the bureaucrats who decided the cases, omitted variable bias is unlikely to be a concern. I employ three different strategies to control for the effect of the personal narrative. First, I combined hand-coding and supervised machine learning to measure and predict whether the narrative contains credible claims of persecution. Second, I developed a list of substantive features of the text that *could* be relevant to the decision. Third, I applied the supervised Indian Buffet Process to a random subset of the data to identify the text features that *are* the most relevant in explaining the decision, which I then predict in the rest of the sample (Fong and Grimmer 2016).

This study yields two main findings. First, I demonstrate evidence of discrimination in the attribution of refugee status in France since 1976. While asylum decisions should only be based on assessments of applicant’s claims that they will be persecuted if forced to return to their country of origin, I instead find that Christians are substantially more likely to be granted refugee status than Muslims. I show that these differences cannot be fully explained by variations in other individual characteristics or by differences in their personal narratives.

---

<sup>4</sup>The French National Archives granted me access on October 9, 2015 (derogation JM/2015/1504) to 10,000 applications filed up to December 31, 2014 and on August 30, 2017 to 500 additional applications filed in 2015 and 2016. The names of the applicants and their family members were redacted to ensure anonymity.

Moreover, these gaps are unlikely to be driven by differences in unobservable characteristics revealed during the interview process or by the selective assignment of particular cases to certain bureaucrats. Second, I show that French bureaucrats stop discriminating based on these characteristics after about a year on the job: Muslims are much more likely than Christians to be discriminated against when an inexperienced bureaucrat is assigned to their case than if an experienced case officer makes the decision. The pattern is similar, though less robust, for educational attainment and skill level. Overall, this suggests that bureaucrats learn on the job not to discriminate.

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.<sup>5</sup> Several studies have examined country-level data and concluded that both humanitarian and strategic interests explain variation in acceptance rates (Holzer, Schneider, and Widmer 2000; Rosenblum and Salehyan 2004; Salehyan and Rosenblum 2008; Fariss and Rottman 2009; Rottman, Fariss, and Poe 2009; Keith, Holmes, and Miller 2013; Neumayer 2005; Schneider and Holzer 2002; Toshkov 2014). However, these country-level analyses are merely suggestive since they do not account for the possibility that variation in the composition of asylum seekers between countries of origin could confound these results.

Second, this study provides credible evidence that on-the-job experience can mitigate discrimination using a fine-grained measure of experience. Arnold, Dobbie, and Yang (2017) uncover a similar pattern when comparing discrimination by bail judges across different courts in the U.S. who have different average levels of experience. Third, the study broadens our understanding of discrimination within bureaucracies. It adds to a growing number of empirical studies which demonstrate that bureaucrats discriminate on the basis of ethnicity and religion (Olsen, Kyhse-Andersen, and Moynihan 2020; White, Nathan, and Faller 2015; Butler and Broockman 2011; McClendon 2016; Hemker and Rink 2017; Neggers 2017). Most

---

<sup>5</sup>Holzer et al. (2000) control for some individual characteristics in their analyses, but these are limited to applicants’ age, gender and marital status. Keith and Holmes (2009) also analyze individual-level data but for a non-representative subset of asylum applications.

notably, by linking self-reported religious affiliation with administrative decisions, a first in the French context, this dataset provides a unique opportunity to study religion-based bureaucratic discrimination in France using real-world data.

## 2 Asylum process in France

To apply for refugee status in France, asylum seekers first need to fill out a standardized application form that elicits demographic and socio-economic information and to provide a personal narrative that describes, in French, their motives for seeking political asylum. Applications submitted to the French asylum office in Paris are directed to the relevant geographic division and assigned to a bureaucrat. Before 2006, this official decided whether to interview the applicant after reading her application; starting in 2006, interviews became mandatory. The bureaucrat advises his supervisor on whether to grant the applicant refugee status, and the supervisor makes the final decision. Those who obtain asylum receive a 10-year renewable residency permit (or a one-year residency permit in the case of subsidiary protection, described below). Those who are rejected can appeal this decision to the National Court of Asylum, a three-judge panel that publicly re-examines asylum cases. If their appeal is rejected, asylum seekers 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? The Geneva Convention require all signatories, including France, 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.”<sup>6</sup> In 1998, France introduced another form of protection called territorial asylum, which was replaced in 2003 by subsidiary protection, which is granted to those who do not meet the Geneva Convention’s definition of persecution but who would be subject to the death penalty, torture or indiscriminate violence in the context of an internal or international armed conflict in their country of origin.

---

<sup>6</sup>Article 1(A)2 of the Geneva Convention.

Bureaucrats working at the French asylum office have substantial discretion to decide who receives refugee status for two main reasons: (1) subjective interpretations of the Geneva Convention’ definition of persecution and (2) applicants may be economic migrants who are falsifying persecution claims. As early as 1970, the asylum office’s 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, when large-scale fraud by applicants from the Republic of Zaire and countries in South East Asia was revealed, bureaucrats at the French asylum office have been advised to use caution when assessing claims of persecution. A former director of the French asylum office (1996-2000) illustrated the challenge associated with 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. The 1978 activity report accessed for this study describes their disastrous working conditions: “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.” Information provided in the 1986 activity report suggests that asylum officers decided on average 3.5 cases per day, twice as many as their German counterparts at the time resolved. While their working conditions have improved over time, these bureaucrats still report working under very stringent time constraints. As recently as 2013, an asylum officer described how a lack of time and documentation significantly impaired her ability to “discover *the truth*” ([Aho Nienne 2013](#)).

Employment at the French asylum office has also been precarious. Until 1993 it had no permanent employees, relying mostly on temporary workers to deal with frequent fluctuations in the number of applications (SI Figure [D.1](#) (pp. 18)). In 1993 the unions successfully negotiated the conversion of temporary contracts into permanent positions. The share of temporary workers dropped radically afterward, but in the early 2000s the office hired more

to deal with a sudden uptick in applications.

In many respects, officials at the French asylum office resemble Michael Lipsky’s notion of street-level bureaucrats, who often respond to tight time constraints by developing “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, this research overwhelmingly relies on experimental methods (Olsen et al. 2020; White et al. 2015; Butler and Broockman 2011; McClendon 2016; Hemker and Rink 2017); very few studies have identified 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. Prior research has shown that citizens in Europe and the U.S. tend to prefer immigrants who are Christian rather than Muslim, high skilled over low skilled, more educated, and who are fluent in the host-country language (Adida, Laitin, and Valfort 2016; Helbling and Trautmüller 2018; Hainmueller and Hiscox 2007, 2010; Hainmueller and Hopkins 2015; Valentino, Soroka, Iyengar, Aalberg, Duch, Fraile, Hahn, Hansen, Harell, Helbling, Jackman, and Kobayashi 2017; Bansak, Hainmueller, and Hangartner 2016; Adida, Lo, and Platas 2019). Yet, these characteristics should not be relevant to decisions about whether to grant asylum, since such determinations should only be based on an assessment of the credibility of the claims of persecution.

### 3 Sources of discrimination

Economists distinguish between two sources of discrimination to explain group-based differential treatment. Differential treatment arises in *taste-based* discrimination because decision-makers derive utility from favoring one group over another (Becker 1971). *Statistical* discrim-

ination stems from decision-makers' beliefs about the distribution of the relevant outcome within a certain group (Phelps 1972; Arrow 1973). Bohren et al. (2019) further distinguish between cases in which the latter is based on accurate beliefs (which they label *accurate statistical discrimination*) or inaccurate beliefs (*inaccurate statistical discrimination*).

I examine how these different sources of discrimination might affect French asylum office decisions as follows. A bureaucrat evaluates the level of persecution faced by an asylum seeker who has observable characteristics (religion, skill level, educational attainment and proficiency in French)  $g = \{A, B\}$  and faces persecution  $\omega$ . The bureaucrat cannot directly observe the level of persecution  $\omega$  that the asylum seeker faces at home; he must rely on the applicant's written narrative. The credibility of the narrative,  $q = \omega + \epsilon$ , is related to the level of persecution  $\omega$ , but with some noise  $\epsilon$  introduced by (1) the fact that the applicant needs to write about her experiences in French (which is unlikely to be her native language) and (2) the fact that the asylum seeker can get help from family, friends and volunteers from migrants' associations. For simplicity, I assume  $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$  is an independent random shock. When reading the narrative, the bureaucrat extracts a noisy signal of its credibility,  $s = q + \eta$ . The noise  $\eta$  comes from the fact that he can only imperfectly assess the narrative's credibility. I model the bureaucrat's ability  $\eta$  to do so as a random noise drawn from a normal distribution  $\mathcal{N}(0, \sigma_\eta^2)$ .

The bureaucrat receives  $-(v - (\omega - c_g))^2$  as the payoff from evaluation  $v \in \mathbb{R}$  for an asylum seeker with characteristic  $g$  and persecution  $\omega$ . The introduction of a taste parameter,  $c_g$ , in the bureaucrat's payoff function allows for the possibility that the official derives utility from favoring one group over another. Normalizing  $c_A = 0$ , a bureaucrat has a taste-based partiality for group  $A$  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.

Statistical discrimination instead arises from *how* bureaucrats reach evaluation  $v$ . I assume that bureaucrats have prior beliefs about the level of persecution faced by group  $g$ ,



which are distributed according to the normal distribution  $\mathcal{N}(\hat{\mu}_g, \sigma_\omega^2)$ . The bureaucrat has a belief-based partiality toward group  $A$  if he considers asylum seekers from that group to be more persecuted than those 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  $\mu_g$ , but they can also be inaccurate ( $\hat{\mu}_g \neq \mu_g$ ).

The bureaucrat updates his beliefs about the level of persecution faced by the asylum seeker after reading their narrative, and chooses the evaluation  $v$  that maximizes his expected payoff with respect to this updated belief (see SI Appendix A (pp. 3) for more details on the derivation of Equation (1)).

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

Equation (1) provides an important insight for the empirical strategy. To estimate the extent of discrimination involved in decisions about whether to grant 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  $\eta$ . Moreover, the estimation inevitably introduces some noise  $\lambda$ . As a result, I control for  $s' = q + \lambda$ . However, as long as  $\lambda$  and  $\eta$  are uncorrelated with group characteristics, I can recover the effect of individual characteristics on the decision. I return to this point 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  $c_A = 0$ ,

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

Equation (2) illustrates that disparate decisions, e.g. in favor of group  $A$ , can arise from two sources: taste-based discrimination, which results from the official's preference for group  $A$  ( $c_B \geq 0$ ), or statistical discrimination, the result of the bureaucrat's belief that group  $A$  is more persecuted on average than group  $B$  ( $\hat{\mu}_B - \hat{\mu}_A \leq 0$ ).

## 4 Data

The sample for this study consists of 4,142 asylum applications filed at the French asylum office between 1976 and 2016. The sampling design was complicated by the fact that the office regularly destroys a subset of rejected applications to free up space in the archives. To deal with this issue, I proceeded in two steps. The French asylum office first extracted a random sample of 100,000 applications of all those filed since 1952, 25,118 of which had already been destroyed. I then selected applications from among those remaining for in-depth data collection using the inverse of the probability of still being in the archives as the selection probability. To correct for remaining imbalance between the data collected and the administrative sample, I used entropy balancing to reweight the sample for the analyses ([Hainmueller 2012](#)). SI Appendix B (pp. 4-9) contains more information on the sampling design 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, 15.2 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 asylum office (Figure 1, dashed line) 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 Geneva Convention is applied in France based solely on examining individual claims of persecution rather than on belonging to a particular group ([Cohen 2000](#)). This practice rules out the possibility that the observed group differences may be driven by the fact that bureaucrats know that some groups are more persecuted, since in theory this information should not matter. 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 applicant’s region of origin, which helps determine

whether she needs protection. In France, asylum cases are examined sequentially, thus only those 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 for the Geneva Convention protection.

[Figure 1 about here.]

