Inferring Individual Preferences from Group Decisions: Judicial Preference Variation and Aggregation in Asylum Appeals

Dominik Hangartner
ETH Zurich and London School of Economics

Benjamin E Lauderdale
University College London

Judith Spirig
University of Zurich

Abstract

While many democracies nominate partisan judges, empirical research has struggled to assess whether such judges adhere to Aristotle's maxim that *like cases should be treated alike*. One fundamental problem hindering empirical research is that many courts only report decisions of panels, not the opinions of individual judges. We propose a methodology that tests which of several decision-theoretic models of group decision-making best fit the panel decisions, infers judges' individual preferences, and quantifies the proportion of cases that would be decided differently if the courts' consensus were consistently applied (an inconsistency rate). Applying this methodology to the Swiss asylum appeal process, where cases are assigned conditionally at random and have a common, unidimensional structure, we find a persistent inconsistency rate of about 5% due to variation in decision-making between judges, and that judges' estimated preferences are correlated with party membership in expected ways.

Dominik Hangartner, Center of Comparative and International Studies, ETH Zurich, Switzerland; Immigration Policy Lab, Stanford University, U.S.A. and ETH Zurich, Switzerland; and Department of Government, London School of Economics and Political Science, U.K. E-mail: dominik.hangartner@gess.ethz.ch.
Benjamin E Lauderdale, Department of Political Science, University College London, U.K. E-mail: b.lauderdale@ucl.ac.uk.
Judith Spirig, Department of Political Science, University of Zurich, and Immigration Policy Lab, Stanford University, U.S.A. and ETH Zurich, Switzerland. E-mail: judith.spirig@uzh.ch.
The usual disclaimer applies.

I. Introduction

In summer 2007, the Swiss Federal Administrative Court (FAC), the highest court reviewing asylum decisions of the Swiss administration, had to decide on two unrelated appeals of rejected asylum seekers from Afghanistan. Both appellants were male, of Hazara ethnicity, Shiite Muslims from Kabul and with family still living there. On July 4, a panel of three judges chaired by a member of the Social Democratic Party ruled that the current situation in Kabul cannot be considered 'reasonable' for removal and that the asylum seeker therefore had to be granted subsidiary protection. Two months later, on September 3, another panel consisting of three different judges chaired by an independent judge rejected the appeal of the other asylum seeker: the panel wrote in their verdict that removal to Afghanistan is both 'reasonable' and 'admissible'. In consequence, the rejected asylum seeker had to leave Switzerland within a few days for an uncertain fate in war-torn Afghanistan. A close reading of the two verdicts[1] reveals few differences in the merits of the cases that could explain the radically different outcomes. This raises a question that touches upon fundamental legal principles: If not the cases' merits, do the different judges on the panels explain the differences between these, and potentially other, verdicts?

Consistency in adjudication is central to the legitimacy of the judiciary and the rule of law (see, e.g., Kornhauser and Sager 1986). Condensed to the Aristotelian maxim that "like cases should be treated alike", consistency features prominently in theories of justice as a safeguard of fair treatment of disputing parties and of the predictability of the court's decisions (see, e.g., Marmor 2007, for an overview). And only a court that applies the same legal rules (and consequences) to the same stipulated set of relevant circumstances (Kornhauser and Sager 1986, p. 104) can fulfill the promise of 'equality before the law'—a legal principle that is fundamental to the rule of law in liberal democracies and frequently enshrined in the constitution (and in the case of the U.S., directly above the main entrance to the Supreme Court). In Switzerland, for example, Article 8(1) of the constitution guarantees that "every person is equal before the law." Since inconsistency in judicial decision making, that is, different decisions for cases of the same merits and

[1] See FAC decisions E-3570/2006 and D-4576/2007.

law, violates the very essence of the isonomy principle of equality before the law, it has been the subject of much empirical scholarship—running the gamut from partisan adjudication of challenges to the U.S. Environmental Protection Agency (Revesz 1997) to ethnic in-group bias in Israeli small claims courts (Shayo and Zussman 2011).

In the context of asylum adjudication, a small but growing literature, primarily focused on the U.S. and Canada, has examined disparities at several stages of the asylum process, from the initial application to the final appeal at the highest courts (Ramji-Nogales, Schoenholtz and Schrag 2007; Rehaag 2007; Fischman 2011). All these studies find substantial heterogeneity in the decisions of asylum officers, judges and appeal courts. Ramji-Nogales, Schoenholtz and Schrag (2007), for instance, show that for Colombian cases before the Miami Immigration Court, judges' grant rate varied between 5% and 88%.

Some of the asylum processes covered by these studies[2] feature (quasi-) random assignment of cases to asylum officers or judges, which allows the authors to causally attribute differences in grant rates to the different decision makers. However, most studies rely on data obtained (often through Freedom of Information requests) from governing institutions, which usually insist on protecting the identity of the decision maker and replacing it with an anonymous identifier. This lack of data renders it impossible to connect estimates of judges' ideal points with their personal characteristics such as ethnicity, gender or partisanship. Thus, most existing research is limited to documenting variation in asylum appeal adjudication, without being able to shed light on how judges' ideology explains this variation.[3]

Another limitation of this literature concerns the courts that have been studied. Most empirical research to date has focused on U.S. and Canadian courts, where—in

[2]Fischman (2011) focuses on asylum appeals at the U.S. federal circuit court level, which features random assignment. The U.S. immigration courts studied by Ramji-Nogales, Schoenholtz and Schrag (2007) also use random assignment of cases *within* courts, but assignment to the 53 courts is based on residence, which likely confounds comparisons across courts.

[3]One exception, focusing on layman judges in Sweden, is Martén (2015).

line with the common law tradition—not only the verdict, but also judges' dissenting opinions are disclosed. In contrast, the principles of collegiality and secrecy of deliberations are more strongly rooted in continental European courts belonging to the civil law system, with the consequence that only the verdict of the entire panel is disclosed. Lacking the methodological tools to infer individual preferences from group decisions, the empirical evidence on judicial preference variation stems mostly from North America.[4] This narrow geographic focus means that our allegedly general theories of judicial behavior and judicial politics are put to test in only a handful of countries rooted in a particular legal tradition. If and how these theories also apply to civil law courts that abstain from disclosing dissenting opinions remains a largely unanswered question.

We overcome these limitations by leveraging a new methodology to analyze panel decisions of the Swiss FAC. As a signatory to the 1951 Refugee Convention, Switzerland grants asylum seekers whose claims are rejected the right to appeal the initial decision. Since 2007, the approximately thirty judges of the FAC centrally review these appeals, and decide on about 3,000 cases per year. In contrast to most other countries and essential for our analysis, judges serving on the Swiss FAC have a publicly known party membership and are voted into judicial office by the members of the national parliament. While not written law, there is an informal agreement that the body of elected judges reflects the relative seat share of the different parties in parliament. These voluntary party quotas ensure that the whole spectrum of political ideology in Swiss society is adequately represented on the court (Kiener 2001). One potential drawback of this heavily politicized nomination process—and the focus of this study—is that judges might be incentivized to reach verdicts in line with the preferences of the party that selects them.

[4]Studies focused on civil law countries cover Israeli courts (Gazal-Ayal and Sulitzeanu-Kenan 2010; Shayo and Zussman 2011; Grossman et al. 2016), the European Court of Human Rights (Carrubba, Gabel and Hankla 2008), the French, German, Portuguese and Spanish constitutional courts (Hönnige 2009; Hanretty 2012) as well as several Latin American courts, in particular in Argentinia, Brazil, Chile and Mexico (see Kapiszewski and Taylor 2008; Helmke and Staton 2011, for overviews).

To analyze judicial behavior on the FAC, we develop a new methodology that allows us to infer individual preferences from group decisions, even when dissenting opinions are not recorded. As described in more detail below, the proposed methodology exploits the quasi-random assignment of judges to thousands of panels to test which model of group decision making best fits the aggregate panel outcomes. We explore the empirical fit of various decision-theoretic models including: majority rule, which implies that the median judge is decisive; unanimity rule to grant (reject) an appeal, which implies that the most restrictive (lenient) judge is decisive; and chair-as-dictator, which implies that the chair judge, who writes the first draft of the verdict, is decisive. Conditional on these aggregation rules, we can then infer and estimate individual preferences using item-response theory models.

Besides the non-disclosure of dissenting opinions, scholars of judicial behavior interested in disparities in adjudication typically face a host of additional inferential challenges. Fortunately, several features of the Swiss asylum appeal process conspire and help us to overcome them. First, all appeals have a common, unidimensional structure, since they exclusively deal with asylum issues. While the assumption that judges' preferences are dominantly unidimensional is often invoked in the empirical study of judicial politics, there is historical (Greenhouse 2007; Jeffries 2001) and statistical (Lauderdale and Clark 2012) evidence that judges' preferences vary across areas of the law. In the present context, however, the unidimensionality assumption seems much more credible, as all verdicts typically involve the appeal of an initial asylum or subsidiary protection decision, or a closely related asylum matter.

Second, the FAC handles all asylum appeals lodged in Switzerland. In other studies (for example, Ramji-Nogales, Schoenholtz and Schrag 2007), differences in the average merit of cases submitted to different regional offices (U.S. Immigration Courts) frequently undermine the comparability of preferences across offices. This is not a concern in our context, where all appeals are centrally processed by the two asylum appeal divisions of the FAC.

Third, at the FAC, a bespoke software program with the sole objective function to minimize workload imbalance assigns cases to panels of three judges. The automated assignment is independent of the merits of a case and judges' characteristics conditional on the language of the asylum decision. This conditionally random allocation of judges

to thousands of panels gives us inferential leverage to causally attribute differences in grant rates to differences in panel composition.

Fourth, not only are judges elected to office by the Swiss parliament, but their identities and party affiliations are also public knowledge. This allows us to go considerably beyond existing studies of asylum adjudication that had to work with anonymized decision-maker identifiers, and to correlate estimated grant rates with judges' ideology, proxied by their party membership.

Analyzing the universe of almost 35,000 verdicts decided by the FAC between 2007–2015, we find that the best-fitting simple aggregation rule is that the panel chair dictates the outcome, closely followed by majority rule for which the median judge is decisive. Turning to more complex aggregation rules, we find that a (Bayesian) mixture model of the chair and median models fits better than either of these simple aggregation rules, indicating that the chair judge can sometimes, but not always, deviate from the median's preference. Based on this mixture model, we show that asylum seekers who submitted comparable appeals in 2007 faced substantially different grant rates, depending on the panel of judges their case was assigned to. This heterogeneity in judges' preferences strongly correlates with party affiliation in expected ways: on average, judges from the right-wing Swiss People's Party have a preferred grant rate of 6.0%. In stark contrast, the preferred grant rate of judges from the left-wing Social Democratic Party is 17.5%, almost three times higher.

We also explore the evolution of the inconsistency in the court's decisions after its inception in 2007. While the inconsistency rate, i.e., the proportion of cases that would have been decided differently if the consensus of the court were consistently applied, fluctuates between 3.0% and 6.4%, we find no evidence that judges' preferences converged over time. The persistence of substantial inconsistency suggests that the issue of preference variation in asylum appeal adjudication is a permanent fixture of this court and its partisan judges.

