Do Judges Have Emotions?

What AI Sees at the European Court of Human Rights

Ezgi Siir Kibris[1] Cantay Caliskan[2,3]

Rochester Institute of Technology University of Rochester

Abstract

It has recently been hypothesized in the political science literature that judges convey their preferences during public hearings (Dietrich et al., 2019a). Using the European Court of Human Rights (ECHR) as our case study, we test whether the leanings of judges can be read from oral arguments by extracting visual cues. This paper analyzes 311 webcasts of hearings, corresponding to 559 hours and 49 minutes of video recordings from 2007 to 2026, that show judges listening to the parties, governments and individual applicants. In 1,851,432 frames, sampled once per second, our pipeline detects 25,076,137 faces and matches 705,155 of them to the judges on the bench. We measure the judges' emotional expressions second by second and relate them to the party that is speaking and to judicial decisions. Our video data is further enriched with a set of manually collected features on judges and cases, which serve as controls. The research contributes to the field by being the first study to systematically measure the emotional expressions of ECHR judges (i), and by combining video with the record of speakers and votes to test sixteen hypotheses about judges in a high court (ii). The results show that visual cues carry information that written judgments do not possess, but not the information we expected. Judges differ from one another in stable ways, down to the emotion that dominates: the classifier reads anger in nearly all men, sadness in most women, and more anger in older judges. Expressions are 3.5 percentage points more negative while the government speaks than while the applicant does, a difference that largely reflects where the judges are looking. We find no evidence of national, regional, or political bias, and expressions do not predict votes.

Keywords: Judicial Behavior; European Court of Human Rights; Artificial Intelligence; Facial Expression Analysis; Judicial Bias.

[1] eskics@rit.edu.

[2] cantay.caliskan@rochester.edu.

[3] We thank Carol Li, Huilin Piao, Leo Pan, and Navya Bhagat for their research assistance.

A good judge must be able "to divest himself of all fear, anger, hatred, love, and compassion."
— Thomas Hobbes, *Leviathan*, Chapter 26

"You can't do human activity—and judging is a human activity—without having human emotions."
— US Supreme Court Justice Sonia Sotomayor

1 Introduction

On March 13, 2024, the European Parliament approved the EU Artificial Intelligence (AI) Act, the world's first comprehensive regulatory framework for AI. Among its provisions, the Act prohibits AI systems that infer emotions in workplaces and educational institutions, with exceptions for medical and safety purposes, and places most other uses of emotion recognition in its high-risk category (European Parliament, 2024a,b; Littler, 2024). These restrictions reflect broader concerns about algorithmic bias, privacy, and the consequences of inaccurate interpretations of human behavior. Such concerns weigh particularly heavily in judicial settings, where decisions can have profound consequences for individuals. At the same time, AI and machine learning offer new opportunities for studying judicial behavior, because they make it possible to analyze audiovisual records that would be difficult to examine by hand.

This paper explores one such application at the ECHR by analyzing video recordings of its public hearings. We examine whether judges display discernible emotional expressions during hearings and whether these expressions reveal systematic patterns in judicial behavior. The Court has recorded its public hearings since 2007 and makes the recordings available online in the original language of the hearing and with interpretation into English and French (European Court of Human Rights, 2026c). Our metadata cover 319 hearings held between 2007 and 2026, 255 before the Grand Chamber and 64 before a Chamber, and we analyze the 311 for which a recording is available, 559 hours and 49 minutes of proceedings. In each video, our computer vision pipeline identifies the judges and classifies their facial expressions once per second. We combine these mea-

sures with a record of who was speaking, which an AI system drafted from automatic transcripts, and with the judgments and votes that followed the hearings.

This study makes two primary contributions. First, it introduces ECHR hearing videos as a new source of data for the study of judicial behavior in Europe. Research that uses audiovisual and vocal information from court proceedings has focused primarily on the United States (US) Supreme Court (Dietrich et al., 2019a; Knox and Lucas, 2021), and comparatively little attention has been given to European and international courts. We extend this line of research to the ECHR and show how methods developed for the US Supreme Court can be adapted to a very different institutional setting. The ECHR is especially valuable for this purpose because of its multinational structure. Judges are elected in respect of the member states, cases set applicants against governments of different countries, and each bench brings together judges with different national, political, and institutional backgrounds. We can therefore ask whether judges respond more favorably to governments with which they share a nationality, a region, or a political system. Research on the Court's votes has found that judges tend to favor their own governments but little evidence of broader cultural or geopolitical bias (Voeten, 2008). We ask whether such ties are visible earlier, in the way a judge listens.

Second, the paper contributes to the emerging literature on emotion recognition in judicial settings. To our knowledge, ECHR hearing videos have not previously been used to analyze judges' facial expressions systematically. We test sixteen hypotheses, and we show both what can be learned from these videos and where automated measures mislead. In our main analysis, judges' expressions are 3.5 percentage points more negative while the respondent government speaks than while the applicant's side speaks, a difference that holds in every robustness check of the usual kind. It does not show, however, that judges look less kindly on governments. Judges turn their heads toward whoever is speaking, and for most of the period the camera saw them more nearly head-on during the government's turn. The classifier scores a turned face as markedly less negative than the face of the same judge looking toward the camera. Once the positions of the eyes and nose are held constant, between 0.5 and 0.9 points of the difference remain, and none of these estimates

can be distinguished from zero.

The hypotheses on national, regional, and political affinity find little support. National judges do not differ from their colleagues in how they appear while their own government speaks, and neither do judges whose states share European Union (EU) membership, a region, or a level of democracy with the respondent state. These tests have little power, because the camera rarely shows a judge while the parties speak. When we condition on the ruling, as the hypotheses are worded, national judges appear slightly more positive when they vote for their own state, but this pattern is not robust. Expressions do not predict votes. In an exploratory test, judges who were more expressive during the government's turn than during the applicant's were less likely to vote for the applicant, but this pattern rests on the 22 benches that divided and weakens when more judge-hearings are included.

What the data show most clearly is that judges differ from one another. The judge accounts for about 48 percent of the variance in negative expression across hearings, and for most judges a single emotion dominates the classifier's readings from one hearing to the next. That emotion is anger for nearly all men and sadness for most women. At the same image quality and in the same hearing, the classifier scores women judges 11.4 points lower on negative expression than men, and it scores older judges higher on anger than younger ones. Region, earlier career, and the part that a judge takes in the hearing make no difference. These comparisons between judges are exploratory, and they describe the classifier's reading of the judges at least as much as the judges themselves.

The remainder of the paper proceeds as follows. First, we review the relevant literature on judicial behavior, emotion, and computational approaches to court proceedings, and we state our research questions and hypotheses. We then introduce the ECHR hearing videos and provide descriptive statistics. Next, we describe the pipeline and the methods used to analyze the videos. We subsequently present the empirical findings and conclude by discussing their implications, limitations, and directions for future research. Throughout, we describe what the classifier returns as an expression that a face displays, and we make no claim about what any judge felt.

2 Literature Review

Research on emotion in the judiciary draws on several strands that developed largely in isolation from one another. One strand asks whether judges experience emotion at all, against a long-standing ideal of the dispassionate jurist. A second asks whether emotion is related to the decisions that judges reach. A third asks how judicial emotion can be detected and measured, moving from written language to voice and, most recently, to video.

The traditional image of judicial authority emphasizes neutrality, rationality, and emotional detachment. A substantial body of research argues instead that emotions are an inherent part of judicial work (Maroney, 2011; Roach Anleu et al., 2015; Snider et al., 2021). For a long time emotion received little attention in research on judging because it was viewed as incompatible with impartiality (Bandes and Blumenthal, 2012). More recent scholarship examines how judges experience and manage emotions as part of their everyday professional responsibilities (Bergman Blix and Wettergren, 2018; Toy-Cronin, 2019). Studies in several countries find that judges and magistrates experience emotions in court and develop strategies for regulating how those emotions are expressed (Mack and Roach Anleu, 2010; Roach Anleu and Mack, 2005; Roach Anleu et al., 2016). The demands are greatest in cases with disturbing subject matter, where exposure to traumatic testimony has been linked to secondary traumatic stress among judges (Chamberlain and Miller, 2009). Two expectations follow for our study. If composure is a professional skill, a judge's face should be quiet most of the time, and whatever varies should vary mainly between individuals. If subject matter is important, cases about death and ill-treatment should leave more of a trace than other cases.

Two findings from psychology bear on what a judge's face shows. Emotional expressivity is a stable trait on which people differ (Gross and John, 1997), which is a second reason to expect that expression belongs to the judge more than to the hearing. The norms that govern the display of emotion also differ for women and men. Women smile more than men, although the difference narrows when both hold the same role (LaFrance et al., 2003), and professional women who express anger are accorded lower status than men who do the same (Brescoll and Uhlmann, 2008).

In German televised debates, Angela Merkel expressed less anger than her male opponents, and audiences punished her displays of anger and rewarded her displays of happiness (Boussalis et al., 2021).

If judges experience emotions during proceedings, the next question is whether those emotions are related to their decisions. Experiments with sitting judges suggest that sympathy or antipathy toward a litigant can move a ruling (Wistrich et al., 2015), in line with psychological research showing that emotional arousal affects judgments that are meant to be objective (Oliver and Griffitt, 1976). Most of the observational evidence comes from the US Supreme Court. Oral arguments there contain information about the eventual outcome of a case (Wasby et al., 1976; Shullman, 2004; Roberts, 2005; Johnson et al., 2006). The side that receives more questions tends to lose (Johnson et al., 2009; Epstein et al., 2010), and so does the side toward which the justices use more unpleasant language (Black et al., 2011). Emotional responses may simply accompany a judge's evaluation of the arguments, and for that reason they may reveal leanings that otherwise remain hidden until the judgment is delivered.

How can such responses be measured? Earlier research relied on written texts, including briefs, transcripts, and parliamentary debates, to identify emotional language (Black et al., 2011; Rheault et al., 2016; Gennaro and Ash, 2022). Text captures only what is put into words, however, and this limitation has led researchers to audio (Dietrich et al., 2019a). Emotional states are reflected in vocal characteristics such as pitch (Laukka et al., 2005), and changes in the voice are difficult to control deliberately (Ekman et al., 1991). Dietrich et al. (2019a) show that Supreme Court justices speak at a higher pitch to the lawyers for the side that they go on to vote against, and that this information improves predictions of their votes. Knox and Lucas (2021) model the audio directly and recover expressions of skepticism that the words alone do not carry. In Congress, vocal pitch tracks the intensity of legislators' commitment to the issues on which they speak (Dietrich et al., 2019b).

Video is the most recent addition to this literature. Automated analysis of faces and movement has been used to study candidates in televised debates (Boussalis et al., 2021; Boussalis and

Coan, 2021), the emotions that politicians express in images (Boussalis et al., 2022), the emotional content of cable news (Ash and Caliskan, 2022), and the social distance between the parties on the floor of the US House of Representatives (Dietrich, 2021). Nearly all of this work concerns national institutions, and nearly all of it studies people while they are speaking. We study an international court, and we study judges mainly while they listen.

Research on emotion also suggests who should display what. Expressivity is a stable trait of the person (Gross and John, 1997). Emotions carry gender stereotypes, with anger ascribed to men and sadness, fear, and happiness to women (Plant et al., 2000), and the same expectations shape how the faces of men and women are read (Hess et al., 2004). Older adults report fewer negative emotions and more control over them (Gross et al., 1997), yet the lines of an older face lead observers to see anger and sadness in it (Hess et al., 2012). Societies differ in the rules that govern the display of emotion (Matsumoto et al., 2008). In deliberating groups, finally, women speak less than men when they are few (Karpowitz et al., 2012), and at the US Supreme Court they are interrupted more often (Jacobi and Schweers, 2017), which raises the question of how voice and expression go together on a bench.

The ECHR is the international court on which questions of bias have been examined most closely. Voeten (2008) found no general cultural or geopolitical bias in the judges' votes. He did find that judges favored their own governments when those governments were respondents, especially judges with insecure career prospects, and that judges from former socialist states were more willing than others to find violations by former socialist governments. Later studies show that judges behaved more independently once their terms could no longer be renewed (Stiansen, 2022) and that the Court became more restrained toward consolidated democracies that had criticized it (Stiansen and Voeten, 2020). Each of these studies starts from the judgment. If national or other affinities exist, they might also be visible earlier, in how a judge listens to a government with which she has a tie. Our hypotheses build on this idea.

3 Research Questions and Hypotheses

3.1 Research Questions

This study examines the emotional expressions of judges during court hearings and investigates whether these expressions vary with contextual, national, and case-specific factors. We ask, first, whether judges exhibit discernible emotional reactions during oral proceedings and how far such reactions differ from one judge to another and between women and men judges. Second, we ask whether judges' expressions differ with the party that is speaking and with the countries involved in a case, and in particular whether judges respond differently when their own country is the respondent. Third, we ask whether judges from EU member states respond differently when the respondent is another EU member state. Fourth, we consider the extent to which cultural and national characteristics, such as region and political regime, shape judges' responses. Fifth, we ask whether patterns of expression vary with the subject matter of the cases being heard. Sixth, we ask whether a judge's expressions during a hearing are related to her eventual vote. Finally, we ask whose emotions dominate on the bench: whether the kind of emotion that a judge displays is her own, and whether it differs with her gender, age, region, earlier career, and part in the hearing.

3.2 Hypotheses

Based on these research questions, we develop a set of hypotheses on whether judges display more positive emotional responses when case outcomes favor countries or applicants with whom they share national, regional, political, or demographic characteristics. These hypotheses build on the expectation that judges' emotional reactions may reflect forms of affinity or identification that are not necessarily captured by their formal votes.

Hypothesis 1 (National Bias): *Judges exhibit a more positive emotional response when their rulings favor their home country.* Judges may feel greater attachment to, familiarity with, or identification with their own country, which could be reflected in their emotional expressions. Social identity theory holds that people favor groups to which they belong (Tajfel and Turner,

1979). On the International Court of Justice, judges favor the states that appoint them (Posner and de Figueiredo, 2005), and ECtHR judges are more likely to vote for their own government (Voeten, 2008).

Hypothesis 2 (EU Bias): *Judges from EU member states exhibit a more positive emotional response when their rulings favor other EU member states.* Shared membership in the EU may create a common political and institutional identity. European integration has given most citizens of EU countries a secondary European identity alongside their national one (Risse, 2010), and judges from EU states may extend to other members some of the affinity they feel for their own.

Hypothesis 3 (Eastern vs. Western Europe): *Judges from Eastern (Western) European countries exhibit a more positive emotional response when their rulings favor Eastern (Western) European countries.* Regional similarities in political history, institutions, and legal traditions may generate greater affinity. Evidence for regional bias among international judges is weak, however (Posner and de Figueiredo, 2005), and Voeten (2008) finds that judges from former socialist states were, if anything, harder on former socialist governments. We therefore treat this hypothesis as two-sided.

Hypothesis 4 (Northern vs. Southern Europe): *Judges from Northern (Southern) European countries exhibit a more positive emotional response when their rulings favor Northern (Southern) European countries.* Cultural, historical, and institutional similarities within these regional groupings may influence how judges respond. National cultures differ in lasting ways (Hofstede, 2001), and judges on the International Court of Justice favor, though weakly, states whose language and religion resemble those of their own state (Posner and de Figueiredo, 2005).

Hypothesis 5 (Democratic Similarity): *Judges from more democratic countries exhibit a more positive emotional response when their rulings favor other highly democratic countries.* Judges may feel greater affinity with states that share similar democratic institutions and political values. Liberal theories of international law hold that relations among liberal democracies differ in kind from relations with other states (Slaughter, 1995), and there is some evidence that international judges favor states with political systems like their own (Posner and de Figueiredo, 2005). We

measure similarity as the closeness of the two states' democracy scores.

Hypothesis 6 (Judge Characteristics): *Female judges exhibit a more positive emotional response when their rulings favor female applicants.* Shared gender may create greater identification or empathy with applicants. In U.S. federal appellate courts, the gender of the judge makes a difference in sex discrimination cases but in few other areas of law (Boyd et al., 2010).

Each of these six hypotheses links an emotional response to a ruling, and we examine both parts. We first compare judges' expressions while the parties plead. If a judge identifies with the respondent state, her expressions should be more favorable (less negative or more positive) during the government's submissions than during the applicant's, relative to her colleagues on the same bench. These comparisons are our main tests (Appendix C). We then examine expressions conditional on the eventual ruling: among judges whose votes favored the respondent state, those sharing a tie with that state should display more positive expressions. These latter tests were introduced after the initial results were examined and are treated as exploratory.

The research questions also lead to five hypotheses that do not depend on a tie between the judge and the respondent state. They concern the judges as individuals, the two parties, the subject matter of the case, the vote, and the gender of the judge.

Hypothesis 7 (Individual Differences): *Judges' emotional expressions differ more between judges than between hearings.* If composure is a professional norm and expressivity a personal trait, what remains visible on the bench should belong to the judge more than to the case, and a judge's level of expression should be stable from one hearing to the next (Gross and John, 1997; Bergman Blix and Wettergren, 2018). For the same reason we expected presiding judges, who conduct the hearing on behalf of the Court, to be the most restrained.

Hypothesis 8 (Parties): *Judges' emotional expressions differ between the submissions of the respondent government and those of the applicant.* At the US Supreme Court, justices treat the two sides differently during oral argument, in the questions they ask, in the language they use, and in their tone of voice (Johnson et al., 2009; Epstein et al., 2010; Black et al., 2011; Dietrich et al., 2019a). Before the ECHR the two sides are unequal as well, with an individual on one side and a

state on the other. Judges may therefore respond differently to the two parties, although we have no directional expectation.

Hypothesis 9 (Subject Matter): *Judges display more negative expressions in cases concerning death, torture, and forced labor under Articles 2 to 4 of the Convention.* Disturbing material is the part of judicial work that is hardest to keep at a distance (Chamberlain and Miller, 2009).

Hypothesis 10 (Votes): *Greater negative expression toward one side is associated with a subsequent vote against that side.* This is the pattern found for questions, language, and vocal pitch at the US Supreme Court (Johnson et al., 2009; Black et al., 2011; Dietrich et al., 2019a).

Hypothesis 11 (Gender of the Judge): *Female judges display less negative emotion than male judges.* Anger is the emotion that professional women are least free to show (Brescoll and Uhlmann, 2008; Boussalis et al., 2021), and judges in a filmed hearing are professionals under observation. We therefore expect women judges to show less negative emotion than men and, if anything, more positive emotion (LaFrance et al., 2003).

The last five hypotheses compare the judges with one another and ask whose emotions dominate.

Hypothesis 12 (Emotional Profiles): *Judges differ in the emotion that dominates their expressions, and a judge's dominant emotion is the same from one hearing to the next.* If expressivity is a trait of the person (Gross and John, 1997), the kind of emotion that a judge shows should be as much her own as the amount.

Hypothesis 13 (Gender and Specific Emotions): *Anger is more prominent in the expressions of male judges, and sadness, fear, and happiness in those of female judges.* These are the emotions that stereotypes assign to men and to women (Plant et al., 2000), and the same expectations shape how faces are read (Hess et al., 2004).

Hypothesis 14 (Age): *Judges' negative expression differs with their age.* Older adults report fewer negative emotions and more control over them (Gross et al., 1997), which predicts less negative expression among older judges. The lines of an older face lead observers to see anger and sadness in it (Hess et al., 2012), which predicts more. We treat the hypothesis as two-sided.

Hypothesis 15 (Culture and Career): *Judges from post-socialist states display less positive emotion, judges from Southern European states are more expressive, and former career judges are more restrained than their colleagues.* Societies differ in the rules that govern the display of emotion, and more individualistic societies endorse more expression, of positive emotions in particular (Matsumoto et al., 2008). The second expectation follows the common contrast between a reserved North and an expressive South. Composure on the bench, finally, is a skill that career judges have practiced longest (Bergman Blix and Wettergren, 2018).

