Gender Bias in Asylum Adjudications: Evidence for Leniency toward Token Women

Alejandro Ecker

University of Mannheim

Laurenz Ennser-Jedenastik and Martin Haselmayer

University of Vienna

Alejandro Ecker, Mannheim Centre for European Social Research, Mannheim Centre for European Social Research;

Laurenz Ennser-Jedenastik, Department of Government, University of Vienna;

Martin Haselmayer, Department of Government, University of Vienna

Earlier versions of this paper were presented at the ECPR General Conference, September 6-9, 2017, University of Oslo (Norway), and at the Conference of Empirical Legal Studies Europe (CELSE), May 31-June 1, University of Leuven (Belgium). Alejandro Ecker gratefully acknowledges financial support from the Austrian Science Fund (FWF) (grant P25490-G22). The authors thank Gerhard Muzak (Department of Constitutional and Administrative Law, University of Vienna) for providing useful background information as to the workings of the Austrian Asylum Court, and Michael Imre for valuable research assistance.

Correspondence concerning this manuscript should be addressed to Alejandro Ecker, University of Mannheim, MZES, 68131 Mannheim, Germany. Email: alejandro.ecker@mzes.uni-mannheim.de

Abstract

Gender is one of the most frequently studied variables in the literature on judicial decision-making. We add to this literature by hypothesizing that the impact of applicant gender is conditional on the gender balance in a judge's caseload. We expect that female applicants receive more favorable decisions from judges whose caseload skews strongly male. Analyzing over 49,000 rulings by the Austrian Asylum Court between 2008 and 2013, we find support for direct gender effects for applicants and judges (yet no significant interaction between the two). We also show that gender balance in the caseload is a strong moderator of applicant gender. Judges with predominantly male caseloads are strongly biased toward female applicants, whereas judges facing a gender-balanced set of applicants display hardly any gender bias at all. These findings tackle essential questions of democratic rule of law and human rights. They indicate that applicants' fundamental rights to a fair and equal trial may have been compromised. We discuss institutional remedies to reduce the potential for gender bias in Austrian asylum adjudication.

Keywords: Political Asylum, Asylum seeking, Gender Gap, Gender Equality, Adjudication, Human rights, Sexism

Gender Bias in Asylum Adjudications: Evidence for Leniency toward Token Women

Gender is one of the most fundamental social variables. It strongly shapes every human domain from private and family life to workplace environments to public life and politics. It is therefore little surprise that scholars of judicial decision-making have often examined the impact of gender on judicial outcomes. The most established hypotheses in this field concern the effects of judges' gender, applicants' gender, and the interaction between the two (Boyd 2016; Etienne 2010; Mustard 2001).

We contribute to this literature by proposing a novel hypothesis. Drawing on tokenism theory (Kanter 1977; King et al. 2010; Yoder 1991), we argue that the impact of applicants' gender is dependent on the social context in which decisions are made. This is because gender becomes more conspicuous as a social characteristic in contexts where the proportions of men and women in a group are strongly skewed. As a simple consequence of such uneven numbers, women will stand out more than men do in male-dominated contexts (and vice-versa), making their gender a more salient feature for their environment. Applying the logic of this social mechanism to the judicial realm, we hypothesize that the effect of applicant gender on judicial decision-making will increase with the gender skew in a judge's caseload. We use data from the Austrian Asylum Court between 2008 and 2013 and exploit the large variation in the gender distribution of refugee populations from different countries of origin to examine the interaction between applicant gender and gender skew in judges' caseloads.

Gender Effects in Judicial Decision-Making

The empirical study of judicial decision-making has been dominated by research focused on the United States. Still, European and other countries have recently become of interest to judicial politics scholars (Bogoch 1999; Hangartner et al. 2016; Hanretty 2012a, b; Voeten 2007). One of the most studied variables, certainly in lower court decision-making, is the gender of both judges and applicants coming before the courts. To this literature, we add a new proposition about gender proportions as social context. Based on tokenism theory and studies of human behavior in groups with strongly skewed proportions of men and women (Kanter, 1977; King et al. 2010; Yoder 1991, 1994), we argue that the gender balance in a judge's caseload moderates the impact of applicant gender.

Effects of Judges' Gender

The literature on gender and judging has produced a number of theoretical arguments, some claiming that we should expect no systematic gender differences (Kritzer and Uhlman 1977; Spohn 1991; Steffensmeier and Hebert 1999) and others arguing that gender effects should be limited to specific (i.e., gender-typed) areas such as abortion, sex discrimination, or sexual violence (Peresie 2005). Other approaches claim that men's and women's experience of and outlook on the world are fundamentally different, such that we should expect gender differences in judicial decision-making to emerge across a wide range of topics (Gilligan 1982). For a concise discussion of these different arguments, see Boyd et al. (2010).

Although much research has been produced over the past decades, the empirical evidence on gender effects in judicial behavior is still mixed. Especially in earlier studies (when judges were still predominantly male), gender differences were often found to be statistically

insignificant (Davis 1986; Gottschall 1983; Kritzer and Uhlman 1977; Myers and Talarico 1987; Spohn 1991). Other studies did find clearer evidence for gender effects, such as women judges being more punitive in sentencing (Gruhl et al. 1981; Steffensmeier and Hebert 1999), but also being more likely to arrive at intra-court settlements than their male colleagues (Boyd 2013). A number of studies have looked specifically at gender differences in strongly gender-typed areas of the law, such as sexual harassment and violence, workplace discrimination, or abortion. Such studies have often found strong evidence of gendered judicial behavior (Boyd 2016; Boyd et al. 2010; Peresie 2005; Songer et al. 1994).

The topic at hand—immigration and asylum law—has not been studied as extensively, yet the studies that exist have typically reported significant gender differences. A common argument in those studies is that female judges are more likely to have personal experience of discrimination and thus find it easier to empathize with immigrants and asylum seekers (Keith et al. 2013). Ramji-Nogales et al. (2007), for instance, found an odds ratio of 1.61 for the variable of female judge (see online appendix to their paper), meaning that the odds of a positive over a negative decision for the asylum applicant increased by over 60% in the presence of a female judge (controlling for a host of applicant and other judge characteristics).

Keith et al. (2013) also reported significant differences in asylum grant rates between male and female judges. Their analysis of over 300,000 decisions of U.S. immigration judges between 1997 and 2004 yielded odds ratios of around 1.43 for female judges. Similarly, Miller et al. (2015) found that female judges grant asylum applications at higher rates than men, with an odds ratio of 1.14. Whereas the available research finds mixed effects of gender, the studies that have examined immigration and asylum adjudication report significant differences in the judicial decision-making of male and female judges. In addition, these effects all point into the same

direction. This leads us to our first hypothesis: Female judges are more likely than male judges to rule in favor of the applicant in asylum-seeking cases (Hypothesis 1).

