

ARTICLE

Inferring Individual Preferences from Group Decisions: Judicial Preference Variation and Aggregation on Collegial Courts

Dominik Hangartner^{1,2} , Benjamin E. Lauderdale³ and Judith Spirig^{2,3,4} 

¹Center of Comparative and International Studies, ETH Zurich, Zurich, Switzerland, ²Immigration Policy Lab, ETH Zurich, Zurich, Switzerland, ³Department of Political Science, University College London, London, UK and ⁴Department of Political Science, University of Zurich, Zurich, Switzerland

Corresponding author: Judith Spirig; Email: j.spirig@ucl.ac.uk

(Received 15 January 2024; revised 16 May 2025; accepted 20 May 2025)

Abstract

Extensive research on judicial politics has documented disparities in adjudication and biases in judging. Yet, lacking statistical methods to infer individual preferences from group decisions, existing studies have focused on courts publishing individual judges' opinions, leaving a gap in understanding collegial courts that report only collective and unanimous ('per curiam') panel decisions. We introduce a statistical methodology to identify the most fitting decision-theoretic models for such collective decisions, infer judges' individual preferences, and quantify the inconsistency in the courts' decisions. This methodology is applicable in various small group decision-making contexts where group assignments are repeated and exogenous. Applying it to the Swiss appellate court for asylum appeals, where decisions are made in three-judge panels, we find that in 45 per cent of cases, the chair-as-dictator rule applies (rather than majority rule). Although judges' preferences vary strongly with partisanship, the partially collective decision making of the panel moderates this heterogeneity.

Keywords: judicial behavior; collegial courts; group decision making; item-response models; asylum adjudication

Introduction

In the Fall of 2007, the Swiss Federal Administrative Court (FAC), the highest court reviewing asylum decisions of the Swiss State Secretariat for Migration (SEM), had to decide on two unrelated appeals of rejected asylum seekers from Guinea. Both appellants were male, claimed to have fled Guinea due to political persecution and instability in their home country, and failed to provide identity or travel documents. The SEM decided not to enter into the substance of both asylum cases, citing the absence of valid documents among other reasons. On 13 September, a panel of three judges, chaired by a judge affiliated with the center-right Free Democratic Party (FDP), upheld the SEM's rejection of the first asylum applicant, arguing that deportation to Guinea is legal despite the tense political situation in the country (FAC decision E-4524/2007). A little more than a month later, on 23 October, the same three judges, but now chaired by a judge from the centrist Christian Democratic People's Party (CVP), reached a markedly different decision for the second case: the panel admitted the appeal and remanded the case to the SEM, concluding that the SEM did not sufficiently consider obstacles to deportation given the political situation and violence in Guinea (FAC decision E-3579/2007). While no two cases are identical, a

look at these decisions reveals few obvious differences in the merits of the cases, and there is little to suggest that changes in Guinea's political situation within the six weeks between the rulings could account for the different outcomes. This example raises larger questions that are fundamental to the social sciences: How do groups translate individual preferences into collective decisions? How does the structure of the decision-making process allow certain decision makers to exert greater influence over the outcome?

In the field of judicial politics, answers to these questions are critical for understanding the behavior of judges and how judicial panels reach joint decisions. While extensive research has illuminated aspects of this process in contexts where individual judges' opinions are published, our grasp of the dynamics in collegial courts that render decisions without detailing individual votes or dissenting opinions remains limited. The development and application of a statistical methodology to bridge this gap is the focus of this study.

Scores of research have estimated judges' preferences, examined their correlation with political ideology and other factors, and documented how diversity in preferences leads to inconsistencies in adjudication (for a recent overview, see Harris & Sen 2019). This research spans a wide array of topics and courts, from challenges to the U.S. Environmental Protection Agency (Revesz 1997) to ethnic in-group bias in Israeli courts (Shayo and Zussman 2011), and disparities in asylum appeal decisions (Ramji-Nogales *et al.* 2007). However, what almost all existing studies have in common is a focus on courts that communicate individual votes or publish individual opinions, from which researchers can estimate judges' preferences (see, for example, Fischman 2011). While most courts (with multi-member decision-making bodies) in common law countries allow for the publication of individual opinions,¹ many courts and decision types lack records of individual voting. This is particularly true in the civil law system, where collegiality and secrecy in deliberations are emphasized, with decisions often made *per curiam*, that is, issued collectively and unanimously without dissenting opinions or indications of individual judges' votes (Ginsburg 1990). This practice is prevalent in numerous civil law countries, including Austria, Belgium, Bulgaria, France, Germany, Hungary, Italy, Latvia and the Netherlands, among others (for an overview, see Raffaelli 2012). In addition, some common law courts, like the U.S. Courts of Appeals and the Supreme Courts of California and Florida, occasionally issue unsigned *per curiam* decisions, including in death penalty appeals (Beim *et al.* 2021). Furthermore, decisions unrelated to the outcome itself, such as the choice to publish or withhold opinions, are sometimes made collectively by panels without disclosing the preferences of individual members in U.S. Courts of Appeals (Merritt and Brudney 2001).

Without the methodological tools to infer individual preferences from group decisions and to understand the processes transforming these preferences into collective rulings, our ability to analyze collegial courts and their decisions is severely limited. In this study, we begin to open this black box by developing a model that leverages the rotating and (quasi-)random assignment of judges to panels.² This model recovers judge-level preferences from panel decisions without dissenting opinions, along with the aggregation rules mapping preferences into collective outcomes. The selection of an aggregation rule has implications for the estimation of individual preferences. Since judges' preferences are estimated conditional on the aggregation rule, model fit depends both on the predictive performance of the aggregation rule and the preferences. For this reason, we test a variety of aggregation rules and estimate associated preferences with an item-response theory (IRT) model, compare their statistical fit, and focus on the one(s) with the best fit.

¹Among the courts in common law countries that do not allow the publication of individual opinions are Maltese courts and the Irish Constitutional Court (see Raffaelli 2012).

²As we discuss in more detail below, our model does not require strict random assignment of judges to panels. Instead, exogenous assignment, conditional on covariates, is sufficient. This weaker requirement broadens to applicability of the methodology advanced here, a point to which we return in the conclusion.

Our model is similar in spirit to the consensus voting model developed by Fischman (2011) to study judicial behavior on courts that publish individual votes and dissenting opinions (two features that are absent in our case). Fischman's model augments the standard sincere voting model with a cost of dissent term that pushes a judge to suppress her disagreement with the other judges on the panel as long as her disutility from voting against her preferred outcome is smaller than the cost of dissent.³ Closest to our approach is the model introduced by Malecki (2012), which identifies individual preferences on the European Court of Justice by assuming that the joint chamber decision is a deterministic function of the *mean* of the judges' preferences in the chamber. Malecki already explains in their paper that the mean rule is not theoretically founded but was adopted because the resulting model is easier to estimate than the median model.⁴ We build on Malecki's seminal work and are able to estimate, *inter alia*, the median model thanks to advances in MCMC sampling.⁵ Our approach also breaks new ground by learning from the data *which* model of group decision making best fits the panel decisions. Thus, our model does not (have to) assume the existence of a particular aggregation rule, but is able to let the data decide which rule, or combination of rules, is most likely to apply. This enables us to explore the empirical fit of various decision rules. We largely focus on those that are vaguely plausible and theoretically grounded in the court's decision-making procedure: 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 decision, is decisive. We also consider more complex aggregation rules that combine these simple decision rules.