Hypothesis 16 (Engagement): *Judges who take an active part in a hearing display more emotion than judges who remain silent.* At the US Supreme Court, the justices' questions carry emotional language and vocal arousal (Black et al., 2011; Dietrich et al., 2019a). If engagement shows on the face, judges who put questions should be more expressive while they listen, and the judges who ask most and are shown most should be the most expressive ones.

Table 1 summarizes the sixteen hypotheses.

4 Data

4.1 The Court and Grand Chamber Cases

The ECHR hears applications against the 46 member states of the Council of Europe, which include the United Kingdom, Turkey, and all EU member states (European Court of Human Rights, 2024). There were 47 member states until Russia ceased to be a member in March 2022. Cases are generally heard by Chambers of seven judges or, when they are particularly important, by the Grand Chamber of seventeen judges. The Grand Chamber includes the President and Vice-Presidents of the Court, the Section Presidents, and the judge elected in respect of the respondent state, who is known as the national judge. The remaining judges are drawn by lot (European Court of Human Rights, 2025). If the national judge cannot sit, the President chooses an ad hoc judge from a list submitted by that state. Decisions are taken by majority vote. A case reaches the Grand Chamber when a Chamber relinquishes jurisdiction or when a panel of five judges accepts

Table 1: Hypotheses and expectations

	Expectation	Comparison
H1 National Bias	More positive response to the judge's home state	National judge vs. colleagues on the same bench
H2 EU Bias	More positive response of judges from EU states to EU respondents	Judge and respondent both from EU states
H3 Eastern vs. Western Europe	More positive response within the same group (two-sided)	Same East/West group
H4 Northern vs. Southern Europe	More positive response within the same group	Same North/South group
H5 Democratic Similarity	More positive response the closer the two states' democracy scores	Distance in democracy scores
H6 Judge Characteristics	More positive response of women judges to women applicants	Women judges in discrimination cases
H7 Individual Differences	Judges explain more variance than hearings; presiding judges are less expressive	Judge-hearing means
H8 Parties	Displays differ with the party speaking (two-sided)	Same judge within a hearing
H9 Subject Matter	More negative display in Article 2–4 cases	Same judge across hearings
H10 Votes	More negative display toward a side predicts a vote against it	Judges on the same bench
H11 Gender of the Judge	Less negative display by women judges than by men	Women and men in the same hearing, at the same image quality
H12 Emotional Profiles	A judge's dominant emotion is the same across hearings	Odd- and even-numbered hearings of each judge
H13 Gender and Specific Emotions	Anger more prominent among men; sadness, fear, and happiness among women	As H11, for each emotion
H14 Age	Negative display differs with age (two-sided)	Judges in the same hearing, at the same image quality
H15 Culture and Career	Less positive display by judges from post-socialist states; more expressivity in the South; less among career judges	As H14
H16 Engagement	More expressivity among judges who ask questions and who are shown most	Same hearing; same judge; across judges

Note: Appendix Table A5 lists the test of each hypothesis.

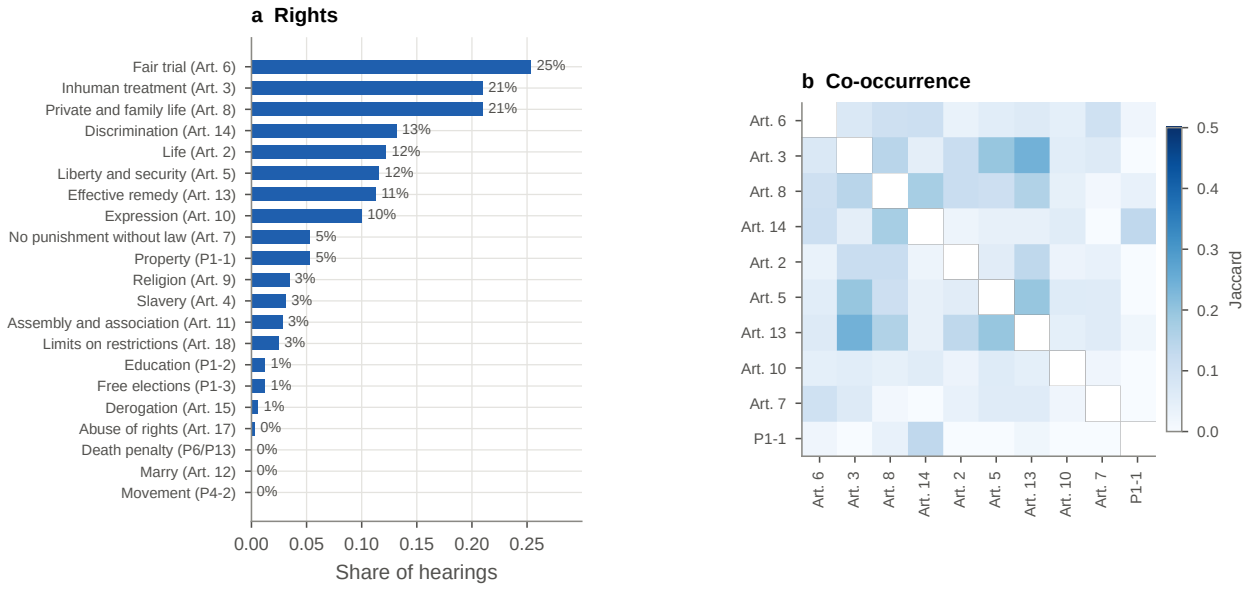

Figure 1: Convention rights at stake. (a) Share of hearings; a hearing can raise several rights. (b) Jaccard overlap between rights.

a party's request for referral after a Chamber judgment. Grand Chamber judgments are final, and in this sense the Grand Chamber functions like a supreme court within the Convention system.

Nearly all applications are decided on the papers. Public hearings are held in most Grand Chamber cases and in a small number of Chamber cases. The applicant's side and the government make their prepared submissions in an order that varies from case to case, the judges put questions, and after a short adjournment each side replies.

The hearings cover a wide range of Convention rights (Figure 1a). A quarter of them concern the right to a fair trial (Article 6), and about a fifth each concern inhuman treatment (Article 3) and private and family life (Article 8). Most hearings raise more than one right, and some rights tend to come up together. Inhuman treatment is often argued together with the right to an effective remedy (Article 13) and the right to liberty (Article 5) (Figure 1b). About a third of the hearings concern Articles 2 to 4, the material we use to test Hypothesis 9.

Table 2: Corpus by period and formation

	Hearings	Hours	Median px	Inter-state	Remote	Judges verified per hearing	Verified seconds/1000	Violation share
2007–2012 Chamber	47	77.7	288	2	0	3.5	32.8	0.86
2007–2012 Grand Chamber	91	145.2	288	1	0	5.0	77.8	0.67
2013–2019 Chamber	12	22.1	720	0	0	5.2	18.5	0.73
2013–2019 Grand Chamber	113	191.9	720	3	0	8.9	174.9	0.67
2020–2026 Chamber	5	9.1	720	0	0	4.6	6.2	0.67
2020–2026 Grand Chamber	51	113.8	720	3	10	6.6	62.2	0.61
All	319	559.8	720	9	10	6.4	372.4	0.69

Note: Verified judge-seconds are measurement frames (one per second) in which a panel judge is identified with high confidence and passes the bench check (Section 5). Violation share among hearings with a matched HUDOC judgment.

4.2 Webcasts of Hearings

Since 2007 the Court has made recordings of its public hearings available through its online archive of webcasts (European Court of Human Rights, 2026c). Each recording can be watched in the original language of the hearing or with interpretation into English or French. We collected the recordings of the hearings held between 2007 and 2026 and use the original-language versions. Our metadata list 319 hearings, of which 255 took place before the Grand Chamber and 64 before a Chamber (Figure 2 and Table 2). We processed 311 recordings, or about 559 hours and 49 minutes of video. The others were missing, truncated, or not hearings.

The recordings show three main groups of participants: the judges, the lawyers who represent the applicants, and the agents and lawyers who represent the respondent governments. The webcast is edited live, and the camera follows whoever is speaking. Figure 3 presents example frames and shows how the mix of shots has changed over time. Picture quality improved over the period. Of the hearings, 147 were recorded in low resolution (288 or 360 pixels high) and 164 in high definition (720 or 1,080 pixels high). In 10 hearings during the COVID-19 pandemic the parties took part by video conference. Russia, the United Kingdom, and Turkey were the most frequent respondents (Appendix Figure A3). About a quarter of the hearings concern the right to a fair trial, and about a third concern Articles 2 to 4 of the Convention (Appendix Figure 1).

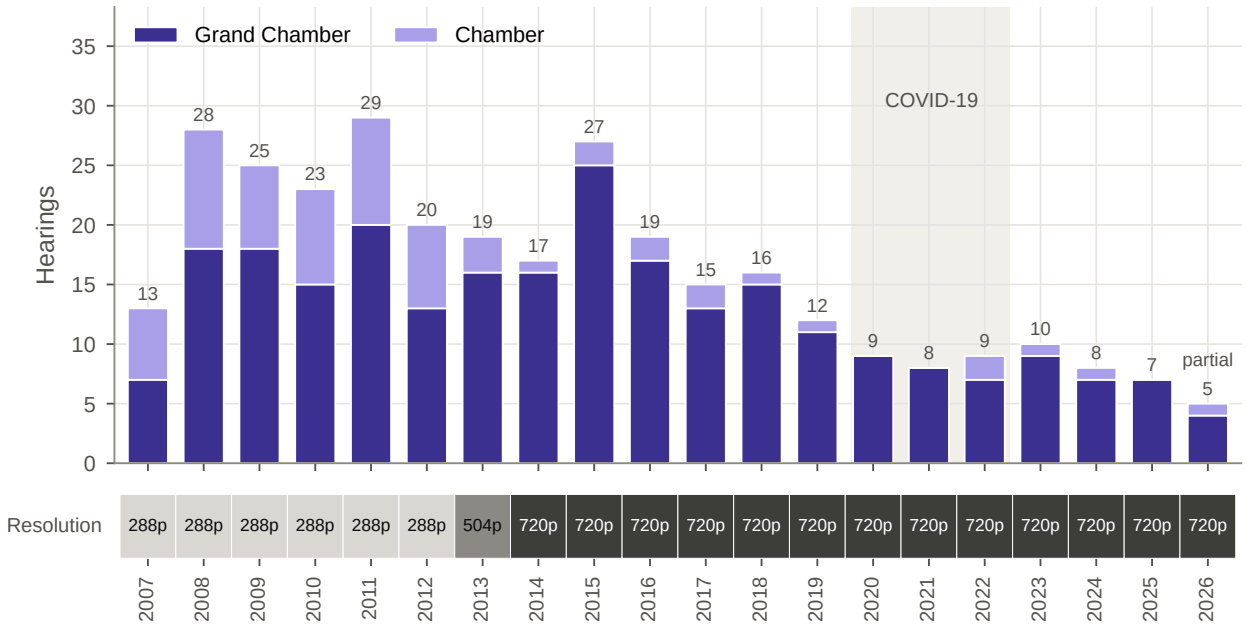

Figure 2: Hearings per year. Note: 2026 runs to November.

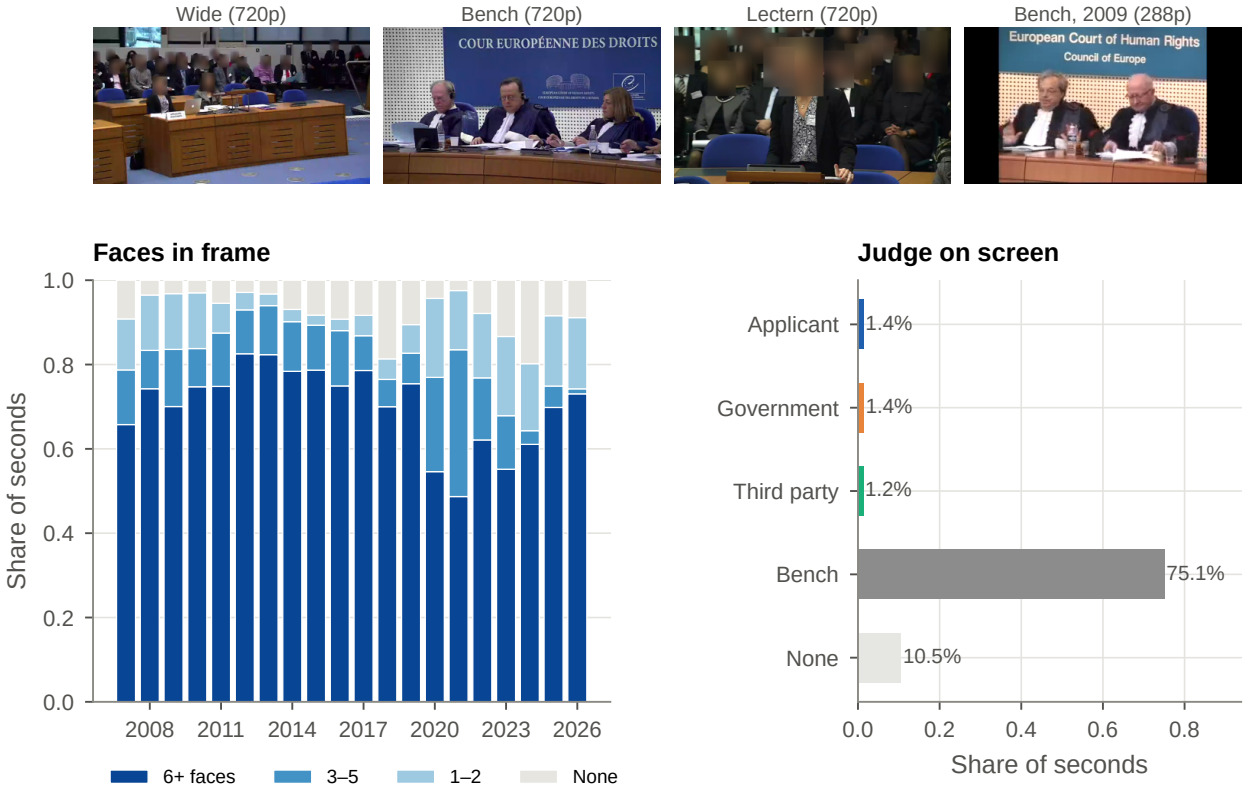

Figure 3: Example frames and camera coverage. Note: Non-judge faces blurred. Frames © ECHR-CEDH.

4.3 Judges

The hearings involve 161 judges in 5,714 seats (Appendix Figure A2). For each judge we recorded the state in respect of which she was elected, which is not always her state of nationality, and whether she sat as national judge, ad hoc judge, President of the formation, or substitute. Gender, year of birth, term, and earlier career come from official records only. Our own biographical file covered 124 judges, and for the judges elected since 2022 we took term dates from the Court's list of judges (European Court of Human Rights, 2026b) and gender from the forms of address in the election records of the Parliamentary Assembly (Parliamentary Assembly of the Council of Europe, 2026). Women are 33 percent of the 145 judges.

To measure the ties named in the hypotheses, we coded for every judge in every hearing whether she is the national judge, whether her state and the respondent state were both EU members at the time of the hearing, whether the two states belong to the same East/West group (post-socialist or not) and to the same North/South group, and how far apart they are on the liberal democracy index of the Varieties of Democracy project (V-Dem Institute, 2026). Appendix Table A1 defines every variable, and Appendix Table A2 gives descriptive statistics.

4.4 Speakers and Hearing Phases

To know who is speaking at each moment, we rely on tables produced for each hearing by ChatGPT 6.1 Sol working as an agent. The system read transcripts of the recordings, which we produced with WhisperX (Bain et al., 2023), the list of participants and, for some hearings, frames from the video. From these it identified speech turns and their speakers, classified procedural phases, and constructed a second-by-second timeline aligned with our frames. The tables cover 312 hearings, and in 307 of the processed hearings they attribute speech to both parties. For the main analyses we use two fields, the side that is speaking and the procedural phase. For the comparisons between judges we also use the judge to whom each question is attributed, a field that we could check against the video. In 99 percent of the 936 questions during which a single face is on screen, that face belongs to the judge named in the record (Appendix F). We drop the seconds within two

seconds of a change of speaker between sides, and we treat gaps of up to three seconds inside one side's speech as part of its turn.

4.5 Judgments and Votes

For each hearing we searched the Court's HUDOC database by application number for the judgment or decision that followed (European Court of Human Rights, 2026a). We recorded the conclusions, the vote on each operative point, and the authors and types of separate opinions. 286 hearings have a matching judgment, 24 have none in HUDOC yet, and we set aside 9 more in which the document retrieved was not the hearing's own decision or came from a different bench. Among the matched cases, 69 percent found at least one violation and 70 percent were divided on at least one point. A judge's vote is coded as favoring the applicant when she is in the majority in a case that found a violation or dissents from a finding of no violation, and as favoring the respondent state otherwise. Partial dissents in cases with mixed outcomes are left uncoded, and substitute judges are excluded.

5 Methods

Every measurement of a face in this paper comes from a computer vision pipeline that we built for the project. Figure 4 gives an overview, and Appendix B documents the algorithms and settings. We began by downloading the hearing webcasts and collecting reference photographs of the judges, which let us tell the judges apart from the lawyers, government agents, applicants, and others who appear on screen. The pipeline then goes through each video twice. The first pass runs once on a graphics processor: it finds every face and does all the work that needs a neural network, including classifying the expression. The second pass decides which judge, if any, each face belongs to. Because it works only from the stored results of the first pass, we could rerun it in seconds whenever we changed a threshold or added a reference photograph.

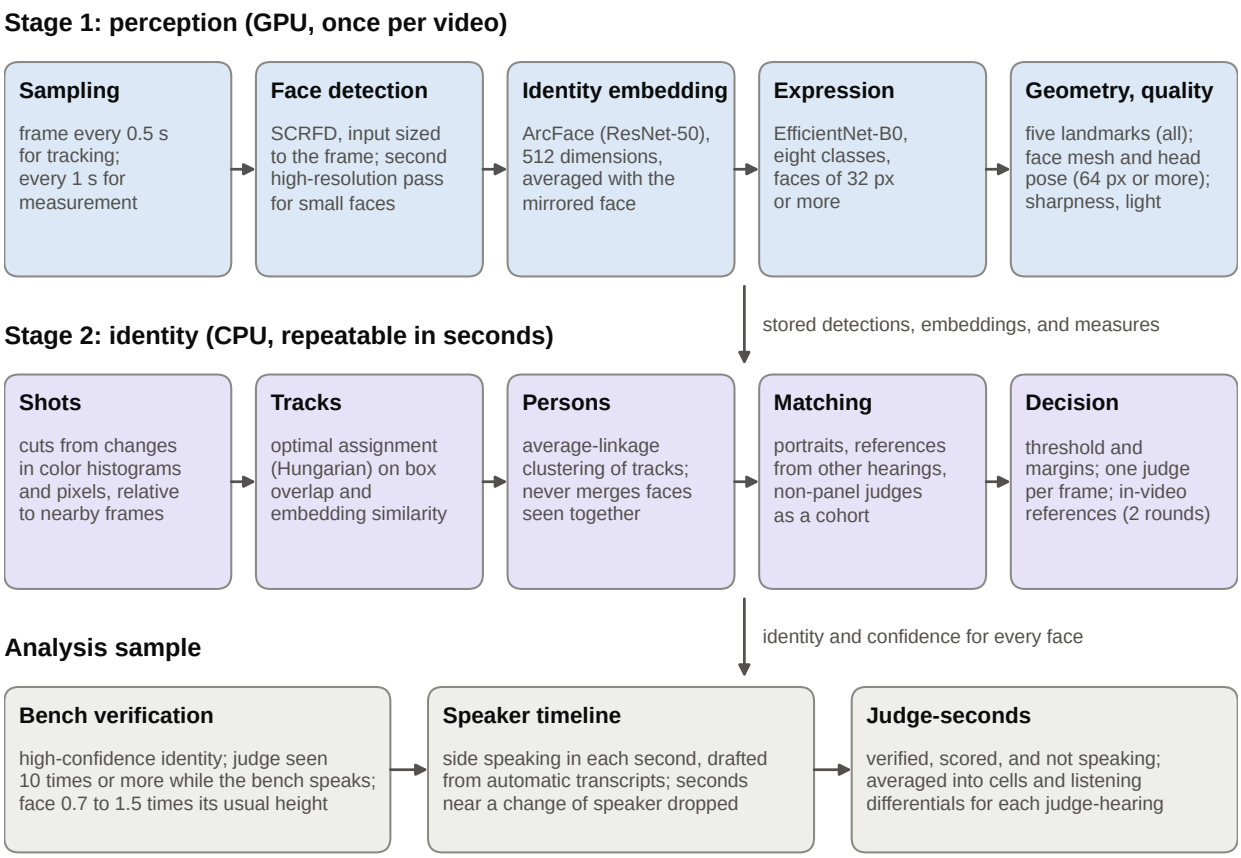

Figure 4: The video pipeline.