Effects of Applicant Gender

Aside from differences between male and female judges, disparities in judicial behavior have also been found between cases involving male and female applicants (Bontrager et al. 2013; Etienne 2010). Spohn (2013) confirmed a number of hypotheses about differences in judicial outcomes for men and women. Male offenders were more likely to be held in pre-trial custody, less likely to receive "substantial assistance departure" (i.e., a lower sentence after providing assistance to prosecutors), and were given longer sentences. Furthermore, these effects were likely to reinforce each other because pre-trial custody and substantial assistance departure also affected sentencing.

In a study of drug offense cases in the early 1990s, Albonetti (1997) found that, controlling for an individual's criminal history and the characteristics of the offense, women were treated with greater leniency than men and that this gender effect varied somewhat across different racial and ethnic groups. The same conclusion can be drawn from Mustard's (2001) examination of over 77,000 U.S. federal sentences. Men are punished more severely and are given more upward and fewer downward adjustments of their sentences. In addition, the effects of offense level and criminal history on sentence length are smaller for female than for male offenders.

Similar effects to those we described were reported by Schanzenbach (2005) and Starr (2014) who demonstrated that much of the disparities in sentencing between male and female offenders was due to differences during charging, plea-bargaining, and fact-finding. Although

extant research is largely confined to U.S. cases, the evidence is strongly in favor of hypothesizing an effect of applicant gender. This proposition is also in line with attitudes in the general population. Female immigrants and asylum seekers are generally evaluated more favorably than male ones (Bansak et al. 2016; Hainmueller and Hopkins 2015). Our second hypothesis captures this expectation: Female applicants are more likely than male applicants to receive favorable decisions (Hypothesis 2).

Gender Interaction Effects

Given that there is evidence for gender effects at the levels of judges and defendants, it is not surprising that there may be interaction effects between the two. After all, if gender affects human behavior, the combination of judge and defendant genders may be an important explanatory factor beyond the direct effects. One typical assertion has been that of male chivalry, paternalism or 'benevolent sexism' (Herzog and Oreg 2008). Male judges may show greater leniency toward female offenders whereas such a difference may not exist (or only to a lesser extent) for female judges.

Evidence for exactly this effect has been reported by Gruhl et al. (1981). These authors found that male judges were considerably less likely to give prison sentences to women, whereas women judges treated male and female defendants equally (and, overall, female judges were slightly more likely to give prison sentences). A similar result emerged from Schanzenbach's (2005) analysis who operationalized judge gender with the proportion of women and men judges in a district. Although the direct effect of defendant gender was in the expected direction (i.e., women receiving shorter sentences), this effect diminished as the proportion of female judges in a district went up. This evidence is consistent with the notion that male judges treat female

defendants more favorably, whereas female judges rule in a more gender-neutral manner. This is the logic underpinning our third hypothesis: The impact of applicant gender is conditional on judge gender. Women applicants are treated more favorably than men are by male judges, but not by female ones (Hypothesis 3).

An alternative theory to explain gender differences in judicial behavior and outcomes is to consider the immediate social context in which judges operate. One way to theorize this social context is to think about the interaction of gender at the individual and the group level (Burns 2007). Over time, the hundreds or even thousands of cases that judges decide not only generate experience, but also produce a frame of reference through which each new case will be evaluated.

Tokenism theory (Kanter 1977; King et al. 2010; Laws 1975; Yoder 1991) provides us with an argument about how this social and experiential context affects individual cases. It posits that there is a relationship between the quantitative distribution of a social characteristic (e.g., gender) in a group and the way this characteristic shapes individuals' behaviors. Tokenism theory has been most prominently applied in occupational sociology, typically to examine women in predominantly male environments or ethnic and racial minority individuals in majority-dominated settings. Being a token woman within a mostly male group increases visibility, social isolation, and role encapsulation (King et al. 2010; Yoder 1991, 1994), thus pushing women into stereotypically female behaviors. Behavioral differences between men and women should therefore be greater when the gender distribution in a group skews strongly male.

We posit that a similar mechanism will be at work for judges who evaluate large numbers of cases. The overall distribution of characteristics in a judge's caseload establishes what a typical case looks like. For example, if 90% of applicants are male, male gender is more likely to

be viewed as a typical characteristic for a group of applicants. By contrast, if 50% of applicants are male, the typical applicant in the perception of a judge will not be clearly defined by gender.

The typical case as established by the statistical distribution of applicants' characteristics thus provides an important point of reference—a baseline against which judges will (unconsciously) evaluate the merits of individuals' claims. To put it in psychological terms, a judge's caseload can be viewed as the frame within which decisions are made. Given the power of framing on human decision-making (Kahneman and Tversky 1984; Tversky and Kahneman 1981), it is reasonable to assume that judges will be influenced by how the social context in which they operate frames individual applications.

As a result, characteristics that are quantitatively dominant are less likely to be viewed as conspicuous (because they are typical of the group of applicants), whereas those that are rare are more likely to stand out. Individuals who possess non-typical characteristics are therefore more likely to be viewed as different from the typical case. Judges may thus find it easier to justify non-typical decisions for such individuals. This logic could, in theory, apply to all types of characteristics (e.g., age, race/ethnicity, religion, language, sexual orientation). However, it is more likely to be empirically relevant when the characteristic in question is (a) highly visible and (b) empirically linked to a specific behavior—in this case the amply documented bias to decide cases in favor of female applicants.

Given that most cases in our analysis were decided negatively (for the asylum applicants), we argue that individuals who stand out more against the other cases decided by the same judge should be more likely to receive the non-typical (i.e., positive) outcome. More specifically, our hypothesis is that the impact of applicant gender is conditional on the gender distribution within a judge's caseload. In addition to a general expectation of greater leniency toward women

applicants, we also assume that women will be treated increasingly more favorably as the proportion of male applicants in a judge's caseload rises. Thus we predict that the impact of applicant gender is conditional on the gender distribution within a judge's caseload. Women applicants will be treated more favorably as the proportion of male applicants in a judge's caseload increases (Hypothesis 4).

Asylum Appeals and the Austrian Asylum Court

We test our expectations about gender bias on cases adjudicated by the Austrian Asylum Court (AC, Asylgerichtshof) which existed between 2008 and 2013. Austria grants asylum in accordance with the 1951 Geneva Convention and the 1967 Protocol Relating to the Status of Refugees. During our observation period, all asylum applications were first handled by the Federal Asylum Office (Bundesasylamt).