Table 1 lists the main independent variables used in the analysis. In this study, I focus on identifying discrimination on the basis of religion, education, skill level and language proficiency. Overall, close to 80 percent of applicants identify as either Christian or Muslim, but there is significant variation over time: Muslims represented 60 percent of applicants in 2016, up from 5 percent in 1976. Half of all applicants had started secondary education, and 14.2 percent reported starting university. Highly skilled jobs including academics, doctors, engineers, high-level executives, lawyers, journalists and students represent 16.6 percent of the applicants. Blue-collar workers, civil servants, mechanics, farmers, guards and other such employees represent 39.7 percent of applicants (*Middle*), and drivers, hairdressers, sales clerks and homemakers 19 percent (*Low*). An applicant was coded as proficient in French if she listed French as a native or spoken language (26.7 percent of applicants did). In all specifications, I also control for age, gender, marital status and time spent in France at the time of the application.

[Table 1 about here.]

For each application, I also transcribed the personal narrative, the bulk of which were handwritten. Of the 4,142 applications included in the final sample, 93 percent submitted a narrative in French. Table 2 provides summary statistics on a number of features of the narratives that, based on my informal discussions with bureaucrats at the French asylum office, should affect how credible they are perceived to be. The number of words (including stop words) and the number of dates and locations mentioned (extracted using Stanford

CoreNLP) are proxies for the narrative’s level of detail. The narratives varied greatly in length, but on average were 778 words long and mentioned seven dates and seven locations. To understand the extent to which the narratives are personal and individualized, I counted the number of first-person pronouns used (“je,” “j’ai,” “me,” “mon,” “mes” and “moi”). On average, the narratives used first-person pronouns 35 times. To control for the originality of the narrative, I also computed each narrative’s average Euclidean distance to the narratives of other asylum seekers from the same country of origin. Moreover, to gain insight into the topics covered in the narratives, I also estimated a structural topic model with 20 topics using country of origin as a covariate (Roberts, Stewart, Tingley, Lucas, Leder-Luis, Gadarian, Albertson, and Rand 2014). See Appendix C (pp. 9-17) for more information on the construction of these features.

[Table 2 about here.]

For confidentiality reasons, I was not authorized to collect any individual-level information about the bureaucrats working at the French asylum office. However, I was able to construct a fine-grained measure of their experience using administrative data on a large random sample of asylum applications. To do this, I complement the list of 100,000 I used to sample asylum applications for archival data collection with an additional random sample of 500,000 applications filed between 1989 and 2014. Together these two administrative samples provide information on 502,997 unique asylum applications filed between 1989 and 2014 — more than half of all applications filed at the asylum office during that period. The information contained in this database is scarce, but for the 309,913 applications filed after 2000, these records include the decision, as well as the date and the identifier of the bureaucrat who made the decision. After excluding bureaucrats who made at least one decision in 2000 (and for whom I cannot rule out the possibility that they started working for the office before that year) and those who made their first decision in 2015 or after, I can infer the start date of each remaining bureaucrat using the date of their first decision and computing the order in which they made each of their decisions. I am then able to match this information with

the main sample, allowing me to use the number of past decisions made by the bureaucrat in charge of the case as a proxy for his level of experience at the time of the decision.

## 5 Research design

This unique individual-level dataset allows me to estimate discrimination in the attribution of refugee status in France. I regress an indicator variable for whether an applicant was granted asylum on her characteristics, while holding constant the credibility of her narrative, her country of origin, and the year of application. 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 became mandatory in 2006 raises the concern that some decisions could have been based on unobservable characteristics revealed during the interview. Fortunately, close to 40 percent of the applicants in the sample were not interviewed, which allows me to check that the results hold in this subsample. Moreover, while we do not know the exact mechanism used to assign cases to bureaucrats, we have good reason to believe it is based on observable characteristics: Cases are first dispatched to different divisions, each of which handles a different set of nationalities. The division head then assigns cases to the officials in his or her division 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 address this concern, I show that the results hold when controlling for bureaucrat fixed effects. Finally, one necessary condition for the estimation is overlap, which is often difficult to achieve 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 (SI Figure D.2 (pp. 19)). While a number of countries cluster on the bottom left and upper right corners of the graph, a non-negligible proportion is spread out on the

diagonal, suggesting substantial overlap.

The efficacy of this research design also hinges on the ability to control for the signal  $s$  that the bureaucrat receives from the narrative, as outlined in Section 3. Moreover, the descriptive patterns reveal that the narratives differ significantly by religion, education, skill level and proficiency in French, further emphasizing the need to control for the narrative’s effect on the decision (SI Tables C.2, C.3, C.4 and C.5 (pp. 11-12)). I use three different methods to control for the narrative. First, in the main specification I control for the credibility of the narrative, which I predict using a combination of human coding and automated classification. Due to the limitations of this method (outlined below), I check that the results are robust to two alternative strategies: controlling for substantive features of the narrative that *should* be relevant in predicting the decision (listed in Table 2) and controlling for features that *are* relevant to the decision (using the supervised Indian Buffet Process). It is important to note here that I use these measures to control for the credibility of the narrative,  $q$ , not the applicant’s level of persecution,  $\omega$ . But this is not a concern since bureaucrats also only observe a signal of persecution, not the persecution itself, before making their decision. I now discuss all three methods in turn.

The first method I use to control for the narrative relies on a combination of hand coding and supervised learning to measure the credibility of the personal narrative — that is, the extent to which the narrative (1) meets the criteria outlined in the definition of refugee in the Geneva Convention and (2) sounds authentic and convincing in its claims of persecution. Three coders read 350 unique narratives, 59 of which were triple coded. These coders were selected from the pool of research assistants who helped with the digitization process and who had master’s degrees in law or political theory. This selection process ensured the confidentiality of the narratives, a first-order parameter in this study, and that coders were knowledgeable about the Geneva Convention. By the time they read and coded these narratives, they had each transcribed several hundred, an important requirement for reliable hand coding (Krippendorff 2004). Importantly, they were not informed of the decision

reached by the bureaucrat.

For each narrative, they first coded whether it mentioned any form of persecution, and if so, what type (race, political opinion, religion, nationality or social group). Second, they assessed whether the claims were believable, convincing, detailed, individualized and coherent, using a 4-point scale for each. Third, they coded whether the narrative referred 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 the question, *“In your opinion, is this person entitled to claim the right to asylum according to the Geneva Convention?”* {Disagree/Somewhat disagree/Somewhat agree/Agree}. On average, coders “agreed” with the statement 15 percent of the time, which is comparable to the average acceptance rate in the sample, and “somewhat agreed” 35 percent of the time (SI Table C.6 (pp. 13)). I coded narratives as credible if research assistants answered either “Agree” or “Somewhat agree” because the inter-coder reliability, as measured by the Krippendorff’s alpha coefficient,<sup>7</sup> is much higher (0.48) when combining these two responses than when considering narratives to be credible only when they were coded as “Agree” (0.19) (SI Table C.7 (pp. 13)).

Using the set of hand-coded narratives, I then 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 credibility measure 76 percent of the time in the left-out sample of 85 narratives, a substantial reduction in error compared to a baseline of 48 percent (SI Table C.9 (pp. 16)), and to ensure that the predicted probabilities are well calibrated (SI Figure C.1 (pp. 16)). This predicted measure correlates as we would expect with the features of the narrative (SI Table C.10 (pp. 17)) and, as detailed in Section

---

<sup>7</sup>Krippendorff  $\alpha$  is computed as  $1 - (\text{number of disagreements actually observed})/(\text{number of disagreements observed by chance})$ . This ratio is used to determine inter-coder reliability, where 0 indicates perfect disagreement and 1 perfect agreement. For more information, see Krippendorff (2004, pp. 221-227). Agreements ranging from 0.4–0.6 are considered a sign of moderate agreement between coders (Landis and Koch 1977).

6, with the decision as well, which suggests that this measure is capturing the credibility of the narrative, even if imperfectly. 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 to summarize, in a single variable, important variation among applicants’ narratives that could explain officials’ decisions whether to grant or withhold refugee status.

Although this procedure inevitably introduces some noise ( $\lambda$ ), the first key identification assumption is that this noise is uncorrelated with the characteristics of asylum seekers I examine (religion, education, skill level and proficiency in French). This would be problematic, for example, if research assistants were more likely to find the same narrative more credible if the asylum seeker is Christian than if she is Muslim. To help mitigate this concern, I check the robustness of the results by controlling for the effect of the narratives in two additional ways. First, I directly control for features of the narratives that *should* be relevant in explaining the decision (listed in Table 2). 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 using a range of parameters to select the model that ranked the highest on a quantitative measure of model fit (Fong and Grimmer 2016). I use this model to infer the latent treatments for the test set (the remaining 70 percent of narratives), which I use as additional controls in the analysis. By using a split-sample design, this procedure solves the identification and estimation problems that arise from using the same documents to discover treatments and estimate causal effects (Egami, Fong, Grimmer, Roberts, and Stewart 2018).

The second identifying assumption is that the bureaucrat’s specific noise that proxies for his ability ( $\eta$ ) — 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 the same narrative when the asylum seeker belongs to one group over another — if, for example, the bureaucrat was more knowledgeable about one group than another such that the same narrative would be appear more credible if it were written by an applicant from a particular group. The fact that bureaucrats specialize by nationality rather than 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 defined in the Geneva Convention. To address this concern, I show that the results on religion are robust to restricting the analysis to the sample of asylum seekers who did not claim persecution on the basis of religion.

## 6 Results

### 6.1 Discrimination in the attribution of refugee status

[Table 3 about here.]

What are the determinants of asylum decisions in France? 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, all specifications include a limited set of controls: age, gender, marital status, time spent in France prior to applying for refugee status and the credibility of the narrative. This analysis reveals that religion is an important predictor of asylum decisions (column 1). Muslim applicants are 6.2 percentage points less likely to obtain refugee status than Christians who are similar across all other characteristics. This represents a substantial difference (41 percent) given that the average acceptance rate was 15 percent during the study period. The effect of educational attainment and skill level on the decision is large as well. Compared to those who reported starting university, those who reported secondary or primary level education were 2.1 and 6.1 percentage points less likely to be granted refugee status, respectively,

though only the difference between primary and post-secondary education is statistically significant. Those with middle or low skill levels were both significantly much less likely (6.6 and 8.5 percentage points, respectively) to obtain refugee status. Compared to those who speak French, not speaking French has a small negative effect (3.3 percentage points) on the decision. However, when restricting the sample to applicants who were not interviewed, even though the estimates of the coefficients are similar in sign and magnitude only the difference between Christians and Muslims remain statistically significant (Table 3, column 4).

To alleviate concerns that the difference between Christians and Muslims is driven by omitted variables, I perform four sets of robustness checks. First, I show in SI Table D.1 (pp. 21) that this result is robust to including additional covariates reported in the application form (column 1), including 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 family member who obtained refugee status in France, whether the applicant completed military service and whether her entry to the territory was legal (see SI Table B.1 (pp. 9) for summary statistics on these additional variables). I also show that this result is robust to including country of origin-year of application interactions (column 2), to pruning the sample using coarsened exact matching (Iacus, King, and Porro 2012) (column 3), and to re-estimating effect sizes using a logistic regression model (SI Table D.2 (pp. 22)). Finally, in column 4, I show that the coefficient is very similar when the credibility measure is omitted, which suggests that the results do not hinge entirely on this particular measure.

Second, I show that the result is robust to alternative ways of controlling for the effect of the personal narrative. In columns 2 and 3 of Table 3, in addition to the credibility of the narrative, I control, respectively, for substantive text features (described in Table 2) and for text features discovered by the supervised Indian Buffet Process in the left-out one-third of the sample. The coefficient is less precise in this last specification, which could be a function of the smaller sample size. Overall, the result is not sensitive to these alternative ways of controlling for the narrative’s effect on the decision.