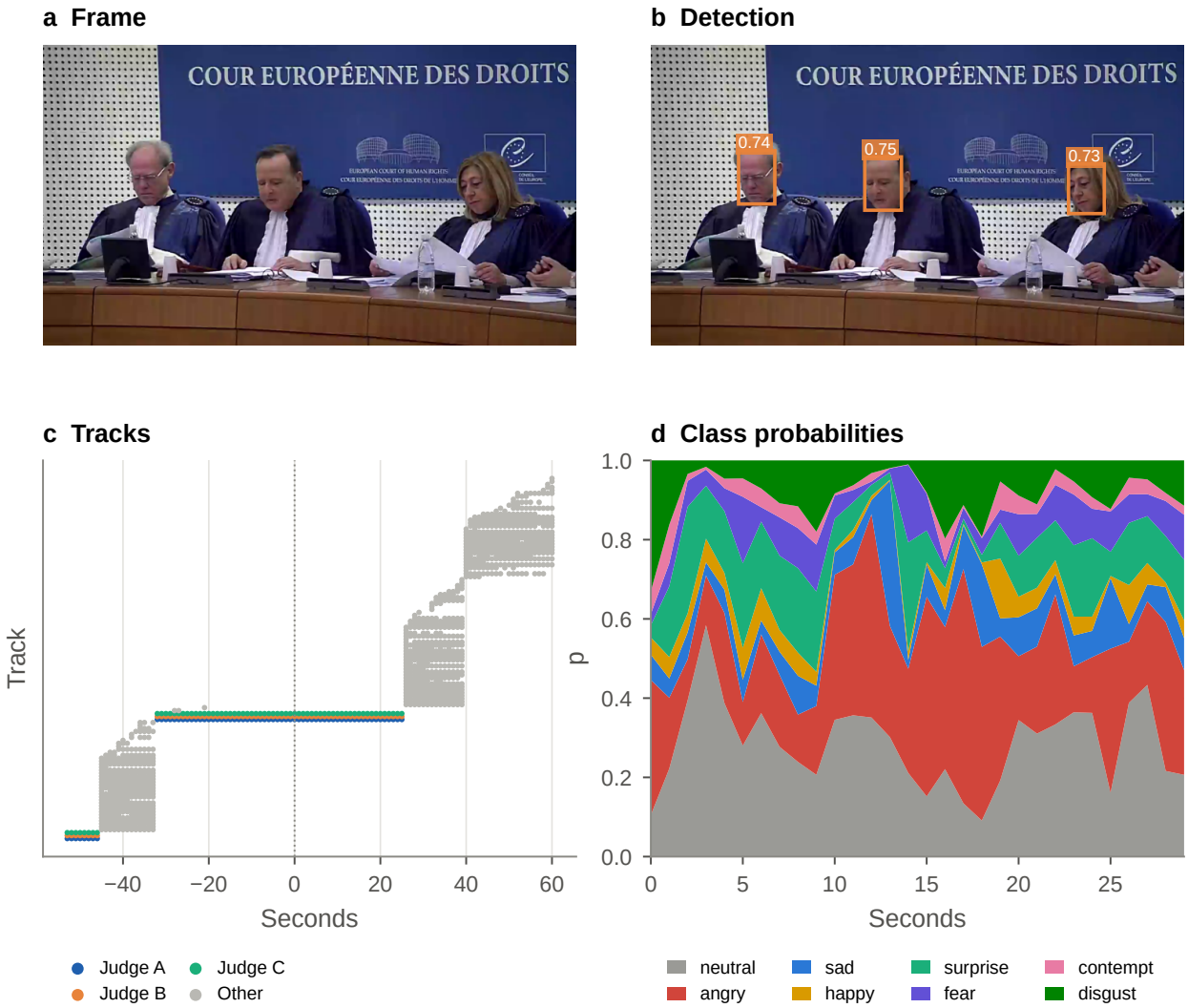

Figure 5: The pipeline applied to one frame (2016 hearing). Note: In (d), the anger share is this face's usual reading. Frame © ECHR-CEDH.

5.1 Face Matching

The pipeline reads a frame every half second to follow faces from one moment to the next, and it measures expressions on one frame per second (Figure 5). It finds faces with SCRFD (Guo et al., 2022), one of the InsightFace `buffalo_l` models. Judges' faces are often only a few dozen pixels high, so the detector's input is fitted to the shape of the frame, low-resolution recordings are enlarged, and the detector looks again at higher resolution whenever it finds only small faces. Every face at least 16 pixels high is then summarized by an ArcFace embedding (Deng et al., 2019), a list of 512 numbers that comes out similar for different images of the same person.

The pipeline identifies people, not single frames. It first finds the camera cuts. Within each shot it links the detections into tracks, pairing faces from one frame to the next by how much their boxes overlap and how alike their embeddings are. It then groups the tracks of a hearing into persons, with the rule that two faces in the same frame can never belong to the same person. The average embedding of each person is compared with reference images of the judges on that day's bench, drawn from their official portraits and from frames in other hearings where they were identified with confidence. It is also compared with the portraits of every other judge of the Court. A person is matched to a judge only when the similarity is high and clearly beats both the next-best judge on the bench and the best judge who was not sitting that day. Once a judge has been identified with confidence, her own faces from that hearing become references too, which recovers shots in profile or looking down. Across the corpus the pipeline found 25,076,137 faces in the measurement frames and matched 705,155 of them to judges on the bench, 72 percent with high confidence.

We keep a face only if three conditions hold: it was matched with high confidence; the judge appears at least ten times in that hearing while a member of the bench is speaking; and the face is between 0.7 and 1.5 times the judge's usual face height on the bench in that hearing. We checked 24 identifications by eye, and all 16 faces that passed these rules were judges on the bench. The rules leave 372,395 judge-seconds. We set them before estimating any of the models that test our hypotheses, and Section 6.7 reports results under looser rules. Because we are interested in how judges react while they listen, we also drop the seconds in which the judge herself is speaking.

5.2 Emotion Classification

In the final stage, faces that are at least 32 pixels high are passed to an expression classifier (EmotiEffLib, model `enet_b0_8_best_vgaf`; Savchenko et al., 2022). This is a convolutional network of the EfficientNet family that was first trained to recognize identities and then tuned on the AffectNet collection of web photographs (Mollahosseini et al., 2019). It receives a square crop of the face from the original frame and, following the basic-emotions framework associated with Ekman (1992), returns probabilities for anger, contempt, disgust, fear, happiness, sadness, and surprise, and for a neutral face. The pipeline scored 511,842 judge faces in this way.

We work with three outcomes. Negative display is the sum of the probabilities of anger, contempt, disgust, fear, and sadness. Positive display is the probability of happiness. Expressivity is one minus the probability of a neutral face. All three describe how a face looks to one classifier, and we call them displays in order not to prejudge what they mean. The happiness probability, for example, correlates only weakly with measured smile width ($r = 0.17$).

The detector also places five landmarks on every face: the eyes, the tip of the nose, and the corners of the mouth. From these the pipeline computes how far the nose sits to the left or right of the midpoint between the eyes, relative to the distance between them. This quantity measures how far the head is turned sideways as the camera sees it, and the pipeline's frontal score, which runs from 0 for a profile to 1 for a face that is square to the camera, is a function of its absolute value. These measures exist for every face. For faces that are at least 64 pixels high, a face-mesh model (Kartynnik et al., 2019) supplies several hundred further landmarks, from which the pipeline estimates head angles in degrees and geometric measures such as smile width. Such faces make up 29 percent of the seconds in which a judge is seen listening to a party.

5.3 Camera Coverage

Because the camera follows the speaker, much of what happens on the bench never reaches the screen. The right-hand panel of Figure 3 shows how seldom judges appear while the parties speak. At least one verified judge is visible in 75 percent of the seconds in which a member of the bench is

speaking, against 1.4 percent of the seconds in which the applicant's side speaks and 1.4 percent of those in which the government speaks. Over the whole corpus, this leaves 33,162 verified seconds of a judge listening to a party.

The listening seconds are also unevenly spread across the bench. The President alone accounts for 40 percent of them, the first three judges in the order of precedence for 89 percent, and national judges for 8 percent. Our statements about listening are therefore mostly statements about presidents and the most senior judges, and about the early years, since 59 percent of listening seconds come from 2007 to 2012. Figure 6 illustrates the pattern for one hearing. During the parties' submissions the camera stays on the lectern, and the display measures arrive in short bursts.

5.4 Estimation

The basic unit of analysis is the judge-second, a measurement frame in which a verified judge's face is scored and the judge is not the person speaking. Averaging the seconds in which one party is speaking gives a cell, the mean of each outcome for judge i in hearing h while side s speaks. We define the listening differential as $d_{ih} = \bar{y}_{ih,\text{gov}} - \bar{y}_{ih,\text{app}}$ and compute it for judge-hearings with at least five seconds on each side. There are 450 such judge-hearings, from 206 hearings and 65 judges. For the analyses that do not distinguish between the parties, we average each judge's non-speaking seconds over the whole hearing and keep judge-hearings with at least ten seconds (1,798).

The main comparisons are made within judge, because the classifier gives different faces very different resting levels (Figure 7). Ranking judges by their raw levels would confuse how they behave with how the model reads their faces, and the comparisons between judges that we report for Hypotheses 11 to 16 must be read with this in mind.

For the comparison between the parties, we regress cell means on an indicator for the government speaking, with a fixed effect for each judge in each hearing and weights equal to the number of seconds:

$$\bar{y}_{ihs} = \beta \mathbb{1}[s = \text{gov}] + \alpha_{ih} + \varepsilon_{ihs}. \tag{1}$$

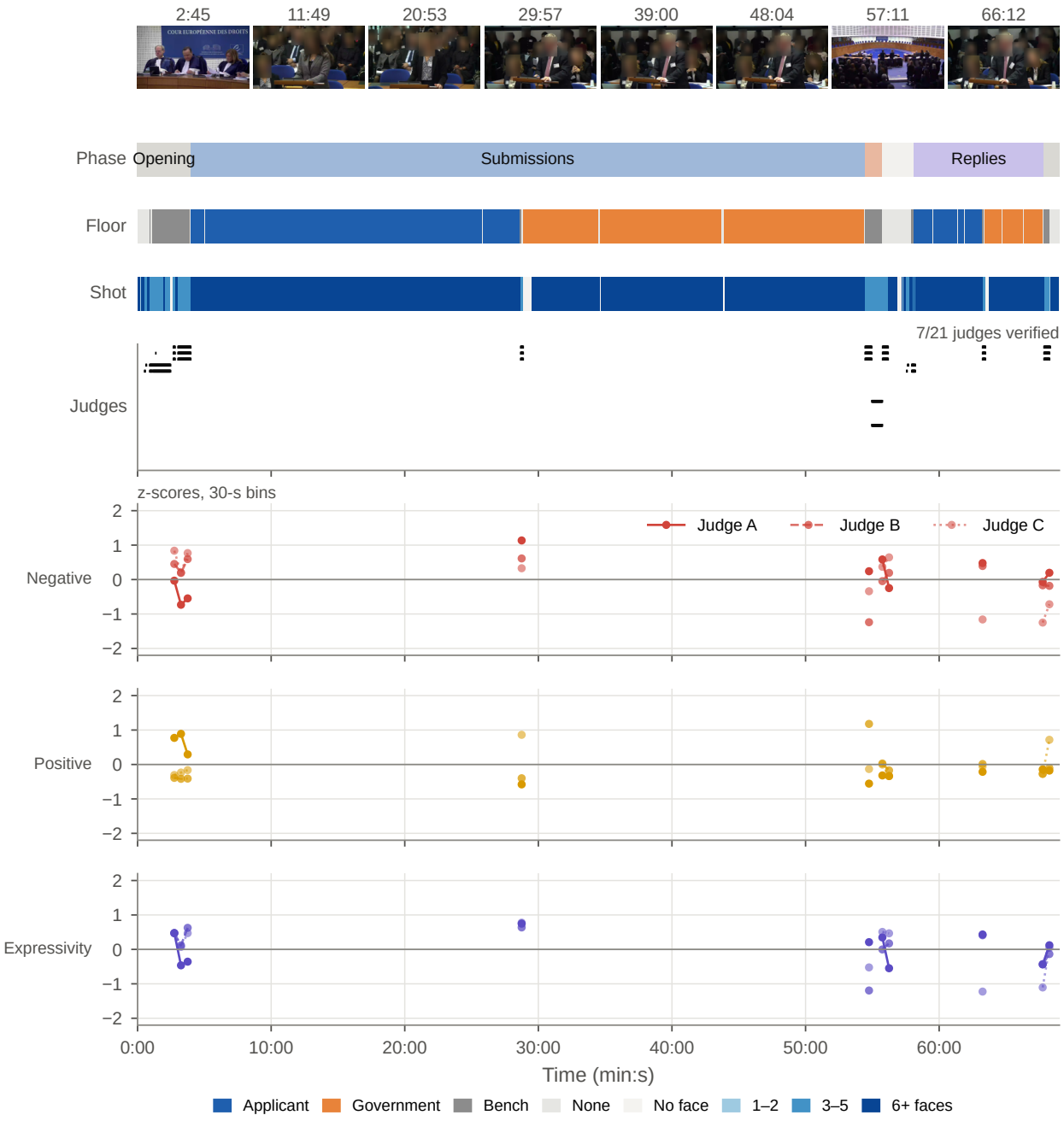

Figure 6: One hearing in detail (Grand Chamber, 2016). Note: Non-judge faces blurred. Frames © ECHR-CEDH.

The fixed effect absorbs whatever is constant for a judge within a hearing, such as her seat and the quality of the recording. For the main tests of Hypotheses 1 to 5, we regress the listening differential on the affinity variable A_{ih} with hearing fixed effects, so that each judge is compared with colleagues who sat on the same bench and heard the same advocates:

$$d_{ih} = \gamma A_{ih} + \delta_h + \varepsilon_{ih}. \tag{2}$$

Judge-hearings are weighted by their number of listening seconds. Standard errors are clustered by judge and by hearing. For the test of national bias we add a wild cluster bootstrap by hearing (Cameron et al., 2008) and a permutation test that reshuffles the national-judge label within hearings.

The exploratory tests that condition on the ruling regress the judge-hearing means on the affinity variable, an indicator that the judge's vote favored the respondent state, and their interaction, with judge and hearing fixed effects. The interaction measures whether a judge with the tie in question looks more positive, relative to her own average and to her colleagues in the same hearing, when her ruling favors the state.

The remaining analyses use standard tools, which Appendix C describes: crossed random-effects models for the variance decomposition, judge fixed effects for subject matter, and linear probability models with hearing fixed effects for votes. We control the false discovery rate within each group of main tests (Benjamini and Hochberg, 1995) and report unadjusted p-values for the other tests.

The comparisons between judges (Hypotheses 11 to 16) use every verified judge-second. We regress the outcome, or the probability that the classifier assigns to a single emotion, on a characteristic of the judge, with hearing fixed effects and controls for the size and geometry of the face, so that judges are compared with colleagues in the same recording at the same image quality. For Hypothesis 12 we take each judge's dominant emotion, the class other than neutral with the highest mean probability, and compare it between her odd-numbered and her even-numbered hearings.

Appendix F gives the details.

5.5 Robustness Checks

We present a series of robustness checks as specification curves (Simonsohn et al., 2020). These vary the minimum number of seconds per side, the guard band around changes of speaker, the minimum face size, the frames and probabilities used, the identity rule, the era, and the mode of hearing (Appendix C).

Additional exploratory analyses examine whether differences between the parties reflect how judges' faces are presented to the camera. We account for facial geometry and image characteristics in three ways (Appendix E). First, we re-estimate the difference between the parties using judge-seconds, introducing the frontal score and then the full set of controls: signed sideways turn, vertical nose position, head roll, face position, brightness, contrast, face size, and shot type. Second, we compare only seconds in which the same judge's head is turned by approximately the same amount in the same direction. Third, we estimate the association between each outcome and these image characteristics using the 202,961 judge-seconds in which neither party is speaking. We subtract the predicted contribution of the image characteristics from each observation and reconstruct the cells and listening differentials from the adjusted values.

6 Results

6.1 Emotional Expressions of Judges

The first research question asks whether judges exhibit discernible emotional reactions and how far these differ from one judge to another. Figure 7 shows that the differences between judges are large. Judges' mean expressivity runs from about 0.5 to above 0.9, and their mean negative display from about 0.4 to 0.9. In a crossed random-effects decomposition of judge-hearing means, judges account for 48 percent of the variance in negative display and 48 percent of the variance

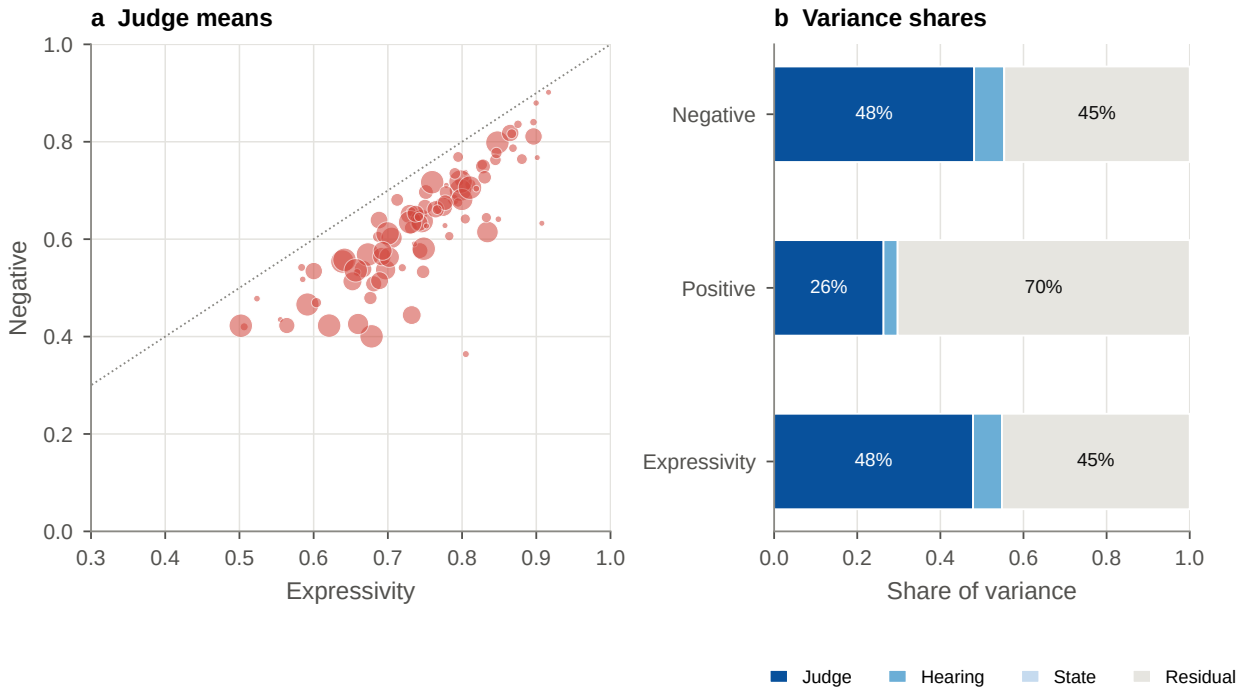

Figure 7: Variation across judges. Note: (a) Judge means; marker area = seconds. (b) Variance shares.

in expressivity, whereas hearings and respondent states together account for 7 and 7 percent. The judge's share remains the larger one after the image adjustment described below, as we expected (Appendix D). Hypothesis 7 is supported.