In case of a negative decision, applicants may appeal against these decisions within 2 weeks. Appeals are then processed by the Asylum Court. The AC was introduced in July 2008 to replace the Independent Federal Asylum Review Board (UBAS, Unabhängiger Bundesasylsenat) and remained the last instance for asylum trials until 31 December 2013 when it was replaced by the Federal Administrative Court (Bundesverwaltungsgericht). A major political objective of the introduction of the AC was to accelerate the legal process by limiting asylum applications to a two-stage process: the Federal Asylum Office and the Asylum Court. Decisions by the AC could not be appealed at the Supreme Administrative Court (Verwaltungsgerichtshof). The Asylum Court thus represented the last resort for refugees applying for asylum in Austria (unless constitutional rights were violated, in which case an appeal could be made to the Constitutional Court). As of 2014, decisions by the Asylum Office can be appealed first at the Federal

Administrative Court and further at the Supreme Administrative Court, thus returning to the three-stage process that was in place before 2008.

Asylum Court and Appeal Procedure

The Austrian president appointed all 81 judges serving at the AC after unanimous nomination by the federal government (§ 2 (2) BGBl 4/2008). Formally, besides Austrian citizenship, judges must hold a law degree and have 5 years of work experience (§ 2 (3) BGBl 4/2008). The AC groups its judges into five regional boards (chambers) broadly representing Africa (A), the Western Balkan states (B), Asia (C), countries of the former Soviet Union (D), and the Middle East and Turkey (E) as countries of origin. A sixth board (chamber S) examines special cases—cases which supposedly fall outside Austrian jurisdiction.

The general assignment of duties follows a predetermined procedure. At the beginning of each year, the AC's general board distributes cases for the following 12 months (§ 13 BGBl 4/2008) (Muzak and Rohrböck 2008). Within the six chambers, cases are assigned to judges according to the applicant's surname and the year of the initial asylum application. For instance, for board A and year Y all applicants with surnames starting with letters A to C will be decided by judge 1, D to E by judge 2, and so on. Beyond that, the allocation of cases seeks to distribute similar workloads to all judges. Because we assume that surname initials are, in principle, uncorrelated with other applicant characteristics, we treat the assignment of cases to judges as a stratified (quasi-) random procedure, conditional on the applicant's country of origin (which determines assignment to one of the AC boards).

Difficulties may arise, however, when characteristics such as ethnic origin or religious affiliation correlate with surnames. Indeed, medical research based on secondary data, for

instance, often exploits surnames as a proxy for unobserved ethnic and racial characteristics. To the extent that particularly persecuted ethnic or religious groups cluster on selected surname initials, this challenges the assumption of quasi-random assignment of cases that is independent of the merits of the case. The different ruling of judges may then simply result from systematic differences of applicant characteristics. Note, however, that this does not immediately bias our estimates of gender effects in the ruling of the Austrian AC. In fact, a systematic bias would additionally require that the judges working cases with surname initials of ethnic or religious minorities are predominantly male or female. Based on qualitative interview evidence, it seems more warranted to assume that the allocation of surname initials is uncorrelated with judge gender. Thus, although the research design may not perfectly control for all unobserved characteristics, we argue that it provides largely unbiased estimates of systematic differences between male and female judges.

There are three different settings for trials: (a) in ordinary procedures, committees of two judges decide unanimously whether to grant or reject asylum and subsidiary protection or to refer cases back to the first instance. If the two judges disagree, (b) the standing order requires the formation of a larger committee of five judges, who decide by simple majority vote. Yet, in the vast majority of all cases, (c) single judges decide about the sort of refugees. These are mainly cases in which applicants appeal against first-instance decision of removal due to the declared safety of a third country (§ 4 AsylG 2005) or the Dublin procedure (§ 5 AsylG 2005), and cases in which the AC decides whether a renewed application is admissible. (“res iudicata” § 10 AsylG 2005). The present paper focuses on these cases.

It is worth noting that the gender balance among Austrian judges has shifted dramatically during the past two decades. Only one in four judges was female in the mid-1990s. Their

numbers have been continuously increasing over the years, and today, the share of female judges stands at well above 50% as reported by the Federal Ministry of Justice (www.justiz.gv.at). Also, Eurostat data for the year 2016 reveal that there is stark variation in the gender make-up of refugee populations arriving in Austria (Eurostat 2019). Overall, one third of applications in 2016 were filed by women, yet among the most important countries of origin (only those with a total of 300 or more applications in 2016 are reported here), this proportion varies from the low single digits (Algeria, Morocco, Pakistan) to over 50% (Russia, Syria, and stateless individuals). This suggests that, depending on the countries or regions of origin that judges specialize in, their caseload will display substantial variation in terms of gender balance.

Method

We tested our theoretical expectations based on 40,980 adjudications by the AC in its 5 years of existence between 1 July 2008 and 31 December 2013. We accessed and extracted all relevant information via the Federal Chancellery's legal database (https://www.ris.bka.gv.at) that provides machine-readable transcripts of all AC adjudications in HTML format. These transcripts are highly standardized and thus allow retrieving valid data on all relevant characteristics via string search and regular expression functions using *R*. The dependent variable indicates whether an appeal against a negative decision on asylum by the Federal Asylum Office with the Austrian AC is successful (coded 1) or unsuccessful (0).

In this context, we excluded all adjudications where appeals have a simple suspensory effect with no ultimate decision on the applicant's residency status. In a similar vein, we did not explore the AC's orders (in contrast to verdicts) because they are primarily concerned with (non-substantial) corrections of initial verdicts, mostly related to the name and country of origin of the

respective applicant. Concerning the general rate of successful appeals and thus the distribution of the dependent variable, we find that approximately 31% of all applicants (12,757 of 40,980) were granted a positive decision.

Regarding the key independent variables, each document provides detailed information on the judge dealing with the respective appeal, which allows exploring any systematic influence of judge characteristics on application grant rates. In total, 81 judges ruled in the Austrian AC between 2008 and 2013 with an average caseload of approximately 90 cases per judge per year. Concerning the first characteristic of theoretical interest—*judge gender*—we observe an almost perfectly balanced gender distribution with 39 of 81 being female judges over the 5-year period. A second relevant characteristic of individual judges is the distribution of *applicant gender* in their yearly caseload. Here, we observe substantial variation with some judges facing exclusively male applicants over the course of one year (no female applicant per year) whereas others face a female-dominated group of applicants (63.6% of female applicants per year). On average, the yearly share of female applicants across judges is 27.8% of all applicants.