We illustrate this methodology with panel decisions to reject or grant asylum appeals of the Swiss Federal Administrative Court (FAC). As a signatory to the 1951 Refugee Convention, Switzerland grants asylum seekers whose claims are rejected the right to appeal the initial decision. In 2007, the year of the FAC's inception and the focus of our study, about thirty asylum judges at the FAC centrally reviewed several thousand appeals, typically in panels of three. Most judges serving on the Swiss FAC have a publicly known party affiliation and all judges are voted into judicial office by the MPs of the upper and lower houses. Voluntary quotas aim to ensure that the body of judges reflects the relative seat share of the different parties in parliament (Kiener 2001; Raselli 2011). A potential drawback of nominating a partisan judiciary is the threat of a heavily politicized court. If judges see themselves primarily as representatives of their party, we might worry that their decisions are not simply derived from applying the 'law' but influenced by their partisanship. These divergent preferences might lead to differences in adjudication, the extent of which depends, among other things, on the procedures by which the three-judge panels aggregate individual preferences into collective decisions. For example, panel decision-making procedures that give disproportional influence to certain judges will typically lead to higher inconsistency compared to processes that aggregate judges' preferences equally. Leveraging the exogenous assignment of cases to judges, the methodology advanced here allows us to trace the relationship between heterogeneity in judges' preferences, preference aggregation mechanisms, and inconsistency in adjudication.

Analyzing 1,739 asylum appeals submitted to the FAC in 2007, 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

³Cost of dissent is a catch-all term for panel effects and in our case will also include the time and effort that the other panel members have to exert in case they propose to revise the decision drafted by the panel chair.

⁴To see why the mean model simplifies estimation, consider a fixed number of judges repeatedly and randomly assigned to three-judge panels. As the number of cases approaches infinity, the fraction of cases granted by the panels on which a given judge sits converges to that judge's preference (cutpoint) under standard case-space assumptions.

⁵Despite the conceptual difference, the mean rule is poised to fit similarly well as the median rule. When fitting the mean rule to our data, the log likelihood value is -708.2 , a tiny bit worse than the median rule, which scores at -707.2 , and significantly worse than the mixture model discussed further below.

that a mixture model of the chair rule and majority rule fits better than either of these simple aggregation rules, indicating that the chair judge can exert disproportionate influence which allows her to deviate from the median's preference in about 45 per cent of cases. Based on this mixture model, we find sizeable disparities in judges' preferred grant rates for comparable appeals. These disparities strongly correlate with partisanship in expected ways: on average, judges from the right-wing Swiss People's Party have a preferred grant rate of about 4.8 per cent. In contrast, the preferred grant rate of judges from the left-wing Social Democratic Party is 20.3 per cent, about four times higher. However, the disparities in preferences are moderated by the other judges on the panel (at least in cases resolved by majority rule), resulting in an inconsistency rate of at least 8.4 per cent. This represents the proportion of cases that would have been decided differently had a different panel been randomly assigned to the same case.

This study contributes to several areas of research. First, our findings have implications for the comparative literature on judicial behavior. Although many studies show large disparities in asylum adjudication between decision makers that potentially face different cases (Law 2004; Ramji-Nogales *et al.* 2007; Taylor 2007), our results provide additional causal evidence that the identity of judges matters when adjudicating cases with, in expectation, the same merit. Beyond asylum adjudication, we contribute to the literature on how judges' identity influences their decisions. Previous research has shown that characteristics such as gender (Peresie 2005; Boyd *et al.* 2010; Glynn and Sen 2015), ethnicity or race (Abrams *et al.* 2012; Gazal-Ayal and Sulitzeanu-Kenan 2010; Grossman *et al.* 2016; Shayo and Zussman 2011), and ideology (Ashenfelter *et al.* 1995; Epstein *et al.* 2013; Sunstein *et al.* 2006) all affect judicial behavior. We add evidence from a multiparty country in the civil law system that judges' political ideology, proxied by partisanship, strongly correlates with preferences over asylum appeal.

Second, our methodology and application add to the theoretical and empirical literature that studies collective decision making on collegial courts (for example, Bonneau *et al.* 2007; Sunstein *et al.* 2006), how panel members influence each other (Cross and Tiller 1998; Fischman 2013; Kastellec 2013; Revesz 1997), and how they aggregate preferences (Fischman 2011; Van Dijk *et al.* 2014) and varying interpretations of legal rules into joint decisions (Landa and Lax 2009). Previous studies that estimate the preferences of judges and their aggregation in collegial decisions with IRT models require information on individual votes (one exception is Malecki 2012 discussed above). The methodology we develop to infer individual preferences from collective decisions opens up the possibility to empirically study the behavior of judges on collegial courts – as well as other small group decision-making bodies – that feature exogenous panel assignment but were heretofore out of methodological reach because they do not publish individual opinions.

Third, our findings also enrich longstanding debates about how judges make decisions. Two central pillars of this debate are the attitudinal model, which posits that judges' decisions reflect their ideological attitudes and values, and the strategic model, which suggests that judges behave strategically to achieve their goals (for an overview, see Epstein and Weinshall 2021; Epstein *et al.* 2022). Our findings suggest that both models contribute to explaining judicial behavior at the FAC. In line with the attitudinal model, we observe significant discrepancies in preferences that are strongly correlated with partisanship. However, if all judges were to always vote according to their attitudes, their panel position (chair, second, third) would be irrelevant, and all cases would be resolved by majority rule. Instead, our results demonstrate that judges sometimes defer to the chair, even when it conflicts with their attitudes.

The rest of this paper is organized as follows. The next section provides background information on decision-making processes and case assignment at the FAC. Section 'Methodology' introduces a case-space model with various decision-theoretic aggregation rules that transform individual preferences into group decisions. We translate this formal model into a likelihood estimator to identify which aggregation rule best fits the panel decisions, infer judges' individual preferences, and estimate the inconsistency rate. Section 'Data: Sample, Outcome Measure and Covariates' introduces the sample, outcome measures, and covariates required to

render case assignment conditionally exogenous. Section ‘Results’ provides the results. The concluding section points to avenues for future research.

Judicial Decision Making at the FAC

Election of Judges by the Swiss Parliament

Judges of the FAC are nominated and voted into office by the joint meeting of the upper and lower house of the Swiss parliament (a contrast to countries like the United States where the executive branch nominates candidates for judicial office). A term lasts for six years, but there is no limit on the number of terms that judges are allowed to serve. Aside from a few candidates who run as independents, most had a known party affiliation and were backed by their party when running for office. While not written law, there is an informal rule that the body of elected judges should reflect the relative seat share of the parties in parliament. The underlying principle motivating this rule is to select a judiciary that is representative of the people it serves (Kiener 2001; Raselli 2011).

Asylum Appeal Procedure and the Structure of Panel Decisions

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 9,577 asylum applications that received a first instance decision in 2007, the SEM granted 21 per cent (BFM 2008). Rejected asylum seekers have the option to appeal the SEM decision. Since its inception in 2007, the two asylum divisions of the FAC handle all such appeals. Because the asylum appeal decisions of the FAC are, in general, not further appealable to the Swiss Federal Supreme Court, the FAC is effectively the court of last resort in the Swiss asylum process (Spirig 2023).

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 the two asylum divisions. A bespoke software program, internally referred to as the ‘Bandlimat’, 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 legal urgency of the appeal, (iii) judges’ mother tongues and (iv) their 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.⁶ 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).