Our study makes several contributions. First, our findings have important implications for the comparative literature that studies disparities in asylum adjudication. While many studies show sizeable disparities in asylum adjudication between decision makers that potentially face different cases (Law 2004; Ramji-Nogales, Schoenholtz and Schrag 2007; Taylor 2007), our results provide clear causal evidence that the identity of

judges matters when facing cases with, in expectation, the same merit.

Second, we add new empirical evidence to the growing literature on the effects of judges' identity on their decisions. Previous research has shown that characteristics such as gender (Boyd, Epstein and Martin 2010; Glynn and Sen 2015; Peresie 2005), ethnicity or race (Abrams, Bertrand and Mullainathan 2012; Gazal-Ayal and Sulitzeanu-Kenan 2010; Grossman et al. 2016; Shayo and Zussman 2011) and ideology (Ashenfelter, Eisenberg and Schwab 1995; Epstein, Landes and Posner 2013; Sunstein et al. 2007) all affect judicial behavior. We provide some of the most direct evidence to date that judges' political ideology, proxied by their party affiliation, shapes preferences over asylum appeals in expected ways.

Third, our study adds to the growing literature on group decision-making (Bonneau et al. 2007; Sunstein et al. 2007), preference aggregation (Fischman 2011; Van Dijk, Sonnemans and Bauw 2014) and interaction among panel members (Fischman 2013; Kastellec 2013). Previous studies (e.g., Fischman 2011) that infer individual preferences using item-response theory-type models rely on the individual votes of decision makers. We explore a context where individual votes are not reported, and only the group's joint verdict is observed. Nonetheless, by leveraging the random and repeated allocation of judges to panels, we are able to recover individual preferences from joint group decisions by fitting a variety of aggregation rules. In addition to providing ideal point estimates for each judge, this new methodology also allows us to test which decision-theoretic model best describes the group decisions. The methodological framework we develop to study panel decisions at the FAC can easily be adapted to other courts, where dissenting opinions are not reported. This opens up the possibility to empirically study the behavior of judges on dozens of civil law courts across continental Europe that were heretofore out of methodological reach. In addition, this framework has the potential for a wide range of applications beyond the study of judicial behavior, since joint decisions without any recorded information on individual votes is the norm in many decision environments relevant for political science and neighbouring disciplines.

The rest of the paper is structured as follows. The next section provides background information about the structure of decision making at the FAC and the election of its judges. Section III describes the proposed methodology based on the case-space model, how we infer individual preferences from group decisions, and how we estimate

the court's inconsistency rate over time. After summarizing the data and measures in Section IV, we detail the results in Section V. In the conclusion, we discuss the legal and political implications of our findings and point to other areas, outside of the realm of judicial politics, where our methodology could be fruitfully applied.

II. Asylum Appeals and the Swiss Federal Administrative Court

Like many other countries, Switzerland grants asylum in accordance with the 1951 Refugee Convention (and the 1967 Protocol). Asylum applications are processed by the State Secretariat for Migration (SEM). Of the 26,000 asylum applications decided in 2012, roughly the mid-point of our study period, only 15% were granted (UNHCR 2013) and more than 10% were appealed.[5] In the event of a negative substantive decision,[6] the applicant has thirty days to lodge an appeal.[7] Since its inception in 2007, the FAC handles all such appeals. Because the verdicts of the FAC are, in general, not appealable to the Swiss Federal Supreme Court, the FAC is effectively the court of last resort in the Swiss asylum process.

[5] Viewed in a comparative perspective, Switzerland received a relatively high share of asylum applications in the last decade (see, e.g., Bansak, Hainmueller and Hangartner 2017). Similar to developments in many other European countries, the increasing number of asylum applications in recent years has made asylum policies, including the appeal process, a heavily contested issue, with right-wing parties, such as the Swiss People's Party, leveraging it as an effective springboard for rallying their voters.

[6] A negative substantive decision is the rejection of an asylum application that the SEM has tried substantively, i.e., one in which the SEM has 'entered into the substance of the case'.

[7] If the asylum request is dismissed without entering into the substance of the case, the appeal window shortens to only five working days.

A. Election of Judges by the Swiss Parliament

The approximately thirty asylum judges belonging to the two divisions of the FAC handle, on average, about 3,000 asylum appeals per year. Unlike in countries where the executive branch nominates candidates for judicial office, in Switzerland, judges of the FAC are nominated and voted into office by the United Federal Assembly, i.e., the joint meeting of the two chambers of the Swiss National Parliament. A term lasts for six years, but there is no limit on the number of terms that judges are allowed to serve. In recent years, all judicial candidates had a known party affiliation and were backed by their party when running for office.[8] While not written law, there is an informal rule that the body of elected judges reflects the relative seat share of the different parties in parliament. The fundamental norm underlying this informal rule is to select a judiciary that is representative of the people it serves(Kiener 2001; Raselli 2011). By enforcing that all parties get the number of judges that is approximately proportional to their strength in parliament, this "Magic Formula" ensures that the whole range of political ideologies is 'adequately' represented at the court.[9]

A potential drawback of electing the judiciary is the threat of a heavily politicized court. If judges see themselves primarily as agents of their respective party, we might worry that their verdicts are not only following the 'law', but are also motivated by their parties' ideology. This could either be the result of a selection process by which parties nominate judges that are ideologically close (or, observationally equivalent, judges joining parties according to their ideology) or a consequence of the re-election process that

[8]In the first election for this court, a few candidates ran as independents. More recently, the only exception was a candidate who is a supporter of the Green Party without being a formal member.

[9]Note, however, that this principle of proportionality is not strictly enforced, if, for example, the judiciary commission cannot find candidates with the relevant qualifications for underrepresented parties. Accordingly, the Swiss People's Party, which holds the most seats in the parliament, has been underrepresented at the FAC, in particular in its early years (see [reference suppressed to protect author anonymity] for details.)

incentivizes judges to reach verdicts in line with the preferences of their party. With both mechanisms, we would expect to see substantial variation in the adjudication of asylum appeals across judges from different parties, and a correlation between judges' grant rates and their parties' general stance on asylum issues.

B. Asylum Appeal Procedure and the Structure of Panel Decisions

When receiving a new asylum appeal, the FAC identifies the language of the asylum decision (German, French, or Italian) and forwards, on an alternating basis, the case to one of the chambers of its two asylum divisions. A bespoke software program called *Bandlimat*, named after the first president of the FAC, assigns the appeal to a three-judge panel and determines judges' roles as chair, second and third judge. When sequentially assigning cases to judges, the *Bandlimat* solely considers (i) the language of the asylum decision, (ii) the urgency of the appeal, (iii) judges' language skills and (iv) their current workload. The assignment of cases is completely mechanical, and non-compliance with the software's assignment has to be justified, logged and entered by the head of the division.[10] The sole objective function of the software is to minimize the imbalance in workload created by case assignment (each case has an identical weight of one) under constraints (i)–(iv). With the exception of the language of the appeal, all constraints are orthogonal to the identity and characteristics of judges, such that in expectation, the assignment of cases by the *Bandlimat* is (conditional on language or origin country) as good as random. We use a series of placebo checks to validate this assumption in the section after next.

The court employs two distinct decision-making procedures to decide on asylum appeals.[11] During the first year of the court's existence, all substantively tried cases were handled through the 'ordinary procedure' that is characterized by the following structure: the chair judge receives the case files, conducts additional investigations if

[10]See, for details, the law guiding the FAC's standard operating procedure, art. 21, par. 1 of the Administrative Court Act.

[11]Note that cases that do not fulfill the formal requirements are 'dismissed without entering into the substance of the case' by the chair judge in a single-judge procedure.

necessary, instructs one of her clerks to draft a decision and forwards all materials including the draft verdict to the second judge. The second judge reads the case files and draft decision, either agrees, or disagrees and proposes changes, and forwards everything to the third judge. The third judge reads the case files, the draft decision and the comments of the second judge, and either agrees, or disagrees and proposes changes, and returns the file and her comments to the chair. In the event of disagreement, the panel further circulates and possibly revises the decision. If the three judges are not able to reach a consensus, the outcome is decided by majority rule.

A partial revision of the Swiss Asylum Law led to the introduction of an alternative 'simplified procedure'. Since January 1, 2008, the simplified procedure allows the chair judge to classify certain cases as either 'clearly with or without merit'. In the vast majority of cases, the simplified procedure is applied to appeals that are 'clearly without merit'. The initial assignment of cases to three-judge panels is the same for both procedures. When the chair judge invokes the simplified procedure, she only needs the second judge to agree with her classification and the verdict. If the second judge agrees with both, the decision-making process ends here, and the file is not forwarded to the third judge. If the second judge disagrees, the process reverts to the ordinary procedure.

Clearly, appeals decided by the simplified procedure are a selective subset of all cases. Furthermore, our analysis, discussed in detail below, reveals that judges vary considerably in their propensity to invoke the simplified procedure. To circumnavigate any selection issues arising from these differing interpretations of what constitutes appeals 'clearly with or without merit', the analysis proceeds by focusing on complementary subsets of the data. First, we focus on the substantively tried appeals that were submitted in 2007—before the introduction of the simplified procedure—and handled by three-judge panels. Building on these cases decided under the ordinary procedure, we test which decision-making rules fit the data best, and estimate judges' preferences and the court's inconsistency rate. Next, we analyze all appeals decided between 2007–2015, independent of the procedure used to decide the case, to explore the evolution of

preference variation over time.[12]

III. Methodology

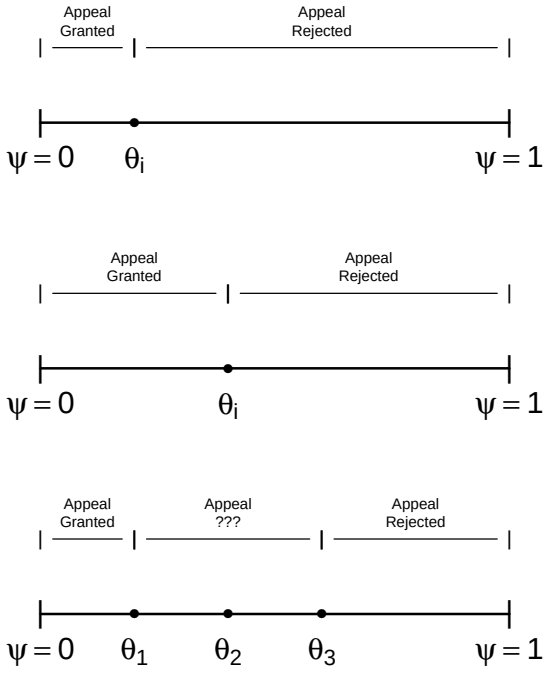
A. *Inferring Individual Preferences from Panel Decisions Using the Case-Space Model*

To estimate individual preferences from the observed aggregate decisions of three-judge panels, we need a framework for modelling preference aggregation. We adopt a unidimensional case-space model (Kornhauser 1992), which allows us to theoretically describe the preferences of judges and to map different preference aggregation rules onto likelihood estimators. Each case j has *facts* that can be described as a location ψ_j. We treat smaller values of ψ_j as indicating stronger appeals (case facts) and larger values of ψ_j as weaker appeals. Each judge i has *preferences* that can be described as a cutpoint θ_i. Each judge, if deciding the case alone, would rule in favor of the appellant if and only if $\psi_j < \theta_i$. Thus, judges with lower cutpoints θ_i are inclined to grant fewer appeals, and judges with higher cutpoints are inclined to grant more appeals. An assumption of this unidimensional model is that all judges agree on the ranking of relative merits of appeals and disagree only on the threshold to apply. While this assumption seems reasonable in the present context where all cases concern asylum appeals, it does have some implications for interpretation, which we discuss below.