Third, by showing that the result holds after controlling for bureaucrat fixed effects, I reject the possibility that these differences result from a process in which asylum applications submitted by Muslims 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 digitized. 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 an additional “missing” category (Table 3, column 5).

Finally, I perform a set of robustness checks to ensure that the difference between Christians and Muslims is not entirely driven by the fact that asylum officers have access to background information regarding the persecution experienced by these different groups. 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 the results hold when I control for whether the applicant belongs to a minority religious group in her home country (Table 3, column 6). The World Religion dataset of the Association of Religion Data Archives estimates the percentage of the population that has identified with Christianity or Islam since 1945 for most countries in the world. Using these estimates, I generate a binary variable that 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.<sup>8</sup>

I then 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 that if Christian applicants are more likely to receive asylum than Muslims because they are more likely to be persecuted than Muslims, then the Muslim gap should not hold in the subsample of applicants who did not claim persecution on religious grounds. Only 6 percent of narratives claimed persecution on religious grounds according to the coding completed

---

<sup>8</sup>A total of 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.

by the research assistants on a representative sample of applicants (SI Table C.6 (pp. 13)), and the proportion of the text dedicated to the topic *Religion* calculated by the structural topic model exceeds 20 percent in only 2 percent of the narratives (SI Table B.1 (pp. 9)). But identifying applicants who claim religious persecution 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 listed in the Geneva Convention), I read 292 narratives, a random sample of the 1,021 narratives that contained at least one religious keyword, and for each I coded whether the applicant was 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 religious persecution, I find that the Muslim penalty holds (Table 3, column 7). 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 are much more likely to be granted asylum than Muslim applicants. In a series of robustness tests, I rule out the possibility that this difference is entirely explained by observable or unobservable omitted variables, by the selective assignment of cases to bureaucrats or by the fact that Muslims are less persecuted than Christians.

## 6.2 Bureaucrats' experience mitigates discrimination

I next examine the influence of bureaucrats' experience (measured as the number of past decisions) on discrimination on the basis of religion, education and skill level. I restrict the sample to the 1,821 asylum applications filed between 2000 and 2014 that were decided by a bureaucrat who started working between 2001 and 2014 for whom I know the number of past decisions. In this sample, the average number of past decisions is 457, the median 312, and the 90th percentile is 1,103 decisions. To estimate the conditional marginal effects, I

use the binning estimator proposed by [Hainmueller, Mummolo, and Xu \(2016\)](#), which has two main advantages over the classical linear multiplicative interaction model. First, the binning estimator allows conditional marginal effects to vary across bins of the moderator by relaxing the linear interaction effect assumption. Second, the binning estimator ensures common support to reliably estimate the conditional marginal effects by constructing bins based on the support, rather than on values of the moderator.

[Figure 2 about here.]

I thus first divide the number of past decisions into three equally sized groups — a low (1–184), middle (185–486) and high (487–3,220) number of past decisions — and pick the median number of past decisions as the evaluation point within each bin. I then estimate a model that includes, in addition to the covariates from the main specification (Table 3, column 1), interactions between the indicator variables for each of the three bins, the individual characteristics of interest, and the number of past decisions minus the evaluation point picked within each bin, as well as triple interactions of these. I conduct this analysis separately for religion (comparing Muslims to Christians), education (comparing secondary and primary to post-secondary) and skill level (comparing middle and low to high). Figure 2 plots the conditional marginal effects of these three characteristics at low, middle and high values of the moderator, with 95 percent 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, suggesting that discrimination is most pronounced at lower levels of experience. The pattern is similar when constructing bins based on the values of the moderator (SI Figure D.3 (pp. 24)) and when estimating the conditional marginal effects using generalized additive models (SI Figure D.4 (pp. 25)).

[Table 4 about here.]

To test whether this reduction in discrimination is statistically significant, I further restrict the sample to the first 486 applications (second tercile of the moderator) of each

bureaucrat in the sample and report in Table 4 estimates from the main specification (column 1) and from the binning estimator described above (columns 2 to 4). While Muslims in this subsample are overall 1.8 percentage points less likely to be granted asylum (standard error: 0.038) compared to Christians (column 1), I find a substantial and statistical difference depending on whether the application was decided by an inexperienced or experienced bureaucrat. Muslims are 6.9 percentage points (standard error: 0.040) 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 Christians' chances of receiving asylum (coefficient: -0.018, standard error: 0.022), it increases Muslims' chances by 10.1 percentage points (standard error: 0.034), such that the difference between Christians and Muslims is small and insignificant among experienced bureaucrats (column 2). To ensure that this pattern is not driven by the changing composition of bureaucrats over time, I further restrict the sample to applications examined by bureaucrats who decided at least 486 applications (columns 4 and 5 in Table 4). The results are robust to restricting the sample in this way and to including bureaucrat fixed effects (column 5).

[Table 5 about here.]

The pattern is similar for educational attainment and skill level (Table 5). Compared to highly educated and highly skilled asylum seekers, those with lower levels of education and skills are less likely to be granted refugee status when their case is examined by an inexperienced bureaucrat (the relevant coefficients are all negative although not all reach statistical significance at conventional levels), but, consistent with a reduction in discrimination among more experienced bureaucrats, the coefficients of the interaction terms are all positive (though only one reaches statistical significance). In SI Table D.3 (pp. 23), I show that applications decided by experienced and inexperienced bureaucrats do not differ systematically, which allows me to mitigate concerns that this result reflects the fact that bureaucrats are assigned different types of decisions during the course of their tenure.

Overall, these patterns suggest that bureaucrats discriminate less as they get more experience, but this reduction is stronger and more robust for religion than for education and skill level. To investigate the time frame during which this reduction is taking place, I also report in SI Table D.3 (pp. 23), for each group of applications, the average number of months the bureaucrat who decided the case had spent on the job at the time of the decision, simply counting the number of months between each decision and his first decision. The reduction in discrimination seems to be taking place over the course of roughly a year: in the first bin, applications are being decided by bureaucrats who had spent roughly 9 months on the job, while in the second bin they had spent at least two years.

## 7 Discussion

What explains this reduction in discrimination as bureaucrats get more on-the-job experience? If discrimination results primarily from the fact that bureaucrats have a preference for different types of applicants, then the reduction could be explained by a change in bureaucrats' preferences on the job, for example if  $c_B$  is positive for inexperienced bureaucrats but close to zero for experienced bureaucrats. Contact theory hypothesizes that personal contact across social lines can reduce prejudice (Allport 1974). According to this logic, as bureaucrats interview more applicants, they could become less prejudiced against Muslims, leading to a reduction in taste-based discrimination. It is important to note, however, that the interview (which is the only point of contact between the bureaucrat and the asylum seeker) does not fit the traditional assumptions of contact theory, which maintain that contact should be cooperative, have a shared goal and be endorsed by communal authority — and that participants should have equal power status. Contact during an interview quite strongly violates all of these assumptions, with the possible exception of communal authority. Indeed, recent research suggests 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 (Lowe 2020).

If discrimination is instead driven by bureaucrats’ prior beliefs about the level of persecution faced by different groups, a change in these beliefs (or in the weight they place on them) over time could also explain the reduction in discrimination observed. Several first-hand accounts of decision-makers at the French asylum office and the appeals court suggest that the narratives are relatively uninformative for bureaucrats when they start working at the asylum office, but that officials hone their ability to identify credible narratives over time. 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,’” one of them testified. As a result, she continued, “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 distinguish themselves through original storytelling and language. As they read more and more applications, bureaucrats become better able to distinguish fake narratives from true ones. A bureaucrat working at the French asylum office powerfully illustrated this intuition in anonymous testimony: “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). Bureaucrats’ experience thus increases the precision — i.e. decreases the variance  $\sigma_\eta^2$  — of the signal they receive when reading narratives. According to equation (1), as  $\sigma_\eta^2$  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  $\left(\frac{\sigma_\epsilon^2 + \sigma_\eta^2}{\sigma_\omega^2 + \sigma_\epsilon^2 + \sigma_\eta^2}\right)$ .



## 8 Conclusion

Are European countries abiding by the Geneva Convention and granting asylum to those who fear persecution for reasons of “race, religion, nationality, membership of a particular group or political opinion”? Until now, the lack of micro-level data on asylum decisions has limited researchers’ ability to answer this question. Yet 10 years ago the French asylum office opened its archives. I exploited this unprecedented degree of transparency to digitize a representative sample of more than 4,000 asylum applications filed over the last 40 years to compile the first in-depth, micro-level dataset of asylum decisions. This dataset facilitates an in-depth examination of whether France is indeed granting asylum to those in need of protection. I compared accepted and rejected applicants to identify the effect of applicants’ personal characteristics on the probability of being granted asylum in France.

This study provides empirical evidence that Christian applicants are much more likely to be granted refugee status than Muslim applicants. Moreover, I show that these effects are unlikely to be driven by either unobserved differences revealed during the interview process or by the selective assignment of cases to bureaucrats. Importantly, the findings also reveal that bureaucrats can self-correct their established discriminatory behaviors. As they gain more on-the-job experience, they discriminate less on the basis of religion, and to a lesser extent based on education and skill level. Anecdotal evidence suggests that bureaucrats learn over time to distinguish authentic narratives, which is consistent with statistical discrimination. While the narratives provided by applicants are relatively uninformative at first, bureaucrats gradually learn how to assess their credibility, allowing them to place more weight on the narratives than on their prior beliefs. However, the available data do not allow me to rule out the possibility that the reduction in discrimination instead comes from a change in bureaucrats’ preferences.

The results of the study have important policy implications for strategies to reduce discrimination at the French asylum office. For instance, exposure to past narratives during bureaucrats’ training could help them learn to better assess the credibility of persecution

claims once they are faced with actual applicants. Moreover, increasing bureaucrats' tenure could reduce the overall level of discrimination. But more research is needed to test the effectiveness of these interventions. 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 within administrations.

These findings unambiguously call for increased scrutiny in the attribution of refugee status in France, and within the EU more broadly. Since the Geneva Convention constitutes the common framework for granting asylum across Europe, asylum officers elsewhere on the continent are afforded similar levels of discretion as those in the French 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 or even the offices directly, should be undertaken to determine whether the discrimination patterns identified in France apply more broadly.