A final trait at the level of individual judges is their assigned chamber. As we discussed, each judge rules exclusively in one of the six different chambers, all of which specialize on certain geographical areas and in which all cases are quasi-randomly assigned (via surname initials) to individual judges. In order to account for the stratification by chamber in the assignment, all empirical models feature chamber fixed effects. To the extent that the merits of the cases before the court are uncorrelated with the assignment of cases within these geographically specialized chambers, this allows us to estimate the effect of individual judge characteristics on grant rates. All empirical findings are robust to several alternative fixed effects model specifications that further take into account the hierarchical structure of the data.

Specifically, Table 2s in our online supplement reports the results of empirical models with chamber-year fixed effects and fixed effects for the country of origin of the applicants.

With regard to the characteristics of an applicant filing an appeal, we are primarily concerned with the *applicant's gender on appeal*. Although all publicly available transcripts of AC adjudications are anonymized due to data privacy, we exploit both definite articles and pronouns in the German language, which differentiate between male and female persons and thus easily allow inferring each applicant's gender. Approximately 700 transcripts which provide either no or ambiguous information and thus do not allow clearly inferring the applicant's gender are excluded from the analysis. According to the data on AC adjudications, approximately 28% (11,373 of 40,980) are female applicants. Finally, we control for the use of *legal assistance* by applicants. In fact, we cannot safely exclude the probability that applicants may anticipate any systematic differences in grant rates between judges and thus seek legal assistance in those instances in which they face a judge who is generally known to be less favorable to their claim. Based on the data on AC adjudications, 14% of the applicants sought legal assistance. Table 1s in the online supplement provides a detailed overview of the distribution of the independent variables.

Concerning the choice of our empirical model, we specified two binary logistic regression models with clustered standard errors at the level of individual judges to account for dependence of observations in the data. The first linear-additive model tested for any direct gender effects. The second multiplicative model examined the interaction effects between judge and applicant gender, as well as applicant gender and the gender distribution within a judge's caseload, which is of primary interest for us.

Results

Hypothesis 1 and 2: Female Judges and Applicants

Table 1 presents the results of our empirical analysis. The columns of Model 1 depict the results of a simple linear-additive model specification. Model 2, in turn, presents the results of a multiplicative model. As postulated (Hypothesis 1), female judges were significantly more likely to repeal an initial negative decision and to grant either subsidiary protection, political asylum, or refugee protection (see Table 1a). It is worth reiterating that the stratified random assignment within legal chambers minimizes the probability that any merits of the case affect this substantial positive effect of female judges on successful appeals before the Austrian AC.

We also observed that female applicants were significantly more likely to be granted any of the three above-mentioned types of legal status, supporting Hypothesis 2 (see Table 1a). Thus, ceteris paribus, females appealing against their first instance asylum denial had a significantly higher likelihood to be successful compared to male applicants. Regarding the empirical controls, we found that applicants with legal assistance were neither more nor less likely to appeal successfully against their initial negative asylum decision. Finally, the results of Model 1 suggest that there were substantial differences in grant rates among the six different legal chambers. Naturally, these differences are to be expected because cases were initially distributed across chambers based on applicants' geographical origin.

In order to assess the substantive effect of the gender of both judges and applicants in appeal cases, Table 2a displays estimated average marginal effects of all covariates included in Model 1. Based on these estimates, the probability of a positive adjudication at appeal increased by approximately 9 percentage points if applicants faced a female in contrast to a male judge. Put

differently, the likelihood of successfully appealing an initial asylum decision (i.e. the predicted probability) increased by approximately 30% for applicants facing a female judge. Similarly, female applicants also were more likely to successfully appeal before the Austrian AC. Their likelihood of a successful appeal was approximately 9 percentage points higher compared to male applicants. In sum, Tables 1a and 2a provide strong empirical support for a significant and substantial effect of both judge and applicant gender on the likelihood of a successful decision. The results of the linear-additive Model 1 thus corroborate both Hypothesis 1 and 2.

Hypothesis 3 and 4: Conditional Effects of Female Judges and Applicants

Having explored the unconditional effect of judge and applicant gender, we now examine potential interaction effects between these two factors, as well as between applicant gender and the overall share of female applicants in each judge's caseload. In fact, our initial results suggest that both the gender of the respective judge and that of the applicant have a considerable influence on applicants' success rates. Yet, the question is whether these two characteristics interact in a meaningful way (i.e., whether male judges are generally less likely to grant asylum status or whether they also discriminate more heavily against male applicants than female judges do). Accordingly, we also explored whether female judges treat male applicants more favorably than female ones.

The interactive Model 2 in Table 1b tests these theoretical propositions on cross-gender effects. Refuting this theoretical assertion specified in Hypothesis 3, the empirical results indicated that applicant gender does not significantly moderate the effect of judge gender. Thus, female judges were more likely to grant asylum status to applicants, but they did not discriminate positively or negatively against applicants of their own gender. Vice versa, male judges were less

likely to repeal an initial negative decision, irrespective of whether they faced a male or female applicant.

At the same time, the results of the interactive Model 2 indicate that the effect of applicant gender is contingent on the overall share of female applicants in each judge's caseload. In line with Hypothesis 4 derived from tokenism theory, the empirical results suggest that the positive effect of being female was substantially reduced to the extent that judges face an increasing share of female applicants. Table 2b shows the decreasing positive effect of being a female applicant across the empirical range of the overall share of female applicants. As apparent from the scheme, being female when facing a judge who very rarely deals with female applicants increased the probability of a positive outcome by approximately 17 percentage points. This positive gender effect decreased with an increasing gender balance in the judge's caseload. As a result, the same female applicant effectively enjoyed no advantage over male applicants if the judge had a balanced applicant gender caseload.

Table 3 summarizes this substantial interaction effect between applicant gender and the gender balance of judge's caseload. Based on the results of Model 2, female applicants enjoyed a considerable advantage over male applicants if a judge normally rules over male applicants (10th percentile; 8% of female applicants). In this context, the likelihood of a positive outcome increased from approximately 0.25 to 0.43. In contrast, in situations where judges were used to ruling over female applicants (90th percentile; 46% of female applicants), there was only a marginal difference in the success rate of male and female candidates (0.34 and 0.39, respectively).

Robustness of Empirical Findings

Overall, our empirical results suggest a substantive bias toward female applicants. Specifically, we found both a strong direct effect and an effect moderated by the gender distribution within a judge's caseload. Yet, one fundamental difficulty in assessing effects of applicant gender is the fact that gender may be correlated with other unobserved characteristics of applicants, most importantly, the actual merits of their case. Whatever the underlying reasons, the level of persecution or hardship necessary to propel an individual to migrate away from their country may be higher for women than for men. As a result, the average female refugee may be able to present a more convincing case before the receiving country's authorities than the average male refugee may. If this were the case, the effects of applicant gender would be spurious and would vanish if it were possible to control perfectly for the merits of each individual application.