The top two axes of Figure 1 show two different hypothetical judges and the decisions they would make if they decided cases alone. However, the cases we are studying are decided jointly by (up to) three judges according to the procedures described in the preceding section, and so the resolution of cases in which the three judges disagree (bottom axis of Figure 1) depends on the aggregation rule that combines their preferences into a decision.

[12]December 31, 2015, is the end point of our study period because the FAC again changed its decision-making procedure, making further comparisons over time more difficult.

Figure 1: Mapping between Preferences and Decisions



Note: The top two axes illustrate the mapping between preferences and hypothetical single-judge decisions. The bottom axis illustrates the three-judge decisions that are actually observed, indicating the range of cases over which decisions depend on which preference aggregation rule most closely matches the court's decision-making.

To map the aggregation rules onto a likelihood estimator, we introduce the following notation. Let $i(j)$ be the indices of the judges hearing case j, so that $\theta_{i(j)}$ is a three-component vector, with the first element $\theta_{1(j)}$ corresponding to the chair, the second $\theta_{2(j)}$ to the second judge and the third $\theta_{3(j)}$ to the third judge. We consider only those aggregation rules that can be described by a function $f(\theta_{i(j)})$ that maps the preferences of the three judges into an effective preference of the panel.[13] This allows us to define

[13]It is difficult to make a substantive argument for the kinds of non-monotonic preference aggregation functions that could not be described as a mapping of the three judges' preferences into a single effective preference on the same scale.

a generic likelihood function for the observable votes:

$$\mathcal{L}(\theta) \quad = \quad \prod_j p\left(\psi_j < f(\theta_{i(j)})\right)^{y_j} \cdot p\left(\psi_j > f(\theta_{i(j)})\right)^{1-y_j} \quad . \tag{1}$$

In the next section, we reparametrize $f(\theta_{i(j)})$ as a function of various aggregation rules.

B. Decision-Theoretic Aggregation Rules

We consider a range of decision-theoretic preference aggregation models in our analysis. While other aggregation rules are certainly possible, we believe that our set of rules comprises the most likely candidates—not just for this application, but also for other contexts where the researcher only observes the group's joint decision, not the individual votes.

If we imagine the panel voting by majority rule internally, we expect the median judge, θ_{med}, to determine the outcome. If we imagine the panel voting with a requirement of a unanimity rule to grant an appeal, θ_{min}, i.e., the most restrictive judge, determines the outcome. If instead unanimity is required to reject an appeal, θ_{max}, i.e., the most lenient judge, determines the outcome. Lastly, if the chair's preference dictates the outcome, we would expect θ_1 to be decisive. Note that for three-judge panels, we cannot point identify, but only bound, the preferences of most extreme judges under certain models. For example, if we assume the median judge's preference determines the outcome, we cannot point identify the preferences of either the judge with the lowest or the highest threshold for asylum appeal cases. Similarly, for the minimum and maximum models, we cannot identify the two highest and the two lowest preferences, respectively. However, we can identify which judges these are and bound their θ with the next most extreme judge's position.

We fit models corresponding to these, as well as three additional and less plausible aggregation functions, in our empirical analysis. In addition to a null model where all judges/panels apply the same threshold, θ_0, that serves as a baseline, we also specify an aggregation rule where the second, θ_2, and third judge, θ_3, respectively, dictate the outcome. We do not expect these last two aggregation rules to fit the data well but include them as plausibility tests. Table 1 provides an overview of the different simple aggregation rules:

Table 1: Overview Decision-Theoretic Rules

Notation	Aggregation Rule
θ_{min}	Most restrictive judge decides
θ_{max}	Most lenient judge decides
θ_{med}	Median judge decides
θ_1	Chair judge decides
θ_2	Second judge decides
θ_3	Third judge decides
θ_0	All judges/panels apply same threshold

Note: Table displays the decision-theoretic rules used to aggregate individual preferences of panel members.

Assumptions over aggregation rules have implications for the estimation of preferences. Since judges' preferences are estimated conditional on the aggregation rule, model fit depends both on the predictive performance of the preferences and the aggregation rule. For this reason, we test a variety of aggregation rules, compare their statistical fit, choose the one(s) with the best fit, and focus on those empirical relationships that are robust to reasonable choices of aggregation rule.

In that spirit, we also consider a mixture model of two of the simple aggregation rules. We do so because the mixture model might better reflect the sequentiality of decision-making and the constraints that judges are facing than either of the simple aggregation rules alone. As noted above, the chair of the panel initially receives the case files, reviews them and sets out a draft decision. Then, the second and third judges get the opportunity to review the file and the draft decision in turn, and only if there are unresolvable disagreements after another round of circulation (and a potential discussion) is a decision by majority rule taken. This is clearly a costly process in terms of time and effort for all judges involved. Yet, whereas the chair judge gets to frame the decision and has to invest time and effort to draft it, there is an incentive for the second and third judges to follow the chair's draft decision, rather than engaging in the effort necessary to determine if they disagree, let alone formulating an alternative to the chair's provisional decision.

One way of thinking about the incentives set up by this process is to consider the cost in terms of time and effort to determine the case facts ψ_j that each judge has to pay. Because the chair judge must pay the cost in any case, but the second and

third judges do not, we can expect that some decisions taken by the chair may be at odds with the preference of the median judge, but the other two judges cannot know which decisions if they do not pay the cost of review. While we do not explicitly formulate a complete game-theoretic model here, we note that the mixture model we estimate approximates the logic of a mixed-strategy equilibrium[14] in an oversight game, while the chair and median models correspond to two different pure strategy equilibria. Which of these we observe depends on the cost of review for the second and third judges relative to the cost of cases being decided "incorrectly" from their perspective. If review costs are low relative to error costs, oversight always occurs and we would expect the median's preferences to always prevail. If review costs are very high relative to error costs, oversight never occurs, and we expect the chair to determine the outcome. The mixture model corresponds to intermediate cases where there is sometimes oversight and the chair is partially, but not completely, able to dictate decision making.

To translate the essence of this idea into a statistical model, we estimate a mixture model of the chair and median aggregation rules.[15] Let λ_1 and λ_{med} be the probability of the chair and the median judge deciding the case, respectively, such that $\lambda_{med} = 1 - \lambda_1$. The corresponding mixture of aggregation functions

$$f(\theta_{i(j)}) = \lambda_1 f(\theta_{1(j)}) + (1 - \lambda_1) f(\theta_{med(j)}) \tag{2}$$

is then plugged into the likelihood function (Equation 1) above.

C. Measuring the Consistency of Decision Making

Given the case-space model, we can also calculate the extent to which random assignment of judges leads to inconsistency in the decisions that the court makes. Let θ_j be the preferences of the judges hearing case j, let $\tilde{\theta}_j$ be the consensus of the court, and let $f(\psi)$ be the distribution of case facts. On average across M cases, the fraction of cases

[14]The existence of such mixed-strategy equilibria depends on assumptions about the utility functions of the judges and the costs of review.

[15]We also tested other mixture models, for example between the median and second judge, but all of these fitted, as expected, significantly worse.

decided differently than how they would if the consensus were consistently applied is

$$\mathcal{E} = \frac{1}{M} \sum_{j=1}^{M} \left| \int_{\theta_j}^{\tilde{\theta}} f(\psi) d\psi \right| \quad . \tag{3}$$

To compute the quantity $\mathcal{E}$, we need a benchmark for what decisions an entirely consistent court ought to make. We estimate $E[\pi|X_j]$, the expected grant rate conditional on covariates X_j, such as country of origin, that predict case strength, but *excluding* which judges heard the case. We then take the mean absolute difference between the predicted probabilities of granted appeals for each judge, $\hat{\pi}_j$, and this benchmark, $E[\pi|X_j]$, to estimate the inconsistency rate:

$$\hat{\mathcal{E}} = \frac{1}{M} \sum_{j=1}^{M} |\hat{\pi}_j - E[\pi|X_j]| \quad . \tag{4}$$

If we would not condition on covariates X_j, $E[\pi|X_j]$ would reduce to the mean appeal grant rate over all cases $E[\pi]$, and the inconsistency rate would reduce to the mean absolute error of the fitted values. This offers some intuition into why this measure captures inconsistency. If the set of judges hearing the case does not matter, then the fitted values, i.e., the predicted probabilities of granted appeals given the identities of the judges, should not vary at all. To the extent they do, this indicates that the judges matter and cases are being decided differently depending on which judges hear those cases. This measure, $\hat{\mathcal{E}}$, assumes again that judges only disagree about thresholds, not the relative merits of cases. That means that $\hat{\mathcal{E}}$ provides a lower bound on the court's inconsistency. If judges also disagree about relative merits, the true inconsistency of the court will be higher than what our estimate of $\hat{\mathcal{E}}$ suggests.

D. Verifying Random Assignment Conditional on Observables

In the previous section, we discussed how the automated case assignment by the *Bandlimat* software takes the language of the appeal and the language skills of the judges into account. This feature of the assignment process has implications for our statistical analysis. For example, French-speaking asylum seekers from Côte d'Ivoire, with typically low asylum grant rates, are likely to submit their asylum application in French-speaking Switzerland (rather than the German-speaking part), which in turn determines the language of both the asylum decision and the eventual appeal. In this case, we would worry

that the judges operating in different official Swiss languages (French, German and Italian) face cases from different origin countries and therefore of differing case strength. In order to account for this threat to internal validity, we adjust for origin country in all our models.

To further substantiate the credibility of the assumption that cases are randomly assigned conditional on origin country, we conduct a series of placebo tests. In a first step, we leverage the available case characteristics, measured pre-treatment, to gauge their predictive power of the appeal outcome. This is similar in spirit to a manipulation check following a randomized experiment (see Frandsen, Lefgren and Leslie 2019, for a similar test). For this, we rely on the coding of the administrative office of the FAC that classifies cases into legal matters. Note that this categorization is conducted before the case is assigned to judges. In addition, we use our own coding of all published verdicts (87% of all verdicts) to code if the appellant was represented by a lawyer or paralegal. After verifying that these case characteristics are indeed predictive of the success of the appeal, we turn to our placebo tests. In line with the two best-fitting simple aggregation rules (discussed below), we estimate the general leniency of the judges on the panel. In particular, we estimate the average grant rate of the chair judge from all panels that she chaired except for the case j under consideration. Similarly, we estimate the leniency of the median judge i as the average grant rate of all panels on which judge i served but for case j under consideration. We would expect that the strength of case j is not predictive of the *general* leniency of the chair or median judge on the panel if the assumption of random assignment conditional origin country is correct. Table 2 confirms that this is indeed the case.