## References

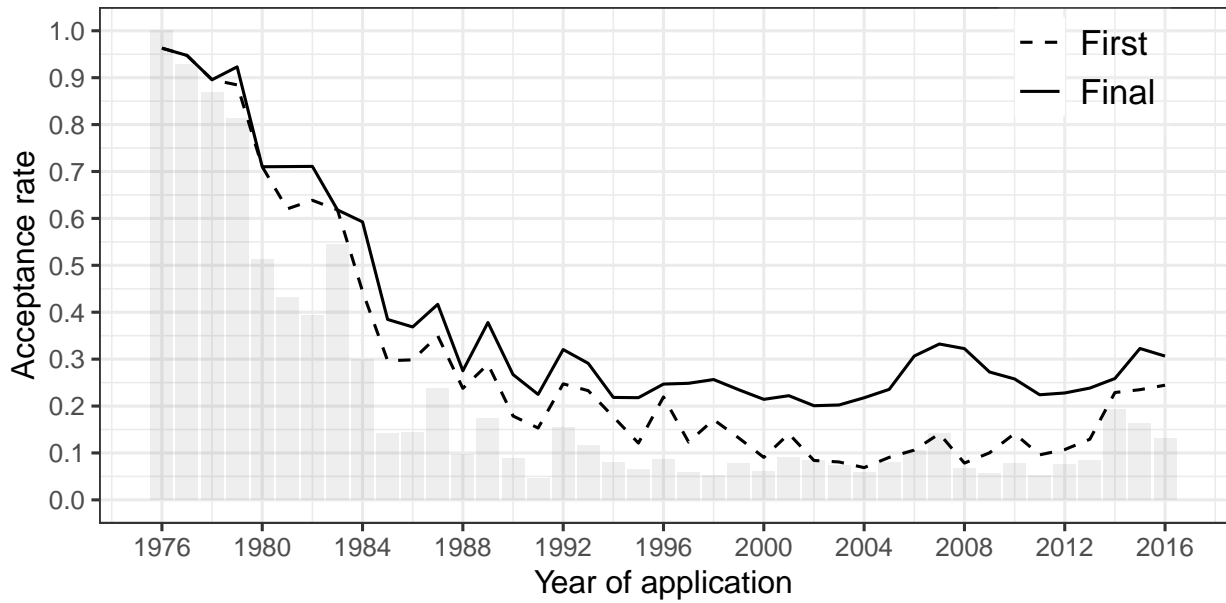
- Adida, C. L., D. D. Laitin, and M.-A. Valfort (2016). *Why Muslim Integration Fails in Christian-Heritage Societies*. Harvard University Press.
- Adida, C. L., A. Lo, and M. R. Platas (2019). Americans preferred Syrian refugees who are female, English-speaking, and Christian on the Eve of Donald Trump’s election. *PLoS ONE* 14(10), 1–18.
- 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.
- Bansak, K., J. Hainmueller, and D. Hangartner (2016). How economic, humanitarian, and religious concerns shape European attitudes toward asylum seekers. *Science* 354(6309), 217–222.
- 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*.
- 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. and M. J. Hiscox (2007). Educated preferences: Explaining attitudes toward immigration in Europe. *International Organization* 61(2), 399–442.
- Hainmueller, J. and M. J. Hiscox (2010). Attitudes toward highly skilled and low-skilled immigration: Evidence from a survey experiment. *American Political Science Review* 104(1), 61–84.
- Hainmueller, J. and D. J. Hopkins (2015). The Hidden American Immigration Consensus: A Conjoint Analysis of Attitudes toward Immigrants. *American Journal of Political Science* 59(3), 529–548.
- Hainmueller, J., J. Mummolo, and Y. Xu (2016). How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice.
- Helbling, M. and R. Trauttmüller (2018). What is Islamophobia? Disentangling Citizens’ Feelings Toward Ethnicity, Religion and Religiosity Using a Survey Experiment. *British Journal of Political Science*.
- 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.
- Iacus, S. M., G. King, and G. Porro (2012). Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis* 20(1), 1–24.
- 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.
- Krippendorff, K. (2004). *Content Analysis: An Introduction to its Methodology*.
- Landis, J. R. and G. G. Koch (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33(1), 159.
- 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.
- Olsen, A. L., J. H. Kyhse-Andersen, and D. Moynihan (2020). The Unequal Distribution of Opportunity: A National Audit Study of Bureaucratic Discrimination in Primary School Access. *American Journal of Political Science*, 1–39.
- 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 europäische 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.
- Valentino, N. A., S. N. Soroka, S. Iyengar, T. Aalberg, R. Duch, M. Fraile, K. S. Hahn, K. M. Hansen, A. Harell, M. Helbling, S. D. Jackman, and T. Kobayashi (2017). Economic and Cultural Drivers of Immigrant Support Worldwide. *British Journal of Political Science* 49(4), 1201–1226.
- 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.

Figure 1: Acceptance rate at the French asylum office (1976–2016)



*Notes:* This figure plots the acceptance rate on the basis of the Geneva Convention and the subsidiary protection after the first examination at the asylum office (dashed), and after examination by the appeals court and subsequent re-examinations by the asylum office (solid). The gray bars plot the acceptance rate based on the Geneva Convention among applicants in the sample.

Table 1: Summary statistics on dependent and independent variables

	N	Mean	Std. Dev.	Min	Max
Got refugee status	4,142	0.152	0.359	0	1
Provided a narrative	4,142	0.933	0.249	0	1
<i>Gender</i>					
Female	4,142	0.309	0.462	0	1
<i>Age</i>					
Less than 20	4,142	0.090	0.286	0	1
Between 20 and 40	4,142	0.771	0.420	0	1
More than 40	4,142	0.139	0.346	0	1
<i>Marital status</i>					
Single	4,141	0.582	0.493	0	1
Married	4,141	0.418	0.493	0	1
<i>Religion</i>					
Christian	4,142	0.376	0.484	0	1
Muslim	4,142	0.396	0.489	0	1
Other	4,142	0.122	0.327	0	1
None/Missing	4,142	0.107	0.309	0	1
<i>Education</i>					
University	4,142	0.142	0.349	0	1
Secondary	4,142	0.495	0.500	0	1
Primary	4,142	0.111	0.314	0	1
None/Missing	4,142	0.251	0.434	0	1
<i>Skill level</i>					
High	4,142	0.166	0.373	0	1
Middle	4,142	0.397	0.489	0	1
Low	4,142	0.186	0.390	0	1
None/Missing	4,142	0.250	0.433	0	1
<i>Speaks French</i>					
Yes	4,142	0.267	0.443	0	1
No/Missing	4,142	0.733	0.443	0	1
<i>Arrival in France</i>					
> 1 year in France	4,142	0.150	0.357	0	1

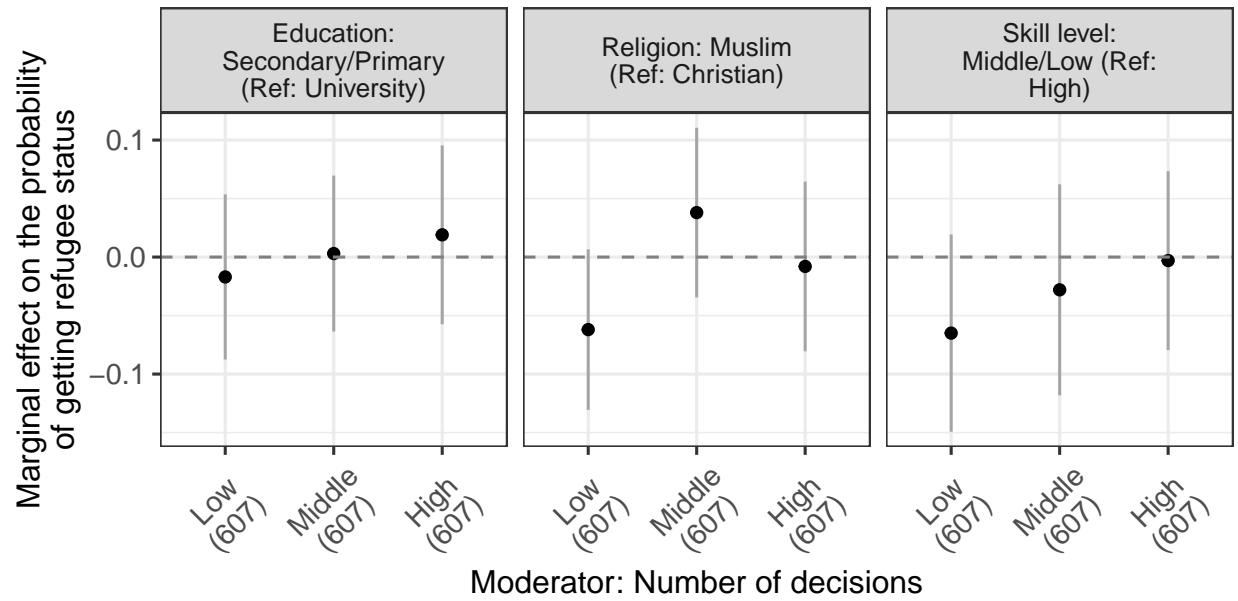
*Notes:* All variables, except the first two, are self-reported.



Table 2: Summary statistics on narrative features

	N	Mean	Std. Dev.	Min	Max
<i>Level of detail</i>					
Number of words	3,932	777.980	765.196	1	7,169
Number of dates	3,932	7.288	8.321	0	95
Number of location	3,932	6.540	7.509	0	84
<i>Degree of personalization</i>					
Number of personal pronouns	3,760	34.575	33.074	0	445
<i>Originality</i>					
Distance	3,927	47.196	23.490	6	347
<i>Topic proportion &gt;.2</i>					
Family members	3,930	0.087	0.282	0	1
Bangladesh	3,930	0.038	0.192	0	1
Student protest	3,930	0.074	0.261	0	1
Escape	3,930	0.092	0.289	0	1
Sri Lanka	3,930	0.040	0.197	0	1
RDC	3,930	0.044	0.204	0	1
Life in France	3,930	0.223	0.417	0	1
Former Yugoslavia	3,930	0.078	0.268	0	1
China	3,930	0.041	0.197	0	1
Ethnic minorities	3,930	0.053	0.224	0	1
Kurds in Turkey	3,930	0.067	0.250	0	1
Political instability	3,930	0.049	0.215	0	1
Political opposition	3,930	0.055	0.228	0	1
Court hearing	3,930	0.020	0.140	0	1
Narrative	3,930	0.150	0.357	0	1
Angola	3,930	0.057	0.232	0	1
Salutation	3,930	0.058	0.234	0	1
Family at risk	3,930	0.043	0.202	0	1
Zaire	3,930	0.034	0.181	0	1
Religion	3,930	0.021	0.143	0	1

Figure 2: Discrimination by 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, and the number of observations in each bin in parentheses. The specifications include covariates and fixed effects for year of application and country of origin.

Table 3: Discrimination in the attribution of refugee status

	Granted refugee status						
	(1) Full Sample	(2) Substantive text features	(3) Discovered text features	(4) Not Interviewed	(5) Bureaucrats fixed effects	(6) Minority status	(7) Not persecuted for religion
Credible narrative	0.058** (0.013)	0.025† (0.015)	0.063** (0.016)	0.047* (0.022)	0.066** (0.015)	0.057** (0.013)	0.033* (0.014)
<i>Religion (Ref: Christian)</i>							
Muslim	-0.062** (0.020)	-0.056** (0.019)	-0.052* (0.023)	-0.081* (0.033)	-0.061* (0.024)	-0.085** (0.024)	-0.045† (0.024)
Other	0.059* (0.027)	0.042 (0.027)	0.075* (0.034)	-0.016 (0.039)	0.031 (0.032)	0.089** (0.028)	0.051† (0.029)
None/Missing	-0.016 (0.022)	-0.006 (0.022)	-0.021 (0.027)	-0.048 (0.030)	-0.008 (0.026)	0.003 (0.022)	0.023 (0.025)
<i>Education (Ref: University)</i>							
Secondary	-0.021 (0.019)	-0.014 (0.020)	-0.006 (0.022)	-0.055 (0.034)	-0.021 (0.023)	-0.018 (0.019)	-0.015 (0.021)
Primary	-0.061* (0.027)	-0.063* (0.028)	-0.059† (0.031)	-0.071 (0.047)	-0.067* (0.033)	-0.056* (0.028)	-0.064* (0.030)
None/Missing	-0.032 (0.022)	-0.025 (0.023)	-0.022 (0.025)	-0.037 (0.039)	-0.039 (0.027)	-0.029 (0.022)	-0.043† (0.024)
<i>Skill level (Ref: High)</i>							
Middle	-0.066** (0.021)	-0.053* (0.021)	-0.062* (0.025)	-0.048 (0.031)	-0.052* (0.023)	-0.068** (0.021)	-0.060* (0.024)
Low	-0.085** (0.022)	-0.072** (0.022)	-0.079** (0.027)	-0.046 (0.032)	-0.064** (0.025)	-0.086** (0.022)	-0.082** (0.024)
None/Missing	-0.028 (0.022)	-0.022 (0.022)	-0.023 (0.025)	0.003 (0.032)	-0.013 (0.024)	-0.034 (0.021)	-0.026 (0.025)
<i>Speaks French (Ref: Yes)</i>							
No/Missing	-0.033† (0.018)	-0.026 (0.018)	-0.038† (0.019)	-0.027 (0.024)	-0.021 (0.021)	-0.035* (0.018)	-0.022 (0.019)
Observations	3,930	3,754	2,772	1,481	3,930	3,898	3,103
$R^2$	0.358	0.383	0.334	0.640	0.508	0.364	0.395
Mean acceptance rate	0.152	0.152	0.152	0.176	0.152	0.152	0.152

Notes: \*\* $p < 0.01$ , \* $p < 0.05$ , † $p < 0.1$ . Point estimates and standard errors in parentheses. 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 marital status), an indicator variable for whether they spent more than one year in France before applying, country of origin and year of application fixed effects.