We thus provide a series of additional analyses in the online supplement to assess the robustness of our results. Specifically, we (a) identify cases with comparable merits based on standardized keywords provided in the judgments, (b) use a reduced set of cassation appeals with procedural issues, and (c) conduct an instrumental variables estimation, with distance between Austria and the applicants' countries of origin as an instrument for applicant gender. This last design exploits the fact that the proportion of female applicants shrinks with distance. The results obtained corroborate the analysis we present here: There is a large and highly significant effect of applicant gender on judges' verdicts, and the interaction model shows the same decrease in the applicant gender effect as a judge's caseload skews less male.

Discussion

Analyzing over 49,000 rulings by the Austrian Asylum Court between 2008 and 2013, we have investigated whether judge and applicant gender impact the success rates of appeals against first-order asylum adjudications. We find support for direct gender effects for applicants and judges, but no significant interaction between the two. We also demonstrate that the gender balance in the caseload is a strong moderator of applicant gender effects. Judges with predominantly male caseloads have a strong bias in favor of female applicants, whereas judges facing a gender-balanced set of applicants display almost no gender preference.

The effect of gender balance is substantive and a number of additional tests corroborate our claim that differences in successful asylum applications are not dependent on variation in the case merits of individual refugees. Our study does not evaluate the quality (or correctness) of individual decisions. Our results suggest neither that women are treated too favorably nor that men receive unjustified sentences according to the gender balance of the judges' workload. Yet, our analyses reveal that men and women are treated unevenly and that an individual's chances to obtain a positive decision do not depend on case merits alone.

Limitations and Future Research

The present paper analyzes asylum adjudications in Austria over a restricted period of time, which obviously limits the generalizability of our results. However, beyond these limitations in time and scope, evidence from other European countries and the United States seems to suggest that gender bias at (high) asylum courts could be a widespread issue that should attract more attention from researchers in the future. Such research could replicate our findings in countries with similar judicial traditions and values to examine whether gender bias at the

courts is a general trend. With regard to our findings on the gender distribution within a judge's workload, a comparative research design could examine if particular decision-making rules or (mixed) panels may tackle gender bias in asylum adjudications. Moreover, qualitative studies, for example based on in-depth interviews with judges, could be helpful to reveal some of the causes that produce the types of biases observed in our study. Such evidence could be particularly helpful in order to elaborate practical solutions to gender bias trough institutional design or gender sensibility training.

Practice Implications

The present findings tackle fundamental questions of democratic rule of law and human rights. Fair trials with independent and impartial judges are fundamental to all democratic constitutions and also guaranteed by Article 6 of the European Convention on Human Rights, which is incorporated in many European constitutions, including in Austria. Thus, democratic countries are committed to designing and providing legal institutions that ensure fair trials. Accordingly, cases with comparable merits (such as applicants from the same country of origin) should yield similar decisions. The AC was the final instance and thus the last resort for individuals seeking asylum in Austria from 2008 to 2013 (unless an applicant's constitutional rights were violated). Negative decisions in trials at the AC typically resorted in expulsion, and thus they had tremendous consequences for the applicants. The results of our analyses indicate that applicants' fundamental rights to a fair and equal trial may have been compromised.

By the end of 2013, the AC was abolished and the Federal Administrative Court newly created, which addressed out some of the fundamental problems pointed out by legal scholars (e.g., by reintroducing the opportunity to appeal at the Supreme Administrative Court). Yet, the

basic organizational rules and the distribution of cases among judges have not been adapted. Also, almost all AC judges went on to join the new Supreme Administrative Court. Our findings illustrate the need for further institutional adaptions to reduce systematic bias in the adjudication of asylum.

The easiest and most obvious remedy would be to formally try to ensure as even as possible a gender balance in the caseload of individual judges. This implies that the pre-distribution of cases among individual judges would have to take into account applicant gender. Our findings also put the allocation of final-instance rulings to single judges into question. Despite the potential acceleration of trials and thus a potential increase in judicial efficiency, which was one of the main motivations for implementing the Austrian Asylum Court, the strong (gender) bias of individual judges provides good arguments against such solutions. The literature on panel effects in judicial decision-making suggests that mixed-gender panels may be one way to mitigate gender bias (Boyd et al. 2010; Farhang and Wawro 2004).

Conclusion

Gender shapes every human domain, from private and family life to workplace environments to public life and politics. Our paper demonstrates that gender is an important factor in asylum adjudications. Confirming prior evidence from the United States, it shows that both judge and applicant gender may strongly impact the sort of individuals seeking asylum. Proposing a novel hypothesis drawing on tokenism theory, we also find that the visibility of gender attributes has a powerful impact: the effect of gender is strongest when the distribution of women in a judge's workload is particular low. In order to provide fair and equal trials in compliance with the European Convention on Human Rights, judicial institutions should address

any potential bias. In the case of gender differences, possible remedies include a case allocation accounting for equal gender balances among judges and decisions by mixed-gender panels.

References

Albonetti, C. A. (1997). Sentencing under the federal sentencing guidelines: Effects of defendant characteristics, guilty pleas, and departures on sentence outcomes for drug offenses, 1991-1992. *Law and Society Review, 31*(4), 789-822. https://doi.org/10.2307/3053987

Bansak, K., Hainmueller, J., & Hangartner, D. (2016). How economic, humanitarian, and religious concerns shape European attitudes toward asylum seekers. *Science,* 354(6309), 217-222. https://doi.org/10.1126/science.aag2147

Bogoch, B. (1999). Judging in a 'different voice': Gender and the sentencing of violent offences in Israel. *International Journal of the Sociology of Law, 27*(1), 51-78. https://doi.org/10.1006/ijsl.1999.0081

Bontrager, S., Barrick, K., & Stupi, E. (2013). Gender and sentencing: A meta-analysis of contemporary research. *Journal of Gender, Race & Justice, 16*(2), 349-372.