Table 2: Manipulation Check and Placebo Tests

	Appeal granted (1)	Leniency Chair (2)	Leniency Median (3)
Enforcement of return	0.288 (0.187)	0.017 (0.056)	0.023 (0.027)
Inadmissibility of request	0.073 (0.025)	0.005 (0.008)	−0.002 (0.004)
Refusal temp. protection	0.856 (0.394)	0.202 (0.118)	−0.030 (0.056)
Revocation temp. protection	0.155 (0.051)	0.037 (0.015)	0.015 (0.007)
Family reunification	0.053 (0.095)	0.021 (0.029)	0.003 (0.014)
Revocation of asylum	0.140 (0.097)	0.034 (0.029)	0.014 (0.014)
Asylum and return (RR)	0.119 (0.051)	0.031 (0.015)	0.017 (0.007)
Return / enforcement (RR)	0.338 (0.263)	0.088 (0.079)	0.036 (0.038)
Asylum request abroad	0.049 (0.052)	0.016 (0.016)	0.005 (0.007)
Asylum procedure (other)	0.299 (0.053)	0.012 (0.016)	0.010 (0.008)
Allocation to canton	0.012 (0.168)	0.062 (0.050)	0.012 (0.024)
Lawyer / paralegal	0.111 (0.022)	−0.002 (0.007)	0.001 (0.003)
Country FE (# 94)	✓	✓	✓
Observations	1,521	1,521	1,521
R^2	0.234	0.128	0.102
Joint F-test	0.000	0.192	0.187

Note: Table shows ordinary least squares regressions of a binary indicator (=1 if the appeal is granted) (Model 1), the average grant rate of the chair judge across all decisions but for case j (Model 2), and the grant rate of the median judge on the panel across all decisions but for case j (Model 3). All models control for 94 country of origin fixed effects. The baseline category is "asylum and return". RR indicates reconsideration requests (following initial rejections). The joint F-test reports the p-value from the null hypothesis that all case strength characteristics, i.e. the legal matter and if the appellant was represented by a lawyer or paralegal, are not predictive of the outcome. In line with the main analysis discussed below, the sample consists of all published cases submitted in 2007 and decided by three-judge panels.

Model 1 shows that case strength characteristics strongly predict the outcome of the appeal, as expected. For example, being represented by a lawyer or paralegal is associated with an 11 percentage point higher probability of winning the appeal. When replacing the dependent variable with the general leniency of the chair judge (Model 2) and of the median judge (Model 3) on the panel, respectively, we find that conditional on country of origin, the case characteristics fail to reach (joint) statistical significance. The p-values from the joint F-tests are both around 0.19, clearly above any conventional level of significance.

IV. Data: Sample, Outcome Measure and Covariates

We obtained the data that form the basis of our analysis directly from the FAC.[16] The key dependent variable measures the outcome of the verdict. While the FAC employs a relatively fine-grained measure of appeal outcomes, we collapse this information into a binary measure, where an appeal is coded as 'granted' if the verdict potentially leads to an improvement of the appellant's situation (independent of whether the first instance decision is reversed or remanded) and 'rejected' otherwise (independent of whether the appeal is rejected or dismissed).

In addition to the outcome measure, the data obtained from the FAC contain the following information: the unique case id, submission date, decision date, the panel composition and the role of the judges, the language of the appeal and the appellant's country of origin. For published decisions, the large majority of cases, this information is also available on the court's online database.[17] We cross-checked the two data sources to confirm that we obtained the complete set of cases and the accuracy of the information. We complement this database with personal information about the judges, most importantly their party affiliation, which we compiled from judges' CVs on the official website of the FAC and supplemented with information from the minutes of the National Council.[18] As part of the data-sharing agreement reached with the FAC, we agreed to abstain from revealing the judges' names. In the following, we replace the name of each

[16]The FAC is obliged by the Administrative Court Act 29, par. 1 to inform the public about its decisions.

[17]See `http://www.bvger.ch/publiws/pub/search.jsf`.

[18]See `https://www.bvger.ch/bvger/en/home/about-fac/judges-and-court-clerks/judges.html` for short bios of sitting judges.

judge with a unique id and an indicator of his or her party affiliation.[19]

Overall, the dataset contains the universe of all 40,506 unique decisions made by a total of 42 asylum judges between January 1, 2007, and December 31, 2015.[20] Of the full set of cases, we drop 13.5% that were either 'written off' or received 'another' decision that cannot be considered as clearly in favor or against the appellant. In addition, we drop 107 cases that are decided by five-judge panels. In the Appendix, we provide detailed descriptive statistics for the 34,926 appeals in our estimation sample, of which, employing our binary outcome measure, 14% are granted and 86% rejected. Table T.1 shows grant rates grouped by panel characteristics, Table T.2 the breakdown of cases by legal category, Table T.3 the caseload by judge, and Table T.4 the proportion of cases by origin country of the appellant.

The analysis proceeds in two parts. First, we focus on substantively tried cases submitted in the first year of the court's existence that were decided by three judges to test which aggregation rule fits the panel decisions best.[21] In the second part, we follow the evolution of judges' preferences and the court's consistency over the years 2007–2015. For this year-by-year analysis, we group appeals by decision year to increase

[19] We, the authors, as well as the FAC, are well aware that this partial anonymization is incomplete at best, and that it would be fairly straightforward to figure out the identity of the judges using publicly available information. Nevertheless, we do believe that reporting anynomyzed results is helpful in focusing the discussion of our findings on structural issues of the court, rather than on the behavior of individual judges.

[20] If several appeals were unified and received a joint decision, we recorded it as one observation (this concerns 451 decisions). Occasionally, judges from other divisions of the FAC serve on asylum appeal panels. We classified all those judges into one category "other".

[21] Note that the cases that do not fulfill the formal requirements are dismissed 'without entering into the substance of the case' by the chair judge in a single-judge procedure. Since these cases are decided by single judges as opposed to panels of judges, we exclude them from this part of the analysis.

the comparability of cases over time.[22]

V. Results

A. Which Aggregation Rule Fits Panel Decisions Best?

In order to understand how individual preferences are aggregated into a joint panel decision, we begin by fitting a series of models using the aggregation rules introduced in the previous section: preference of the most restrictive judge (min), most lenient judge (max), median, chair, and the mixture model combining the latter two rules. In addition, we include the preference of second and third judge and the null model as plausibility tests, but do not expect them to fit well. Therefore, they provide us with a check that our estimation approach has statistical power against implausible alternatives. All of these models have one degree of freedom per judge, except for the null model, which only fits a constant, and the mixture model which adds the mixing parameter.

During our extensive test runs, we found that for some models, the maximum likelihood estimates are somewhat dependent on the starting values, indicating that we might only find local, not global maxima. Hence, we resort to Bayesian Markov chain Monte Carlo (MCMC) to estimate our models, which is better suited to explore the entire posterior density. We add hierarchical random effects priors for each judge, but the substantive results are the same with flat priors. All models control for appellants' country of origin. To facilitate the comparison across models, we employ the Deviance Information Criteria (DIC), a Bayesian generalization of the AIC, to assess model fit.

[22]Because of the difference in classifying cases according to their submission date (for the first part of the analysis) or decision date (for the second part of the analysis), the preference estimates and inconsistency rate for 2007 will slightly differ between the two parts.

Table 3: Fit Statistics for Preference Aggregation Rules

	DIC	LL	Inconsistency Rate	95% Credible Interval	Parameters
mixture	1430.4	-705.1	0.06	0.05, 0.08	119
chair	1440.1	-710.8	0.06	0.04, 0.07	118
median	1441.3	-709.1	0.06	0.04, 0.08	118
max	1453.0	-716.4	0.09	0.05, 0.13	118
min	1460.6	-719.9	0.09	0.04, 0.16	118
second	1486.4	-733.3	0.03	0.01, 0.05	118
null	1491.7	-739.5	0.00	—, —	91
third	1492.5	-738.2	0.02	0.00, 0.03	118

Note: Table of fit statistics for Bayesian estimates of judges' preferences in 2007 under the mixture, chair, median, max (most lenient), min (most restrictive), second, and third judge, and the null model. Models are sorted by Deviance Information Criteria (DIC) shown in Column 1. Column 2 shows the log likelihood statistic. Columns 3 and 4 show the inconsistency rate and associated 95% credible interval. Note that the inconsistency rate for the null model is by definition zero. Column 5 shows the number of estimated parameters.

Table 3 shows the results. We first discuss the simple, i.e., non-mixture, aggregation rules. The best-fitting simple aggregation rule is the chair judge, closely followed by the median model. The difference in the DIC between these two non-nested models with the same number of parameters is 1.2, indicating that chair model fits the data only marginally better than the median model. All of the other non-mixture models fit considerably worse than these two according to the DIC. That (one of) the best-fitting simple aggregation rule posits that the chair judge decides as dictator is a theoretically compelling result, given the structure of the decision procedure followed by the court. Because the chair sees the case first and writes the initial draft of the decision, she has an opportunity to frame the decision, while the second and third judges have an incentive to not investigate the case as thoroughly as they would if they were the chair. However, the results also indicate that the preferences of the second and third judge matter to some extent. If the other judges exerted no constraint on the chair, we would expect a larger and statistically meaningful difference between the chair and the median model in terms of the DIC.

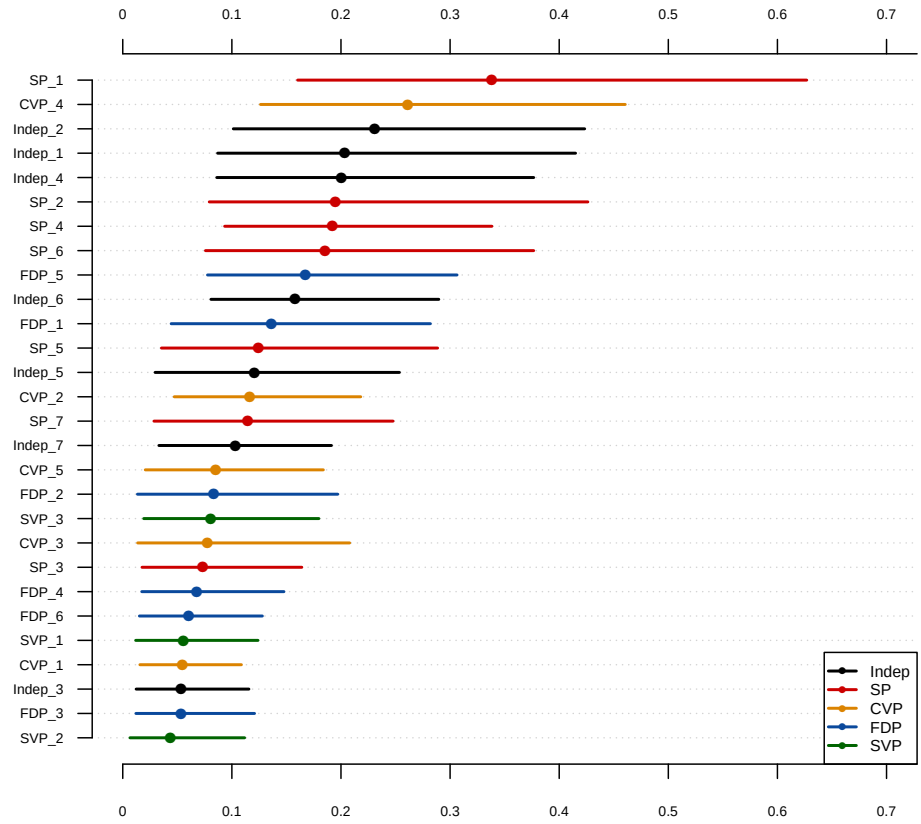
Next, we turn to the results from the mixture model to more explicitly investigate the trade-off faced by second and third judge between paying the cost for review and letting the chair decide. The comparison between the mixture and the chair or median models

is aided by their nested structure (c.f. equation 2): the chair model, θ_1, is a special case of the mixture with $\lambda_1 = 1$, and the median model is a special case of the mixture with $\lambda_1 = 0$. Table 3 shows that the mixture model performs significantly better than the 'pure' chair and median models. The estimated mixing parameter λ_1 is 0.45 (95% CI 0.17-0.77), indicating that slightly less than half the cases are decided by the chair, and slightly more than half by majority rule. The fact that the DIC is almost 10 points lower (better) for the mixture model, and that λ_1 is neither close to zero (in which case the mixture would simplify to the median model) nor close to one (in which case the mixture would simplify to the chair model), indicates that even when penalizing for higher model complexity, a mixture of the chair and median judge models substantially outperforms an aggregation rule in which the chair can always act as dictator or all cases are decided by majority rule. Substantively, this implies that the chair judge has disproportionate, but not absolute, control over the panel's decision.