If not accounted for, we would be concerned that the language of the case would render the assignment non-random. This concern arises because the language used in the first instance asylum decision, which also determines the language of the asylum appeal procedure and decision, is typically based on the language of the Swiss region (French, German and Italian) where the asylum seeker initially applies for asylum. The choice of this region might, in turn, be influenced by the asylum seeker’s country of origin. For instance, French-speaking asylum seekers from Côte d’Ivoire (with typically low asylum grant rates) might be more likely to submit their asylum application in French-speaking Switzerland rather than the German-speaking part (Martén et al. 2019). This

⁶A recent study covering the years 2008–18 raises concerns that division heads manually modified the initial, automatic assignment for a significant number of cases (Büchel et al. 2022). The primary reason stated for deviations from automatic assignment is to allocate judges to cases in their mother tongue, a factor our analysis takes into account; see below. Furthermore, for a limited number of cases (129 judge-to-panel assignments, or 2.5 per cent of the $3 \times 1,739$ judge-to-panel assignments in our estimation sample) we observe in our data for 2007 that the judge serving on the panel deviates from the initially assigned judge. For those cases where this information is available, our analyses use the initially assigned judge. This alleviates concerns about endogeneity and implies that the resulting intention-to-treat estimates provide a lower bound for the true heterogeneity in judges’ preferences.

choice influences the language used in both the initial asylum decision and the subsequent appeal. Consequently, judges operating in different Swiss languages may encounter cases from specific origin countries and therefore of differing strength. To address this, our analysis controls for the appellants' country of origin, under the assumption that, conditional on this factor, the assignment of cases by the 'Bandlimat' is as good as random. We confirmed the operational details of the 'Bandlimat' with one of its lead programmers and use a series of placebo checks, discussed below, to validate the assumption of conditionally exogenous assignment.

The FAC's asylum divisions almost always make decisions through file reviews, with hearings being an extremely rare occurrence. In 2007, 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 necessary, instructs one of her clerks to draft a decision and forwards all materials including the draft decision 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 chair amends the draft, further circulates it and possibly revises the decision. If the three judges are not able to reach a consensus, the outcome is decided by majority vote.

A partial revision of the Swiss Asylum Law introduced an alternative 'simplified procedure' per 1 January 2008, which provides judges with the power to decide on cases in panels of two.⁷ Since the decision to invoke the simplified procedure is also a function of the assigned judge, appeals decided by the simplified procedure are a selective subset of all cases (Spirig 2024). To circumvent these selection issues, our analysis focuses on the substantively tried appeals that were submitted in 2007 – before the introduction of the simplified procedure – and handled by three-judge panels. For these cases, we test which decision-making rules fit the data best, and estimate judges' preferences and the court's inconsistency rate.

Methodology

Inferring Individual Preferences from Panel Decisions in the Case-Space Model

To estimate individual preferences from the observed aggregate decisions of three-judge panels, we need a framework for modeling 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 is often invoked in the empirical study of judicial politics, there is historical (Greenhouse 2005; Jeffries 2001) and statistical (Lauderdale and Clark 2012) evidence that judges' preferences can vary across areas of the law. Reducing those preferences to a single left–right dimension can lead to biased inferences and implies that the estimated variance in judges' preferences is a lower bound for the

⁷The simplified procedure allows the chair judge to classify certain cases as either 'clearly with or 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 decision. 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.

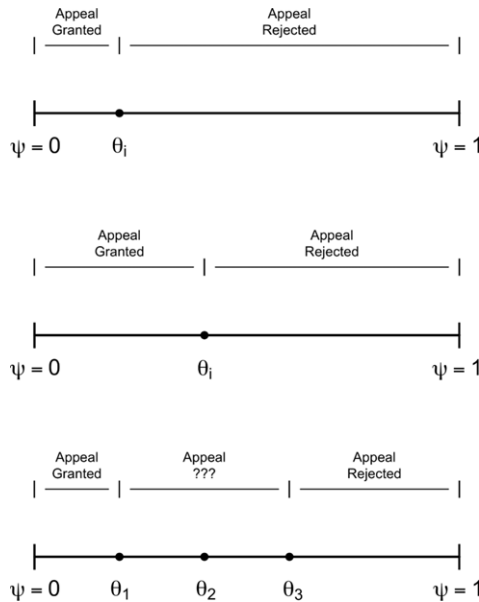


Figure 1. Mapping between preferences and decisions.

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 resolves the panel's decision making.

true variance if judges disagree not only on the threshold but also on the relative merits. In the present context, where all cases concern asylum appeals, the unidimensionality assumption seems likely to hold at least approximately. However, as we show below, our framework also allows us to examine how relaxing the unidimensionality assumption – by considering the possibility of no consensus on the relative merits – implies higher inconsistency rates between differently composed panels.

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 collectively by three judges according to the procedures described in the preceding section, and so the resolution of cases in which the three judges disagree depends on the aggregation rule that combines their preferences into a decision (bottom axis of Figure 1).

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.⁸ This allows us to define a generic likelihood function for the observable decision of the panel to grant ($y_j = 1$) or reject ($y_j = 0$) the appeal:

$$\mathcal{L}(\theta) = \prod_j \pi_j^{y_j} (1 - \pi_j)^{1-y_j}, \quad \text{with } \pi_j = p(\psi_j < f(\theta_{i(j)})). \quad (1)$$

⁸While 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 are possible (see Landa and Lax 2009 for a formal treatment), such complexities are beyond the scope of our statistical model.

In the next section, we reparametrize $f(\theta_{i(j)})$ as a function of various aggregation rules. In addition to estimating the model with judge-specific preferences, we also fit a model for which we replace θ_i in Equation (1) with a single cutpoint for each party. This model is essentially equivalent to a weighted-average of the preferences of each judge aggregated by party, but provides party-level estimates of uncertainty and facilitates comparisons across parties.

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 a group's collective decision, not 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} , that is, the most restrictive judge, determines the outcome. If instead unanimity is required to reject an appeal, θ_{max} , that is, the most lenient judge, determines the outcome. Lastly, if the chair's preference dictates the outcome, we would expect θ_1 to be decisive.⁹

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 as well as a mixture model of the chair and majority rule, which we discuss next.

We also consider mixture models of these simple aggregation rules. One reason for doing so is that a mixture of the chair and majority rule 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 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 (see Clark *et al.* 2018 for a similar argument on the effects of leisure incentives on judicial performance).

One way of thinking about the incentives created 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 if they do not pay the cost of review. While we do not explicitly formulate a game-theoretic model here, we note that the mixture model approximates the logic of a mixed-strategy equilibrium¹⁰ in an oversight game, while the chair and majority rule models

⁹Note that for three-judge panels, we cannot point identify, but only bound, the preferences of the 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.

¹⁰The existence of such mixed-strategy equilibria depends on assumptions about the utility functions of the judges and the cost of review.

Table 1. Overview of decision-theoretic rules

Notation	Aggregation rule
θ_{min}	Unanimity rule to grant: most restrictive judge decides
θ_{max}	Unanimity rule to reject: most lenient judge decides
θ_{med}	Majority rule: 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
λ_1	Mixture model of chair and majority rule

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

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. As noted above, the resulting mixture model is similar in spirit to Fischman’s (2011) consensus voting model that features a ‘cost of dissent’ disutility for a judge whose vote differs from the other judges on the panel. Indeed, if the cost of dissent goes to zero, both Fischman’s and our model would simplify to majority rule.¹¹ The argument concerning the chair’s partial dictator power also resonates with the model by Lax and Cameron (2007). In their model, the opinion writer can steer decisions away from the median justice because the bargaining protocol used by the U.S. Supreme Court confers a certain degree of monopoly power on the opinion writer.¹²

To translate the essence of this idea into a statistical model, we estimate a mixture model of the chair and majority rules. Let λ_1 be the probability the chair rule governs the case and $\lambda_{med} = 1 - \lambda_1$ for majority (median) rule. The corresponding mixture of aggregation functions

$$\pi_j = \lambda_1 p(\psi_j < \theta_{1(j)}) + (1 - \lambda_1) p(\psi_j < \theta_{med(j)}) \quad (2)$$

is then plugged into the likelihood function (Equation 1). We also tested other mixture models, for example between the median and second judge, but all of these fitted, as expected, significantly worse.