Boyd, C. L. (2013). She'll settle it? *Journal of Law and Courts, 1*(2), 193-219. https://doi.org/10.1086/670723

Boyd, C. L. (2016). Representation on the courts? The effects of trial judges' sex and race. *Political Research Quarterly, 69*(4), 788-799. https://doi.org/10.1177/1065912916663653

Boyd, C. L., Epstein, L., & Martin, A. D. (2010). Untangling the causal effects of sex on judging. *American Journal of Political Science, 54*(2), 389-411. https://doi.org/10.1111/j.1540-5907.2010.00437.x

Burns, N. (2007). Gender in the aggregate, gender in the individual, gender and political action. *Politics & Gender, 3*(1), 104-124. https://doi.org/10.1017/S1743923X07221014

Davis, S. (1986). President Carter's selection reforms and judicial policymaking: A voting analysis of the United States courts of appeals. *American Politics Quarterly, 14*(4), 328-344. https://doi.org/10.1177/1532673X8601400404Etienne, M. (2010). Sentencing women: Reassessing the claims of disparity. *Journal of Gender, Race and Justice, 14*(1), 73-84.

Eurostat. (2019). Asylum and first time asylum applicants by citizenship, age and sex Annual aggregated data (rounded) (migr_asyappctza). [Data explorer]. Retrieved from http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=migr_asyappctza&lang=en

Farhang, S., & Wawro, G. (2004). Institutional dynamics on the US court of appeals: Minority representation under panel decision making. *Journal of Law, Economics, and Organization, 20*(2), 299-330. https://doi.org/10.1093/jleo/ewh035

Gilligan, C. (1982). *In a different voice*. Cambridge, MA: Harvard University Press.

Gottschall, J. (1983). Carter's judicial appointments: The influence of affirmative action and merit selection on voting on the US Courts of Appeals. *Judicature, 67*, 164-173. https://heinonline.org/HOL/Page?handle=hein.journals/judica67&div=42&g_sent=1&casa_token=&collection=journals

Gruhl, J., Spohn, C., & Welch, S. (1981). Women as policymakers: The case of trial judges. *American Journal of Political Science, 25*(2), 308-322. https://doi.org/10.2307/2110855

Hainmueller, J., & Hopkins, D. J. (2015). The hidden American immigration consensus: A conjoint analysis of attitudes toward immigrants. *American Journal of Political Science, 59*(3), 529-548. https://doi.org/10.1111/ajps.12138

Hangartner, D., Lauderdale, B. E., & Spirig, J. (2016). *Refugee roulette revisited: Judicial preference variation and aggregation on the Swiss Federal Administrative Court 2007-2012*. Working Paper. Retrieved from https://ssrn.com/abstract=2629290

Hanretty, C. (2012a). The decisions and ideal points of British Law Lords. *British Journal of Political Science, 43*(3), 703-716. https://doi.org/10.1017/S0007123412000270

Hanretty, C. (2012b). Dissent in Iberia: The ideal points of justices on the Spanish and Portuguese Constitutional Tribunals. *European Journal of Political Research, 51*(5), 671-692. https://doi.org/10.1111/j.1475-6765.2012.02056.x

Herzog, S., & Oreg, S. (2008). Chivalry and the moderating effect of ambivalent sexism: Individual differences in crime seriousness judgments. *Law & Society Review, 42*(1), 45-74. https://doi.org/10.1111/j.1540-5893.2008.00334.x

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*(4), 341. https://doi.org/10.1037/0003-066X.39.4.341

Kanter, R. M. (1977). Some effects of proportions on group life: Skewed sex ratios and responses to token women. *American Journal of Sociology, 82*(5), 965-990. https://www.jstor.org/stable/2777808

Keith, L. C., Holmes, J. S., & Miller, B. P. (2013). Explaining the divergence in asylum grant rates among immigration judges: An attitudinal and cognitive approach. *Law & Policy, 35*(4), 261-289. https://doi.org/10.1111/lapo.12008

King, E. B., Hebl, M. R., George, J. M., & Matusik, S. F. (2010). Understanding tokenism: Antecedents and consequences of a psychological climate of gender inequity. *Journal of Management, 36*(2), 482-510. https://doi.org/10.1177/0149206308328508

Kritzer, H. M., & Uhlman, T. M. (1977). Sisterhood in the courtroom: Sex of judge and defendant as factors in criminal case disposition. *Social Science Journal, 14*(2), 77-88.

Laws, J. L. (1975). The psychology of tokenism: An analysis. *Sex Roles, 1*(1), 51-67. https://doi.org/10.1007/BF00287213

Miller, B., Keith, L. C., & Holmes, J. S. (2015). Leveling the odds: The effect of quality legal representation in cases of asymmetrical capability. *Law & Society Review, 49*(1), 209-239. https://doi.org/10.1111/lasr.12123

Mustard, D. B. (2001). Racial, ethnic, and gender disparities in sentencing: Evidence from the US federal courts. *The Journal of Law and Economics, 44*(1), 285-314. https://doi.org/10.1086/320276

Muzak, G., & Rohrböck, J. (2008). *Der Asylgerichtshof [The Asylum Court]*. Verlag Österreich, Vienna.

Myers, M. A., & Talarico, S. M. (1987). *The social contexts of criminal sentencing*. New York: Springer Verlag.

Peresie, J. L. (2005). Female judges matter: Gender and collegial decisionmaking in the federal appellate courts. *Yale Law Journal, 114*(7), 1759-1790. https://www.jstor.org/stable/4135764

Ramji-Nogales, J., Schoenholtz, A. I., & Schrag, P. G. (2007). Refugee roulette: Disparities in asylum adjudication. *Stanford Law Review, 60*(2), 295-411. https://ssrn.com/abstract=983946

Schanzenbach, M. (2005). Racial and sex disparities in prison sentences: The effect of district-level judicial demographics. *The Journal of Legal Studies, 34*(1), 57-92. https://doi.org/10.1086/425597

Songer, D. R., Davis, S., & Haire, S. (1994). A reappraisal of diversification in the federal courts: Gender effects in the courts of appeals. *The Journal of Politics, 56*(2), 425-439. https://doi.org/10.2307/2132146

Spohn, C. (1991). Decision making in sexual assault cases: Do Black and female judges make a difference? *Women & Criminal Justice, 2*(1), 83-105. https://doi.org/10.1300/J012v02n01_06

Spohn, C. (2013). The effects of the offender's race, ethnicity, and sex on federal sentencing outcomes in the guidelines era. *Law & Contemporary Problems, 76*(1), 75-104. https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=4347&context=lcp

Starr, S. B. (2014). Estimating gender disparities in federal criminal cases. *American Law and Economics Review, 17*(1), 127-159. https://doi.org/10.1093/aler/ahu010

Steffensmeier, D. & Hebert, C. (1999). Women and men policymakers: Does the judge's gender affect the sentencing of criminal defendants? *Social Forces, 77*(3), 1163-1196. https://doi.org/10.1093/sf/77.3.1163

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453-458. https://doi.org/10.1126/science.7455683