B. Heterogeneity in Judges' Preferences and Inconsistency of the Court

Having identified the best-fitting aggregation rule, this section explores the heterogeneity in grant rates between judges in 2007, and how this impacts the consistency of decision-making at the court. Figure 2 shows the estimated preferences of the judges from the mixture model that controls for origin country and uses hierarchical priors on the judges, setting the country of origin effect to its average value. Table T.5 in the Appendix shows the underlying numerical estimates.

Figure 2: Heterogeneity in Judges Preferences



Note: Estimated preferences of judges from the best-fitting mixture model that controls for origin country and uses hierarchical priors on judges, setting the country of origin effect to its average value. Sample consists of all three-judge panel decisions on cases submitted in 2007. Mixture probability for chair model: $\lambda_1 = .45$. Posterior means and 95% credible intervals. Parties are abbreviated as follows: Christian Democrats (CVP); Free Democratic Party (FDP); non-partisan (Indep); Social Democrats (SP); Swiss People's Party (SVP).

Figure 2 reveals three striking features. First, there is substantial heterogeneity in the preferences of judges, with preferred grant rates ranging between 5% and 34%. Second, this heterogeneity is driven by variance both within and across parties. Third, despite relevant within-party variation, there is a clear association between the preferences of judges and their political affiliation in the expected direction. Judges affiliated with the leftist Social Democratic Party (SP) are, on average, among the most favorable toward

asylum seekers. Non-partisan judges and those affiliated with the centrist Christian Democracts (CVP) exhibit the most intra-party variance, but both groups are on average close to the court's median. Judges affiliated with the center-right Free Democratic Party (FDP) are the second most restrictive, while judges affiliated with the populist right-wing Swiss People's Party (SVP) are the least favorable toward asylum seekers. Figure F.1 in the Appendix shows the voting behavior of MPs, averaged across parties, in the Swiss National Council for the study period 2007 to 2015. We find that the estimated preference ordering of the asylum judges on the FAC is entirely consistent with their parties' general stance on asylum policies.

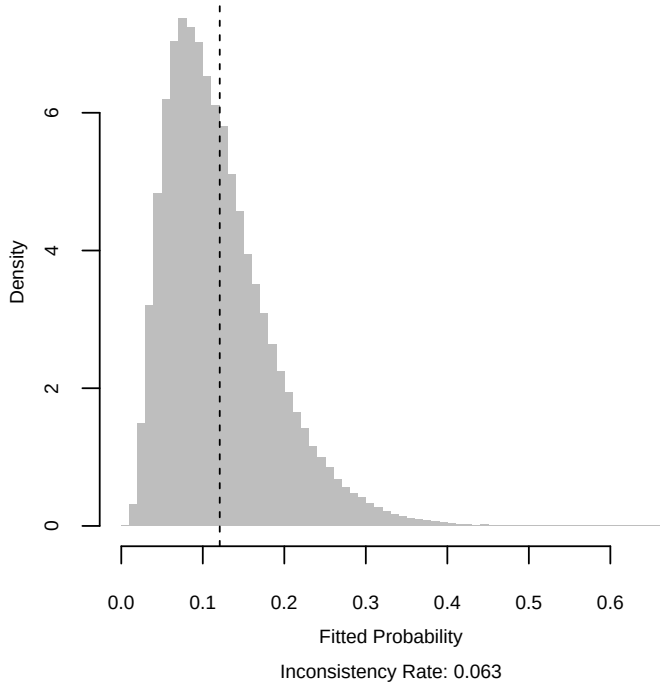
The differences in preferences are substantively meaningful. For example, the most lenient judge (SP, preferred grant rate = 33.8%: 95% CI: 16.0–62.7%) would, if s/he could decide alone, achieve a grant rate that is almost eight times higher than the rate of the most restrictive judge (SVP, preferred grant rate = 4.4%: 95% CI: 0.7–11.2%). Figure 2 makes clear that this preference variation is not restricted to the two most extreme judges. For example, the difference in preferred grant rates for the five most lenient (average grant rate = 21.0%) and five most restrictive judges (average grant rate = 5.2%) are similarly sizeable and estimated with sufficient precision to be statistically meaningful.

How robust are these findings? Figures F.2 and F.3 in the Appendix show the corresponding preference estimates from the (non-mixture) Bayesian chair and median model, respectively. While there are some differences with regard to the point estimates for the preferences and the implied ordering of judges from lenient to restrictive, the general findings of a substantial variation in preferences, and their association with party membership, are also clearly evident in those simpler models. In sum, we find across a variety of models that the political ideology of judges, proxied by their party affiliation, is a robust predictor of their preferred grant rate.

What does this heterogeneity in the preferences of the judges imply for the consistency with which the court applies the law? To answer this question, we predict the probability of a successful appeal for each composition of panels, as observed in 2007, and based on the preference estimates from the best-fitting mixture model. If panel composition had no effect on the success of appeals, we would predict a constant probability for all cases (at the court's average grant rate of 12.1% for 2007). In this case,

the inconsistency rate would be zero.

Figure 3: Distribution of Predicted Probability of Appeal Success as a Function of the Judges Serving on the Panel



Note: Empirical distribution of the predicted probability of appeal success for each case under the mixture model, given the observed compositions of three-judge panels for cases submitted in 2007 and controlling for origin country (and setting the country of origin effect to its average value). The court's average predicted grant rate, indicated by the dashed black line, is 12.1%.

Figure 3 shows that this is clearly not the case. In contrast, we find considerable variation in the predicted grant probabilities solely due to different judges serving on panels. Overall, this heterogeneity in adjudication results in an inconsistency rate of 6.3% (95% CI 4.6–8.2%), indicating that more than one in sixteen cases is decided differently than how it would be if the court's consensus were consistently applied. Column 3 in Table 3 shows that inconsistency rates are virtually identical for the chair and median models.

C. Does the Court Become More Consistent Over Time?

Two related questions that emerge from the substantial preference variation reported in the previous section concern the possible convergence of these preferences over time: is the inconsistency rate observed in 2007 merely an artefact of the first year of the court's existence and do judges' preferences converge over subsequent years? Or is it a rather permanent feature of a politicized court whose judges are affiliated with political parties and voted into office by the parliament?

In order to answer these questions, we now turn to the analysis of all asylum appeal decisions between 2007 and 2015. As discussed above, on January 1, 2008, the court introduced the simplified procedure that allows the chair judge to avoid three-judge panels for appeals that she deems, with the approval of the second judge, 'clearly with or without merit'. Among other things, this means that we do not observe the identity of the third judge for this selective subset of cases over the entire study period.[23] Methodologically, the missing data on the potential third judge is an issue if the chair's classification of cases clearly with(out) merit also depends on the preferences of the third judge. For example, a restrictive judge might be more likely to rule that an appeal is clearly without merit if the third judge is lenient, and thereby avoid the more complex procedure of determining the case's merits in a full panel, and running the risk of having to grant it against her preferences.

While we cannot directly test the hypothesis that judges' propensity to invoke the simplified procedure depends on the second and (unobserved to us) third judge, we can show that they disagree about what constitutes a case that is clearly with(out) merit. Appendix Figure F.4 shows the Bayesian estimates from regressing the binary decision to invoke the ordinary procedure (as opposed to a shorter procedure) on judge and country of origin fixed effects using all cases decided between 2008 and 2015. We not only find considerable variation across judges in the likelihood of invoking the ordinary procedure,

[23] We do observe the potential third judge even for cases decided under the simplified and single-judge procedure for cases submitted after 2011, when the FAC implemented a software update that saves the initial assignment of all three judges, independent of whether or not they were involved in deciding the case.

ranging between 58 percent (95% CI 36–79) and 8 percent (95% CI 6–11), but also see that this decision is strongly correlated with party membership in the expected direction (i.e., more conservative judges are less likely to invoke the ordinary procedure).

Thus, we see two principled ways to deal with those cases for which the simplified procedure is used. Our preferred option is to impute the missing third judge with the court's average preference (computed without the preferences of the chair and second judge serving on the particular case) and then apply the mixture model. The imputed average is a consistent estimator of the preference of the unobserved third judge and will therefore not bias the estimates for the other judges.[24] An alternative option is to focus on the second best-fitting chair model, for which the identity of the third (and second) judge is by definition not relevant. Appendix Figure F.5 shows that the evolution of the court's preference and inconsistency estimated from the chair model is very similar to estimates from the mixture model with imputation.

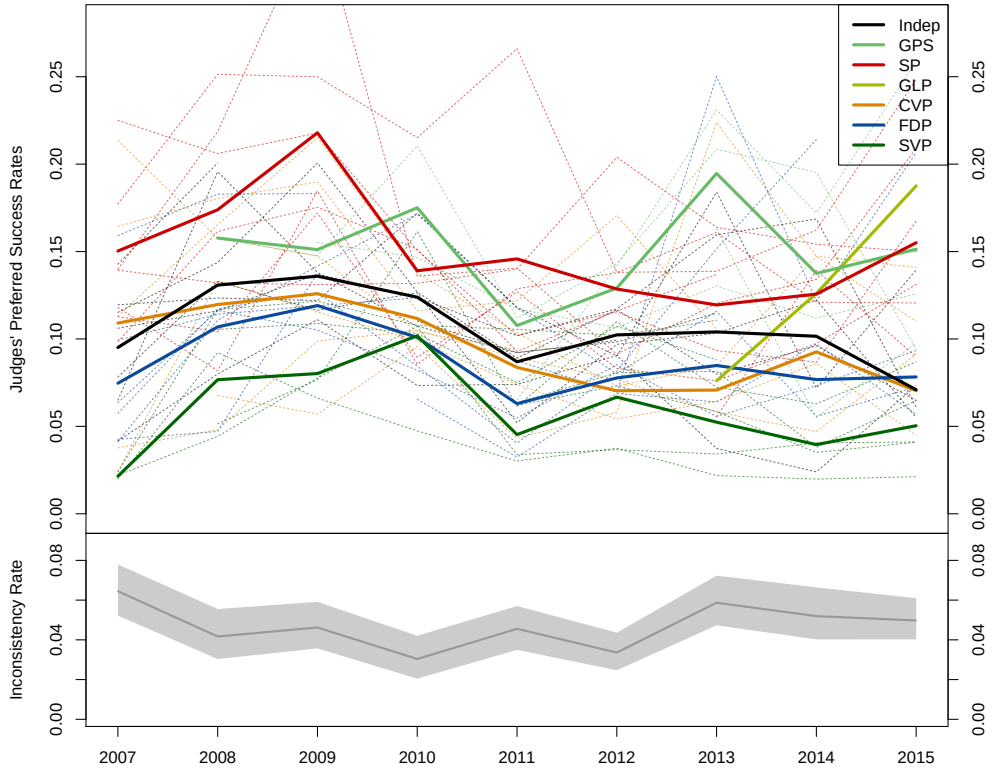
Another caveat concerns the comparison of dynamic preferences over time. While the random assignment of judges to panels allows us to attribute differences in grant rates to differences in judges' ideology at any given point in time, it does not allow us to causally attribute changes in grant rates of the same judge over time to changes in his or her ideology, since we cannot rule out that the average merit of asylum appeals also varies. We expect that controlling for country of origin adjusts for most, but probably not all, of the unobserved variation in case strength across years. Without further and strong assumptions about constancy of case merits (conditional on country of origin), we cannot make absolute comparisons of judges' preferences across years. In effect, this means that we are not able to distinguish the hypothesis that cases become systematically easier

[24] We expect, however, that this imputation procedure will somewhat underestimate the weight of the median model (relative to the chair model), since the preferences of the actual third judge will typically be farther away from the median than the imputed mean suggests. Consistent with the preceding analysis of judges' propensity to invoke the ordinary procedure, our sample also includes single judge decisions that are solely decided by the chair judge without involving a second or third judge. Thus, we analyze those cases with the chair model.