Table 4: Discrimination by bureaucrats' experience (religion)

	Granted refugee status			
	(1)	(2)	(3)	(4)
More than 184		-0.018 (0.022)	0.010 (0.032)	-0.099 <sup>†</sup> (0.055)
<i>Religion (Ref: Christian)</i>				
Muslim	-0.018 (0.038)	-0.069 <sup>†</sup> (0.040)	-0.030 (0.052)	-0.060 (0.072)
Other	0.101* (0.041)	0.023 (0.037)	0.025 (0.062)	0.047 (0.115)
None/Missing	0.029 (0.044)	0.044 (0.057)	0.130 (0.094)	0.087 (0.135)
<i>Interactions</i>				
More than 184 × Muslim		0.101** (0.034)	0.127** (0.047)	0.170* (0.068)
More than 184 × Other		0.147* (0.067)	0.094 (0.084)	0.100 (0.130)
More than 184 × None/Missing		-0.025 (0.049)	-0.090 (0.084)	-0.042 (0.122)
Bureaucrats fixed effects	No	No	No	Yes
$R^2$	0.229	0.252	0.308	0.501
Observations	1,185	1,185	656	656

Notes: \*\*  $p < 0.01$ , \*  $p < 0.05$ , <sup>†</sup> $p < 0.1$ . This table shows point estimates and standard errors clustered by bureaucrats in parentheses from ordinary least squares (OLS) regressions with individual covariates and fixed effects for year of application and country of origin. The sample is restricted to the first 486 decisions of bureaucrats in columns 1 and 2 and to the first 486 decisions of bureaucrats who made at least 486 decisions in columns 3 and 4.

Table 5: Discrimination by bureaucrats' experience (education and skill level)

	Education				Skill level			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
More than 184		-0.008 (0.043)	0.015 (0.071)	-0.146 (0.101)		0.001 (0.056)	0.044 (0.090)	-0.133 (0.128)
<i>Characteristics (Ref: High)</i>								
Medium	0.003 (0.026)	-0.016 (0.036)	-0.027 (0.054)	-0.081 (0.074)	-0.045 (0.035)	-0.066 (0.046)	-0.088 (0.076)	-0.080 (0.113)
Low	0.007 (0.038)	-0.033 (0.046)	-0.076 (0.075)	-0.165 (0.118)	-0.071* (0.034)	-0.094* (0.041)	-0.086 (0.073)	-0.121 (0.110)
None/Missing	0.002 (0.037)	-0.071† (0.040)	-0.116† (0.059)	-0.191* (0.095)	-0.047 (0.032)	-0.062 (0.040)	-0.072 (0.070)	-0.081 (0.105)
<i>Interactions</i>								
More than 184 × Medium		0.029 (0.048)	0.035 (0.073)	0.099 (0.095)		0.046 (0.064)	0.032 (0.094)	0.080 (0.138)
More than 184 × Low		0.058 (0.063)	0.100 (0.100)	0.218 (0.142)		0.049 (0.063)	0.027 (0.100)	0.134 (0.135)
More than 184 × None/Missing		0.140* (0.060)	0.129 (0.087)	0.236† (0.128)		0.033 (0.063)	0.007 (0.094)	0.110 (0.127)
Bureaucrats fixed effects	No	No	No	Yes	No	No	No	Yes
$R^2$	0.229	0.245	0.291	0.496	0.229	0.242	0.294	0.501
Observations	1,185	1,185	656	656	1,185	1,185	656	656

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 bureaucrats' first 486 decisions in columns 1 and 2 and to their first 486 decisions if they made more than that in columns 3 and 4.

Learning to be Unbiased:  
Evidence from the French Asylum Office

*Supplemental Information*

Mathilde Emeriau<sup>‡‡</sup>

July 31, 2021

---

<sup>‡‡</sup>Department of Government, London School of Economics and Political Science, and Immigration Policy Lab, Stanford University, Stanford, CA 94305-6044, USA. m.emeriau@lse.ac.uk

<b>Appendix A Model</b>	<b>3</b>
<b>Appendix B Sampling</b>	<b>4</b>
B.1 Representativeness of the sampling frame . . . . .	4
B.2 Sampling design and weights . . . . .	5
B.3 Data Collection . . . . .	7
<b>Appendix C Measuring the credibility of the narrative</b>	<b>9</b>
C.1 Construction of narrative features . . . . .	9
C.2 Comparing the narratives by religion . . . . .	10
C.3 Hand coding the credibility of the narrative . . . . .	12
C.4 Predicting the credibility of the narrative . . . . .	15
<b>Appendix D Additional tables and figures</b>	<b>18</b>

## A Model

In this section, I provide additional details on the derivation of equation (1). The bureaucrat receives  $-(v - (\omega - c_g))^2$  as payoff from evaluation  $v$  for an asylum seeker with characteristic  $g$  and persecution  $\omega$ . He chooses the evaluation  $v$  that maximizes his expected payoff with respect to his posterior belief about the applicant's level of persecution.

$$v(s, g)^* \equiv \arg \max \mathbb{E} [-(v - (\omega - c_g))^2 \mid s, g]$$

His optimal evaluation is  $v(s, g)^* = \mathbb{E}[\omega \mid s, g] - c_g$ . To see this, let  $F$  be the conditional density function of  $\omega$  and rewrite the bureaucrat's maximization problem as

$$v(s, g)^* = \arg \max \int_{-\infty}^{+\infty} -(v - (\omega - c_g))^2 dF(\omega \mid s, g)$$

and take the first-order condition of the objective function with respect to  $w$  and set it equal to 0.

$$\begin{aligned} & \int_{-\infty}^{+\infty} -2(v - (\omega - c_g)) dF(\omega \mid s, g) = 0 \\ \iff & -2(v + c_g) + 2 \int_{-\infty}^{+\infty} \omega dF(\omega \mid s, g) = 0 \\ \iff & -2(v + c_g) + 2 \mathbb{E}[\omega \mid s, g] = 0 \\ \iff & -(v + c_g) + \mathbb{E}[\omega \mid s, g] = 0 \\ \iff & v = \mathbb{E}[\omega \mid s, g] - c_g \end{aligned}$$

To derive  $\mathbb{E}[\omega \mid s, g]$ , note that the posterior distribution  $p(\omega|s)$  is proportional to the signal distribution  $(s|\omega \sim \mathcal{N}(\omega, \sigma_\epsilon^2 + \sigma_\eta^2))$  multiplied by the prior distribution for the applicant's level of persecution  $(\omega \sim \mathcal{N}(\hat{\mu}_g, \sigma_\omega^2))$ . Given that both the prior and the signal are normally distributed, the posterior distribution is also normally distributed with a mean equal to  $\frac{\sigma_\omega^2}{\sigma_\omega^2 + \sigma_\epsilon^2 + \sigma_\eta^2} s + \frac{\sigma_\epsilon^2 + \sigma_\eta^2}{\sigma_\omega^2 + \sigma_\epsilon^2 + \sigma_\eta^2} \hat{\mu}_g$  and variance  $\frac{\sigma_\omega^2 (\sigma_\epsilon^2 + \sigma_\eta^2)}{\sigma_\omega^2 + \sigma_\epsilon^2 + \sigma_\eta^2}$ .

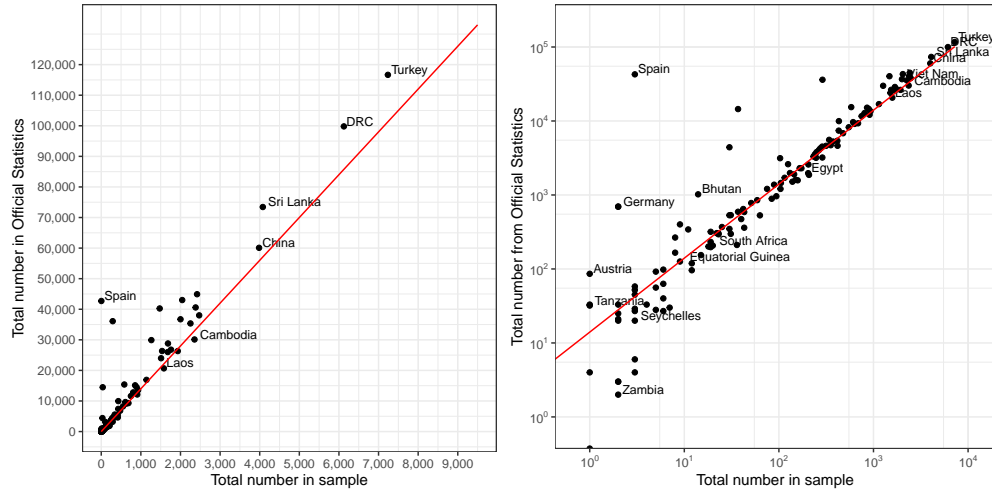


## B Sampling

### B.1 Representativeness of the sampling frame

In July 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 2001). I compiled a dataset of the total number of first-time applications by country of origin and year of application. In Figure B.1, I compare 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).

Figure B.1: Total number of first-time applications by country of origin (1952–2014)



*Notes:* Each dot represents a country of origin in the administrative sample. The  $x$  axis represents the number of applications from that country 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 using a linear scale on the left and using a log scale on the right.

## B.2 Sampling design and weights

From the list of 100,000 asylum applications, I selected 10,000 for in-depth data collection. I first excluded three perfect duplicates and 1,572 applications from the 81 nationalities that comprised fewer than 100 applications overall. I then split the dataset into four subsamples depending on (1) the year of application and (2) the database in which the application was recorded.

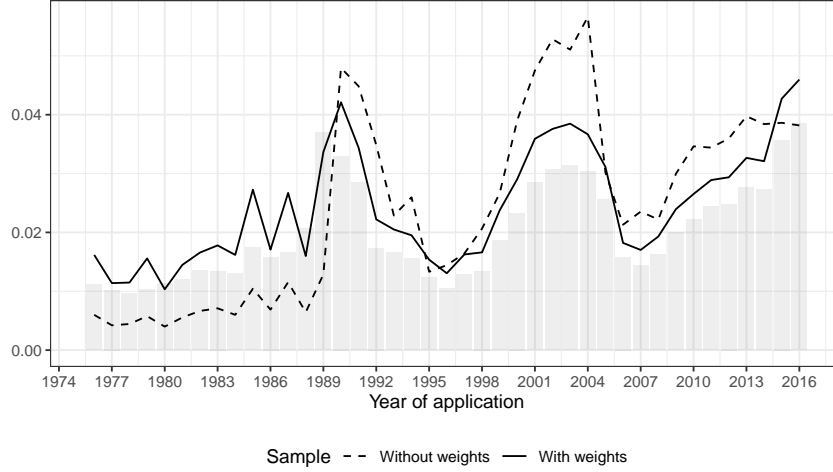
1. Applications from *Garonor* filed before January 1, 1990 (N=26,433). None of these had been destroyed by July 2015. For these applications, I only know the nationality of the applicant.
2. Applications from *Inerec* filed before January 1, 1990 (N=14,259). 40% of these had been destroyed by July 2015.
3. Applications from *Inerec* filed between January 1, 1990 and December 31, 2004 (N=35,398). 54% of these had been destroyed by July 2015.
4. Applications from *Inerec* filed between January 1, 2005 and December 31, 2014 (N=22,335). Only 13 had been destroyed July 2015.