Measuring the Consistency of Decision Making

Given the case-space model, we can also calculate the extent to which (quasi-random) assignment of judges leads to inconsistency in the decisions that the court makes. Here, we apply the logic of test–retest reliability, as advocated by Fischman (2014). A consistent legal process is one which always yields the same outcome for a given set of case facts, regardless of which judges hear the case. Thus, as a measure of inconsistency, we are interested in what proportion of the observed cases would have been decided differently, had they been assigned to a different panel of judges. To calculate this, for each case in our data, we calculate $E[\pi|X_j]$, the expected grant rate conditional on X_j , the country of origin, for all observed panels in the dataset. For each of these, we calculate the probability that this hypothetical alternative panel would have decided differently

¹¹If the cost of dissent goes to infinity, all decisions in Fischman’s model would be unanimous and our model would simplify to chair rule.

¹²We thank an anonymous reviewer for pointing out these connections.

than the observed panel did, given the case-space assumptions described above. We then take the average of these probabilities for all alternative panels for each case and then across all cases.

This measure of inconsistency is a lower bound, as it relies on the assumption that judges only disagree about the threshold for making a decision, but not the relative merits of cases. If judges also disagree about relative merits, the true inconsistency of the court will be higher than what our estimate suggests. The nature of our data is such that we cannot empirically assess whether judges disagree about the relative merits of appeals, because we never observe information about the individual judges' preferences as distinct from the panel decision of a case.¹³ We can, however, estimate alternative models under independence assumptions, which assumes that judges' views about the relative merits of cases have no correlation (conditional on the observed country of origin). We present these estimates as an upper bound on the inconsistency rate.

For the reason mentioned above, we cannot distinguish between the case-space assumption of perfect correlation in judges' relative merits judgments and the independence assumption of zero correlation empirically with our data (when estimated, the log likelihoods are the same under either assumption). The truth is likely to lie somewhere in between, although we suspect closer to the case-space assumption since we expect judges to agree on at least some of the merits. An important avenue for future research is to use panel judgment models like the ones we have considered here to estimate the extent to which judges agree on the relative merits of cases, in empirical contexts where the observable data are capable of revealing such information.

Data: Sample, Outcome Measure and Covariates

We obtained the data that form the basis of our analysis directly from the FAC. The key dependent variable measures the outcome of the case (the equivalent of the disposition in common law). 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 decision 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). See SM Table S1 for more details on the coding.

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.¹⁴ 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.¹⁵ Finally, for the subset of 1,519 published decisions, we are able to add information on whether or not the appellant had legal representation. 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 judge with a unique id and an indicator of their party affiliation.¹⁶

¹³If we knew which cases were decided 3–0 and which were 2–1, even without knowing which judges voted in the majority in the latter cases, we would be able to estimate the degree of correlation across judges in assessments of the relevant merits, but our data lack even this information.

¹⁴For published decisions, the large majority of cases, this information is also available on the court's online database (<https://www.bvger.ch/en/jurisprudence/judgments-database>).

¹⁵See <https://www.bvger.ch/en/about-fac/employees?category=Judges> for short bios of sitting judges.

¹⁶We, the authors, as well as the FAC, are well aware that this anonymization is incomplete at best, and that it is possible to figure out the identity of the judges using publicly available information. Nevertheless, we do believe that reporting anonymized results is helpful in focusing the discussion of our findings on structural issues of the court, rather than on the behavior of individual judges.

Overall, the dataset contains the universe of all 3,919 unique decisions of asylum appeals submitted in 2007, made by a total of 36 asylum judges between 2007 and 2013.¹⁷ We exclude a total of 2,180 cases: we drop 673 cases because they were ‘written off’, which means that they did not receive a regular decision. We remove 48 cases that received decisions that cannot be considered as clearly in favor of or against the appellant. Furthermore, we exclude all cases that were not decided by three-judge panels (11 are decided by five-judge panels, 942 by a single judge, and 145 under the simplified procedure). The reason for excluding cases that were decided by fewer than three judges is that for most of these cases we do not know who the (initially assigned) second and third judges were.¹⁸ Finally, we drop cases that ended up being handled by judges who were not yet at the court in the year when the case was submitted (361). All in all, these restrictions leave us with a sample of 1,739 cases, which represents 44.4 per cent of all appeals submitted to the FAC in 2007.¹⁹ In the Supplementary Material, we provide detailed descriptive statistics for the appeals, of which, employing our binary outcome measure, 19.4 per cent are granted and 80.6 per cent rejected. Table S2 shows grant rates grouped by panel characteristics, Table S3 the breakdown of cases by legal category, Table S4 the caseload by judge, and Table S5 the proportion of cases by origin country of the appellant.

Results

Verifying Exogenous Case Assignment

Before we proceed with estimating judge-level preferences and their aggregation, we have to substantiate the assumption that the case assignment of ‘Bandlimat’ can be considered exogenous once we condition on the appellant’s origin country. To assess this, we conduct a series of placebo tests. In a first step, we leverage the available case characteristics to gauge their predictive power of the appeal outcome. This test is similar in spirit to a manipulation check following a randomized experiment (see Frandsen et al. 2023). To account for relevant predictors, we rely on the coding of the FAC that classifies cases by legal issue right at the opening of the proceedings. In addition, we use our own coding of all published decisions (87 per cent of all decisions) 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 (majority and chair rule; see 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 does not predict the *general* leniency of the chair or median judge on the panel if the assumption of quasi-random assignment conditional on origin country holds.

The linear regression models in Table 2 confirm that this is indeed the case. Model 1 shows that the case characteristics are highly predictive of the outcome of the appeal, as expected. For example, being represented by a lawyer or paralegal is associated with an eleven percentage point higher probability of winning the appeal. When predicting 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

¹⁷If several appeals were unified and received a joint decision, we recorded it as one observation (this concerns 34 decisions). Note that cases that are directly related to previously filed appeals are allocated to the judges that handled that appeal.

¹⁸Note that dropping these cases will, if anything, lead us to underestimate the differences in judges’ ideal points because, in these cases, the chair judge’s influence is likely to be more pronounced.

¹⁹For the balance tests reported below, we focus on the subset of 1,519 decisions that were published, which allows us to code additional information about whether the appellant was represented by a lawyer or paralegal.