(harder) to decide consistently from the hypothesis that the judges converge (diverge) in their behavior. Nonetheless, our flexible model specification, where the preferences of the judges, and the distributions of cases from each country, vary from year to year enables us to determine, without invoking any further assumptions, whether overall consistency improved or worsened over time.

The top panel of Figure 4 shows the evolution of judges' preferences estimated from the mixture model from 2007 to 2015. The dashed lines show the preferred grant rate of the individual judges, and the solid lines show the mean grant rate, grouped by parties. While there is some variation in preferences from year to year, the overall pattern is very clear: the substantial variance of judges' preferred grant rates, and the correlation with party affiliation, remains fairly constant over time.

Figure 4: Evolution of Preference Estimates and Inconsistency Rate 2007–2015



Note: Top panel: Year-by-year preference estimates for all judges serving on the court between 2007 and 2015, with party mean trajectories. Judges' preferences are estimated based on all cases decided in a given year. Party names: Conservative Democrats (BDP); Christian Democrats (CVP); Free Democratic Party (FDP); Green Liberal Party (GLP); Green Party (GPS); non-partisan (Indep); Social Democrats (SP); Swiss People's Party (SVP). Bottom panel: Evolution of inconsistency rate (dark grey line) and corresponding 95% credible intervals (light gray) estimated from the same data. All estimates are based on the mixture model.

Next, we explore the evolution of the inconsistency rate more explicitly. The bottom panel of Figure 4 shows the year-to-year inconsistency rate estimated from the mixture model for 2007-2015. While there is considerable fluctuation in the inconsistency rate across years, ranging between 3.0 and 6.4 percent, the narrow credible intervals show that it is always significantly larger than zero. When tracing the evolution of the inconsistency rate over the study period, we find no evidence that judges' preferences converged over

time. In sum, these findings suggest that preference variation in asylum adjudication on the FAC is by no means confined to the early years of the court's existence, but rather a persistent feature. The following conclusion explores the legal and political implications of this finding.

VI. Conclusion

Several studies of asylum adjudication show considerable variation between decision makers (see, e.g., Fischman 2011; Ramji-Nogales, Schoenholtz and Schrag 2007; Rehaag 2007). A common limitation faced by existing studies is that researchers often only have access to anonymized information about the identity of the decision maker, which renders correlating estimated preferences with decision maker characteristics such as political ideology impossible. We overcome this challenge by focusing on asylum appeal decisions of the Swiss FAC, where judges' identity and ideology, proxied by party affiliation, is public knowledge. A second limitation of existing studies is their almost exclusive focus on common law courts, where judges' individual opinions are disclosed. In contrast, in many civil law courts, only the verdict of the entire panel is published. To date, this has made it impossible to estimate individual preferences of judges serving in those courts. We tackle this second challenge by developing a methodology that allows us to test different models of preference aggregation and infer individual preferences from repeated group decisions.

Our analysis of the universe of asylum appeals decided between 2007 and 2015 by the FAC demonstrates that judges' preferences vary substantially with regard to grant rates. Furthermore, judges' preferred grant rates correlate strongly with their political ideology in expected ways. Since cases are, conditional on appellants' origin country, randomly assigned to judges, the disparities between judges cannot be explained by differences in case merits. Overall, this leads to substantial inconsistency in the court's decision making, and we find no evidence that the resulting arbitrariness in adjudication is confined to the early years of the court. These persistent disparities violate Aristotle's maxim that 'like cases be treated alike' and the very essence of article 8(1) of the Swiss Constitution, which stipulates that "every person is equal before the law."

Our findings have important implications for several audiences. For the comparative literature on disparities in asylum adjudication, our paper provides some of the most

direct evidence to date that judges' political ideology influences their preferences over asylum appeals, even in the context of a high-stakes appeal court of last resort. While there is considerable within-party variation, the disparity between the preferred grant rate of judges affiliated with the most left-wing and the most right-wing parties are both statistically significant and substantially relevant. The resulting inconsistency rate is, however, smaller than what has been found in previous studies in the U.S. and Canada (see Fischman 2011; Ramji-Nogales, Schoenholtz and Schrag 2007; Rehaag 2007).

Our study also has direct policy implications for the Swiss FAC. Our background research and meetings with members of the court revealed that the court is aware of, and concerned by, allegations of disparities in adjudication that we quantitatively confirm here. The FAC has two institutional features that are designed to increase consistency in decision-making. First, when new legal questions arise or the situation in a country of origin changes, five-judge panels issue leading decisions that have *stare decisis*-like implications for subsequent decisions on similar appeals. Second, judges from different chambers are assigned to serve together on panels in non-urgent cases to facilitate exchange and consistency. While our analysis cannot speak to the success or failure of these measures, the overall trend gives little reason for optimism: When exploring the evolution of disparities in appeal decision over the study period, we find little evidence that the court converged in more recent years.

If the goal is to minimize the variation in grant rates within the current partisan selection procedure, the FAC would have to look at the design of the decision-making procedure. Abolishing the simplified procedure would likely have only minor effects on the overall inconsistency (while significantly increasing the workload), since the structure of the ordinary decision-making procedure also grants considerable power to the chair judge. Based on the findings of this study, we believe that panels are only able to effectively moderate inconsistency if all three judges have to simultaneously review the appeal and independently draft a verdict. While more work than under the existing procedure, only such a redesigned decision-making process promises to unleash the full power of Condorcet's jury theorem (Kornhauser and Sager 1986; Lorenz et al. 2011).

Lastly, our study also has broad methodological implications for scholars concerned with inferring individual preferences from group decisions. We show that in a context where decision makers are repeatedly and randomly allocated to groups, our methodol-

ogy can recover the aggregation rule (or mixture of aggregation rules) that best fits the decision-making process, and the individual preferences, from group decisions—without observing individual votes. We expect that this methodology can also be fruitfully applied in a variety of other contexts of repeated group interactions, where joint decisions or performance indicators without any information on individual votes or contributions are the norm. Beyond judicial behavior, particularly promising examples for future applications are, for example, the estimation of preferences of MPs who are repeatedly allocated to serve on various committees, or the abilities of students working on group projects with rotating group membership. We hope that sharing the code that implements the estimator proposed in this study (add `dataverse doi` upon acceptance) facilitates the adoption of this methodology in political science and neighbouring disciplines concerned with decision-making and preference aggregation of groups with varied interests.

References

Abrams, David S, Marianne Bertrand and Sendhil Mullainathan. 2012. "Do Judges Vary in Their Treatment of Race?" *The Journal of Legal Studies* 41(2):347–383.

Ashenfelter, Orley, Theodore Eisenberg and Stewart J Schwab. 1995. "Politics and the Judiciary: The Influence of Judicial Background on Case Outcomes." *Journal of Legal Studies* 24:257–281.

Bansak, Kirk, Jens Hainmueller and Dominik Hangartner. 2017. "Europeans Support a Proportional Allocation of Asylum Seekers." *Nature Human Behaviour* 1(7):0133.

Bonneau, Chris W, Thomas H Hammond, Forrest Maltzman and Paul J Wahlbeck. 2007. "Agenda Control, the Median Justice, and the Majority Opinion on the U.S. Supreme Court." *American Journal of Political Science* 51(4):890–905.

Boyd, Christina L, Lee Epstein and Andrew D Martin. 2010. "Untangling the Causal Effects of Sex on Judging." *American Journal of Political Science* 54(2):389–411.

Carrubba, Clifford J, Matthew Gabel and Charles Hankla. 2008. "Judicial Behavior under Political Constraints: Evidence from the European Court of Justice." *American Political Science Review* 102(4):435452.

Epstein, Lee, William M Landes and Richard A Posner. 2013. *The Behavior of Federal Judges: A Theoretical and Empirical Study of Rational Choice*. Harvard University Press.

Fischman, Joshua B. 2011. "Estimating Preferences of Circuit Judges: A Model of Consensus Voting." *Journal of Law and Economics* 54(4):781–809.

Fischman, Joshua B. 2013. "Interpreting Circuit Court Voting Patterns: A Social Interactions Framework." *Journal of Law, Economics, & Organization* 31(4):808–842.

Frandsen, Brigham R, Lars J Lefgren and Emily C Leslie. 2019. "Judging Judge Fixed Effects.".

Gazal-Ayal, Oren and Raanan Sulitzeanu-Kenan. 2010. "Let My People Go: Ethnic In-Group Bias in Judicial Decisions: Evidence from a Randomized Natural Experiment." *Journal of Empirical Legal Studies* 7(3):403–428.

Glynn, Adam N and Maya Sen. 2015. "Identifying Judicial Empathy: Does Having Daughters Cause Judges to Rule for Women's Issues?" *American Journal of Political Science* 59(1):37–54.

Greenhouse, Linda. 2007. *Becoming Justice Blackmun: Harry Blackmun's Supreme Court Journey*. Macmillan.

Grossman, Guy, Oren Gazal-Ayal, Samuel D Pimentel and Jeremy M Weinstein. 2016. "Descriptive Representation and Judicial Outcomes in Multiethnic Societies." *American Journal of Political Science* 60(1):44–69.

Hanretty, Chris. 2012. "Dissent in Iberia: The Ideal Points of Justices on the Spanish and Portuguese Constitutional Tribunals." *European Journal of Political Research* 51(5):671–692.

Helmke, Gretchen and Jeffrey K Staton. 2011. "The Puzzling Judicial Politics of Latin America." *Courts in Latin America* pp. 306–31.

Hönnige, Christoph. 2009. "The Electoral Connection: How the Pivotal Judge Affects Oppositional Success at European Constitutional Courts." *West European Politics* 32(5):963–984.

Jeffries, John Calvin. 2001. *Justice Lewis F Powell, Jr.* Fordham Univ Press.

Kapiszewski, Diana and Matthew M Taylor. 2008. "Doing Courts Justice? Studying Judicial Politics in Latin America." *Perspectives on Politics* 6(4):741–767.