These differences are stable over time. Among the 90 judges who sat in at least four hearings, a judge's mean in one half of her hearings correlates with her mean in the other half at $r = 0.83$ for negative display and $r = 0.85$ for expressivity. Stability does not tell us, however, whether these are habits of behavior or habits of the model in reading particular faces. The difference between women and men that we report below is one reason to suspect that part of it is the model.

Within a hearing, expressions change very little, and hardly more during questions and replies than during the prepared submissions (Appendix D). Contrary to our expectation, presiding judges are more expressive than their colleagues in the same hearing, by 4.3 points (95% CI 1.6 to 6.9). When we compare the same judge when presiding and when not presiding, however, the difference shrinks to 0.9 points ($p = 0.44$), so the contrast seems to reflect who becomes President more than what presiding does to a face.

6.2 Women and Men Judges

Hypothesis 11 predicts that women judges display less negative emotion than men. Raw levels cannot be compared across judges, because recordings differ in resolution and judges differ in how the camera shows them. We therefore compare women and men who appear in the same hearing, and we hold constant the size of the face and its geometry: how nearly it is turned toward the camera, the sideways turn, the vertical position of the nose, and the tilt of the head. These measures come from the five landmarks that the pipeline places on every face, so the comparison can use every verified judge-second for which the judge's gender is documented. There are 372,170 such seconds, of 37 women and 75 men in 308 hearings, 243 of which show both women and men. Women's negative display is 11.4 points lower than men's (95% CI -15.9 to -6.9; $p < .001$), and their expressivity is 6.3 points lower (95% CI -9.7 to -2.9; $p < .001$). Their positive display is 0.92 points higher (95% CI 0.22 to 1.62; $p = 0.010$), a small difference in the direction that research on smiling would predict (LaFrance et al., 2003).

The result does not depend on these choices. A second version of the comparison uses head angles that exist only for faces of at least 64 pixels, and so covers 211,944 seconds of 34 women and 68 men. It gives a difference of 11.2 points in negative display (95% CI -16.9 to -5.4), 6.7 points in expressivity, and 0.8 points in positive display ($p = 0.070$). With one mean per judge and hearing, women's negative display is 13.7 points lower than that of men on the same bench, and on image-adjusted outcomes the difference is 13.8 points (Appendix D).

Hypothesis 11 is supported statistically, with two reservations. The test is exploratory, because we formulated the hypothesis after the results were known. It also cannot separate behavior from measurement. Expression classifiers can perform unevenly across gender, age, and race (Xu et al., 2020), and a model that reads women's faces at rest as more neutral than men's would produce the same result. We return to this point in Section 6.10.

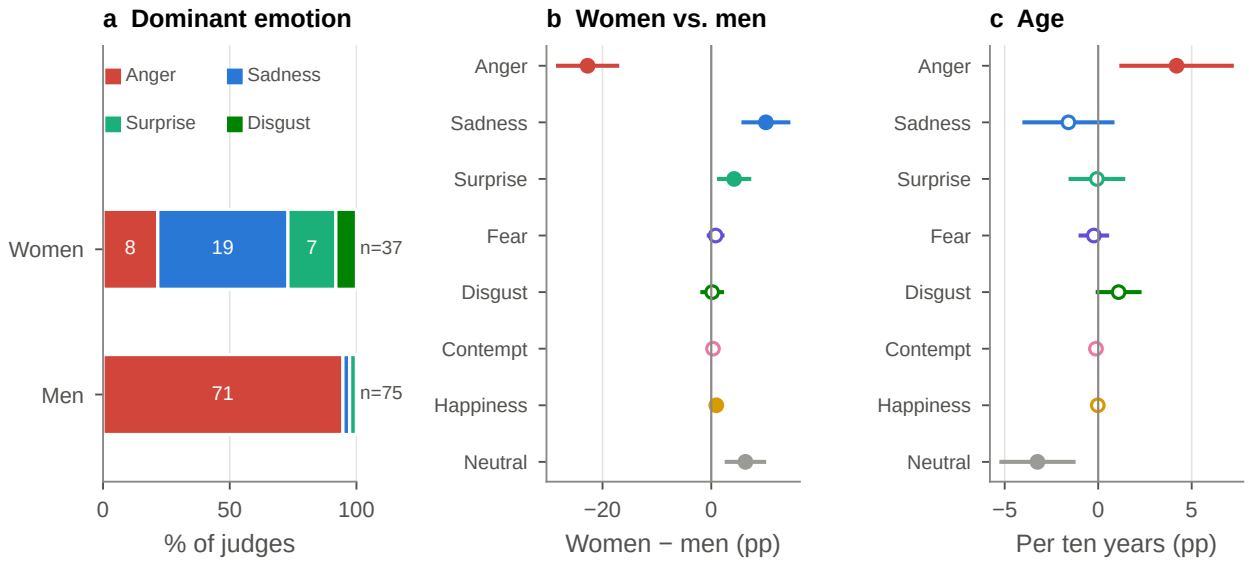

Figure 8: Whose emotions. (a) Most frequent non-neutral reading. (b), (c) Same hearing, same image quality; 95% CIs; filled: $p < .05$.

6.3 Whose Emotions Dominate

Hypotheses 12 to 16 compare the judges with one another. They use every verified judge-second, 370,876 seconds of 113 judges, and like the comparison of women and men they are made within hearings and at the same image quality.

For most judges the classifier's readings are dominated by one emotion, and it is the judge's own (Figure 8). Leaving the neutral class aside, anger is the most frequent reading for 79 of the 113 judges, sadness for 22, surprise for 9, and disgust for 3. Among the 106 judges who appear in at least two hearings, the dominant emotion in the odd-numbered hearings is also the dominant emotion in the even-numbered ones for 88 percent, against 57 percent if the two were unrelated ($q < .001$). The shares of the single emotions are stable as well, with a correlation between the two halves of .88 for anger and .79 for sadness. Hypothesis 12 is supported.

Whose emotion is which depends strongly on gender (Table 3). In the same hearing and at the same image quality, the probability of anger is 22.7 points lower for women than for men (95% CI -28.2 to -17.3), that of sadness is 10.0 points higher (95% CI 5.9 to 14.2), and that of happiness is 0.92 points higher ($q = 0.027$). Fear does not differ ($q = 0.37$). Hypothesis 13 is supported for three of the four emotions. The lower negative display of women (Hypothesis 11) is therefore a

matter of anger alone, since women score higher than men on sadness.

Table 3: Single emotions by gender and age of the judge

	Mean (%)		Women − men		Per ten years of age		Split-half
	Women	Men	Estimate	p	Estimate	p	r
Anger	16.3	40.5	−22.7 (2.7)	<.001	4.2 (1.5)	.006	0.88
Sadness	24.4	14.9	10.0 (2.1)	<.001	−1.6 (1.2)	.183	0.79
Surprise	14.5	9.1	4.2 (1.4)	.003	−0.1 (0.7)	.923	0.73
Fear	5.9	4.7	0.8 (0.6)	.210	−0.2 (0.4)	.516	0.62
Disgust	3.1	2.8	0.2 (0.9)	.857	1.1 (0.6)	.056	0.77
Contempt	2.5	2.0	0.3 (0.3)	.286	−0.1 (0.1)	.398	0.71
Happiness	2.9	1.9	0.9 (0.4)	.011	0.0 (0.1)	.875	0.69
Neutral	30.5	24.1	6.3 (1.7)	<.001	−3.2 (1.0)	.001	–

Note: Mean probability that the classifier assigns to each class, in percent, over all verified judge-seconds. Differences in percentage points with standard errors in parentheses, same hearing and same image quality; the age model also holds gender constant. Split-half r: correlation across judges between the class share in the odd and in the even hearings of a judge, net of hearing and image quality.

Older judges score higher on negative display, by 3.3 points for every ten years of age (95% CI 0.8 to 5.9; $q = 0.027$). The difference lies in anger (4.2 points, $q = 0.022$) and not in sadness ($p = 0.18$). This is what research on the reading of older faces predicts, and the opposite of what research on emotional aging predicts. The estimate is smaller and not significant when only the seconds in which a judge is silent are used, or when judges are compared by their hearing means (Appendix F). We therefore regard Hypothesis 14 as supported in the main specification only.

The remaining expectations are not borne out. Judges from post-socialist states do not display less positive emotion (0.04 points, $p = 0.88$), judges from Southern European states are not more expressive (−2.2 points, $p = 0.22$), and former career judges are not more restrained (−2.0 points, $p = 0.28$), so Hypothesis 15 is not supported. Neither is Hypothesis 16. A judge who asks a question is no more expressive while she listens than her silent colleagues (0.1 points, $p = 0.92$) or than she is herself in hearings in which she asks nothing ($p = 0.44$). Across judges, expressivity is unrelated to how often a judge asks ($r = .12$, $p = 0.45$) and to how much the camera shows her ($r = .02$, $p = 0.90$).

These comparisons share the reservation that applies to Hypothesis 11. They compare faces with one another and not a face with itself, and a classifier that reads the resting faces of men and of older judges as angry would produce the same pattern (Xu et al., 2020; Hess et al., 2012). For

this reason we do not rank individual judges by their expressions.

Who takes the floor and whom the camera shows can be established more firmly, and both follow a judge's role (Appendix F). National judges put a question in 54 percent of their hearings and other judges in 23 percent, a difference of 31 points within the same hearing (95% CI 23 to 40) that remains when the same judge is compared across hearings (28 points). Women ask as often as men (-2.5 points, $p = 0.39$). The camera gives the presiding judge 5.6 percent of a hearing, the two judges beside her 3.6 percent each, and the others 0.7 percent.

6.4 Applicants and Governments

Within the same judge and hearing, the classifier's output differs with the party that is speaking. While the government speaks, negative display is 3.5 points higher (95% CI 2.2 to 4.8), positive display is 0.29 points lower, and expressivity is 2.5 points higher than while the applicant's side speaks. All three differences pass the false-discovery correction ($q < .005$), and the difference in negative display appears in every specification of the main test (Section 6.7). Additional checks agree with this result: the difference is 3.8 points in hearings in which the government spoke first and 3.3 points in the others, and it is 3.7 points when we keep only speaker attributions marked with medium or high confidence. By the usual standards, this is a robust finding, and Hypothesis 8 is supported in the main tests.

6.5 Head Orientation

The difference between the parties is nevertheless, for the most part, a difference in the direction in which the judges were facing. Panels (a) and (b) of Figure 9 show the three judges from the hearing described in the Introduction, once during the applicant's turn and once during the government's turn. In the applicant's turn their heads are turned further away from the camera.

The landmarks allow us to measure this for every face. We use the horizontal offset of the nose from the midpoint between the eyes, in units of the distance between the eyes. It is zero for a face that is square to the camera. Among the seconds in which a judge is seen listening to a

party, 92 percent have a negative offset, which means that the camera is nearly always on the same side of the judges' line of sight. The offset averages -0.27 in the applicant's turn and -0.17 in the government's turn (panel c). Within judge and hearing the difference is 0.10 (95% CI 0.08 to 0.12), or 0.6 standard deviations, and it has the same sign in 84 percent of the 450 judge-hearings in the main sample. In short, judges hold their heads in two different directions for the two parties, as people do when they look at whoever is speaking, and one of those directions was closer to the direction of the camera.

The classifier is sensitive to that direction. Panel (d) uses only the seconds in which neither party is speaking and compares each judge with herself within a hearing. Relative to a face that is square to the camera, negative display is about 9 points lower at an offset of -0.4 and 19 points lower for the most turned faces. The more a face is turned, the more often the classifier calls it neutral.

Together these two facts account for most of the difference between the parties (panel e). Holding the frontal score constant leaves 1.3 points ($p = 0.054$). The frontal score discards the direction of the turn, however, and a head turned one way need not look the same to the classifier as a head turned the other way. With the signed offset, the vertical position of the nose, and head roll added, 0.5 points remain (95% CI -0.4 to 1.5), and 0.6 points remain when the position, brightness, contrast, and size of the face and the type of shot are held constant as well. Comparing only seconds in which the head is turned by about the same amount leaves 1.9 points, and 0.9 points once nose height and roll are also held constant ($p = 0.061$). When the cells are rebuilt from adjusted seconds, the difference is 0.9 points (95% CI -0.4 to 2.2; sign-flip $p = 0.070$), and among nearly frontal faces it is -0.3 points (95% CI -2.1 to 1.5). The estimates that use the full set of landmarks thus put the remainder between 0.5 and 0.9 points, about a quarter of the main estimate or less, and none of them can be distinguished from zero. The difference in expressivity falls in the same way, from 2.5 points to 0.2.

The history of the recordings offers a check that does not depend on these controls (Appendix Figure A9). Judges turn their heads by about the same amount between the two parties' turns in

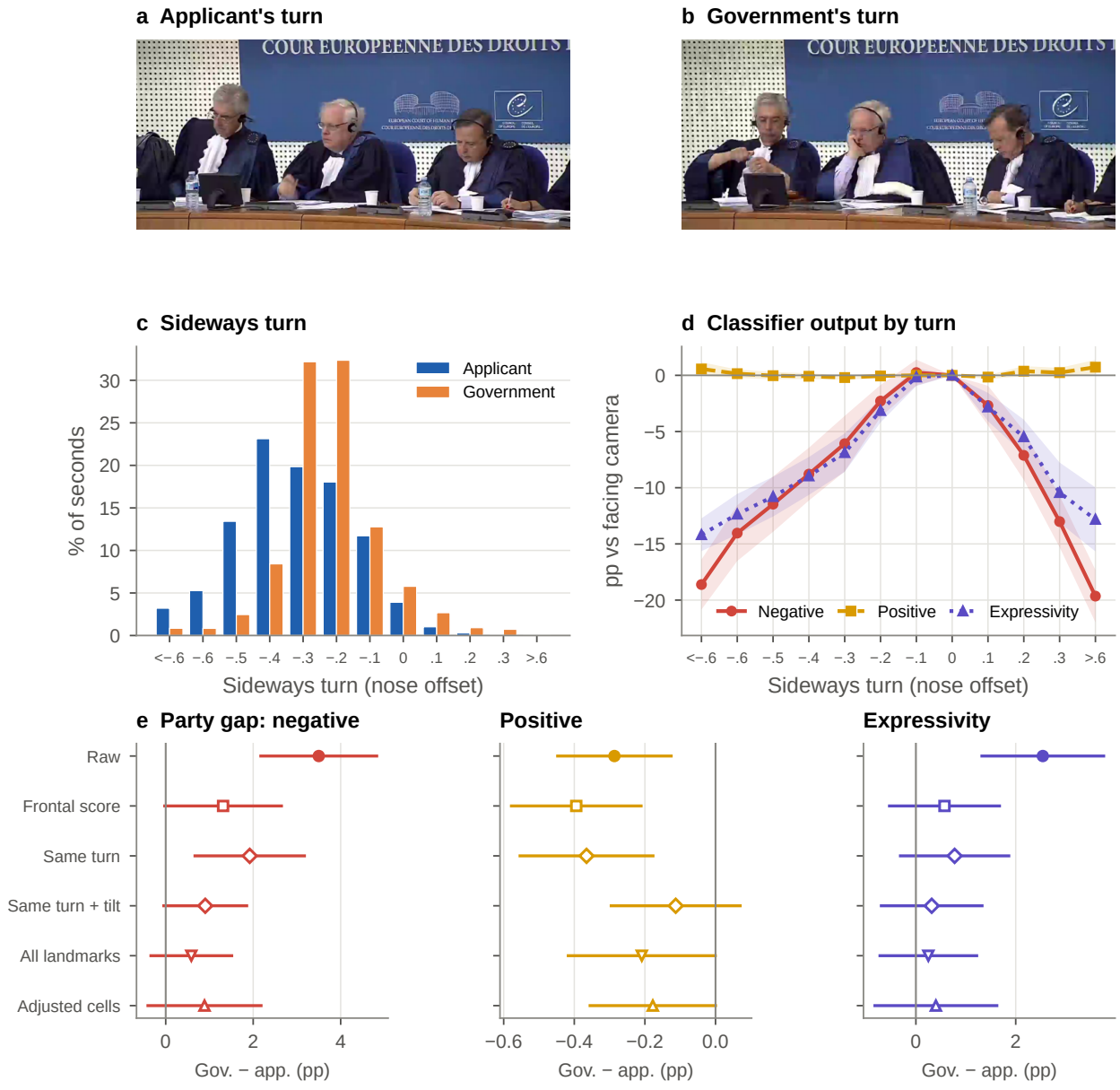

Figure 9: Head orientation and the difference between parties (exploratory). Note: (a, b) Same judges, one 2015 hearing. (c) Listening seconds. (d) Within judge-hearing, seconds with neither party speaking; bands = 95% CI. (e) Estimates with 95% CIs. Frames © ECHR-CEDH.

every period, but the position of the camera changes. Until 2016 the government's turn is the more frontal one. From 2020, heads turn to either side of a camera that now faces the bench, and neither turn is more frontal. The difference in negative display is 3.9 and 3.6 points in 2007–2012 and 2013–2016, 1.2 points in 2017–2019, and -0.5 points from 2020. The last two periods hold few of the listening seconds, so this evidence corroborates our account without proving it.

Three results qualify this account, and Appendix E reports them in detail. For presiding judges, the full set of landmark controls removes the difference (0.3 points, $p = 0.67$), but the adjusted cells leave 2.1 points ($p = 0.035$). Among faces that are large enough for geometric measurement, the eyes look more closed during the government's turn even with the landmark controls (0.19 standard deviations, $p < .001$), which may mean only that the judges are reading. The small difference in positive display is also more persistent (-0.21 points with the full set of controls, $p = 0.049$), although measured smile width does not differ between the turns ($p = 0.79$).

A head that is turned away from the camera is not a less negative expression. On this evidence, the difference between the turns in the classifier's output is mainly a difference in what the camera could see. Hypothesis 8 therefore holds for what the classifier reports, and we do not read it as evidence that judges respond more negatively to governments.