Table 2. Manipulation check and placebo tests

	Appeal granted	Leniency chair	Leniency median
	(1)	(2)	(3)
Allocation to canton	0.013 (0.168)	0.062 (0.050)	0.012 (0.024)
Asylum and return (RR)	0.134 (0.051)	0.030 (0.015)	0.017 (0.007)
Asylum procedure (other)	0.299 (0.053)	0.012 (0.016)	0.010 (0.008)
Asylum request abroad	0.049 (0.052)	0.015 (0.016)	0.005 (0.007)
Enforcement of return	0.284 (0.187)	0.021 (0.056)	0.025 (0.027)
Family reunification	0.053 (0.095)	0.020 (0.028)	0.002 (0.014)
Inadmissibility of request	0.073 (0.025)	0.005 (0.008)	−0.002 (0.004)
Refusal temporary protection	0.856 (0.393)	0.210 (0.118)	−0.028 (0.056)
Return/ enforcement (RR)	0.340 (0.263)	0.088 (0.079)	0.036 (0.038)
Revocation of asylum	0.140 (0.097)	0.033 (0.029)	0.014 (0.014)
Revocation temporary protection	0.156 (0.051)	0.036 (0.015)	0.015 (0.007)
Lawyer/paralegal	0.111 (0.022)	−0.002 (0.007)	0.001 (0.003)
Country FE (# 83)	✓	✓	✓
Observations	1,519	1,519	1,519
R ²	0.235	0.126	0.101
Joint <i>F</i> -test	0.000	0.211	0.218

Note: table shows ordinary least squares regressions of the binary appeal outcome ($= 1$ if the appeal is granted) on case characteristics (legal category and legal representation) in Model 1; of the average grant rate of the chair judge across all decisions except case j on case characteristics in Model 2; and of the grant rate of the median judge on the panel across all decisions except case j on case characteristics in Model 3. All models control for country of origin fixed effects (FE). The baseline legal 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 included case strength characteristics are not predictive of the outcome. The sample consists of all 1,519 published three-judge panel decisions (granted/rejected) on cases submitted in 2007.

significance. The two *p*-values from the joint *F*-tests are between 0.21 and 0.22, well above conventional level of significance. This suggests that, conditional on country of origin, cases are indeed exogenously assigned to panels of judges for the study sample. This allows us to infer the aggregation rules from quasi-randomly varying panel compositions and to causally interpret differences in judges’ preferences, which we turn to estimating in the next sections.

Which Aggregation Rule Fits Panel Decisions Best?

In order to understand how individual preferences are aggregated into a collective 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 (majority rule), 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. In addition to the eighty-three origin countries’ fixed effects, all of these models use 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 Criterion (DIC), a Bayesian generalization of the AIC, to assess model fit.

Table 3 shows the results. We first discuss the simple, that is, non-mixture, aggregation rules. The best-fitting simple aggregation rule is the chair judge, closely followed by the majority rule model.

Table 3. Fit statistics for preference aggregation rules

Model	DIC	LL	Parameters
Mixture chair and median	1427.2	-703.6	118
Chair judge	1436.7	-708.8	117
Median judge	1437.1	-707.2	117
Most lenient judge	1450.5	-714.8	117
Most restrictive judge	1457.0	-718.4	117
Second judge	1480.9	-730.7	117
Third judge	1487.4	-736.5	117
Null model	1488.9	-738.0	90

Note: table of fit statistics for Bayesian estimates of judges' preferences under the mixture, chair, median (majority rule), max (most lenient judge), min (most restrictive judge), 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 (LL) statistic. Column 3 shows the number of estimated parameters. The sample consists of 1,739 three-judge panel decisions (granted/rejected) on cases submitted in 2007.

The difference in the DIC between these two non-nested models with the same number of parameters is 0.4, indicating that the chair model fits the data only marginally better than the majority rule model. All of the other simple models fit considerably worse than these two according to the DIC. That 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 majority rule 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 the second and third judge between paying the cost for review and letting the chair decide.²⁰ The comparison between the mixture and the chair or majority rule 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 majority rule 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 majority rule models. The estimated mixing parameter λ_1 is 0.45 (95 per cent 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 ten points lower (better) for the mixture model, and that λ_1 is neither close to zero nor close to one, 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 decision.

Our framework also allows us to develop and test hypotheses about the conditions under which we expect the chair judge to have more (or less) dictatorial power to steer decisions away from the median justice. One expectation is that the second and the third judges are more likely to defer to the chair even when the draft decision is not in line with their preferences if they perceive the chair as particularly competent and experienced (see, for example, Eisenberg et al. 2013). To explore this expectation, we re-parametrize Equation 2 to allow the mixture parameter to vary by seniority and test whether more senior chairs have a higher λ_1 value – indicating that cases are more frequently

²⁰Our mixture model mixes relatively slowly through the posterior space and requires fairly long runs. In the models presented here, we use 1,000 burn-in iterations followed by 4,000 draws, which was sufficient to achieve ergodicity of the Markov chains.

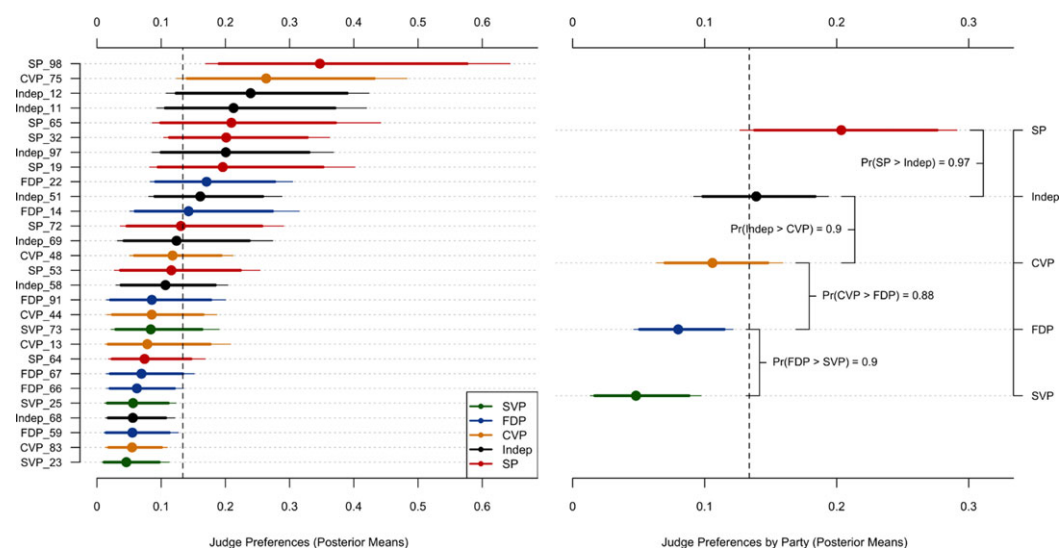


Figure 2. Heterogeneity in judges' preferences.

Note: left panel: estimated preferences of judges (posterior means) from the best-fitting mixture model that controls for origin country and uses hierarchical priors on judges. Right panel: same model but with party- rather than judge-specific estimates. Comparisons between neighboring parties show the probability that the party above is more lenient than the party below. The sample consists of 1,739 three-judge panel decisions (granted/rejected) on cases submitted in 2007. Both panels show posterior means along with 90 per cent (bold lines) and 95 per cent (thin lines) credible intervals. Mixture probability for chair model: $\lambda_1 = 0.45$. Black dashed line indicates the average judge preference of 13.4 per cent (country of origin effect set to 'Sri Lanka'). Party acronyms: Christian Democrats (CVP; 351 cases); Free Democratic Party (FDP; 408 cases); non-partisan (Indep; 487 cases); Social Democrats (SP; 334 cases); Swiss People's Party (SVP; 159 cases).

determined by the chair – than more junior judges. To code seniority, we median split the judge sample by the number of years they served on the FAC and the 'Schweizerische Asylrekurskommission', the predecessor of the FAC's asylum divisions. In the more senior group, judges have an average of 14.8 years of service, compared to 4.8 years in the more junior group. In line with our expectation, we find that the mixture parameter for the chair's power is 0.66 (95 per cent CI: 0.24–0.97) for senior compared to 0.31 (95 per cent CI: 0.05–0.65) for junior chairs. This implies that, with senior chairs, the chair-as-dictator rule applies in about two out of three decisions (rather than majority rule), and that this rate drops to less than one out of three for junior chairs. While the difference in the mixture parameters is sizeable, the uncertainty around these parameters is also considerable. Consistent with this pattern, a one-sided Bayesian hypothesis test shows a 92 per cent probability that the mixture parameter for the chair's power is higher for senior judges than for junior judges.