Kastellec, Jonathan P. 2013. "Racial Diversity and Judicial Influence on Appellate Courts." *American Journal of Political Science* 57(1):167–183.

Kiener, Regina. 2001. *Richterliche Unabhängigkeit: Verfassungsrechtliche Anforderungen an Richter und Gerichte*. Stämpfli.

Kornhauser, Lewis A. 1992. "Modeling Collegial Courts. II. Legal Doctrine." *Journal of Law, Economics, & Organization* 8:441–470.

Kornhauser, Lewis A and Lawrence G Sager. 1986. "Unpacking the Court." *Yale Law Journal* 96:82–117.

Lauderdale, Benjamin E and Tom S Clark. 2012. "The Supreme Court's Many Median Justices." *American Political Science Review* 106(04):847–866.

Law, David S. 2004. "Strategic judicial Lawmaking: Ideology, Publication, and Asylum Law in the Ninth Circuit." *University of Cincinnati Law Review* 73:817–866.

Lorenz, Jan, Heiko Rauhut, Frank Schweitzer and Dirk Helbing. 2011. "How Social Influence Can Undermine the Wisdom of Crowd Effect." *Proceedings of the National Academy of Sciences* 108(22):9020–9025.

Marmor, Andrei. 2007. *Law in the Age of Pluralism*. Oxford University Press.

Martén, Linna. 2015. "Political Bias in Court? Lay Judges and Asylum Appeals." *Uppsala University Department of Economics Working Paper Series 2* .

Peresie, Jennifer L. 2005. "Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Appellate Courts." *Yale Law Journal* pp. 1759–1790.

Ramji-Nogales, Jaya, Andrew I Schoenholtz and Philip G Schrag. 2007. "Refugee Roulette: Disparities in Asylum Adjudication." *Stanford Law Review* 60:295–411.

Raselli, Niccolò. 2011. "Richterliche Unabhängigkeit." *Justice—Justiz—Giustizia* 3.

Rehaag, Sean. 2007. "Troubling Patterns in Canadian Refugee Adjudication." *Ottawa Law Review* 39:335–365.

Revesz, Richard L. 1997. "Environmental regulation, ideology, and the DC Circuit." *Virginia Law Review* 83(8):1717–1772.

Shayo, Moses and Asaf Zussman. 2011. "Judicial Ingroup Bias in the Shadow of Terrorism." *The Quarterly Journal of Economics* 126(3):1447–1484.

Sunstein, Cass R, David Schkade, Lisa M Ellman and Andres Sawicki. 2007. *Are Judges Political?: An Empirical Analysis of the Federal Judiciary*. Brookings Institution Press.

Taylor, Margaret H. 2007. "'Refugee Roulette' in an Administrative Law Context: The "Déjà vu" of Decisional Disparities in Agency Adjudication." *Stanford Law Review* 60:475–501.

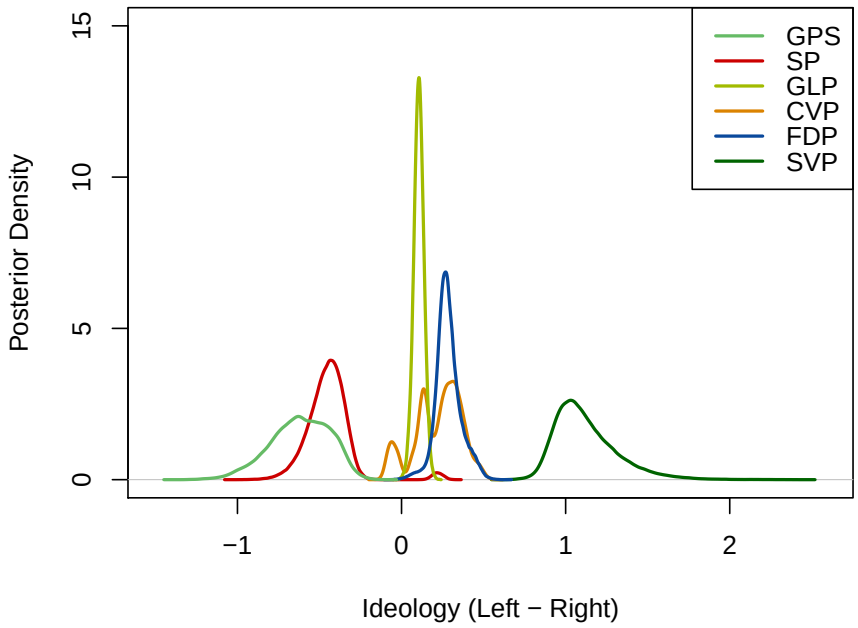
UNHCR. 2013. "Asylum Levels and Trends in Industrialized Countries – 2012." UNHCR Report.

Van Dijk, Frans, Joep Sonnemans and Eddy Bauw. 2014. "Judicial Error by Groups and Individuals." *Journal of Economic Behavior & Organization* 108:224–235.

Online Appendix

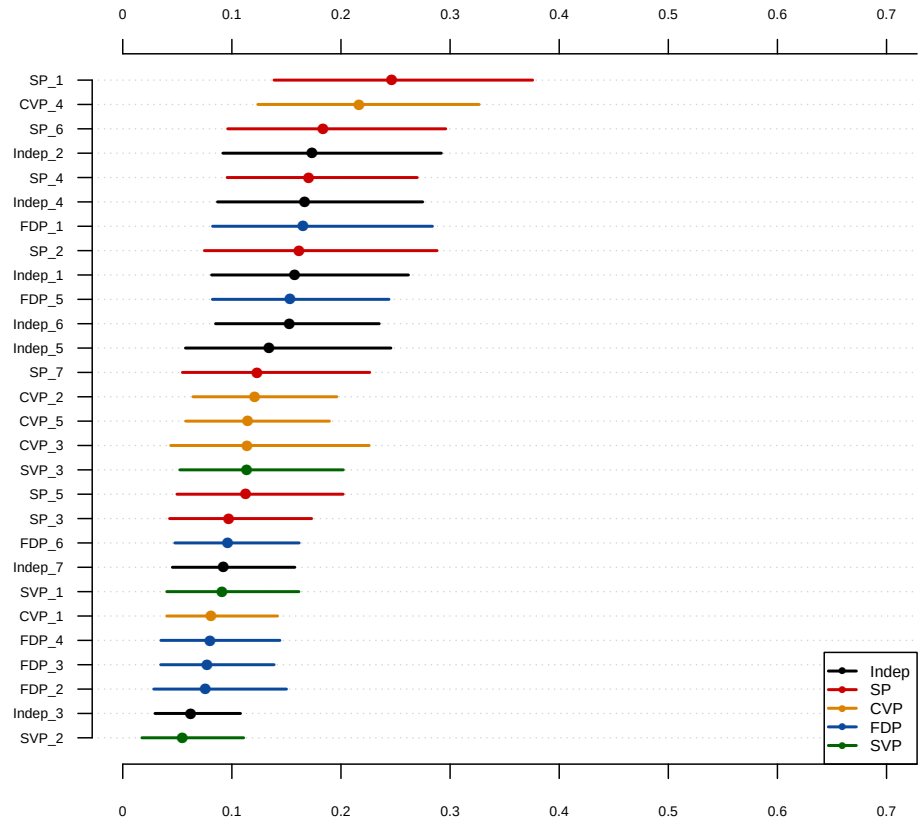
Additional Figures

Figure F.1: Party Aggregates of MPs' Preferences on Asylum Issues



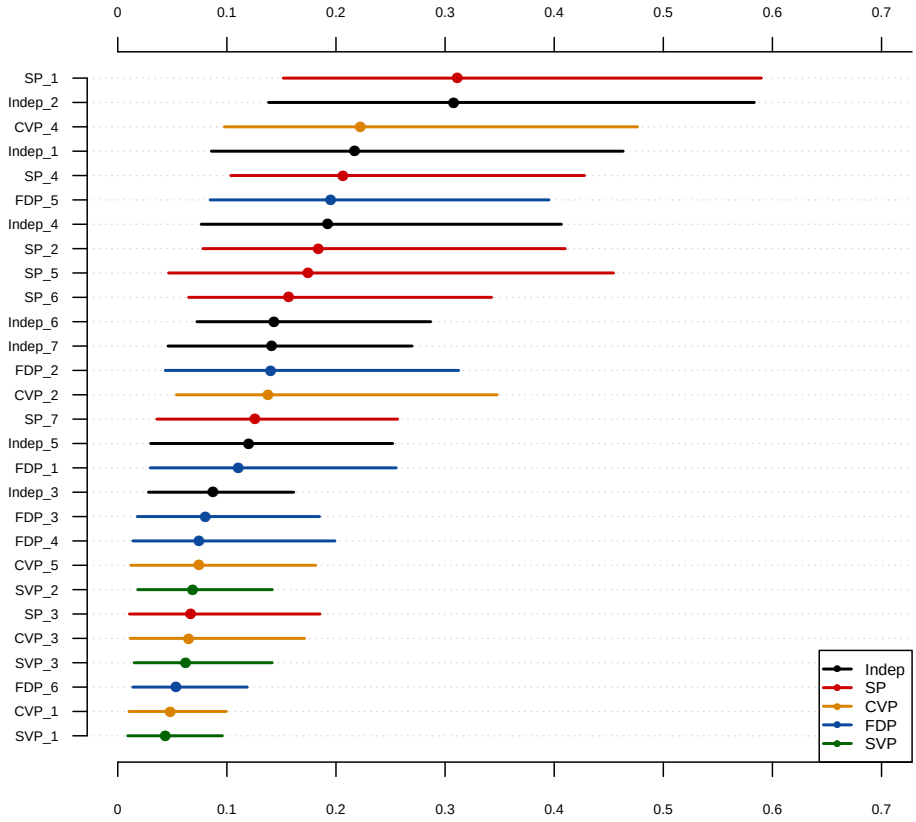
Note: Figure displays item-response theory (IRT) estimates of parties' ideal points from pro- vs. anti-asylum roll call votes cast by MPs in the National Council over the study period, 2007–2015.