6.6 National, Regional, and Political Affinity

Hypotheses 1 to 5 predict that judges respond more favorably to governments with which they share a tie. The main tests, which compare the listening differentials of judges on the same bench, provide no support for them (Figure 10). For Hypothesis 1, national judges have a listening differential in negative display of 1.1 points relative to their colleagues (95% CI -3.4 to 5.6; wild bootstrap $p = 0.67$; permutation $p = 0.62$), -0.11 points in positive display, and 0.2 points in expressivity. The estimate for negative display is 1.2 points without weights and 1.4 points on adjusted outcomes. The estimates for shared EU membership (Hypothesis 2), for the same East/West and North/South groups (Hypotheses 3 and 4), and for the distance in democracy scores (Hypothesis 5) cannot be distinguished from zero either. The smallest q-value among these tests is 0.87,

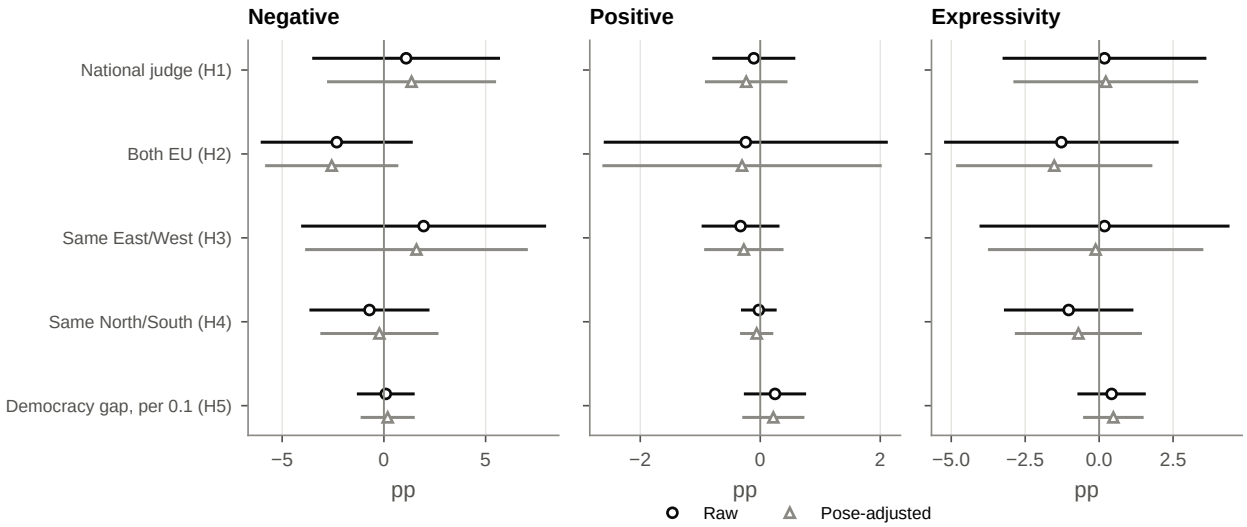

Figure 10: Affinity estimates. Hearing fixed effects; 95% CIs.

and the results are the same when national judges are controlled for or excluded.

These tests have little power. A national judge was seen listening to both parties in only 31 hearings, and in 26 of them alongside a colleague with whom she can be compared. The smallest difference that the design could detect with 80 percent power is about 6.4 points, more than half a standard deviation of the differential. A national bias of moderate size could therefore be present without our detecting it. Some specifications lean in one direction or the other without changing this conclusion (Appendix D). Only 7 national judges come from the period in which re-election was still possible, too few to test whether career incentives strengthen national bias. For Hypothesis 6, the proxy comparison of women and men judges in hearings that concern discrimination gives 9.1 points ($p = 0.32$).

The hypotheses are worded in terms of rulings, whereas the listening differential ignores how a judge went on to vote. We therefore added exploratory tests that condition on the vote (Appendix Table A13). They use each judge's mean display over all the seconds in which she is on screen and not speaking, which is available for many more judges than the listening differential: 1,400 judge-hearings with a coded vote, including 92 observations of national judges. For Hypothesis 1 the estimates point in the expected direction. Relative to their colleagues, national judges' positive display is 0.79 points higher (95% CI 0.00 to 1.59; $p = 0.051$) in the 39 instances in which they

voted for their own state than in the 53 instances in which they voted against it. This pattern is not robust, however. The estimate falls to 0.51 points after the image adjustment ($p = 0.18$) and to 0.46 points when judge-hearings are weighted equally ($p = 0.40$). For Hypothesis 2 the estimates run against the hypothesis, since judges from EU states look more negative, not more positive, when they vote for an EU respondent (2.9 points, $p = 0.028$). None of the 20 tests passes a false-discovery correction (smallest $q = 0.27$). We therefore regard the result for national judges as a pattern that deserves a test in other data and not as support for Hypothesis 1.

6.7 Robustness Checks

Figure 11 shows the difference between the parties and the national-judge estimate across 13 specifications of the usual kind and across the specifications that we added to account for head orientation. The first set consists of the main specification and twelve variants, each of which changes one choice. Three vary how long a judge must be seen during each party's turn, requiring at least 1, 10, or 20 seconds instead of 5. Two vary the guard band, the seconds dropped around each change of speaker, from 2 to 0 or 5. Two keep only the better images, either faces at least 64 pixels high or recordings in high definition. One leaves out the hearings held by video during the COVID-19 pandemic. Two change how the classifier's output is used, keeping only the seconds in which it is confident or averaging its readings over neighboring seconds. The last two change the identity rule, keeping only faces matched to a judge with high confidence or also those matched with medium confidence. Across this first set, the difference in negative display ranges from 2.6 to 4.5 points and is distinguishable from zero every time.

The second set accounts for the direction in which the judges are looking. These specifications control for how nearly the face points at the camera, compare only seconds in which the judge's head is turned to the same degree, additionally hold the tilt of the head constant, control for the positions of the eyes, nose, and mouth, remove the effect of pose and image quality from each second before averaging, keep only nearly frontal faces, or use only the recordings since 2020, which were made with a different camera arrangement. In all of them the difference falls below

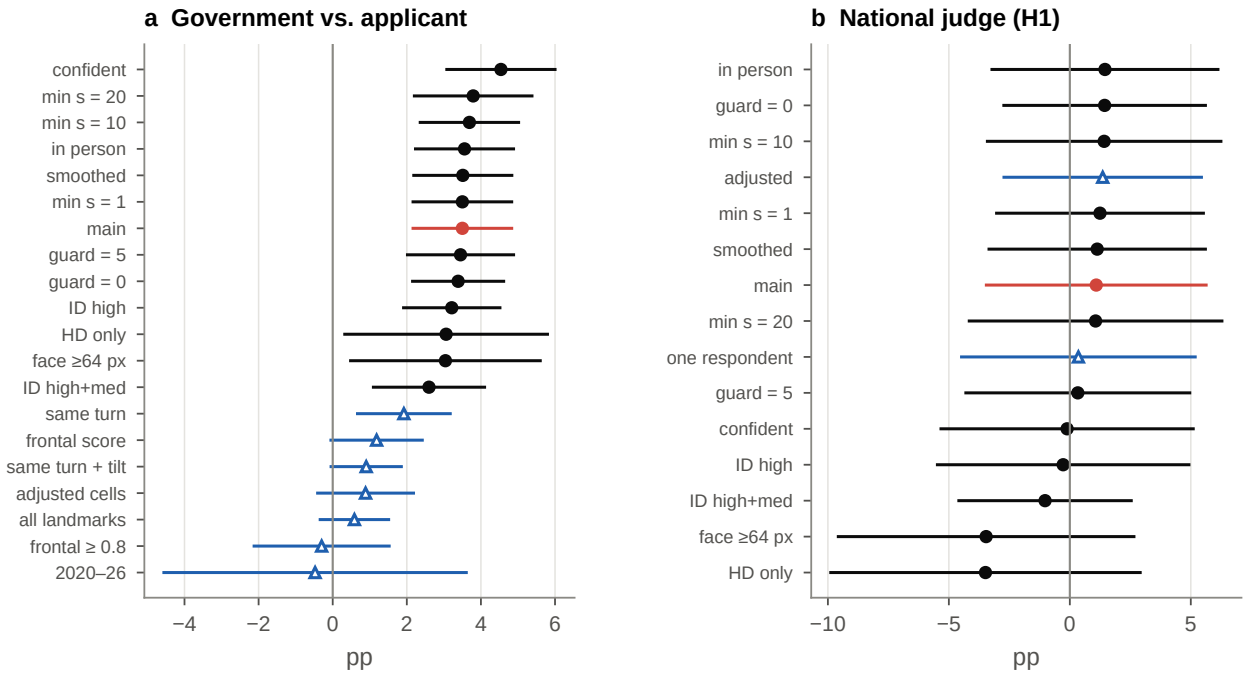

Figure 11: Specification curves. Note: Negative display. Red: main; black: variants; blue: added.

2 points, and in most it cannot be distinguished from zero. The national-judge estimate ranges from -3.5 to 1.5 points across the first set and is never significant for negative display (smallest $p = 0.26$). Appendix Table A11 lists the coefficients of the main tests and Appendix Table A12 the exploratory ones.

6.8 Subject Matter of Cases

We find little evidence that expressions vary with the subject matter of a case. Hearings under Articles 2 to 4 do not produce more negative display than other hearings for the same judge (0.4 points, $p = 0.61$), so Hypothesis 9 is not supported. Inter-state hearings show slightly lower positive display (-0.67 points, $p = 0.005$), but there are only 9 such hearings.

6.9 Facial Expressions and Votes

Votes seldom split within the benches that we observe. Of the 181 hearings with codable votes in the main sample, 22 have sampled judges who voted differently from one another, with 55 votes in

all. The main test regresses a vote for the applicant on the listening differential in negative display, and separately in positive display, with hearing fixed effects. Neither differential is related to the vote (-0.8 percentage points per 10-point difference in negative display, $q = 0.54$; for positive display $q = 0.38$). Hypothesis 10 is not supported. The conclusion is the same when we use every judge-hearing in which the judge was seen during both parties' turns, however briefly (477 judge-hearings in 204 hearings; -0.1 points, $p = 0.91$), and when we relate the vote to a judge's mean negative display over the whole hearing, which is available for 1,400 judge-hearings ($p = 0.22$).

The same model with the differential in expressivity, which was not one of our main measures, gives -4.1 points per 10-point difference (95% CI -6.8 to -1.3; $p = 0.005$; wild bootstrap $p = 0.004$). The estimate changes little on adjusted outcomes (-4.3), without national judges (-4.1), or when any one hearing is left out (-4.3 to -3.5; Appendix D). It is smaller in the 477 judge-hearings that include every judge seen during both turns, among them 33 divided benches (-2.6 points, 95% CI -5.1 to -0.2), and a judge's mean expressivity over the whole hearing is not related to her vote ($p = 0.12$). This result should therefore be read with caution. Unanimous benches say nothing about how votes divide, and among the 55 votes in divided benches alone the slope is -43.5 points per 10, which is far too large to be taken literally from 22 benches. The direction is also the reverse of the pattern in vocal pitch at the US Supreme Court, where justices showed more arousal toward the side that they voted against (Dietrich et al., 2019a). We record it as an observation for other data to test. Judges' expressions are related to their votes in this one measure, but not in the way that Hypothesis 10 predicts.

Expressivity does not distinguish judges who dissent or write separately (smallest $p = 0.32$), and display measures do not help to predict votes. Neither the models with covariates only nor the models that add display measures do better than chance when whole hearings are held out (AUC 0.47 and 0.46 for penalized logit, 0.55 and 0.47 for boosted trees; Figure 12).

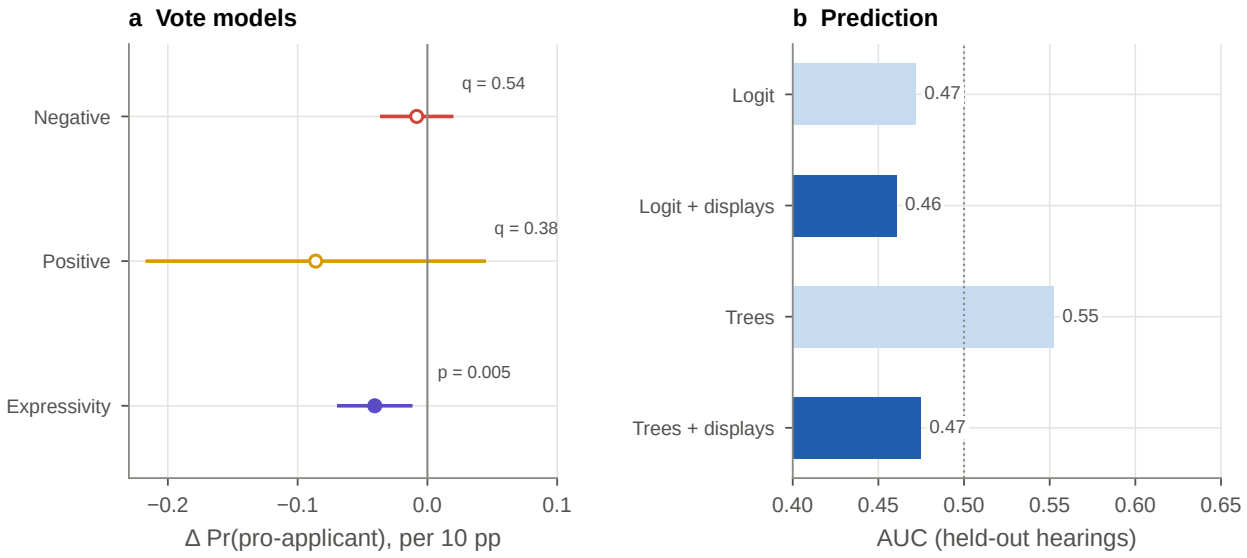

Figure 12: Displays and votes. Note: (a) Hearing fixed effects, all benches; expressivity is exploratory.

6.10 Measurement Error

Other properties of the measurement are relevant to all of the results above (Figure 13). Larger faces look more expressive in the raw data, but not when faces of the same judge are compared at the same resolution and head angles (0.1 points, $p = 0.95$). At the same image quality and within the same hearing, the classifier scores women judges 11.2 points lower on negative display (95% CI -16.9 to -5.4) and 6.7 points lower on expressivity than men. We cannot say whether this reflects how women judges behave or how the model reads their faces. It is the kind of difference that would be reported as a gender effect if levels were compared across judges. The camera is selective as well. While the parties speak, the President is on screen more often than other judges (0.9 points, $p < .001$), whereas national judges are not (0.03 points, $p = 0.53$).

6.11 Summary of Results

Table 4 summarizes the results for the sixteen hypotheses. None of Hypotheses 1 to 6 is supported by the main tests, and the one pattern in the expected direction, for national judges conditional on their ruling, is not robust. Of the other ten hypotheses, six are supported statistically. The stability of individual differences is borne out, both in the amount of emotion (Hypothesis 7) and

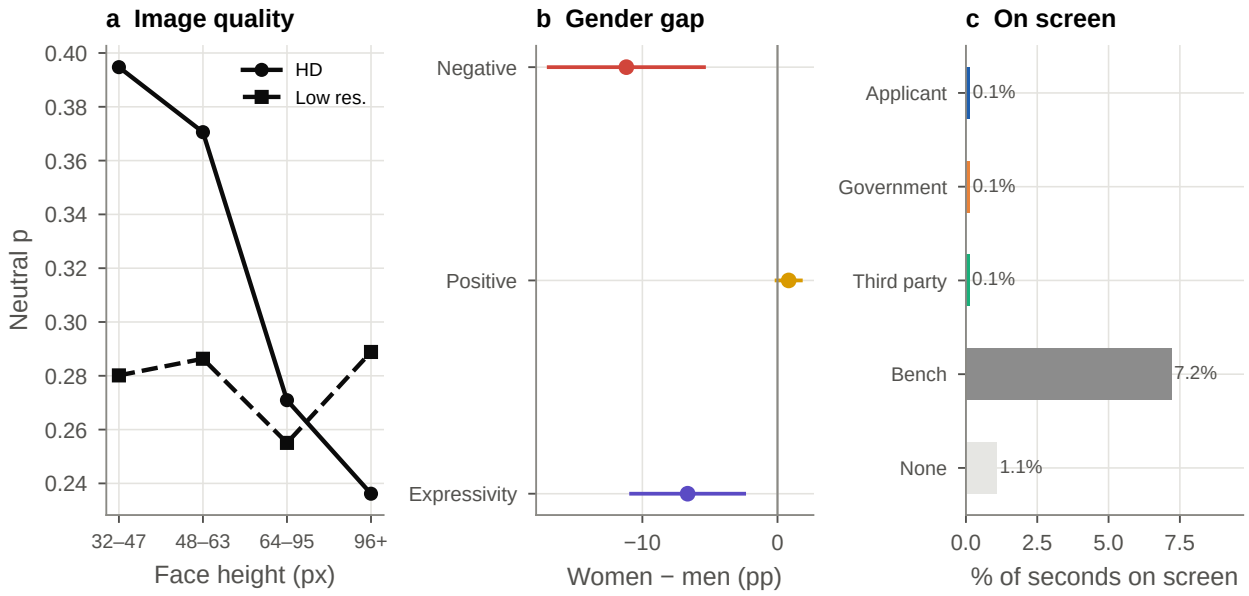

Figure 13: Sources of measurement error.

in its kind (Hypothesis 12). Expressions differ with the party that is speaking in every main test (Hypothesis 8), although most of that difference reflects head orientation. Women judges score lower on negative display than men (Hypothesis 11), because they score much lower on anger, and they score higher on sadness and happiness (Hypothesis 13). Older judges score higher on negative display in the main specification (Hypothesis 14). The last three results come from exploratory comparisons between judges, which cannot separate behavior from measurement. Hypotheses 9, 10, 15, and 16 are not supported. The tests of Hypotheses 7 to 16 use all the observations that each design allows: 1,798 judge-hearings of 113 judges for Hypotheses 7 and 9, the 587 judge-hearings in which a judge is seen during both parties' turns for Hypothesis 8, the 477 of these that have a coded vote for Hypothesis 10, 372,170 judge-seconds of 112 judges for Hypothesis 11, and the 370,876 verified judge-seconds of 113 judges for Hypotheses 12 to 16.

7 Discussion

We set out to learn whether the faces of judges at the ECHR reveal emotional reactions that are related to their national, regional, and political affinities and to their votes. With these recordings

Table 4: Summary of results

	Main estimate	Conclusion
H1 National Bias	1.1 (−3.4 to 5.6)	Not supported. By ruling: 0.79 points of positive display ($p = 0.051$), not robust
H2 EU Bias	−2.3 (−6.0 to 1.3)	Not supported. By ruling: opposite sign
H3 Eastern vs. Western Europe	2.0 (−4.0 to 7.9)	Not supported
H4 Northern vs. Southern Europe	−0.7 (−3.6 to 2.2)	Not supported
H5 Democratic Similarity	0.1 per 0.1 of distance (−1.3 to 1.4)	Not supported
H6 Judge Characteristics	9.1 ($p = 0.32$), proxy	Not supported; no direct test
H7 Individual Differences	Judges 48% of variance, hearings and states 7%; split-half $r = 0.83$	Supported. Presiding judges are more expressive, not less
H8 Parties	3.5 (2.2 to 4.8); 0.5 to 0.9 with head-orientation controls	Supported in the main tests. Difference found, largely explained by head orientation
H9 Subject Matter	0.4 ($p = 0.61$)	Not supported
H10 Votes	−0.8 per 10 points ($q = 0.54$); expressivity −4.1 (−6.8 to −1.3), exploratory, and −2.6 with all judge-hearings	Not supported; displays do not improve prediction. Exploratory: expressivity is related to votes, in the opposite direction
H11 Gender of the Judge	−11.4 (−15.9 to −6.9), exploratory	Supported statistically. Cannot be separated from how the classifier reads women's faces
H12 Emotional Profiles	Same dominant emotion in odd and even hearings for 88% of judges (chance 57%)	Supported. Anger dominates for 79 of 113 judges
H13 Gender and Specific Emotions	Women − men: anger −22.7, sadness 10.0, happiness 0.92, fear 0.80	Supported for anger, sadness, and happiness. Cannot be separated from how the classifier reads faces
H14 Age	3.3 per ten years (0.8 to 5.9); anger 4.2	Supported in the main specification; weaker in two checks
H15 Culture and Career	Post-socialist 0.04 (positive display); South −2.2, career judges −2.0 (expressivity)	Not supported
H16 Engagement	Judges who ask 0.1 (expressivity); across judges $r = .12$	Not supported

Note: Negative display, percentage points; 95% confidence intervals in parentheses. For H1 to H6, difference in the listening differential between judges with and without the tie on the same bench; a negative value favors the government. For H11 to H16, differences between judges in the same hearing at the same image quality.

and this instrument, the answer is largely negative, and the one pattern that at first looked like an answer turned out to describe the position of the camera.