Heterogeneity in Judges' Preferences and Inconsistency in Decision Making

Having identified the best-fitting aggregation rule with the mixture model, we can explore the heterogeneity in preferences between judges, and how this impacts the consistency of decision making at the court.

Figure 2, left panel, shows the estimated preferences of the judges from the mixture model that controls for origin country and uses hierarchical priors on the judges. To generate the figure, the country of origin fixed effects are set to 'Sri Lanka'. SM Table S6 shows the underlying numerical estimates. Figure 2, right panel, shows the same model but with party- rather than judge-specific preferences. To facilitate comparisons across parties, we conduct one-sided Bayesian hypothesis

tests for all sets of neighboring parties to estimate the probability that the cutpoint of the pair's more lenient party is higher.

Figure 2 reveals several striking features. First, the left panel shows that there is substantial heterogeneity in the preferences of judges. For example, the most lenient judge (from the Social Democrats (SP), preferred grant rate = 34.7 per cent) would, if they could decide alone, achieve a grant rate that is more than seven times higher than the rate of the most restrictive judge (from the Swiss People's Party (SVP), preferred grant rate = 4.6 per cent). This difference is legally meaningful and statistically significant: a one-sided Bayesian hypothesis test shows that the probability that the cutpoint for judge SP_98 is higher than the cutpoint of judge SVP_23, that is, $\Pr(\theta_{SP98} > \theta_{SVP23})$, is 1. This preference variation is not limited to the two most extreme judges. For example, the difference in preferred grant rates for the tenth most lenient judge (Indep_51, 16.1 per cent) and the tenth most restrictive judge (SVP_73, 8.4 per cent) is sizeable and the probability that the cutpoint of the more lenient judge is higher than that of the more restrictive judge is still 91 per cent.²¹ Second, as the left panel of Figure 2 shows, this heterogeneity is driven by variance both within and across parties. Third, the two panels show that despite relevant within-party variation, there is a clear association between judges' preferences and their partisanship in the expected direction. Judges affiliated with the leftist SP have, on average, a preferred grant rate of 20.3 per cent. Non-partisan judges (preferred grant rate 13.9 per cent) and those affiliated with the centrist Christian Democrats (CVP, preferred grant rate 10.6 per cent) exhibit the most intraparty variance. Judges affiliated with the center-right Free Democratic Party (FDP, preferred grant rate 8 per cent) are the second most restrictive, while judges affiliated with the hard-right SVP (preferred grant rate 4.8 per cent) are the least favorable towards asylum seekers. Our Bayesian hypothesis tests displayed in the right panel of Figure 2 show that even among neighboring parties, the probability that the judges from the slightly more lenient party have a higher cutpoint ranges between 88 per cent and 97 per cent (the probability always increases to (almost) 100 per cent when we compare the cutpoints of more distant parties such as the leftist SP and the centrist CVP; see SM Table S8 for detailed results on all interparty hypothesis tests). In sum, we find statistically significant and legally meaningful differences in how judges from different parties prefer to adjudicate similar appeals.

How robust are these findings? Figures S1a and S1b in the Supplementary Material show the corresponding preference estimates from the (non-mixture) Bayesian majority rule and chair models, 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 partisanship, are also clearly evident in those simpler models. In sum, we find across a variety of models that the political ideology of judges is a robust predictor of their preferred grant rate.

How do the preferences of judges compare to the voting behavior of MPs from the same party on asylum issues? It is important to note that comparing the opinions of judges and MPs involves assessing them across different contexts. Judges express their opinions within a case-space that typically limits their ability to consider factors beyond the specifics of the case (with important exceptions, see Spirig 2023). In contrast, MPs vote within a broader policy-space that usually allows them greater flexibility to incorporate other issues, such as their general attitude towards immigration, into their decisions. SM Figure S3 shows the estimates from a Bayesian IRT model based on roll-call data 2007–2015 of MPs in the Swiss lower house, who are responsible for

²¹Note that the wide credible intervals for the preference estimates from our mixture model provide a somewhat different visual impression of statistical precision than the Bayesian hypothesis tests directly comparing preferences estimates. The reason for this apparent discrepancy is that within the same MCMC iteration, cutpoint estimates for different judges are often positively correlated. This implies that when comparing estimates for different judges from the same MCMC iteration, the cutpoints of more lenient judges frequently dominate the cutpoints of more conservative ones, which translates to Bayesian hypothesis tests that are close to 1. Also note that Bayesian estimates are, unlike those based on other estimation methods such as ordinary least squares, not necessarily normally distributed, which can result in asymmetric credible intervals.

electing the judiciary (together with the upper house). Despite the aforementioned differences between the case-space and policy-space, we find that the ordering of the preferences of the asylum judges on the FAC, grouped by party, is perfectly consistent with the general stance of the party's MPs on asylum policies. Consequently, a simple randomization test rejects the null hypothesis of no correlation between the voting behavior of MPs and the preferences of judges, averaged by party, for the four parties that nominated judges in 2007 with the minimal achievable p -value of $\frac{1}{4!} = 0.04$.

What does this substantial heterogeneity in judges' preferences 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. We hold country of origin fixed and use 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. In this case, the inconsistency rate would be zero. SM Figure S2 shows that this is not the case. We find that the predicted grant probabilities vary with the different judges serving on panels, leading to a 'test-retest' inconsistency rate of 8.4 per cent (95 per cent CI: 6.6–10.2 per cent). This indicates that more than one in twelve cases is decided differently than how it would be if an alternative panel had been drawn from those observed in the data.²²

As noted earlier, these estimates rely critically on the case-space assumption that judges only disagree about the threshold for granting appeals, not the relative merits of cases. When we instead adopt an independence assumption – where judges' views on the relative merits of cases are uncorrelated (conditional on the appellants' country of origin) – we obtain a significantly higher inconsistency estimate: 26.9 per cent (95 per cent CI: 25.9–28.0 per cent) under the mixture model. While we cannot empirically distinguish between the case-space and independence assumptions, we believe the true inconsistency rate is closer to the lower bound provided by the case-space assumption, as judges are likely to share some agreement on case merits.

While there is some uncertainty about the true inconsistency rate, it appears relatively modest in light of the substantial heterogeneity in judges' preferences. An important reason for this is the moderating effect of the three-judge panel – at least for those cases where majority rule prevails. Indeed, if we repeat the above 'test-retest' calculation using the same, mixture model based, preference estimates but counterfactually assume that the chair is able to dictate the outcome in all cases, the inconsistency rate would rise from 8.4 per cent to 10.3 per cent (95 per cent CI: 7.8–12.9 per cent), an increase of 22.4 per cent. In other words, even the current procedure, which confers the chair judge disproportionate powers, considerably reduces the inconsistency rate compared to a court where the chair – or single judges – can dictate all outcomes.

Discussion and Conclusion

Scores of research have documented how the background and ideology of judges shape their decisions across a wide range of legal areas (see Harris & Sen 2019 for an overview). The vast majority of these studies have focused on common law courts, where researchers can readily apply IRT models to estimate individual preferences from decisions that disclose the opinion of every judge on the panel. We introduce a flexible methodology that allows students of judicial behavior to extend this literature to courts rendering collegial decisions without information on dissenting opinions. Our methodology opens the black box of such *per curiam* decisions by testing a variety of decision-theoretic models of preference aggregation and inferring judge-level preferences from repeated group decisions with varying members.