To oversample applications filed after 1989, I drew 5% of the applications from subsamples 1 and 2 and 14% from subsamples 3 and 4. To account for the non-random destruction of folders (the asylum office only destroyed rejected applications), I proceeded as follows for each subsample. For each application, I first predicted the probability of still being in the archives 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 drew several hundred samples, using the inverse of the predicted probability of still being in the archives as the selection probability. I used 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,

overall rejection rate, yearly rejection rate, and the proportion of the top 10 nationalities. Finally, for each model, I selected the sample that minimizes overall deviation. This strategy did not yield enough rejected applications for the period 1990–1992 since more than 68% of applications filed in these 3 years had already been destroyed. In May 2016, I requested an additional sample of 50,000 asylum applications filed between 1990 and 1992 and sampled an additional 500 folders from these 3 years. Finally, in June 2016, I requested an additional sample of 8,000 asylum applications filed in 2015 and 2016, and sampled an additional 500 applications.

In the analysis, I use weights to adjust for the fact that the sampling correction was not perfect, and 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 from 2016. To illustrate, Figure [B.2](#) displays the distribution of the year of application in the administrative sample (gray bar) to the density in the unweighted data (dashed) and in the weighted dataset (solid).

Figure B.2: Number of first-time applications (1976–2016)



*Notes:* This figure displays the number of first-time applications by year in the administrative sample (gray bar), in the unweighted sample (dashed line) and in the weighted sample (solid line).

### B.3 Data Collection

The final sample included 10,995 applications. I randomly ordered applications and searched for the first 5,421 applications and found 5,220 of them (142 had been destroyed between the sampling and the data collection, and 59 were lost). Of these 5,220, 447 applications were excluded because they did not meet the criteria of being a first-time application for refugee status with an application form in French and a decision. With the help of a team of research assistants, I digitized 4,773 applications over the course of 21 months. The application form changed a few times over the study period but the questions remained similar. The application form contains questions about the applicant's origin (country of origin, nationality and ethnicity, which was added in the 1990s), marital status 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 the country of origin), languages spoken, military service (date, role and country), and official documents (passport or diplomatic laissez-

passer). Only a few questions were removed from the form. Among them were: “Are you planning on staying permanently in France?” and “Are you registered in the consulate of your country of origin?” 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 further restrict the sample to 4,172 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 Geneva Convention and opened the asylum process to non-European nationalities for events that happened after 1951 (Figure D.1). Moreover, between 1952 and 1976, the acceptance rate at the French asylum office was close to 100 percent.

Table B.1 presents summary statistics on additional variables used in the analysis reported in Table D.1 (column 2). The variable “Accelerated procedure” indicates whether the application was expedited or not. The Prefecture determines whether an application follows the normal or accelerated procedure; the Prefecture 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. “Number of children” corresponds to the total number of children the applicant listed on the form. The variable “A family member is refugee in France” indicates whether the applicant listed a family member currently residing in France as a refugee. Finally, the variables “Military service” and “Arrival in France” code self-reported information by the applicant.

Table B.1: Summary statistics on additional independent variables

	N	Mean	Std. Dev.	Min	Max
Accelerated procedure	4,142	0.072	0.258	0	1
Passport (reported)	4,142	0.254	0.435	0	1
Laissez-Passez (reported)	4,142	0.033	0.179	0	1
Number of children	4,142	0.000	0.000	0	0
A family member is refugee in France	4,142	0.078	0.268	0	1
<i>Military service</i>					
No	4,142	0.687	0.464	0	1
Yes	4,142	0.162	0.369	0	1
Missing	4,142	0.151	0.358	0	1
<i>Arrival in France (reported)</i>					
Irregular	4,142	0.545	0.498	0	1
Regular	4,142	0.158	0.364	0	1
Missing	4,142	0.298	0.457	0	1

*Notes:* This table presents summary statistics on variables used as additional controls in the analysis reported in Table D.1 column 2.

## C Measuring the credibility of the narrative

### C.1 Construction of narrative features

In this section, I describe the construction of the distance measure and the topic proportions reported in Table 2. I start by converting 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 the term frequency in each document weighted by the inverse document frequency (tf-idf). Using this document-term matrix which represents each narrative as a vector of weighted term frequencies for 2,842 different unigrams, I can compute the Euclidean distance (defined as the square root of the sum of squares of differences between corresponding vector elements)

between each narrative and all other narratives from applicants from the same country of origin and take their average, which I use as a measure of the narrative’s originality.

I also use this document-term matrix to estimate a structural topic model with 20 different topics and allow topic prevalence to vary by country of origin. This model associates words with each topic (Table C.1) and estimates the frequency with which each topic is discussed in each narrative. I then code narratives as discussing a topic substantially if the proportion of the narrative discussing the topic exceeds 0.20. I use these estimates to control for the content of the narrative in the empirical strategy.

Table C.1: Topical content

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

*Notes:* This table presents, for each of the 20 topics in the structural topic model, the eight words that scored the highest on the FREX metric (Roberts et al. 2014).

## C.2 Comparing the narratives by religion

In this section, I compare the narratives of Christians vs. Muslims (Table C.2), those with a post-secondary education vs. a secondary or primary education (Table C.3), those who are highly skilled vs. middle or low skill levels (Table C.4), and those who are proficient in French vs. those who are not (Table C.5).

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

	Christians			Muslims			t-test	
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Diff.	p
<i>Level of detail</i>								
Number of words	836.716	823.601	1,426	769.191	735.579	1,679	-67.53	0.034
Number of dates	7.470	8.805	1,435	7.121	7.641	1,681	-0.35	0.311
Number of location	6.796	8.250	1,435	6.452	6.941	1,681	-0.34	0.232
<i>Personal</i>								
Number of personal pronouns	35.641	35.293	1,359	36.261	32.620	1,611	0.62	0.653
<i>Originality</i>								
Distance	49.276	24.412	1,423	48.118	22.016	1,679	-1.16	0.197
<i>Topics</i>								
Family members	0.090	0.287	1,426	0.100	0.300	1,677	0.01	0.375
Bangladesh	0.002	0.040	1,426	0.074	0.262	1,677	0.07	0.000
Student protest	0.086	0.280	1,426	0.061	0.240	1,677	-0.02	0.036
Escape	0.071	0.257	1,426	0.136	0.343	1,677	0.06	0.000
Sri Lanka	0.020	0.138	1,426	0.005	0.068	1,677	-0.01	0.000
RDC	0.112	0.315	1,426	0.000	0.021	1,677	-0.11	0.000
Life in France	0.220	0.415	1,426	0.184	0.387	1,677	-0.04	0.038
Former Yugoslavia	0.048	0.213	1,426	0.123	0.328	1,677	0.07	0.000
China	0.021	0.145	1,426	0.007	0.083	1,677	-0.01	0.008
Ethnic minorities	0.086	0.280	1,426	0.028	0.165	1,677	-0.06	0.000
Kurds in Turkey	0.003	0.057	1,426	0.137	0.344	1,677	0.13	0.000
Political instability	0.088	0.283	1,426	0.023	0.149	1,677	-0.07	0.000
Political opposition	0.023	0.150	1,426	0.107	0.309	1,677	0.08	0.000
Court hearing	0.023	0.150	1,426	0.020	0.141	1,677	-0.00	0.629
Narrative	0.176	0.381	1,426	0.148	0.355	1,677	-0.03	0.045
Angola	0.087	0.282	1,426	0.032	0.176	1,677	-0.05	0.000
Salutation	0.051	0.220	1,426	0.063	0.243	1,677	0.01	0.210
Family at risk	0.041	0.199	1,426	0.056	0.230	1,677	0.01	0.056
Zaire	0.074	0.261	1,426	0.006	0.076	1,677	-0.07	0.000
Religion	0.022	0.146	1,426	0.020	0.141	1,677	-0.00	0.782

Table C.3: Differences between the narratives of highly educated applicants vs. applicants with secondary or primary education

	Post-secondary			Secondary/Primary			t-test	
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Diff.	p
<i>Level of detail</i>								
Number of words	1105.956	1070.318	558	807.142	731.839	2,367	-298.81	0.000
Number of dates	10.296	11.588	560	7.521	7.895	2,388	-2.77	0.000
Number of location	8.026	9.109	560	6.815	7.587	2,388	-1.21	0.005
<i>Personal</i>								
Number of personal pronouns	44.140	41.701	530	36.361	33.395	2,277	-7.78	0.001
<i>Originality</i>								
Distance	51.699	24.489	558	48.213	23.795	2,363	-3.49	0.008



Table C.4: Differences between the narratives of applicants who are highly skilled vs. middle or low levels of skills

	Highly skilled			Middle and Low			t-test	
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Diff.	p
<i>Level of detail</i>								
Number of words	904.296	955.471	564	774.403	743.505	2,307	-129.89	0.012
Number of dates	7.873	9.362	575	7.130	7.682	2,321	-0.74	0.110
Number of location	6.905	8.289	575	6.394	7.397	2,321	-0.51	0.209
<i>Personal</i>								
Number of personal pronouns	35.913	35.805	542	35.471	33.917	2,212	-0.44	0.828
<i>Originality</i>								
Distance	45.592	23.855	564	47.354	23.563	2,302	1.76	0.179

Table C.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
<i>Level of detail</i>								
Number of words	773.338	890.253	1,044	779.811	710.033	2,888	6.47	0.857
Number of dates	6.429	8.764	1,044	7.521	8.105	2,916	1.09	0.004
Number of location	6.378	8.179	1,044	6.507	7.212	2,916	0.13	0.666
<i>Personal</i>								
Number of personal pronouns	32.817	36.155	1,001	35.272	31.750	2,759	2.45	0.098
<i>Originality</i>								
Distance	45.733	25.174	1,042	47.773	22.771	2,885	2.04	0.038

### C.3 Hand coding the credibility of the narrative

In this section, I provide additional information regarding the hand coding of the credibility of the narratives. Table C.6 reports summary statistics on all variables collected by the research assistants. 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. I assess the inter-coder reliability in Table C.7 using Krippendorff’s alpha coefficient. In Table C.8, I show that a positive correlation between different features of the text (as coded by a research assistant) and a binary indicator of whether the coders “agreed” or “somewhat agreed” that

the asylum seeker is entitled to claim the right to asylum, suggesting reasonable internal validity of this measure.

Table C.6: Summary statistics on the narratives

	N	Mean	Std. Dev.	Min	Max
<i>Persecution</i>					
Is persecuted	459	0.797	0.402	0	1
<i>For reasons related to</i>					
Race	459	0.172	0.378	0	1
Political opinion	459	0.418	0.494	0	1
Religion	459	0.063	0.244	0	1
Nationality	459	0.048	0.214	0	1
Social group	459	0.070	0.255	0	1
None	459	0.346	0.476	0	1
<i>Narrative was coded as</i>					
Believable	459	2.597	0.831	1	4
Convincing	459	2.307	0.817	1	4
Detailed	459	2.235	0.992	1	4
Individualized	459	2.399	0.815	1	4
Coherent	459	2.593	0.752	1	4
<i>The narrative mentions</i>					
A historical event	459	0.455	0.499	0	1
Family in France	459	0.109	0.312	0	1
<i>Credibly claims the Geneva Convention</i>					
No	459	0.209	0.407	0	1
Somewhat no	459	0.292	0.455	0	1
Somewhat yes	459	0.353	0.478	0	1
Yes	459	0.146	0.353	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 C.7: Assessment of inter-coder agreement for the measure of the credibility of the narrative using Krippendorff's alpha coefficient

	Indicator if Agree (0-1)	Indicator if Agree or Somewhat Agree (0-1)
Between coder 1 and 2	0.291	0.381
Between coder 1 and 3	0.036	0.623
Between coder 2 and 3	0.120	0.428
Between all three coders	0.186	0.483

Table C.8: Assessing the internal validity of the credibility measure in the set of hand-coded narratives

	(1) Credible Narrative
Is persecuted	0.113* (0.055)
Probable	0.120** (0.036)
Convincing	0.200** (0.039)
Detailed	0.053† (0.028)
Singular	-0.047 (0.030)
Coherent	0.083* (0.034)
Constant	-0.603** (0.069)
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 an ordinary least squares (OLS) regression of the main credibility variable collected in the narrative hand coding on other leading questions focused on features of the text such as the level of detail. The unit of observation is narrative/coder.