Voeten, E. (2007). The politics of international judicial appointments: Evidence from the European Court of Human Rights. *International Organization, 61*(4), 669-701. https://doi.org/10.1017/S0020818307070233

Yoder, J. D. (1991). Rethinking tokenism: Looking beyond numbers. *Gender & Society, 5*(2), 178-192. https://www.jstor.org/stable/189483

Yoder, J. D. (1994). Looking beyond numbers: The effects of gender status, job prestige, and occupational gender-typing on tokenism processes. *Social Psychology Quarterly*, *57*(2), 150-159. https://www.jstor.org/stable/2786708

Table 1

Gender Effects in Austria Asylum Adjudications

Hypotheses	Variables	(a) Model 1 *B*	(a) Model 1 Clustered *SE*	(b) Model 2 *B*	(b) Model 2 Clustered *SE*
H1 (+)	Female judge	0.44**	0.15	0.40**	0.15
H2 (+)	Female applicant	0.45***	0.06	0.95***	0.16
	Legal assistance	0.10	0.11	0.07	0.11
	% female applicants in judge's caseload			1.13*	0.49
H3 (-)	Female judge × female applicant			-0.01	0.10
H4 (-)	Female applicant × % female applicants in judge's caseload			-1.65***	0.38
AC chamber					
	A (Africa) (reference category)				
	B (Western Balkan states)	-0.49	0.27	-0.53*	0.26
	C (Asia, unless included in other chambers)	-0.41	0.27	-0.45	0.26
	D (Post-Soviet states)	-0.11	0.23	-0.27	0.27
	E (Turkey, Middle East, Armenia, Azerbaijan)	-0.90**	0.31	-0.94**	0.31
	S (special cases)	-1.07***	0.17	-1.11***	0.17
Intercept		-0.69***	0.21	-0.91***	0.21
	Observations	40,980		40,980	
	Log-likelihood	-24,459		-24,398	
	AIC	48,935		48,819	
	% correctly predicted	68.91		68.87	

Note. Cell entries report binary logistic regression coefficients with clustered standard errors at the level of individual judges. (+) indicates a hypothesized positive effect and (-) denotes a hypothesized negative effect. Model 1 tests for direct gender effects; Model 2, for interaction effects between judge and applicant gender and between applicant gender and the gender distribution within a judge's caseload.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 2

Gender Effects in Appeals Decisions

	(a) Model 1	
	Average marginal effect	95% confidence interval
Female judge	0.09	[0.03, 0.15]
Female applicant	0.09	[0.07, 0.12]
Legal assistance	0.02	[-0.02, 0.07]
AC chamber		
A (Africa) (reference category)		
B (Western Balkan states)	-0.11	[-0.23, 0.01]
C (Asia, unless included in other chambers)	-0.09	[-0.21, 0.03]
D (Post-Soviet states)	-0.03	[-0.13, 0.08]
E (Turkey, Middle East, Armenia, Azerbaijan)	-0.19	[-0.31, -0.07]
S (special cases)	-0.22	[-0.30, -0.13]

	(b) Model 2	
	Average marginal effect of female applicant	95% confidence interval
% female applicants in judge's caseload		
5%	0.17	[0.11, 0.22]
15%	0.14	[0.10, 0.17]
25%	0.11	[0.09, 0.14]
35%	0.08	[0.06, 0.10]
45%	0.04	[0.02, 0.07]
55%	0.00	[-0.03, 0.04]

Note. Cell entries report average marginal effects with the corresponding 95% confidence intervals. Model 1 tests for direct gender effects; Model 2, for interaction effects between judge and applicant gender and between applicant gender and the gender distribution within a judge's caseload.

Online supplement for Ecker, A., Ennser-Jedenastik, L. and Haselmayer, M. (2019). Gender bias in asylum adjudications: Evidence for leniency toward token women. *Sex Roles*.
Alejandro Ecker, University of Mannheim. Email: alejandro.ecker@mzes.uni-mannheim.de

S1 Descriptive statistics

Table 1s Distribution of independent variables

	Mean	Minimum	Maximum
Female judge	0.43	0	1
Female applicant	0.28	0	1
% female applicants in judge's caseload	27.81	0	63.64
Legal assistance	0.14	0	1
AC chamber			
A	0.14	0	1
B	0.13	0	1
C	0.21	0	1
D	0.22	0	1
E	0.19	0	1
S	0.12	0	1

All figures are based on 40,980 adjudications by the AC between 1 July 2008 and 31 December 2013.

S2 Fixed effects model specifications

Table 2s Alternative model specifications with fixed effects at the AC chamber-year and country of origin level

	Model 1 Coefficients	SEs	Model 2 Coefficients	SEs	Model 3 Coefficients	SEs	Model 4 Coefficients	SEs	Model 5 Coefficients	SEs	Model 6 Coefficients	SEs
Female judge	0.44***	0.02	0.40***	0.03	0.47***	0.02	0.40***	0.03	0.54***	0.03	0.50***	0.03
Female applicant	0.45***	0.02	0.95***	0.07	0.47***	0.03	1.00***	0.07	0.42***	0.03	0.98***	0.08
Legal assistance	0.10**	0.03	0.07*	0.03	0.19***	0.03	0.16***	0.03	0.26***	0.04	0.25***	0.04
% female applicants in judge's caseload			1.13***	0.12			1.53***	0.13			1.12***	0.15
Female judge × female applicant			-0.01	0.05			-0.01	0.05			0.01	0.05
Female applicant × % female applicants in judge's caseload			-1.65***	0.19			-1.74***	0.19			-1.72***	0.21
AC chamber fixed effects	*included*											
AC chamber-year fixed effects					*included*				*included*			
Country of origin fixed effects									*included*			
Intercept	-0.69***	0.03	-0.91***	0.04	-1.01***	0.09	-1.34***	0.10	0.70***	0.13	0.45***	0.14
Observations	40,980		40,980		40,980		40,980		40,980		40,980	
Log-likelihood	-24,459		-24,398		-23,738		-23,654		-20,751		-20,707	
AIC	48,935		48,819		47,554		47,394		41,811		41,729	
% correctly predicted	68.91		68.87		70,82		71,00		76,27		76,53	

Cell entries depict binary logistic regression coefficients. * $p < .05$. ** $p < .01$. *** $p < .001$.