We deploy this methodology to illuminate panel decisions of the FAC. Analyzing the universe of asylum appeals submitted in 2007, we find that a mixture of the chair and majority rule models

²²The estimated inconsistency rates are virtually identical if estimated under the chair (7.8 per cent; 95 per cent CI 5.8–9.6 per cent) and majority rule models (7.7 per cent; 95 per cent CI 5.7–9.4 per cent).

best fits the panel decisions. We find that slightly more than half the cases are decided by majority rule, while for the other cases the second and third judges on the panel prefer to defer to the chair rather than paying the cost for review. This finding – that the structure of the procedure grants disproportionate power to the judge drafting the decision – may help explain the contrasting outcomes of the two asylum cases from Guinea discussed in the introduction, where the same panel of three judges reached markedly different conclusions within weeks of each other, depending on who chaired the panel. Similar dynamics may also occur at other courts. For instance, opinion assignment at the U.S. Supreme Court provides the author with leverage to shift an opinion away from the median (Lax and Cameron 2007). It seems plausible that, as at the FAC, more senior Supreme Court justices are better positioned to exploit this leverage.

Conditional on this best-fitting mixture model, we discover substantial variation in judges' preferences over appeal decisions and a strong correlation with their partisanship. Since cases are, conditional on appellants' origin country, exogenously assigned to judges, the differences between judges' preferences cannot be explained by differences in case merits. The heterogeneity in preferences has implications for the consistency with which the court applies the law. The lower bound on the estimated inconsistency rate of 8.4 per cent suggests that more than one in twelve cases would have been decided differently if an alternative panel had been assigned to the case (this rate increases to one in five if we replace the case-space with the independence assumption). Although any inconsistency violates Aristotle's fundamental legal principle that 'like cases be treated alike', it is important to note that this rate compares favorably to other courts. What explains this relatively low inconsistency rate? Three factors stand out. First, the FAC's asylum divisions are adjudicating a high number of relatively routine cases, where we would expect smaller differences in judging compared to courts that focus on different areas of the law and frequently face non-routine cases (see, for example, Posner 2010). Second, FAC judges exhibit less preference variation than comparable asylum courts, for example in Canada or the United States (Ramji-Nogales et al. 2007; Rehaag 2007; Fischman 2011), which mechanically reduces inconsistency.²³ The third reason is rooted in the court's collective decision making: unlike their colleagues at single-judge courts, where differences in preferences directly translate into differences in decisions, asylum judges in Switzerland decide as a panel. Our analysis shows that even though majority rule is only used in slightly more than half of the cases, this reduces the differences in adjudication by close to 20 per cent compared to a counterfactual in which the chair (or a single judge) would decide alone.

We hope that the methodology introduced here facilitates more research on collegial courts to better understand preference aggregation and variation in other countries and other areas of the law. Courts that abstain from disclosing individual opinions (for all or some decisions) include collegial courts in Austria, France, Germany (North Rhine-Westphalia), and Italy (Engst et al. 2017; Raffaelli, 2012), US appellate courts (for example, Beim et al. 2021), and the European Court of Justice (Carrubba et al. 2008; Malecki 2012). To apply our method, two conditions need to be met: the decisions have to be made by a subset of judges in panels or chambers with rotating membership, and feature exogenous assignment of cases and judges to panels.²⁴ Note that our methodology can be applied to courts where case assignment is truly random, exogenous conditional on observable factors (like in the FAC), or partially exogenous, for example when assignment is based on case load balancing or defendant's last name. To exploit the latter

²³Comparing judges' preferred grant rates across studies is not straightforward. But we can create benchmarks by comparing the most lenient to the most restrictive judge. For example, Ramji-Nogales et al. (2007) document that in four of the largest US asylum courts, the most liberal judge has a grant rate that is between ten and eighteen times higher than the most restrictive judge (controlling for court and nationality of the asylum seeker). In contrast, we estimate that the preferred grant rate of the FAC's most lenient judge is about seven times that of the most restrictive judge.

²⁴For details on the panel and case assignment for the courts mentioned above, see Ecker et al. (2020) and Fabri and Langbroek (2007).

assignment mechanism, our framework can be adapted to incorporate instrumental variable strategies (Fabri and Langbroek 2007).

We expect that our methodology can also be fruitfully applied in a variety of other contexts of repeated small-group interactions, where collective decisions or performance indicators without information on individual votes or contributions are the norm. Examples of potential applications include 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 membership. By interacting the mixture parameter of our model with various dimensions of decision makers' identities (examples other than seniority include party affiliation, gender, or race), our model can be easily adapted to identify which individuals exert disproportionate influence on group decisions or performance. We hope that sharing the code that implements the estimators proposed in this study (available on the Harvard Dataverse Network, at: <https://doi.org/10.7910/DVN/X22XKN>) facilitates the adoption of this methodology in political science and neighboring disciplines.

This research has implications for the Swiss FAC. Our background research and meetings with members of the court revealed that the court is concerned about allegations of disparities in adjudication. While our analysis shows that the heterogeneity in judges' preferred grant rates is substantial, the three-judge panels moderate this heterogeneity substantially. If the goal of panel judgment is to provide some limited additional oversight to reduce inconsistencies that would arise under single-judge decisions, without large additional workload for judges, we find evidence that the observed system is potentially meeting that goal. If, however, the goal is instead to further decrease the inconsistency while keeping the partisan selection procedure, the FAC could look at the design of the decision-making process. The findings of this study suggest two options. First, given the moderating effect of majority rule, it is worth considering interventions to discourage judges from deferring to the chair in case of dissent. A second, more extreme version is to require all three judges to simultaneously review the appeal and independently draft a decision (Kornhauser and Sager 1986; Lorenz et al. 2011), ideally without information about the identities of the other judges on the panel – this might further minimize differences in adjudication. Both options promise to better unleash the power of Condorcet's jury theorem, but would also entail significantly more work. Without additional resources and judges, this risks creating a significant backlog of cases – a trade-off that must be carefully considered, especially given the already lengthy processing time.

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.1017/S0007123425100574>.

Data availability statement. Replication data for this article can be found in Harvard Dataverse at: <https://doi.org/10.7910/DVN/X22XKN>.

Acknowledgements. We thank the Swiss Federal Administrative Court for sharing data. We thank Paul Nulty for his support with the natural language processing of the decisions and Ethan Koch for excellent research assistance. We are grateful to all academic colleagues, judges, and legal experts who have provided input, feedback, and background information.

Financial support. J.S. was supported by a grant from the Swiss National Science Foundation through their Doc.CH funding scheme (Grant No. 159161).

Competing interests. None.

References

- Abrams DS, Bertrand M and Mullainathan S (2012) Do judges vary in their treatment of race? *The Journal of Legal Studies* 41(2), 347–83.
- Ashenfelter O, Eisenberg T and Schwab SJ (1995) Politics and the judiciary: The influence of judicial background on case outcomes. *Journal of Legal Studies* 24, 257–81.