On Hypotheses 1 to 5 the evidence consists of imprecise zeros. This is compatible with the broadly reassuring reading of the Court's votes in Voeten (2008), and equally with the national favoritism that he did find, because our design could detect only a large difference. The constraint is the camera: a national judge is on screen during both parties' turns in only 31 hearings.

The strongest results concern the judges as individuals. Judges differ in stable ways (Hypothesis 7), as research on the management of emotion by judges and on expressivity as a personal trait would lead one to expect (Bergman Blix and Wettergren, 2018; Gross and John, 1997), and the differences extend to the kind of emotion that dominates (Hypothesis 12). The differences between women and men (Hypotheses 11 and 13) and between older and younger judges (Hypothesis 14) have the direction that research on gender, age, and emotion predicts (Brescoll and Uhlmann, 2008; Boussalis et al., 2021; Hess et al., 2012). They are exploratory, however, and a model that reads anger into the resting faces of men and of older judges would produce the same results (Xu et al., 2020). Only human ratings of the same faces could separate the two readings.

The difference between the parties (Hypothesis 8) resembles the unequal treatment of the two sides at the US Supreme Court (Johnson et al., 2009; Black et al., 2011), but it matters mainly for what it shows about method. The estimate was significant after correction for multiple testing and stable across 13 specifications, and none of this protected us against a confound that we had not anticipated. A specification curve built from variations that share a flaw reproduces it each time. What exposed the problem was a descriptive question, namely whether the faces compared in the two conditions look alike in ways that are unrelated to expression. We therefore suggest that researchers treat the geometry of the face, its size, and the type of shot as covariates from the start and report estimates with and without them. A few patterns survive our checks, among them a relation between expressivity and votes in the 22 benches that divided, and they should be observed in other data before anything is built on them.

The psychological literature warns that faces do not map onto feelings (Barrett et al., 2019).

Our data add that classifier output does not map cleanly onto faces either. The European restrictions on emotion recognition, which have applied since February 2, 2025, single out workplaces and schools (European Parliament and Council of the European Union, 2024), and our results support the caution behind them. Used without identity checks, within-person comparisons, and an analysis of head geometry, the same software would have reported, with a tight confidence interval, that judges scowl at governments.

8 Conclusion

This study combines the webcasts of ECHR hearings with computational measures of facial expression, and so adapts to a multinational court an approach that was developed for oral argument at the US Supreme Court.

The results are more cautious than the hypotheses with which we began. Of our sixteen hypotheses, six are supported statistically. Judges' faces differ in stable ways, in the amount of emotion and in its kind. The apparent difference between applicants and governments is mostly a matter of where the judges were looking, and the differences by gender and age may owe as much to the classifier as to the judges. We detect neither favoritism toward particular governments nor a link between expressions and votes. Our evidence cannot say whether the judges in Strasbourg feel as little as Hobbes would wish or as much as Justice Sotomayor reports. It does show that their faces, as the webcasts record them, reveal very little of it.

The project also yields a dataset that links hearing videos to individual judges, their expressions, the speaker in each second, and the judgment that followed. A first look at the judges' questions shows where it could lead: national judges put questions far more often than their colleagues, and women as often as men. Future work could examine the content of these questions, interruptions, and the judges' voices. Courtroom videos contain information that written decisions do not, and it is there, rather than in the judgments themselves, that judges' personal feelings may be found.

References

Ash, E. and Caliskan, C. (2022). Fight or flight on fox? partisan fear responses on u.s. cable news shows. *SSRN Electronic Journal*.

Bain, M., Huh, J., Han, T., and Zisserman, A. (2023). WhisperX: Time-accurate speech transcription of long-form audio. In *Proceedings of Interspeech 2023*, pages 4489–4493.

Bandes, S. A. and Blumenthal, J. A. (2012). Emotion and the law. *Annual Review of Law and Social Science*, 8:161–181.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., and Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, 20(1):1–68.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300.

Bergman Blix, S. and Wettergren, Å. (2018). *Professional Emotions in Court: A Sociological Perspective*. Routledge.

Black, R. C., Treul, S. A., Johnson, T. R., and Goldman, J. (2011). Emotions, oral arguments, and Supreme Court decision making. *The Journal of Politics*, 73(2):572–581.

Boussalis, C. and Coan, T. G. (2021). Facing the electorate: Computational approaches to the study of nonverbal communication and voter impression formation. *Political Communication*, 38(1-2):75–97.

Boussalis, C., Coan, T. G., and Holman, M. R. (2022). Mad, sad, but mostly glad: How men and women in politics communicate using emotions in images. *SocArXiv*.

Boussalis, C., Coan, T. G., Holman, M. R., and Müller, S. (2021). Gender, candidate emotional expression, and voter reactions during televised debates. *American Political Science Review*, 115(4):1242–1257.

Boyd, C. L., Epstein, L., and Martin, A. D. (2010). Untangling the causal effects of sex on judging. *American Journal of Political Science*, 54(2):389–411.

Brescoll, V. L. and Uhlmann, E. L. (2008). Can an angry woman get ahead? Status conferral, gender, and expression of emotion in the workplace. *Psychological Science*, 19(3):268–275.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427.

Chamberlain, J. and Miller, M. K. (2009). Evidence of secondary traumatic stress, safety concerns, and burnout among a homogeneous group of judges in a single jurisdiction. *Journal of the American Academy of Psychiatry and the Law*, 37(2):214–224.

Deng, J., Guo, J., Xue, N., and Zafeiriou, S. (2019). ArcFace: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4690–4699.

Dietrich, B. J. (2021). Using motion detection to measure social polarization in the U.S. House of Representatives. *Political Analysis*, 29(2):250–259.

Dietrich, B. J., Enos, R. D., and Sen, M. (2019a). Emotional arousal predicts voting on the U.S. Supreme Court. *Political Analysis*, 27(2):237–243.

Dietrich, B. J., Hayes, M., and O'Brien, D. Z. (2019b). Pitch perfect: Vocal pitch and the emotional intensity of Congressional speech. *American Political Science Review*, 113(4):941–962.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3–4):169–200.

Ekman, P., O'Sullivan, M., Friesen, W. V., and Scherer, K. R. (1991). Face, voice, and body in detecting deceit. *Journal of Nonverbal Behavior*, 15(2):125–135.

Epstein, L., Landes, W. M., and Posner, R. A. (2010). Inferring the winning party in the Supreme Court from the pattern of questioning at oral argument. *The Journal of Legal Studies*, 39(2):433–467.

European Court of Human Rights (2024). Composition of the court. `https://www.echr.coe.int/composition-of-the-court`. Accessed: 2026-08-16.

European Court of Human Rights (2025). Rules of court. `https://www.echr.coe.int/documents/d/echr/rules_court_eng`. 15 September 2025. Accessed: 2026-08-16.

European Court of Human Rights (2026a). HUDOC database of the case-law of the European Court of Human Rights. `https://hudoc.echr.coe.int`, accessed 11 November 2026.

European Court of Human Rights (2026b). List of judges of the court since 1959. `https://www.echr.coe.int/documents/d/echr/list_judges_since_1959_bil`, accessed 11 November 2026.

European Court of Human Rights (2026c). Webcasts of hearings. `https://www.echr.coe.int/webcasts-of-hearings`. Accessed: 2026-08-15.

European Parliament (2024a). Artificial intelligence act: Meps adopt landmark law. `https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law`.

European Parliament (2024b). EU AI Act. `https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf`.

European Parliament and Council of the European Union (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (artificial intelligence act). Official Journal of the European Union, L series, 12 July 2024.

Gennaro, G. and Ash, E. (2022). Emotion and reason in political language. *The Economic Journal*, 132(643):1037–1059.

Gross, J. J., Carstensen, L. L., Pasupathi, M., Tsai, J., Götestam Skorpen, C., and Hsu, A. Y. C. (1997). Emotion and aging: Experience, expression, and control. *Psychology and Aging*, 12(4):590–599.

Gross, J. J. and John, O. P. (1997). Revealing feelings: Facets of emotional expressivity in self-reports, peer ratings, and behavior. *Journal of Personality and Social Psychology*, 72(2):435–448.

Guo, J., Deng, J., Lattas, A., and Zafeiriou, S. (2022). Sample and computation redistribution for efficient face detection. In *International Conference on Learning Representations*.

Hess, U., Adams, Jr., R. B., and Kleck, R. E. (2004). Facial appearance, gender, and emotion expression. *Emotion*, 4(4):378–388.

Hess, U., Adams, Jr., R. B., Simard, A., Stevenson, M. T., and Kleck, R. E. (2012). Smiling and sad wrinkles: Age-related changes in the face and the perception of emotions and intentions. *Journal of Experimental Social Psychology*, 48(6):1377–1380.

Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*. Sage, Thousand Oaks, CA, 2 edition.

Jacobi, T. and Schweers, D. (2017). Justice, interrupted: The effect of gender, ideology, and seniority at supreme court oral arguments. *Virginia Law Review*, 103(7):1379–1485.

Johnson, T. R., Black, R. C., Goldman, J., and Treul, S. (2009). Inquiring minds want to know: Do justices tip their hands with questions at oral argument in the U.S. Supreme Court? *Washington University Journal of Law & Policy*, 29:241–261.

Johnson, T. R., Wahlbeck, P. J., and Spriggs, J. F. (2006). The influence of oral arguments on the U.S. Supreme Court. *American Political Science Review*, 100(1):99–113.

Karpowitz, C. F., Mendelberg, T., and Shaker, L. (2012). Gender inequality in deliberative participation. *American Political Science Review*, 106(3):533–547.

Kartynnik, Y., Ablavatski, A., Grishchenko, I., and Grundmann, M. (2019). Real-time facial surface geometry from monocular video on mobile GPUs. arXiv:1907.06724.

Knox, D. and Lucas, C. (2021). A dynamic model of speech for the social sciences. *American Political Science Review*, 115(2):649–666.

LaFrance, M., Hecht, M. A., and Paluck, E. L. (2003). The contingent smile: A meta-analysis of sex differences in smiling. *Psychological Bulletin*, 129(2):305–334.

Laukka, P., Juslin, P. N., and Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition & Emotion*, 19(5):633–653.

Littler (2024). European parliament adopts the world's first comprehensive ai law. `https://www.littler.com/publication-press/publication/european-parliament-adopts-worlds-first-comprehensive-ai-law`.

Mack, K. and Roach Anleu, S. (2010). Performing impartiality: Judicial demeanor and legitimacy. *Law & Social Inquiry*, 35(1):137–173.

Maroney, T. A. (2011). Emotional regulation and judicial behavior. *California Law Review*, 99(6):1485–1556.

Matsumoto, D., Yoo, S. H., Fontaine, J., et al. (2008). Mapping expressive differences around the world: The relationship between emotional display rules and individualism versus collectivism. *Journal of Cross-Cultural Psychology*, 39(1):55–74.

Mollahosseini, A., Hasani, B., and Mahoor, M. H. (2019). AffectNet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 10(1):18–31.

Oliver, E. and Griffitt, W. B. (1976). Emotional arousal and 'objective' judgment. *Bulletin of the Psychonomic Society*, 8:399–400.

Parliamentary Assembly of the Council of Europe (2026). Election of judges to the European Court of Human Rights: Voting results and press releases, 2022–2026. `https://pace.coe.int` and `https://rm.coe.int`, accessed 11 November 2026.

Plant, E. A., Hyde, J. S., Keltner, D., and Devine, P. G. (2000). The gender stereotyping of emotions. *Psychology of Women Quarterly*, 24(1):81–92.

Posner, E. A. and de Figueiredo, M. F. P. (2005). Is the international court of justice biased? *The Journal of Legal Studies*, 34(2):599–630.

Rheault, L., Beelen, K., Cochrane, C., and Hirst, G. (2016). Measuring emotion in parliamentary

debates with automated textual analysis. *PLOS ONE*, 11(12):e0168843.

Risse, T. (2010). *A Community of Europeans? Transnational Identities and Public Spheres*. Cornell University Press, Ithaca, NY.

Roach Anleu, S., Bergman Blix, S., and Mack, K. (2015). Researching emotion in courts and the judiciary: A tale of two projects. *Emotion Review*, 7(2):145–150.

Roach Anleu, S. and Mack, K. (2005). Magistrates' everyday work and emotional labour. *Journal of Law and Society*, 32(4):590–614.

Roach Anleu, S., Rottman, D. B., and Mack, K. (2016). The emotional dimension of judging: Issues, evidence, and insights. *Court Review*, 52(2):60–71.

Roberts, Jr., J. G. (2005). Oral advocacy and the re-emergence of a Supreme Court bar. *Journal of Supreme Court History*, 30(1):68–81.

Savchenko, A. V., Savchenko, L. V., and Makarov, I. (2022). Classifying emotions and engagement in online learning based on a single facial expression recognition neural network. *IEEE Transactions on Affective Computing*, 13(4):2132–2143.

Shullman, S. L. (2004). The illusion of devil's advocacy: How the justices of the Supreme Court foreshadow their decisions during oral argument. *Journal of Appellate Practice and Process*, 6:271–293.

Simonsohn, U., Simmons, J. P., and Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, 4(11):1208–1214.

Slaughter, A.-M. (1995). International law in a world of liberal states. *European Journal of International Law*, 6:503–538.

Snider, K. M., Devereux, P. G., and Miller, M. K. (2021). Judges' emotion: An application of the emotion regulation process model. *Psychiatry, Psychology and Law*, 29(2):256–273.

Stiansen, Ø. (2022). (non)renewable terms and judicial independence in the European Court of Human Rights. *The Journal of Politics*, 84(2):992–1006.

Stiansen, Ø. and Voeten, E. (2020). Backlash and judicial restraint: Evidence from the European Court of Human Rights. *International Studies Quarterly*, 64(4):770–784.

Tajfel, H. and Turner, J. C. (1979). An integrative theory of intergroup conflict. In Austin, W. G. and Worchel, S., editors, *The Social Psychology of Intergroup Relations*, pages 33–47. Brooks/Cole, Monterey, CA.

Toy-Cronin, B. (2019). Leaving emotion out: Litigants in person and emotion in new zealand civil courts. *Oñati Socio-Legal Series*, 9(5):684–701.

V-Dem Institute (2026). V-Dem country–year dataset, accessed through the `vdemdata` R package. `https://github.com/vdeminstitute/vdemdata`, accessed 11 November 2026.

Voeten, E. (2008). The impartiality of international judges: Evidence from the European Court of Human Rights. *American Political Science Review*, 102(4):417–433.

Wasby, S. L., D'Amato, A. A., and Metrailer, R. (1976). The functions of oral argument in the U.S. Supreme Court. *Quarterly Journal of Speech*, 62(4):410–422.

Wistrich, A. J., Rachlinski, J. J., and Guthrie, C. (2015). Heart versus head: Do judges follow the law or follow their feelings? *Texas Law Review*, 93(4):855–923.

Xu, T., White, J., Kalkan, S., and Gunes, H. (2020). Investigating bias and fairness in facial expression recognition. In *Computer Vision – ECCV 2020 Workshops*, pages 506–523. Springer.

Online Appendix

A Variables, Definitions, and Sources

Table A1: Variables, definitions, and sources

Variable	Definition	Source
Negative display	Sum of anger, contempt, disgust, fear, and sadness probabilities	EmotiEffLib `enet_b0_8_best_vgaf`, faces $\geq$32 px
Positive display	Happiness probability	as above
Expressivity	One minus the neutral probability	as above
Sideways turn	Horizontal offset of the nose from the midpoint between the eyes, divided by the distance between the eyes; signed; all faces	video pipeline (`kps_yaw_proxy`; Appendix B)
Frontal score	$1 - \lvert\text{sideways turn}\rvert/0.6$, limited to the range 0 to 1; all faces	video pipeline
Nose height, roll	Vertical position of the nose between eyes and mouth; tilt of the eye line in degrees; all faces	video pipeline
Head yaw, smile width, brow and eye-closure proxies	Angle in degrees; geometric ratios; faces $\geq$64 px	video pipeline
Image-adjusted outcome	Outcome minus the part predicted by frontal score (quadratic), sideways turn (cubic), nose height and roll (quadratic), face position, brightness, contrast, log face height, faces in frame, and shot type, each as a deviation from its mean over listening seconds; slopes from seconds with neither party speaking	video pipeline
Bench-verified face	High-confidence identity cluster; judge seen $\geq$10 s while a member of the bench is speaking; face 0.7–1.5$\times$ the judge's median bench face height	video pipeline
Speaker side	Applicant side, respondent government, bench, third party, unattributed; gaps $\leq$3 s within one side filled	ChatGPT 6.1 Sol
Guard band	Seconds to the nearest change of speaker between sides	ChatGPT 6.1 Sol
Phase	Opening, submissions, questions, replies, recess, closure	ChatGPT 6.1 Sol
National judge (H1)	Judge elected in respect of a respondent state	case metadata
Ad hoc, President, substitute	Role on the bench	case metadata
Woman judge	Official form of address or biographical record	ECHR; PACE
Re-election possible	Hearing held before 1 June 2010, when Protocol No. 14 introduced single nine-year terms	Protocol No. 14
Both EU (H2)	Judge's state and respondent state both EU members at the hearing date	EU accession dates
Same East/West (H3)	Both or neither post-socialist	authors' coding
Same North/South (H4)	Same group; South = UN M49 Southern Europe and Western Asia	UN M49
Democracy distance (H5)	$\lvert\text{LDI}_{\text{judge}} - \text{LDI}_{\text{respondent}}\rvert$ in the hearing year (latest available for 2026)	V-Dem (V-Dem Institute, 2026)
Vote for the applicant	Majority in a violation judgment, or dissent from a no-violation judgment; otherwise a vote for the respondent state	HUDOC (European Court of Human Rights, 2026a)
Dissent / separate opinion	Author or co-signatory of a (partly) dissenting or separate opinion	HUDOC
Articles 2–4	Hearing concerns the right to life, ill-treatment, or forced labor	case metadata topics
Inter-state	Applicant named in the title is a member state	case title

B The Video Pipeline: Technical Details

The face measurements are produced by the ECHR judge visual pipeline, a Python package that we wrote for this project. In our recordings, a single frame of a judge typically reaches a cosine similarity of only 0.2 to 0.45 with that judge's portrait, so those releases left most judges' faces unidentified. Release 16 separates the work into two stages (Figure 4) and makes identity decisions for a person over a whole hearing. Averaging a person's embeddings over all of his or her

Table A2: Descriptive statistics (main sample)

	N	Mean	SD	Min	Max
Negative display, government speaking	450	0.655	0.148	0.128	0.974
Negative display, applicant speaking	450	0.621	0.147	0.137	0.987
Listening differential, negative (gov − app)	450	0.034	0.109	-0.352	0.449
Positive display, government speaking	450	0.015	0.018	0.000	0.184
Positive display, applicant speaking	450	0.018	0.032	0.000	0.516
Listening differential, positive	450	-0.004	0.027	-0.428	0.072
Expressivity, government speaking	450	0.743	0.110	0.337	0.976
Expressivity, applicant speaking	450	0.721	0.110	0.348	0.988
Listening differential, expressivity	450	0.022	0.093	-0.372	0.417
Seconds observed, government speaking	450	33	30	5.000	196
Seconds observed, applicant speaking	450	33	29	5.000	203
National judge	450	0.069	0.254	0.000	1.000
President of formation	450	0.369	0.483	0.000	1.000
Ad hoc judge	450	0.002	0.047	0.000	1.000
Woman	450	0.164	0.371	0.000	1.000
Term began before 2010	396	0.806	0.396	0.000	1.000
Both states in the EU	450	0.569	0.496	0.000	1.000
Same East/West group	450	0.598	0.491	0.000	1.000
Same North/South group	450	0.567	0.496	0.000	1.000
Democracy distance (V-Dem LDI)	420	0.200	0.238	0.000	0.783
Grand Chamber	450	0.824	0.381	0.000	1.000
Inter-state case	450	0.027	0.161	0.000	1.000

Note: Judge-hearings with at least five bench-verified, expression-scored seconds while each side was speaking (N = 450; 206 hearings; 65 judges). Display measures are mean class probabilities; differentials are government-turn minus applicant-turn means.

appearances raises the similarity for the right judge to roughly 0.5 to 0.85, while unrelated people stay below about 0.2.