Figure F.2: Chair Model: Heterogeneity in Judges Preferences



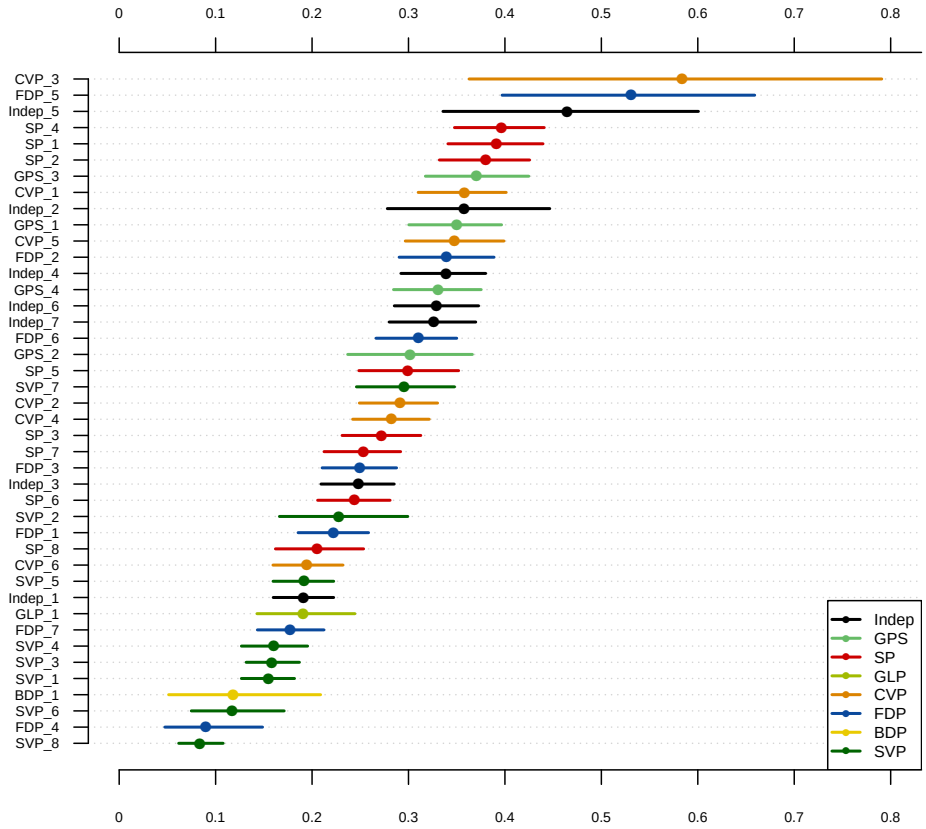
Note: Estimated preferences of judges from chair model that controls for origin country and uses hierarchical priors on judges, setting the country of origin effect to its average value. Sample consists of all three-judge panel decisions on cases submitted in 2007. Posterior means and 95% credible intervals. Parties are abbreviated as follows: Christian Democrats (CVP); Free Democratic Party (FDP); non-partisan (Indep); Social Democrats (SP); Swiss People's Party (SVP).

Figure F.3: Median Model: Heterogeneity in Judges Preferences



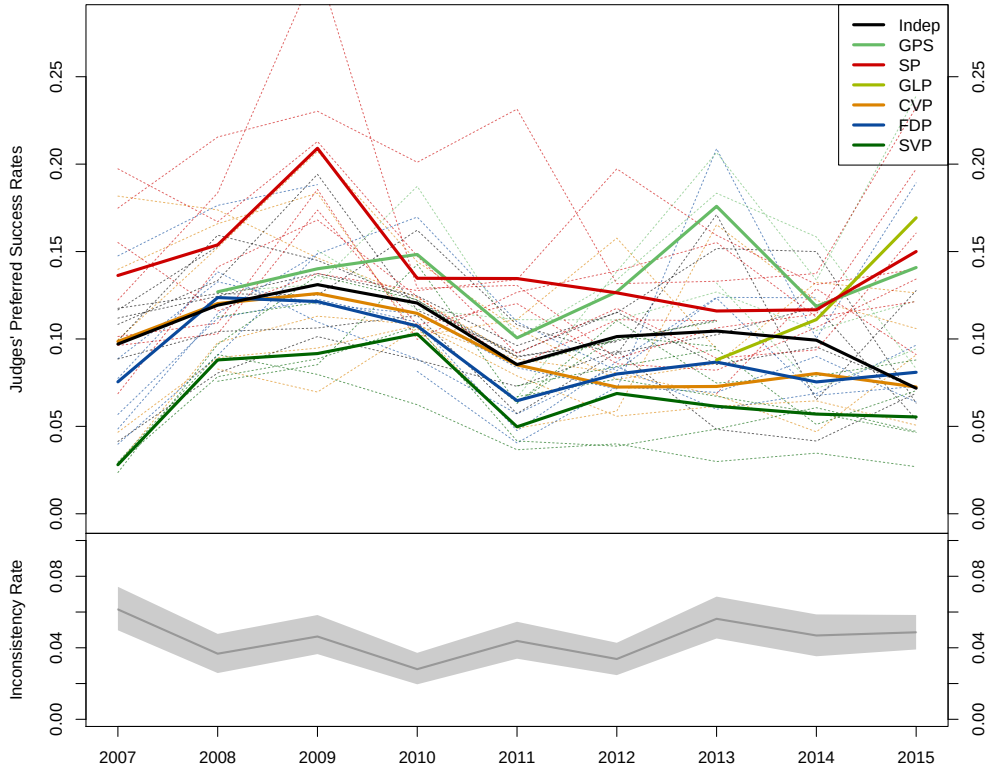
Note: Estimated preferences of judges from median model that controls for origin country and uses hierarchical priors on judges, setting the country of origin effect to its average value. Sample consists of all three-judge panel decisions on cases submitted in 2007. Posterior means and 95% credible intervals. Parties are abbreviated as follows: Christian Democrats (CVP); Free Democratic Party (FDP); non-partisan (Indep); Social Democrats (SP); Swiss People's Party (SVP).

Figure F.4: Probability of Deciding Case by Ordinary Procedure



Note: The graph displays the probability that a chair judge invokes the ordinary procedure, as opposed to the simplified (for cases clearly with(out) merit) or single-judge procedure (without entering into the substance of the case). Point estimates and corresponding confidence intervals are Bayesian estimates from regressing the binary decision to invoke the ordinary procedure (as opposed to one of the shorter procedures) on chair judge and country of origin fixed effects using all cases decided between 2008 and 2015. Party names: Christian Democrats (CVP); Free Democratic Party (FDP); Green Liberal Party (GLP); Green Party (GPS); non-partisan (Indep); Social Democrats (SP); Swiss People's Party (SVP).

Figure F.5: Chair Model: Evolution of Preference Estimates and Inconsistency Rate



Note: Top panel: Year-by-year preference estimates for all judges serving on the court between 2007 and 2015, with party mean trajectories. Judges' preferences are estimated based on all cases decided in a given year. Party names: Conservative Democrats (BDP); Christian Democrats (CVP); Free Democratic Party (FDP); Green Liberal Party (GLP); Green Party (GPS); non-partisan (Indep); Social Democrats (SP); Swiss People's Party (SVP). Bottom panel: Evolution of inconsistency rate (dark grey line) and corresponding 95% credible intervals (light gray) estimated from the same data. All estimates are based on the chair model.

Additional Tables

Table T.1: Grant Rates Grouped by Panel Characteristics

	Grant Rate
All cases	0.14
Ordinary Procedure	0.26
Simplified Procedure	0.09
Single-Judge Decisions	0.02
German	0.15
French	0.12
Italian	0.05
CVP	0.12
FDP	0.11
GLP	0.22
GPS	0.19
Independent	0.13
SP	0.20
SVP	0.08

Note: Grant rate averaged across all 34,926 cases decided between 2007 and 2015 and grouped according to the procedure invoked (ordinary, simplified, or single), the language of the case (German, French, Italian), and the party affiliation of the chair judge.

Table T.2: Descriptive Statistics for Legal Categories

Legal Category	Proportion of Cases
Asylum and return	0.393
Inadmissibility of request	0.349
Asylum request abroad	0.060
Asylum and return (RR)	0.045
Enforcement of return	0.027
Asylum	0.026
Asylum procedure (other)	0.025
Return / enforcement (RR)	0.025
Revocation temporary protection	0.016
Family reunification	0.012
Allocation to canton	0.007
Airport asylum and return	0.006
Revocation of asylum	0.005
Airport entry refusal	0.001
Airport Inadmissibility	0.001
Deprivation of refugee status	0.001
Detention review	0.001
Costs	0.001

Note: Table shows the proportion of cases in each legal category for all categories with at least 0.1% of cases. RR indicates reconsideration requests (following initial rejections). The table includes all 34,926 cases decided between 2007 and 2015. Note that the balance tests for 2007 only include legal categories with at least one case in that year.

Table T.3: Caseload by Judge

Chair Judge	Proportion of Cases
FDP_6	0.048
SVP_1	0.047
SVP_3	0.043
CVP_2	0.042
Indep_1	0.042
Indep_4	0.041
Indep_3	0.040
CVP_1	0.039
Indep_7	0.038
SP_4	0.038
CVP_4	0.037
SP_2	0.037
SVP_5	0.035
Indep_6	0.035
SP_6	0.031
SP_1	0.030
FDP_1	0.029
SP_3	0.028
FDP_3	0.028
SP_7	0.025
FDP_7	0.024
CVP_6	0.024
GPS_4	0.022
GPS_1	0.022
CVP_5	0.022
FDP_2	0.021
GPS_3	0.018
SVP_8	0.017
SVP_4	0.016
SP_5	0.014
SVP_7	0.013
SP_8	0.011
FDP_4	0.009
GLP_1	0.006
GPS_2	0.006
SVP_2	0.005
Indep_2	0.005
FDP_5	0.004
SVP_6	0.004
Indep_5	0.004
CVP_3	0.003
BDP_1	0.001

Note: Table shows the proportion of cases by chair judge for the 34,926 cases decided between 2007 and 2015.

Table T.4: Proportion of Cases by Top-20 Origin Country

Origin Country	Proportion of Cases
Sri Lanka	0.084
Serbia	0.084
Nigeria	0.078
Eritrea	0.077
Turkey	0.058
Iraq	0.050
Afghanistan	0.040
Syrian Arab Republic	0.035
Congo, the Democratic Republic of the	0.032
Ethiopia	0.030
Iran, Islamic Republic of	0.027
unknown	0.026
Georgia	0.022
Russian Federation	0.019
Guinea	0.018
Bosnia and Herzegovina	0.018
Algeria	0.017
Somalia	0.017
missing	0.016
Tunisia	0.016

Note: Table shows the proportion of cases by appellants' origin country (restricted to top 20) for the 34,926 cases decided between 2007 and 2015.

Table T.5: Mixture Model: Judges' Preference Estimates for 2007

Judge	Point Estimate	2.5%	97.5%
SP_1	0.338	0.160	0.627
CVP_4	0.261	0.126	0.461
Indep_2	0.231	0.101	0.423
Indep_1	0.204	0.087	0.415
Indep_4	0.200	0.086	0.377
SP_2	0.195	0.079	0.426
SP_4	0.192	0.093	0.338
SP_6	0.186	0.076	0.377
FDP_5	0.167	0.078	0.306
Indep_6	0.158	0.081	0.290
FDP_1	0.136	0.044	0.282
SP_5	0.124	0.035	0.289
Indep_5	0.121	0.030	0.254
CVP_2	0.116	0.047	0.218
SP_7	0.114	0.029	0.248
Indep_7	0.103	0.033	0.191
CVP_5	0.085	0.021	0.184
FDP_2	0.083	0.013	0.197
SVP_3	0.081	0.019	0.180
CVP_3	0.078	0.014	0.208
SP_3	0.073	0.018	0.164
FDP_4	0.068	0.017	0.148
FDP_6	0.060	0.015	0.128
SVP_1	0.055	0.012	0.124
CVP_1	0.055	0.016	0.109
Indep_3	0.053	0.012	0.116
FDP_3	0.053	0.012	0.121
SVP_2	0.044	0.007	0.112

Note: Table displays estimated preferences of the judges from the best-fitting mixture model that controls for origin country and uses hierarchical priors on the judges, setting the country of origin effect to its average value. Sample consists of all three-judge panel decisions on cases submitted in 2007. Mixture probability for chair model: $\lambda_1 = .45$. Posterior means alongside with 95% credible intervals. Parties are abbreviated as follows: Christian Democrats (CVP); Free Democratic Party (FDP); non-partisan (Indep); Social Democrats (SP); Swiss People's Party (SVP).