S3 Robustness checks on substantial effect of applicant gender

A first approach to account for potential systematic differences in the merits of the case between male and female applicants is exploiting the variety of standardized keywords used to classify each case.[1] The primary objective is to combine these keywords with the observed gender of the applicant in order to identify a robust set of keywords that indicates gender-related causes of migration. For this purpose, we first identify 106 keywords in the 99.9 percentile of the keyword distribution, which are used across a sufficiently large number of verdicts (thus excluding rarely used keywords). For each keyword, we then then estimate the observed ratio of female to male applicants, resulting in a list of keywords associated predominantly with verdicts on female or male applicants, respectively. Keywords associated predominantly with male applicants are, for instance, 'blood feud', '(objection to) military service', and 'forced recruitment'. The list of 'female' keywords, on the other hand, includes 'family unit', 'non-protectability', and 'pro-Western orientation'.

Based on the female to male applicant ratio across keywords, we then exclude all verdicts that include any keyword that belongs either to the top or the bottom 15% of the keyword distribution (N = 20,365) and thus classifies as a typically 'female' or 'male' keyword. The remaining set of observations (N = 20,615) includes verdicts with no gender-related keywords. As such, we expect gender-specific root causes of migration to be less relevant and consequently applicant gender to be largely uncorrelated with the actual merits of the case.

A second robustness check also draws on a keyword-based pre-selection of verdicts. Specifically, we use the subset of cassation appeals (N = 4,583), which according to Austrian law are appeals against procedural errors, nullity pleas and objections against "incorrect" decisions based on the facts of the case or the legal analysis. Most importantly for our

[1] The median number of keywords used is 5 with a minimum number of 1 keyword and a maximum number of 25 keywords.

purposes, these decisions do not involve any (re)-examinations of the merits of the case. Rather, judges have to review the formal correctness of the original jurisdiction. Thus, individual factors, such as applicants' gender and origin will have less impact on the verdict. The prevalence of formal elements in the judges' decisions designates these appeals for a more rigorous test of our theoretical argument that gender effects are uncorrelated with the merits of the case.

Table 3s Gender effects in Austrian asylum court adjudications

	Model 1 Coefficients	SEs	Model 2 Coefficients	SEs	Model 7 Coefficients	SEs	Model 8 Coefficients	SEs	Model 9 Coefficients	SEs	Model 10 Coefficients	SEs
Female judge	0.44***	0.02	0.40***	0.03	0.44***	0.03	0.42***	0.04	0.42**	0.14	0.26	0.16
Female applicant	0.45***	0.02	0.95***	0.07	0.11**	0.04	0.46***	0.11	0.74***	0.19	1.99***	0.55
Legal assistance	0.10**	0.03	0.07*	0.03	0.04	0.05	0.03	0.05	0.01	0.25	-0.02	0.25
% female applicants in judge's caseload			1.13***	0.12			0.47**	0.17			1.62*	0.75
Female judge × female applicant			-0.01	0.05			0.02	0.08			0.41	0.38
Female applicant × % female applicants in judge's caseload			-1.65***	0.19			-1.22***	0.31			-4.67**	1.45
AC chamber fixed effects	*included*											
Intercept	-0.69***	0.03	-0.91***	0.04	-0.59***	0.04	-0.69***	0.05	2.21***	0.13	1.96***	0.16
Observations	40,980		40,980		20,615		20,615		4,583		4,583	
Log-likelihood	-24,459		-24,398		-11,104		-11,095		-909		-902	
AIC	48,935		48,819		22,225		22,213		1,837		1,828	
% correctly predicted	68.91		68.87		74.67		74.67		94.34		94.34	

Cell entries depict binary logistic regression coefficients.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 3s presents the results of these two additional robustness checks. Models 7 and 8 test the hypothesized relationships based on the subset of verdicts without gender-related keywords. Models 9 and 10 repeat the empirical exercise based solely on cassation appeals. Overall, these robustness checks corroborate the initial results. Specifically, we observe a persistent positive effect of both female judges and female applicants on the likelihood of a successful appeal before the Austrian AC. In a similar vein, the multiplicative interactive models 8 and 10 confirm the substantive negative moderating effect of the share of female applicants in each judge's caseload. As before, the positive effect of being a female applicant is substantially reduced to the extent that a judge faces an increasing share of female applicants.

As a final approach to account for a potential spurious effect of applicant gender, we propose an instrumental variable (IV) estimation that uses distance from Austria as an instrument for applicant gender. The rationale here is that the distance between the country of origin and the receiving country affects the gender composition of the refugee population (the further away, the more male), but is uncorrelated with the actual merits of an individual's case. Therefore, the only way through which distance influences judicial decisions is by affecting the gender composition of arriving applicants.

We therefore collect data on the distance between Vienna and the capital in the 40 most frequent countries of origin in our sample. Specifically, this information was gathered by using the Google Maps API to calculate the walking distance between Vienna and the other capitals. We use distance to Vienna as an IV in a first-stage regression to predict applicant gender. The second-stage regression then predicts the Asylum Court's decision.

The results in Table 4s suggest that the effect of applicant gender holds up well, even if instrumented with geographical distance. The coefficient in model 11 is large (0.84) and highly significant. In addition, the coefficients from the interaction terms in model 12 corroborate the findings presented in the main text: As the gender balance in a judge's

caseload becomes less skewed, the impact of applicant gender diminishes. The effect is strongest when the gender balance in a judge's caseload is heavily male.

Table 4s Instrumental variable estimation

		Model 1 Coefficients	SEs	Model 2 Coefficients	SEs	Model 11 Coefficients	SEs	Model 12 Coefficients	SEs
Second stage	Female judge	0.44***	0.02	0.40***	0.03	0.23***	0.02	0.16***	0.02
	Female applicant	0.45***	0.02	0.95***	0.07	0.84***	0.13	1.80***	0.09
	Legal assistance	0.10**	0.03	0.07*	0.03	0.06**	0.02	0.02	0.02
	% female applicants in judge's caseload			1.13***	0.12			0.74***	0.07
	Female judge × female applicant			-0.01	0.05			-0.05	0.03
	Female applicant × % female applicants in judge's caseload			-1.65***	0.19			-1.20***	0.10
	AC chamber fixed effects	*included*							
	Intercept	-0.69***	0.03	-0.91***	0.04	-0.60***	0.03	-0.75***	0.03
First stage	Distance to Vienna (km)					-0.06***	0.01	-0.06***	0.01
	Female judge					0.20***	0.01	0.20***	0.01
	Legal assistance					0.04*	0.02	0.04*	0.02
	AC chamber fixed effects					*included*		*included*	
	Intercept					-0.72***	0.04	-0.73***	0.04
	Observations	40,980		40,980		39,674		39,674	
	Log-likelihood	-24,459		-24,398		-46,059		-45,975	
	Correlation parameter ρ					-0.34***		-0.70***	

Cell entries depict bivariate probit regression coefficients.

* $p < .05$. ** $p < .01$. *** $p < .001$.