The package runs under Python 3.11 with InsightFace 2.0.0 (model pack `buffalo_l`, executed with ONNX Runtime), EmotiEffLib 1.1.1 (PyTorch), MediaPipe 0.10.21, OpenCV 4.11, and SciPy 1.13. The first stage needs a graphics processor and is run once per recording. It stores its results, including the embedding of every face, so that the second stage can be repeated on an ordinary processor in a few seconds per hearing whenever a threshold, a reference photograph, or a manual correction changes. Every hearing's output records a hash of the settings and inputs of each stage, which ties each table to the configuration that produced it. Table A3 lists the methods and the settings used, and the following paragraphs explain the steps that matter most for the analysis.

Table A3: Methods and settings of the video pipeline (release 16)

Step	Method	Settings
Stage 1: perception		
Sampling	Frames decoded at fixed intervals	One frame every 0.5 s for tracking; every second such frame (one per second) is a measurement frame
Face detection	SCRFD (Guo et al. 2022), InsightFace `buffalo_l`	Input matched to the shape of the frame, short side 1,088 px, with enlargement of up to four times for small recordings; second pass at 1,440 px when the first finds no face or a face under 40 px; detection score at least 0.35; overlapping detections from the two passes merged (non-maximum suppression at an overlap of 0.40); faces of at least 8 px kept; at most 60 per frame

Step	Method	Settings
Identity embedding	ArcFace with a ResNet-50 backbone (Deng et al. 2019)	Faces of at least 16 px; face aligned to 112 × 112 px from five landmarks; embedding of the face and of its mirror image summed and normalized to unit length; length of the raw embedding kept as a measure of how recognizable the face is
Expression	EfficientNet-B0 (Tan and Le 2019), model `enet_b0_8_best_vgaf` of EmotiEffLib (Savchenko et al. 2022)	Measurement frames; faces of at least 32 px; square crop with a 10 percent margin taken from the original frame; eight class probabilities; a face is flagged as low-confidence if the largest probability is below 0.40, exceeds the second by less than 0.10, or the normalized entropy is above 0.85; smoothed series are centered moving averages over five measurement frames within a track
Landmark geometry	Five landmarks of the detector	Roll: angle of the line between the eyes. Sideways turn: horizontal offset of the nose from the midpoint between the eyes, divided by the distance between the eyes, after removing roll. Nose height: vertical offset of the nose divided by that of the midpoint of the mouth. Frontal score: one minus the absolute sideways turn divided by 0.6, limited to the range 0 to 1
Face mesh and head pose	MediaPipe Face Mesh (Kartynnik et al. 2019); perspective-n-point fit	Measurement frames; faces of at least 64 px; six mesh points fitted to a generic three-dimensional face model; angles kept if the mean reprojection error is at most 18 px; eye and mouth aspect ratios and proxies for eye closure, smile width, and brow raise
Image quality	OpenCV	Sharpness (variance of the Laplacian), brightness, contrast, and share of the frame; quality score with weights 0.35 for size, 0.30 for sharpness, 0.20 for contrast, and 0.15 for brightness
Stage 2: identity		
Shot detection	Change in color histograms and pixels between consecutive tracking frames	Frames reduced to 160 × 90 px; score equal to 0.65 times the change in the hue and saturation histograms plus 0.35 times the mean absolute change in pixels; a cut is recorded when the score is at least 0.10, or at least 0.06 and three times the median of the nine surrounding frames
Tracking	Optimal assignment (Hungarian algorithm; Kuhn 1955) within shots	Affinity equal to 0.6 times box overlap plus 0.4 times cosine similarity when the overlap is at least 0.15; links by position alone need an overlap of 0.30; links by embedding alone need a similarity of 0.40 and a similar size; faces with a similarity below 0.05 are never linked; gaps of up to 1.5 s are bridged
Persons	Average-linkage clustering of track templates with a co-occurrence constraint	A template is the weighted mean embedding of a track, with weights that increase with recognizability, face size, and frontal score; two groups merge when their mean similarity is at least 0.42; groups that share a frame never merge

Step	Method	Settings
References	Portrait galleries, references from other hearings, and a comparison group	Portraits embedded at full size and reduced to face widths of 28 and 48 px to resemble video; portraits unlike the rest of a judge's gallery (similarity below 0.20) dropped; portrait score equal to the mean of two quantities, the average of the two best portrait similarities and the similarity to the mean embedding of the gallery; up to 40 references per judge from high-confidence appearances in other hearings (score at least 0.45); portraits of all judges who are not on the bench form the comparison group
Decision	Threshold and margins	Score s equal to the larger of the portrait score and the score from other hearings minus 0.20; a person is assigned when $s \geq 0.30$, s exceeds the second-best judge on the bench by 0.07 and the best judge outside it by 0.05, and at least two embeddings are available; persons are assigned in descending order of s, and a judge cannot be assigned to two persons who share a frame
In-video references	Two further rounds	Persons assigned with $s \geq 0.42$ and visible for at least 2 s become references for the same hearing; an unassigned person is accepted at a similarity of 0.50 to such a reference, or at 0.35 when the portrait score is at least 0.22, with a margin of 0.10 over the next judge; weak assignments that a reference contradicts are withdrawn
Confidence	Labels	High: $s \geq 0.45$ with margins of 0.12 over the bench and 0.10 over the comparison group, or a similarity of 0.60 to an in-video reference. Medium: $s \geq 0.35$ with a margin of 0.10, or a similarity of 0.50 to an in-video reference. Low: all other assignments

Detection is the step at which earlier releases lost most faces. In a wide shot of the bench, a judge's face may be 20 pixels high, and in the recordings made before 2013 the whole frame has only 288 lines. The detector therefore receives the frame at its own aspect ratio and enlarged, and a second pass at higher resolution is added whenever the first pass finds nothing or only small faces. All faces of at least 8 pixels are kept for tracking, faces of at least 16 pixels receive an identity embedding, and expression is scored only from 32 pixels upward.

Tracking and clustering turn single detections into persons. Tracks never cross a camera cut. Cuts in these webcasts change the picture only moderately, because robes, wood, and lighting are similar from shot to shot, so a cut is recorded not only when the frame-to-frame change is large but also when it is clearly larger than the change in the surrounding frames. Within a shot, the detections of consecutive frames are matched by solving an assignment problem in which the gain from linking two faces rises with the overlap of their boxes and the similarity of their embeddings. The tracks of a hearing are then merged by average-linkage clustering of their mean embeddings. Two tracks that appear in the same frame can never be merged, since one person cannot be in two places.

Let e_c be the unit-length mean embedding of person c and let s_{cj} be its similarity to the references of judge j. With P the set of judges on the bench of the hearing and j^* the judge in P with the highest score, the pipeline assigns c to j^* if

$$s_{cj^*} \geq 0.30, \qquad s_{cj^*} - \max_{j \in P,\, j \neq j^*} s_{cj} \geq 0.07, \qquad s_{cj^*} - \max_{k \notin P} s_{ck} \geq 0.05. \tag{3}$$

The third condition uses the portraits of all judges who are not sitting in the hearing. A person who looks more like one of them than like any judge on the bench is most likely a lawyer, a member of the registry, or a judge whose portrait has been filed under the wrong name, and is left unidentified. Table A4 gives the complete sequence of steps for one hearing.

Table A4: Identity resolution for one hearing (stage 2)

1	Compute the change score between consecutive tracking frames and divide the hearing into shots.
2	Within each shot, link detections into tracks by optimal assignment on box overlap and embedding similarity.
3	Compute a weighted mean embedding (template) for every track with at least one reliable embedding.
4	Merge tracks into persons by average-linkage clustering of the templates, never merging tracks that share a frame.
5	Score every person against the portrait gallery and the references from other hearings of each judge on the bench, and against the portraits of all other judges.
6	Mark as eligible the persons that pass the threshold and both margins and have at least two embeddings.
7	Assign eligible persons to judges in descending order of score, skipping an assignment if the judge already has a face in one of the same frames.
8	Build an in-video reference for each judge from the confident assignments, accept further persons that match it, withdraw weak assignments that it contradicts, and repeat once.
9	Label each assignment as high, medium, or low confidence, and write the tables and a review sheet for the hearing.

In the run used here the pipeline detected 50.2 million faces on the tracking frames, 89 percent of them large enough for an embedding. It formed 2.03 million tracks and 198,674 persons, and it assigned 6,345 persons to judges. A typical hearing has about 6,011 tracks and 610 persons, of whom 18 are assigned. Official portraits were available for 122 of the 161 judges, which covers 90 percent of all seats. The hearing records spell the names of eight judges in more than one way. Where two spellings led to the same portrait folder, the pipeline withheld the portrait from both until the match had been reviewed. As a result, seven judges who sat often were left without a portrait in 377 seats, 7 percent of the total, and none of them is verified in any hearing (Table A9). On average 15.9 of the 18.0 judges on a bench have a portrait and 15.5 have a reference from another hearing, and 11.3 are found at least once (63 percent). Of the 705,155 judge faces on measurement frames, 72 percent have high, 11 percent medium, and 16 percent low confidence. 81 percent were assigned through portraits, 17 percent through references from other hearings, and 2 percent through in-video references. The expression classifier scored 73 percent of the judge faces, and it flagged 37 percent of the scored faces as low-confidence. In 284 hearings at least one person on screen resembled a judge who was not on the bench and was left unidentified for that reason.

The package provides a review sheet for each hearing, in which every person is shown next to the portrait to which he or she was matched, and a routine that compares the automatic decisions with reviewed cases. No manual corrections were applied in this run, and the accuracy of the identity decisions against human review has not yet been measured. For that reason the analysis uses only assignments with high confidence that also pass the bench check described in Section 5, and it reports results under looser rules as robustness checks. The classifier's eight probabilities are used as they are returned. Faces flagged as low-confidence are kept in the main analysis and excluded in the specification called "confident frames only," and the smoothed probabilities are used in the specification called "smoothed."

The record of who is speaking is built separately from the audio. We transcribed the recordings with WhisperX (Bain et al. 2023), which adds word-level time stamps and speaker segments to

the Whisper speech-recognition model. The transcripts were then read by ChatGPT 6.1 Sol, which drafted the speech turns and assigned them to the applicant's side, the government, the bench, or third parties (Appendix J). The two streams are joined on the one-second measurement frames, so that every judge face carries the side that was speaking in that second.

Additional references Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1–2):83–97.

Tan, M. and Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, pages 6105–6114.

C Empirical Specifications and Exploratory Analyses

This section gives the specification of each test. The main tests were chosen before any model was estimated. Analyses that we added afterwards are exploratory and are listed at the end of the section.

Two definitions should be noted. The renewable-term variable for C3 refers to hearings held before Protocol No. 14 took effect on 1 June 2010, and the minimum-seconds rule in the specification curve applies to C1 as well as C2.

The model specifications are as follows. For C1, cell means are regressed on an indicator for the government speaking, with judge-hearing fixed effects and weights equal to the number of seconds. Judge-hearings observed during only one side's turn contribute nothing, leaving 1,174 cells. For C2 to C8, the listening differential is regressed on the affinity variable with hearing fixed effects. Hearings with a single sampled judge drop out, leaving 418 judge-hearings. We also report versions with judge fixed effects and with both sets of fixed effects. The permutation test for C2 reshuffles the national-judge label within hearings 1,000 times. The variance decomposition (A2) uses crossed random effects for judge, hearing, and respondent state. Case content (B1, B2) is tested on judge-hearing means with judge fixed effects and controls for year, resolution, formation, average face size, and the share of seconds in which a member of the bench was speaking. The vote models (D1) are linear probability models with hearing fixed effects, checked with conditional logit, which uses only benches whose members voted differently, and with a wild bootstrap. D2 models dissent, or any separate opinion, as a function of mean expressivity. D3 compares penalized logistic regression with and without display features, using five-fold cross-validation that holds out whole hearings. The measurement models (E1 to E3) use judge-seconds with judge or hearing fixed effects. The false discovery rate is controlled within each family of main tests across its outcomes. For D1, these are negative and positive display only; the expressivity analysis is exploratory.

The robustness checks vary the minimum number of seconds per side (1, 5, 10, and 20), the guard band around changes of speaker (0, 2, and 5 seconds), the minimum face size (64 pixels), the frames used (confident frames only), the probabilities used (smoothed), the identity rule (bench-verified, high confidence only, high and medium confidence), the era (high definition only), and the mode of hearing (in person only).

Tests are labeled A1 to G5. Table A5 maps these identifiers to the sixteen hypotheses: C2 and C3 test Hypothesis 1, C4 Hypothesis 2, C5 Hypotheses 3 and 4, C6 Hypothesis 5, and C8

Table A5: Hypotheses, expectations, and empirical tests

	Expectation	Test	ID
H1 National Bias	More positive response to the judge's home state	National judge vs. colleagues on the same bench; larger where re-election was possible	C2, C3
H2 EU Bias	More positive response of judges from EU states to EU respondents	As H1; judge and respondent both from EU states	C4
H3 Eastern vs. Western Europe	More positive response within the same group (two-sided)	As H1; same East/West group	C5
H4 Northern vs. Southern Europe	More positive response within the same group	As H1; same North/South group	C5
H5 Democratic Similarity	More positive response the closer the two states' democracy scores	As H1; distance in democracy scores	C6
H6 Judge Characteristics	More positive response of women judges to women applicants	Proxy: women judges in discrimination cases	C8
H7 Individual Differences	Judges explain more variance than hearings; the presiding judge is less expressive; departures from a judge's baseline concentrate in questions and replies	Judge-hearing means; judge-seconds	A1–A3
H8 Parties	Displays differ with the party speaking (two-sided)	Within judge and hearing	C1
H9 Subject Matter	More negative display in Article 2–4 cases; more expressivity in inter-state cases	Judge-hearing means, within judge	B1, B2
H10 Votes	More negative display toward a side predicts a vote against it; dissenters are more expressive; displays improve prediction	Votes within bench; held-out hearings	D1–D3
H11 Gender of the Judge	Less negative display by women judges than by men (formulated after the results were known)	Women and men in the same hearing, at the same image quality	E2
H12 Emotional Profiles	A judge's dominant emotion is the same across hearings	Dominant emotion in the odd- and even-numbered hearings of each judge	G1
H13 Gender and Specific Emotions	Anger more prominent among men; sadness, fear, and happiness among women	As E2, for the probability of each emotion	G2
H14 Age	Negative display differs with age (two-sided); more anger and sadness in older faces	Judges in the same hearing, at the same image quality	G3
H15 Culture and Career	Less positive display by judges from post-socialist states; more expressivity in the South; less among career judges	As G3	G4
H16 Engagement	More expressivity among judges who ask questions and who are shown most	Same hearing; same judge; across judges	G5
Measurement	Image quality, judge gender, and camera selection affect the measures	Judge-seconds	E1–E3

Note: Test identifiers correspond to those used in the appendix tables. The main tests are A2, B1, C1, C2, C4 to C6, and D1 (D1 for negative and positive display). Three further tests, of victim narratives (B3), legal tradition (C7), and applicant groups (C9), need codings that we do not have. The tests of H1 to H6 conditional on the ruling, the head-orientation analyses, and G1 to G5 are exploratory. E2 was first used as a check on the measurement.

Hypothesis 6 by proxy. A2 tests Hypothesis 7, B1 Hypothesis 9, and D1 Hypothesis 10. C1 compares the two parties and tests Hypothesis 8. Hypothesis 11 uses E2, the comparison of women and men at the same image quality, which we first used as a check on the measurement. G1 to G5 test Hypotheses 12 to 16 (Appendix F).

The following analyses were added after the first results had been examined and are treated as exploratory: the tests of Hypotheses 1 to 6 conditional on the judge's ruling (Table A13); the comparisons between judges for Hypotheses 12 to 16 (Appendix F); the reading of the comparison between women and men (E2) as a test of Hypothesis 11, and the versions of the tests of Hypotheses 10 and 11 on all available observations; the event study around changes of speaker (Figure A7); the C1 estimates by speaking order, phase, time elapsed, and speaker-attribution confidence; all analyses in Section 6.5, including the landmark measures, image-adjusted outcomes and models estimated on them, comparisons within turn bins and above frontal thresholds, estimates for presiding and other judges, period comparisons (Figure A9), geometric proxies, and the relationship between party differences in displays and pose; the class-by-class decomposition (Figure A8); the smile-width comparison; the split-half reliability of judge means; C4 to C6 with a control for the national judge; the minimum detectable effect and the 90% interval for C2; the single-respondent and mixed-model versions of C2; the vote model with the expressivity differential, its leave-one-hearing-out range, and its estimate on divided benches alone; boosted trees in the prediction exercise; and the relationship between the listening differential and the case outcome. Table A12 reports these estimates. The variance decomposition for A2 is reported as shares without intervals.

The corrections to the outcome coding described in Appendix **??** affected only the vote models. Before correction, the within-bench models rested on 53 votes in 21 benches, and the D1 estimate for expressivity was -3.6 points per 10-point difference. After correction, they rest on 55 votes in 22 benches, and the estimate is -4.1.

D Additional Results

This section reports results that are summarized in the main text.

On individual differences, the standard deviation of judges' mean negative display is 13.3 points, and for many judges the most probable class is not the neutral one. The variance shares are REML estimates from a model with crossed random effects for judge, hearing, and respondent state, and the optimum was located from several starting values. The respondent state accounts for none of the variance once judges and hearings are in the model. After the image adjustment, the judge's share of the variance is 46 percent for negative display and 46 percent for expressivity, and the share of hearing and state is 8 and 5 percent. For positive display the judge's share is 26 percent, and most of the variation is residual. Within a hearing, the average absolute deviation from a judge's own mean is 0.18 points larger during questions and replies than during the prepared submissions for positive display ($p = 0.005$) and no different for the other two measures. The split-half correlations of judge means are almost the same after the image adjustment (0.83 for negative display and 0.85 for expressivity). The difference between presiding judges and their colleagues is unchanged by the adjustment (3.8 points), and presidents are not shown at a different angle from other judges ($p = 0.18$).

On women and men judges (Hypothesis 11), the main text reports the comparison on all veri-

fied seconds and a second version restricted to larger faces. The difference in negative display is 11.0 points when the brightness and contrast of the face are added to the controls, 13.2 points in the 240,982 seconds in which the judge is not speaking, 13.7 points on judge-hearing means with hearing fixed effects (1,787 judge-hearings), and 13.8 points on image-adjusted outcomes, each with $p < .001$. The corresponding differences in expressivity are 6.1, 7.9, 7.8, and 8.5 points. Women's positive display is higher by 0.88, 0.99, 1.10, and 1.13 points in the same four comparisons (each $p < .02$).