## C.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. 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 hand-coded documents into the labeled training set (75 percent) and the labeled test set (25 percent).
- I trained three different algorithms using tenfold cross validation to tune the model and used the tf-idf of unigrams and additional narrative covariates as predictors (e.g., length, number of pronouns, number of dates and locations mentioned, and an indicator for whether one of the 20 topics covers more than 20 percent of a narrative).
- For each of these models, I calculated the predicted probabilities and computed AUC, Brier Score and accuracy (Table C.9), as well as the calibration plots (Figure C.1).

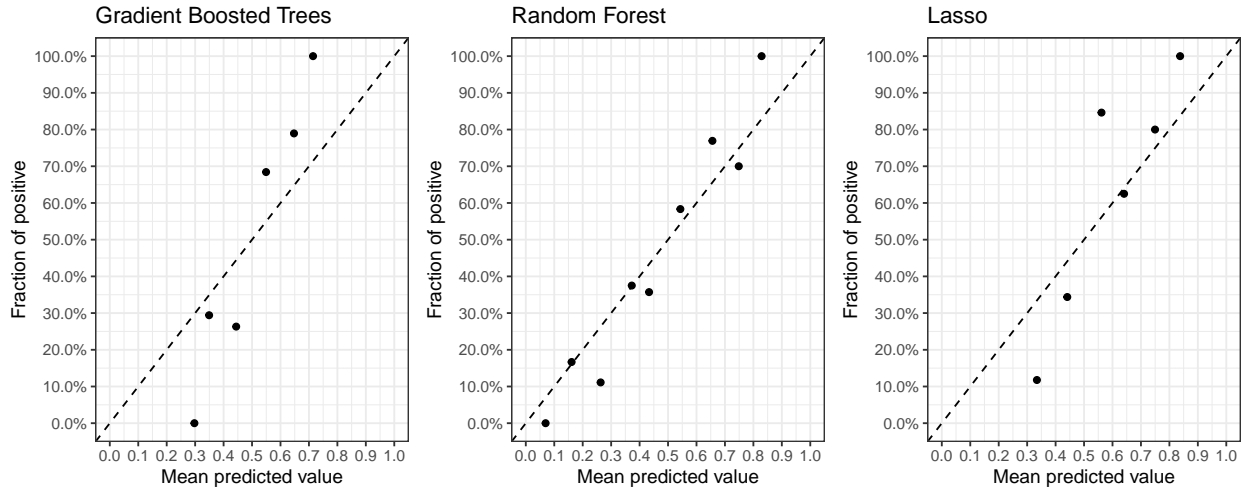
All three algorithms perform relatively well on standard metrics (Table C.9), but predicted probabilities from Random Forest are better calibrated (Figure C.1). Using Random Forest, I am able to accurately predict the quality measure 76 percent of the time in the left-out sample of 81 narratives, a substantial reduction in error compared to a baseline of 48 percent. In addition, Table C.10 shows that the predicted measure of credibility correlates as expected with extracted features of the text.

Table C.9: Model performance

AUC	Accuracy	Brier Score	Model
0.817	0.762	0	Random Forest
0.823	0.750	0	Gradient Boosted trees
0.830	0.762	0	Lasso

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

Figure C.1: Calibration plots



*Notes:* This figure plots the calibration of predicted probabilities from three different algorithms in a left-out sample of 85 narratives. After dividing the predicted probabilities into 10 equally spaced intervals, I compute, within each bin, the average of the predicted probabilities ( $x$ -axis) and the fraction of observations in that bin with a true positive ( $y$ -axis).

Table C.10: Assessing the correlation between the credibility measure and features of the narratives

	Credibility of the narrative	
	(1) Hand Coded	(2) Predicted
Number of words ('000)	0.211* (0.089)	0.255** (0.024)
Distance ('000)	0.259 (1.057)	1.359** (0.299)
# of dates mentioned ('000)	13.996* (6.209)	9.679** (1.331)
# of places mentioned ('000)	-1.544 (3.966)	2.510* (1.220)
Number of personal pronouns ('000)	-1.405 (1.447)	0.254 (0.419)
Constant	0.274** (0.053)	0.186** (0.016)
Observations	320	3472

*Notes:* \*\* $p < 0.01$ , \* $p < 0.05$ ,  $^{\dagger}p < 0.1$ . 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.

## D Additional tables and figures

Figure D.1: Number of first-time applications filed at the French asylum office (1952–2016)

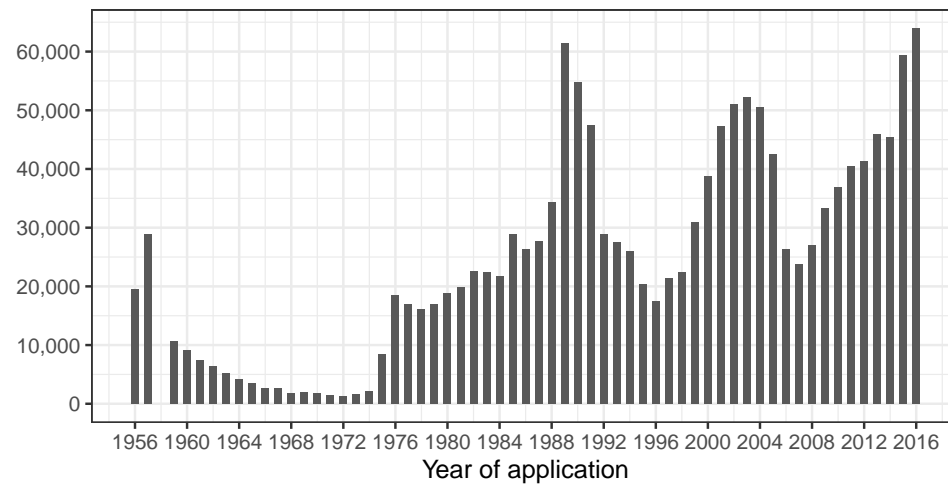
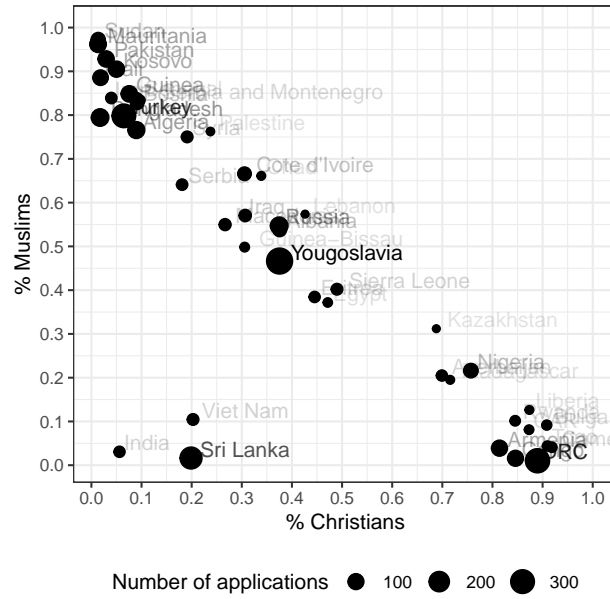


Figure D.2: Proportion of Muslim and Christian applicants by country of origin (1976–2016)





In Table D.1 column 4, I report estimates from the main regression on a matched sample. I used coarsened exact matching (Iacus et al. 2012) to first group applicants from the same country of origin (exact matching) who arrived within up to 4 years of each other (coarsened exact matching). The algorithm next discards all applicants from strata with only Christians or only Muslims, keeping 1,577 applicants from strata with both Muslims and Christians. For each of the remaining units, the algorithm assigns a weight of 1 to Muslims and a weight of  $\frac{m_C}{m_T} \frac{m_T^s}{m_C^s}$  to Christians, with  $m_C$  ( $m_T$ ) being the number of Christians (Muslims) in the matched sample, and  $m_C^s$  ( $m_T^s$ ) the number of Christians (Muslims) in the stratum. I then use a weighted OLS regression to estimate the effect of being a Muslim on the decision, controlling, as before, for all standard covariates except the country of origin.

Table D.1: Determinants of the attribution of refugee status (robustness tests)

	Granted refugee status				
	(1) Main specification	(2) Additional covariates	(3) Country- year interaction	(4) Matched Sample	(5) Omitting credibility
Credible narrative	0.058** (0.013)	0.057** (0.013)	0.050** (0.014)	0.096** (0.017)	
<i>Gender (Ref: Male)</i>					
Female	0.018 (0.014)	0.005 (0.014)	0.001 (0.014)	0.014 (0.020)	0.017 (0.014)
<i>Age (Ref: Less than 20)</i>					
Between 20 and 40	-0.083** (0.024)	-0.064** (0.024)	-0.084** (0.026)	-0.090** (0.028)	-0.086** (0.024)
More than 40	-0.066* (0.028)	-0.057* (0.028)	-0.082** (0.030)	-0.102** (0.035)	-0.061* (0.028)
<i>Marital status (Ref: Single)</i>					
Married	0.061** (0.013)	0.051** (0.013)	0.054** (0.014)	0.114** (0.017)	0.058** (0.013)
<i>Time in France (Ref: &lt; 1 year)</i>					
> 1 year in France	0.003 (0.016)	0.008 (0.016)	0.011 (0.017)	0.022 (0.022)	0.001 (0.015)
<i>Religion (Ref: Christian)</i>					
Muslim	-0.062** (0.020)	-0.050* (0.020)	-0.059** (0.022)	-0.104** (0.017)	-0.055** (0.020)
Other	0.059* (0.027)	0.058* (0.027)	0.078** (0.029)		0.045† (0.027)
None/Missing	-0.016 (0.022)	-0.019 (0.021)	-0.014 (0.024)		-0.013 (0.022)
<i>Education (Ref: University)</i>					
Secondary	-0.021 (0.019)	-0.021 (0.019)	-0.014 (0.018)	-0.024 (0.028)	-0.011 (0.019)
Primary	-0.061* (0.027)	-0.055* (0.027)	-0.067* (0.027)	-0.088* (0.036)	-0.059* (0.027)
None/Missing	-0.032 (0.022)	-0.021 (0.022)	-0.019 (0.022)	-0.022 (0.031)	-0.032 (0.022)
<i>Skill level (Ref: High)</i>					
Middle	-0.066** (0.021)	-0.064** (0.020)	-0.045* (0.021)	-0.020 (0.029)	-0.066** (0.021)
Low	-0.085** (0.022)	-0.083** (0.021)	-0.069** (0.022)	-0.105** (0.031)	-0.090** (0.021)
None/Missing	-0.028 (0.022)	-0.025 (0.021)	-0.005 (0.022)	0.018 (0.030)	-0.033 (0.021)
<i>Speaks French (Ref: Yes)</i>					
No/Missing	-0.033† (0.018)	-0.028 (0.017)	-0.015 (0.018)	-0.054* (0.022)	-0.032† (0.018)
Observations	3,930	3,930	3,930	1,525	4,141
$R^2$	0.358	0.378	0.610	0.226	0.351
Mean acceptance rate	0.152	0.152	0.152	0.129	0.152

Notes: \*\* $p < 0.01$ , \* $p < 0.05$ , † $p < 0.1$ . 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 country-of-origin and year-of-application fixed effects.