- Beim D, Clark TS and Lauderdale BE (2021) Republican-majority appellate panels increase execution rates for capital defendants. *The Journal of Politics* 83(3), 1163–7.
- BFM (2008) Asylstatistik 2007. <https://www.news.admin.ch/news/message/attachments/10884.pdf>.
- Bonneau CW, Hammond TH, Maltzman F and Wahlbeck PJ (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 CL, Epstein L and Martin AD (2010) Untangling the causal effects of sex on judging. *American Journal of Political Science* 54(2), 389–411.
- Büchel K, Kiener R, Lienhard A and Roller M (2022) Automated assignment of judges to court panels: Principles and empirical findings based on the Swiss Federal Administrative Court. *Revue française d'administration publique* 184(4), 1001–13.
- Carrubba CJ, Gabel M and Hankla C (2008) Judicial behavior under political constraints: Evidence from the European Court of Justice. *American Political Science Review* 102(4), 435–52.
- Clark TS, Engst BG and Staton JK (2018) Estimating the effect of leisure on judicial performance. *The Journal of Legal Studies* 47(2), 349–90.
- Cross FB and Tiller EH (1998) Judicial partisanship and obedience to legal doctrine: Whistleblowing on the Federal Courts of Appeals. *The Yale Law Journal* 107(7), 2155–76.
- Ecker A, Enns-Jedenastik L and Haselmayer M (2020) Gender bias in asylum adjudications: Evidence for leniency toward token women. *Sex Roles* 82, 117–26.
- Eisenberg T, Fisher T and Rosen-Zvi I (2013) Group decision making on appellate panels: Presiding justice and opinion justice influence in the Israel Supreme Court. *Psychology, Public Policy, and Law* 19(3), 282.
- Engst BG, Gschwend T, Schaks N, Sternberg S and Wittig C (2017) Zum Einfluss der Parteinähe auf das Abstimmungsverhalten der Bundesverfassungsrichter—eine quantitative Untersuchung. *JuristenZeitung* 72(17), 816–26.
- Epstein L, Šadl U and Weinshall K (2022) The role of comparative law in the analysis of judicial behavior. *The American Journal of Comparative Law* 69(4), 689–719.
- Epstein L, Landes WM and Posner RA (2013) *The Behavior of Federal Judges: A Theoretical and Empirical Study of Rational Choice*. Cambridge, MA: Harvard University Press.
- Epstein L and Weinshall K (2021) *The Strategic Analysis of Judicial Behavior: A Comparative Perspective, Elements in Law, Economics and Politics*. Cambridge, MA: Cambridge University Press.
- Fabri M and Langbroek PM (2007) Is there a right judge for each case - a comparative study of case assignment in six European countries. *European Journal of Legal Studies* 1(2), 292–315.
- Fischman JB (2011) Estimating preferences of circuit judges: a model of consensus voting. *Journal of Law and Economics* 54(4), 781–809.
- Fischman JB (2013) Interpreting circuit court voting patterns: a social interactions framework. *Journal of Law, Economics, & Organization* 31(4), 808–42.
- Fischman JB (2014) Measuring inconsistency, indeterminacy, and error in adjudication. *American Law and Economics Review* 16(1), 40–85.
- Frandsen B, Lefgren L and Leslie E (2023) Judging judge fixed effects. *American Economic Review* 113(1), 253–77.
- Gazal-Ayal O and Sulitzeanu-Kenan R (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), 403–28.
- Ginsburg RB (1990) Remarks on writing separately. *Washington Law Review* 65, 133.
- Glynn AN and Sen M (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 L (2005) *Becoming Justice Blackmun: Harry Blackmun's Supreme Court Journey*. New York: Times Books, Henry Holt.
- Grossman G, Gazal-Ayal O, Pimentel SD and Weinstein JM (2016) Descriptive representation and judicial outcomes in multiethnic societies. *American Journal of Political Science* 60(1), 44–69.
- Hangartner D, Lauderdale BE and Spirig J (2025) Replication Data for: Inferring Individual Preferences from Group Decisions: Judicial Preference Variation and Aggregation on Collegial Courts, <https://doi.org/10.7910/DVN/X22XKN>, Harvard Dataverse, V1.
- Harris AP and Sen M (2019) Bias and judging. *Annual Review of Political Science* 22(1), 241–59.
- Jeffries JC (2001) *Justice Lewis F Powell, Jr.* New York: Fordham University Press.
- Kastellec JP (2013) Racial diversity and judicial influence on appellate courts. *American Journal of Political Science* 57(1), 167–83.
- Kiener R (2001) *Richterliche Unabhängigkeit: Verfassungsrechtliche Anforderungen an Richter und Gerichte*. Bern: Stämpfli.
- Kornhauser LA (1992) Modeling collegial courts. II. Legal doctrine. *Journal of Law, Economics, & Organization* 8, 441–70.
- Kornhauser LA and Sager LG (1986) Unpacking the court. *Yale Law Journal* 96, 82–117.
- Landa D and Lax JR (2009) Legal doctrine on collegial courts. *The Journal of Politics* 71(3), 946–63.
- Lauderdale BE and Clark TS (2012) The Supreme Court's many median justices. *American Political Science Review* 106(04), 847–66.

- Law DS** (2004) Strategic judicial lawmaking: Ideology, publication, and asylum law in the ninth circuit. *University of Cincinnati Law Review* 73, 817–66.
- Lax JR and Cameron CM** (2007) Bargaining and opinion assignment on the US supreme court. *The Journal of Law, Economics, & Organization* 23(2), 276–302.
- Lorenz J, Rauhut H, Schweitzer F and Helbing D** (2011) How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences of the United States of America* 108(22), 9020–5.
- Malecki M** (2012) Do ECJ judges all speak with the same voice? Evidence of divergent preferences from the judgments of chambers. *Journal of European Public Policy* 19(1), 59–75.
- Martén L, Hainmueller J and Hangartner D** (2019) Ethnic networks can foster the economic integration of refugees. *Proceedings of the National Academy of Sciences of the United States of America* 116(33), 16280–5.
- Merritt DJ and Brudney JJ** (2001) Stalking secret law: What predicts publication in the United States Courts of Appeals. *Vanderbilt Law Review* 54(1), 71–121.
- Peresie JL** (2005) Female judges matter: Gender and collegial decisionmaking in the federal appellate courts. *Yale Law Journal* 114(7), 1759–87.
- Posner RA** (2010) *How Judges Think*. Cambridge, MA: Harvard University Press.
- Raffaelli R** (2012) Dissenting Opinions in the Supreme Courts of the Member States. In *Study for the European Parliament*. Directorate General for Internal Policies-Policy Department C: Citizens' Rights and Constitutional Affairs. <https://www.europarl.europa.eu/document/activities/cont/201304/20130423ATT64963/20130423ATT64963EN.pdf>
- Ramji-Nogales J, Schoenholtz AI and Schrag PG** (2007) Refugee roulette: Disparities in asylum adjudication. *Stanford Law Review* 60, 295–411.
- Raselli N** (2011) Richterliche Unabhängigkeit. In *Justice—Justiz—Giustizia*, 3.
- Rehaag S** (2007) Troubling patterns in Canadian refugee adjudication. *Ottawa Law Review* 39, 335–65.
- Revesz RL** (1997) Environmental regulation, ideology, and the DC Circuit. *Virginia Law Review* 83(8), 1717–72.
- Shayo M and Zussman A** (2011) Judicial ingroup bias in the shadow of terrorism. *The Quarterly Journal of Economics* 126(3), 1447–84.
- Spirig J** (2023) When issue salience affects adjudication: Evidence from Swiss asylum appeal decisions. *American Journal of Political Science* 67(1), 55–70.
- Spirig J** (2024) *Changing Procedures: How Political Actors Shape Judicial Outcomes*. Working Paper.
- Sunstein CR, Schkade D, Ellman LM and Sawicki A** (2006) *Are Judges Political?: An Empirical Analysis of the Federal Judiciary*. Washington DC: Brookings Institution Press.
- Taylor MH** (2007) Refugee roulette in an administrative law context: The “Déjà vu” of decisional disparities in agency adjudication. *Stanford Law Review* 60, 475–501.
- Van Dijk D, Sonnemans JS and Bauw E** (2014) Judicial error by groups and individuals. *Journal of Economic Behavior & Organization* 108, 224–35.