Table D.2: Determinants of the attribution of refugee status (logistic regression)

	Average Predicted Probability	[95% Conf. Interval]	
Religion			
<i>Christian</i>	0.153	0.127	0.178
<i>Muslim</i>	0.106	0.088	0.124
Education			
<i>University</i>	0.167	0.132	0.201
<i>Secondary</i>	0.141	0.124	0.157
<i>Primary</i>	0.108	0.079	0.138
<i>None/Missing</i>	0.132	0.109	0.156
Skill level			
<i>High</i>	0.187	0.153	0.222
<i>Middle</i>	0.122	0.104	0.140
<i>Low</i>	0.101	0.077	0.125
<i>None/Missing</i>	0.156	0.133	0.179

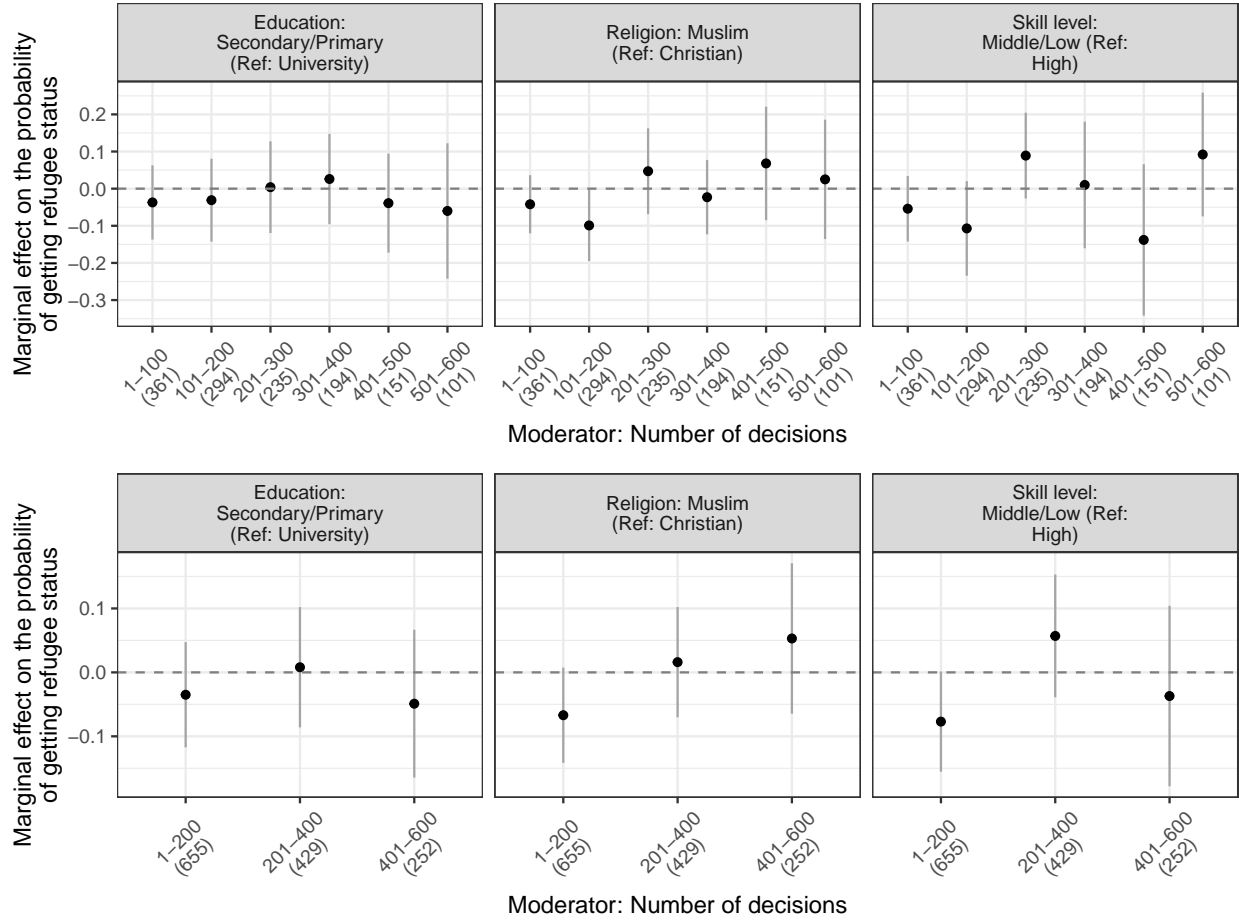
*Notes:* This table reports the average predicted probability of being granted refugee status for different subgroups using estimates from a logit regression model that includes all individual covariates and fixed effects from the main specification. For each of the characteristics listed in the table, I predict the probability of being granted asylum for each applicant in the sample holding all characteristics fixed except the one listed, assigning in turn all applicants to be Christian, Muslim, having some university, etc.

Table D.3: Comparing applications examined by experienced vs. inexperienced bureaucrats

	Inexperienced Bureaucrats			Experienced Bureaucrats			t-test	
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Diff.	<i>p</i>
<i>Level of experience</i>								
Months since first decision	9.065	10.900	607	26.423	15.152	607	17.358	0.000
<i>Age</i>								
Less than 20	0.058	0.234	607	0.056	0.230	607	-0.002	0.856
Between 20 and 40	0.781	0.414	607	0.767	0.423	607	-0.014	0.564
More than 40	0.161	0.368	607	0.178	0.383	607	0.017	0.452
<i>Marital status</i>								
Single	0.546	0.498	607	0.482	0.500	607	-0.065	0.026
Married	0.454	0.498	607	0.518	0.500	607	0.065	0.026
<i>Education</i>								
University	0.163	0.369	607	0.153	0.360	607	-0.010	0.645
Secondary	0.564	0.496	607	0.553	0.498	607	-0.012	0.689
Primary	0.088	0.284	607	0.127	0.333	607	0.039	0.031
None/Missing	0.185	0.388	607	0.167	0.373	607	-0.018	0.419
<i>Religion</i>								
Christian	0.388	0.488	607	0.359	0.480	607	-0.029	0.301
Muslim	0.406	0.492	607	0.417	0.493	607	0.011	0.701
Other	0.092	0.290	607	0.101	0.302	607	0.009	0.595
None/Missing	0.113	0.317	607	0.122	0.328	607	0.009	0.632
<i>Skill level</i>								
High	0.135	0.342	607	0.117	0.322	607	-0.018	0.367
Middle	0.348	0.477	607	0.389	0.488	607	0.040	0.150
Low	0.222	0.416	607	0.225	0.418	607	0.004	0.883
None/Missing	0.295	0.456	607	0.269	0.444	607	-0.026	0.309
<i>Proficiency in French</i>								
Yes	0.261	0.439	607	0.221	0.415	607	-0.040	0.110
No/Missing	0.739	0.439	607	0.779	0.415	607	0.040	0.110
<i>Narrative</i>								
Provided a narrative	0.978	0.146	607	0.975	0.156	607	-0.003	0.698
Credibility of the narrative	0.651	0.477	594	0.671	0.470	591	0.019	0.483
Number of words	988.837	933.337	594	964.368	772.253	591	-24.469	0.631
Distance	52.574	23.560	594	51.219	20.535	591	-1.355	0.300
Number of dates	9.115	8.820	595	9.531	8.188	593	0.417	0.409
Number of location	8.090	9.365	595	8.038	8.376	593	-0.053	0.920

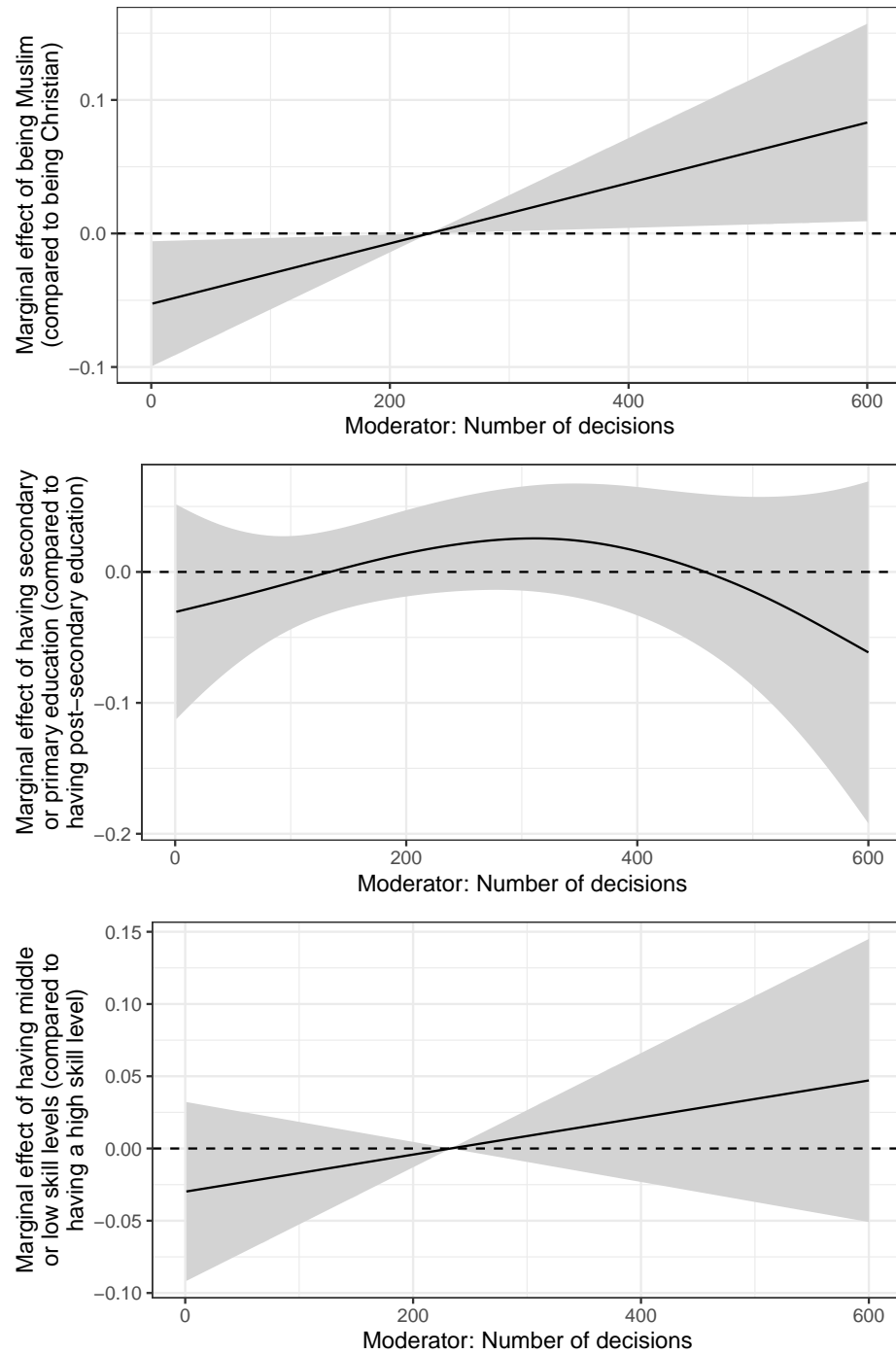
*Notes:* This table presents, for a selection of individual characteristics, the difference in means between applications decided by inexperienced bureaucrats (less than 185 decisions) and experienced bureaucrats (more than 185 decisions) in the subsample of the first 486 applications decided by bureaucrats.

Figure D.3: Discrimination by bureaucrats' experience (robustness)



*Notes:* This figure shows the estimated conditional marginal effect, and 95 percent confidence intervals based on standard errors clustered by bureaucrats, and the number of observations in each bin in parentheses. These conditional marginal effects are estimated by interacting a bin indicator (every 100 cases on the top and every 200 cases on the bottom) with the individual characteristics of interest. All specifications include covariates and fixed effects for year of application and country of origin.

Figure D.4: Discrimination by bureaucrats' experience (robustness)



*Notes:* This figure shows the estimated conditional marginal effect and 95 percent confidence intervals. These conditional marginal effects are estimated using three general additive models that include all the covariates and fixed effects from the main specification and two smooth functions of the number of past decisions, one for each value of each characteristic under consideration.