On the comparison between the parties, the main difference in negative display is about 0.32 of a standard deviation of the listening differential. It is unchanged by a control for the time elapsed in the hearing, it is 3.8 points in the prepared submissions alone, and 3.7 points (95% CI 1.8 to 5.6) in the 11,639 seconds whose speaker attribution has medium or high confidence. We also examined the 1,006 moments at which one party stopped speaking and the other began (Figure A7), but the intervals are too wide to say when any difference opens up.

On Hypothesis 1, the 90% confidence interval for the national-judge estimate runs from -2.6 to 4.8 points, and on adjusted outcomes the 95% interval runs from -2.7 to 5.4. The smallest p-value among the 15 affinity tests on adjusted outcomes is 0.11. Dropping one judge at a time moves the national-judge estimate between -2.7 and 2.8 points, and dropping one respondent state at a time moves it between -1.5 and 2.6 points. Restricting the sample to hearings with a single respondent state gives 0.4 points, and a model with crossed random effects for judge and hearing gives 1.2 ($p = 0.56$). Some specifications lean in one direction or the other. For positive display, two of the robustness specifications give national judges a differential that favors the government, as Hypothesis 1 predicts: 0.69 points among faces at least 64 pixels high ($p = 0.005$) and 0.83 points in high-definition recordings ($p = 0.010$). The other specifications are close to zero, and the two samples overlap heavily. With judge fixed effects, national judges look slightly more negative during their own government's turn (1.3 points, $p = 0.06$), which is the opposite of favoritism. For the period before Protocol No. 14 took effect in June 2010, when re-election was still possible, the estimate for national judges is -0.5 points ($p = 0.89$), and only 1 ad hoc judge appears in the sample at all.

On subject matter, the estimate for hearings under Articles 2 to 4 is 0.1 points after the image adjustment, and inter-state hearings show no difference in expressivity.

On votes (Hypothesis 10), the main sample requires at least five seconds of each party's turn. With every judge-hearing in which the judge was seen during both turns (477 judge-hearings with a coded vote in 204 hearings, 33 of them with divided benches and 98 votes), the estimates per 10-point difference are -0.1 points for negative display ($p = 0.91$), -13.5 for positive display (95% CI -27.9 to 0.8; wild bootstrap $p = 0.007$), and -2.6 for expressivity (95% CI -5.1 to -0.2; wild bootstrap $p = 0.030$). When the vote is regressed on the judge's mean display over the whole hearing, with hearing fixed effects, none of the three measures is related to it in the 1,400 judge-hearings with a coded vote ($p = 0.22$, 0.84, and 0.12), with or without judge fixed effects. In the main sample, the estimate for the expressivity differential is -4.3 on adjusted outcomes ($p = 0.006$) and -4.1 without national judges, it stays between -4.3 and -3.5 when any one hearing is left out, and its conditional-logit p-value is 0.014. Unanimous benches hold 91 percent of the within-bench variance in the regressor and so pull a linear estimate toward zero; among divided benches alone the 95% interval for the slope runs from -61.0 to -26.0. Within judge, the listening differential in negative display is no larger in cases that ended in a finding of a violation (0.9 points, $p = 0.61$). In the prediction exercise, a split by time, with training on hearings up to

2012 and testing on later ones, gives AUCs of 0.48 without and 0.46 with display measures.

On measurement, face size limits what can be scored. In 2007–2012, 69 percent of identified judge faces were large enough to score, against 87 percent afterward (Figure A4). The average probability of a neutral face in high-definition recordings falls from about 0.40 for faces 32 to 47 pixels high to about 0.24 for faces of 96 pixels or more, a pattern that the within-judge model does not confirm. An individual judge is on screen and verified in 7.2 percent of the seconds in which a member of the bench is speaking and in 0.14 percent of those in which the applicant's side speaks. Of the 96 judges who appear in at least one listening second, ten supply 70 percent of the total, 64 percent of listening seconds come from recordings with fewer than 480 lines, and only 36 percent of seats yield even one second in which the judge is identified, verified, and on screen.

E Head Orientation: Details

This analysis uses the signed sideways turn, the vertical position of the nose, head roll, the five landmarks, the brightness and contrast of the face, and its position in the frame, for every judge face in all 311 processed hearings. The frontal score equals one minus the absolute sideways turn divided by 0.6, limited to the range from zero to one, so it carries the amount of the turn and not its direction.

The image adjustment is estimated on the 202,961 verified judge-seconds in which the judge is not speaking and neither party is speaking (about four fifths of them while a member of the bench is speaking). For each outcome we regress the judge-second value on the frontal score and its square, a cubic in the sideways turn, quadratics in nose height and roll, the position of the face, its brightness and contrast, the logarithm of face height in pixels, the number of faces in the frame (capped at eight) and indicators for the type of shot, with a fixed effect for each judge in each hearing. The adjusted value of a second is its raw value minus the fitted contribution of the deviation of these variables from their means over all listening seconds. Cells, differentials and judge-hearing means are then rebuilt exactly as for the raw outcomes. Because the variables are centered on common means, the adjustment changes judge-hearing means as well as within-hearing contrasts. The slope of negative display on the frontal score is 20 points per unit while a member of the bench is speaking and 15 within the parties' turns, which is why we also report adjusted cells with slopes estimated inside the turns.

For the comparison within turn bins, the fixed effect is the judge-hearing crossed with the twentieth of the range of sideways turn, so that a second from the government's turn is compared only with seconds from the applicant's turn in which the same judge in the same hearing was turned by about the same amount in the same direction. By band of the frontal score, the raw difference between the parties is -1.4 points below 0.2, -0.1 between 0.2 and 0.4, 4.2 between 0.4 and 0.6 ($p =< .001$), 1.9 between 0.6 and 0.8 ($p = 0.016$) and -0.3 at 0.8 and above. Across judge-hearings, a weighted regression of the differential in negative display on the differential in the frontal score, with standard errors clustered by judge and hearing, has a slope of 1.4 points per tenth of the score ($p < .001$) and an intercept of 1.7 points (95% CI 0.0 to 3.3). The main C1 estimate was also recomputed directly from the differentials, with the same result, and it remains significant when cell means are transformed to the logit scale ($p < .001$).

Presiding judges and their colleagues turn their heads between the two parties' turns to the same extent (0.10 and 0.09). The frontal score is 0.14 higher in the government's turn, whichever party

spoke first (0.13 and 0.15). Head roll differs between the turns by about four degrees, whereas the vertical position of the nose, the place of the face in the frame, its size, and its brightness do not differ. The comparison within turn bins has a 95% interval from 0.7 to 3.2 points, the adjusted cells give 0.6 points when the slopes are estimated inside the parties' turns, and the comparison among nearly frontal faces rests on 7,121 seconds. The sign-flip test for the adjusted difference uses 5,000 permutations. Figure A8 shows where the raw difference came from: the classifier reports more anger and sadness and less neutrality during the government's turn, and each of these components shrinks by half or more once the frontal score alone is held constant.

By period (Figure A9), the difference in sideways turn between the two parties' turns is 0.09 in 2007–2012, 0.12 in 2013–2016, 0.11 in 2017–2019, and 0.08 in 2020–2026. Until 2016, 97 percent of faces in these shots are turned to the same side, and the average offsets are -0.30 in the applicant's turn and -0.20 in the government's turn. From 2020 they are -0.08 and 0.01, with absolute offsets of 0.14 and 0.15, so that neither turn is more frontal than the other. The last two periods hold only 6 and 8 percent of the listening seconds. We have not checked the layout of the hearing room or the position of the cameras against any source other than the recordings.

For presiding judges the raw difference between the parties is 5.0 points, against 2.4 points for other judges, and with the frontal score held constant it is still 2.6 points ($p = 0.007$), against 0.4. The full set of landmarks brings it to 0.3 points ($p = 0.67$). In the adjusted cells, where the slopes come from seconds in which neither party speaks, it is 2.1 points ($p = 0.035$), against 0.0 for other judges. We cannot say whether presiding judges show something that their colleagues do not or whether slopes learned outside the parties' turns fit their shots less well.

Where faces are large enough for the pipeline's other measurements, measured yaw and pitch remove about a quarter of the difference (from 3.5 to 2.6 points), and the landmark measures remove most of it (to 0.8 points, 95% CI -1.1 to 2.7). The two kinds of measurement correlate at 0.60. On the same faces, two measures that do not involve the classifier differ between the turns: the brow proxy by -0.33 standard deviations and the eye-closure proxy by 0.43. With the landmark controls the first difference disappears (-0.05), and the second is halved but remains (0.19, $p < .001$).

The difference in positive display is -0.39 points with the frontal score held constant and -0.11 within turn bins with nose height and roll ($p = 0.22$). It comes from judges other than the President (-0.33 points, against 0.06 for presidents). Among the large faces for which smile width is measured, neither smile width ($p = 0.79$, and 0.60 with landmark controls) nor the classifier's positive display (-0.22 points, $p = 0.25$) differs reliably between the turns.

F Comparisons Between Judges

This section documents the tests of Hypotheses 12 to 16 and describes who asks questions and who is on camera. All of it is exploratory. Judges are counted as persons, and the three hearings whose recorded bench belongs to another hearing (Table A10) are left out.

Displays: The tests use all 370,876 verified judge-seconds of 113 judges. Each regresses an outcome, or the probability that the classifier assigns to one emotion, on a characteristic of the judge, with hearing fixed effects and controls for the size of the face, the frontal score, the sideways turn, the vertical position of the nose, and the tilt of the head, as in the comparison of women and men for Hypothesis 11. The models for age, region, and career also hold gender constant. Standard errors are clustered by judge and by hearing. Age is documented for 101 of these judges and earlier career for 52. For Hypothesis 12, a judge's dominant emotion is the class other than neutral with the highest mean probability. We order each judge's hearings by date, compute the dominant emotion separately in the odd-numbered and the even-numbered ones, and compare the share of judges for whom the two agree with its distribution when the second half is reshuffled across judges 5,000 times. All 106 judges with verified seconds in at least two hearings enter the test. Among the judges with at least 300 seconds in each half the agreement is higher still. For Hypothesis 16, the across-judge correlations relate a judge's mean expressivity while she is not speaking, net of hearing and image quality, to her asking rate, net of hearing and role, and to her camera time, net of hearing, role, asking, and the two places beside the presiding judge. Each judge is weighted by the number of seconds behind her mean, and the p-values come from permutations.

Table A6 lists the fifteen tests with their expected signs, and Table 3 gives the estimates for every emotion. Three results qualify the summary in the main text. The relation between age and negative display is 2.1 points per ten years ($p = 0.13$) in the seconds in which the judge is not speaking and 1.7 points ($p = 0.23$) on judge-hearing means, against 3.3 points on all verified seconds. Women's readings are also higher on surprise (4.2 points, $p = 0.003$), which was not among our expectations. And although judges who ask a question are no more expressive than others, their positive display is slightly higher in the hearings in which they ask than in the hearings in which they do not (0.35 points, $p = 0.001$).

Questions: The speaker record contains 2,727 questions from the bench, of which it attributes 2,526 to a named judge. Because the attributions carry low or medium confidence labels, we checked them against the video. A question can be checked when the judge named in the record is one whom the face pipeline verifies in that hearing and when at least three seconds of the question show a single face. This is the case for 936 questions in 131 hearings. In 99 percent of them the face on screen is that of the judge named in the record, and the share is the same for attributions labeled as low confidence (99 percent) and for questions attributed to national judges (99 percent). The two records were produced independently, since the speaker record names judges from what is said in the hearing and the pipeline from their faces. The record remains incomplete, because questions that it leaves unattributed are missing from every count below.

A judge is recorded as asking in a hearing when at least one question is attributed to her. The sample consists of the 289 hearings with at least one attributed question, with 5,181 judge-hearings of 156 judges and 1,203 question turns. Asking is a habit of the judge. Among the 106 judges with at least ten hearings, the share of hearings with a question runs from 5 percent at the tenth percentile to 53 percent at the ninetieth, with a median of 22 percent, and 6 of the 118 judges with at least five hearings never ask (Figure A1). A judge's asking rate in her odd-numbered hearings correlates

Table A6: Hypotheses 12 to 16: tests

	Test	Exp.	Estimate	95% CI	p	q	N
H12	Same dominant emotion in odd and even hearings (share of judges)	+	88%	chance 57%	<.001	<.001	106
H13	Women − men: anger	−	−22.7	−28.2 to −17.3	<.001	<.001	370,651
	Women − men: sadness	+	10.0	5.9 to 14.2	<.001	<.001	370,651
	Women − men: fear	+	0.80	−0.46 to 2.07	.210	.367	370,651
	Women − men: happiness	+	0.92	0.22 to 1.62	.011	.027	370,651
H14	Age, per ten years: negative display	±	3.3	0.8 to 5.9	.011	.027	367,042
	Age, per ten years: sadness	+	−1.6	−3.9 to 0.8	.183	.367	367,042
	Age, per ten years: anger	+	4.2	1.2 to 7.1	.006	.022	367,042
H15	Post-socialist state: positive display	−	0.04	−0.53 to 0.62	.877	.923	370,651
	Southern state: expressivity	+	−2.2	−5.8 to 1.4	.220	.367	370,651
	Career judge: expressivity	−	−2.0	−5.7 to 1.7	.277	.415	180,289
H16	Asked a question: expressivity, same hearing	+	0.1	−2.2 to 2.4	.923	.923	231,454
	Asked a question: expressivity, same judge	+	−0.3	−1.2 to 0.5	.445	.560	231,454
	Asking rate and expressivity, across judges (r)	+	0.12	−0.19 to 0.42	.448	.560	113
	Camera time and expressivity, across judges (r)	+	0.02	−0.27 to 0.28	.899	.923	113

Note: Exp.: expected sign (±: two-sided). Differences in percentage points, from regressions on all verified judge-seconds with hearing fixed effects and controls for image quality (and for the judge's gender in the rows on age, region, and career); standard errors clustered by judge and hearing. The rows on asking use the seconds in which the judge is not speaking, in hearings with attributed questions. Correlations are across all judges, weighted by the seconds behind each judge's mean. q: Benjamini–Hochberg over the fifteen tests. N: seconds, or judges for the first and the last two rows.

with her rate in the even-numbered ones at $r = .67$ (95% CI .54 to .78; 115 judges), net of hearing and role, and the tenth of judges who ask most account for 36 percent of all question turns. The clearest difference is one of role. National judges ask in 54 percent of their hearings, other sitting judges in 23 percent, presiding judges in 22 percent, and substitute judges in 4 percent. Women ask in 23 percent of their hearings and men in 26 percent, a difference that is not significant, and judges ask somewhat less often the longer they have served (-5.9 points per ten years, $q = 0.065$). Speaking time as a whole belongs to the presiding judge, who speaks for 5.5 minutes in an average hearing against 0.5 minutes for another judge.

Camera time: A judge's camera time is the number of seconds in which the pipeline verifies her on screen, as a percentage of the hearing. The sample consists of the 113 judges whom the pipeline verifies at least once, in 306 hearings (4,747 judge-hearings and 103 hours on screen). In 57 percent of these judge-hearings the judge is never verified on screen. The presiding judge receives 5.6 percent of a hearing and 27 percent of all the seconds in which any judge is verified. The two judges who follow her in the order of precedence receive 3.6 percent each, because 82 percent of their time on screen is spent in the frame with her, against 4 percent for the other judges. With hearing fixed effects, these two places are worth 2.8 points (95% CI 2.0 to 3.6), and 2.4 points when the same judge is compared across hearings. Further down the order, a place makes almost no difference (-0.01 points per place), and neither does it in the seconds in which a judge is alone on screen ($p = 0.93$). A judge who asks a question gains 1.3 points. Women do not receive less camera time than men in the same role (-0.13 points, 95% CI -0.41 to 0.14).

Table A7 relates asking, camera time, and the three displays to eight characteristics of the judge, one at a time, and Table A8 names the ten judges with the most question turns and the ten with the most camera time. The second list consists largely of the Presidents and Vice-Presidents of the Court. We name judges only on these two counts, which are matters of record, and not on any measure of expression.

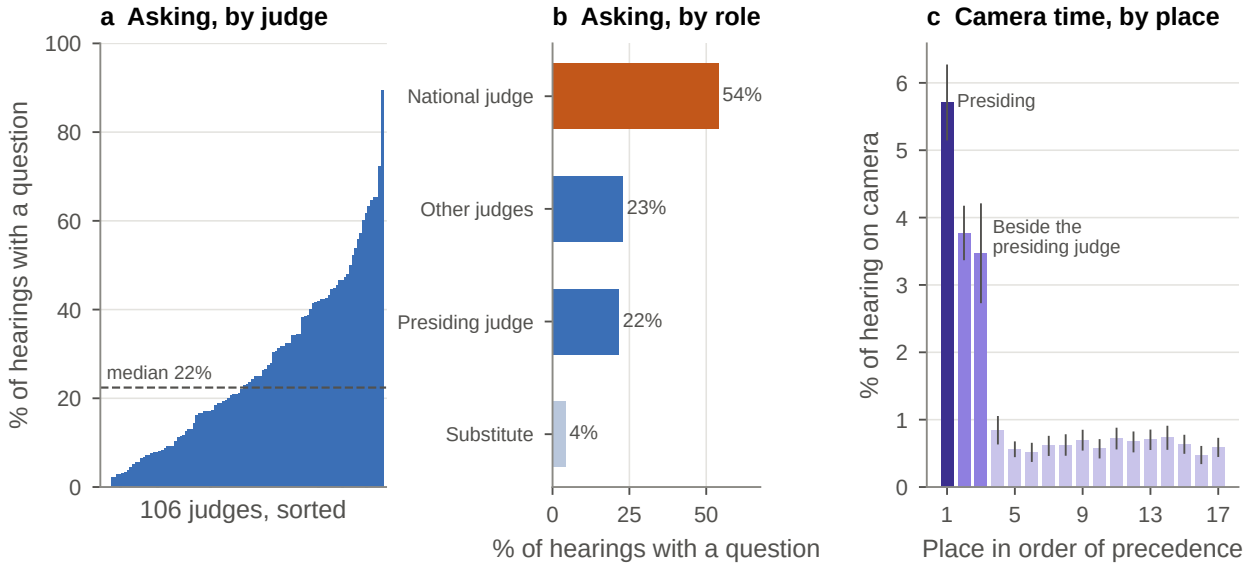

Figure A1: Who asks and who is shown. Note: (a) Judges with ten or more hearings. (c) Grand Chamber; 95% CIs.

Table A7: Asking, camera time, and displays by characteristics of the judge (exploratory)

	Asked a question	Camera share	Negative display	Positive display	Expressivity
Woman	−2.5 (2.9)	−0.13 (0.14)	−11.4 (2.3)	0.92 (0.35)	−6.3 (1.7)
	[.553]	[.525]	[<.001]	[.053]	[.004]
Age (per ten years)	−1.9 (1.7)	0.07 (0.07)	3.3 (1.3)	−0.02 (0.14)	3.2 (1.0)
	[.501]	[.525]	[.053]	[.877]	[.009]
Years on the Court (per ten)	−5.9 (2.4)	−0.60 (0.36)	−2.8 (1.6)	0.31 (0.27)	−0.6 (1.3)
	[.065]	[.316]	[.316]	[.501]	[.776]
Post-socialist state	−2.7 (2.9)	−0.27 (0.12)	3.2 (2.5)	0.04 (0.29)	1.6 (2.0)
	[.532]	[.119]	[.501]	[.877]	[.577]
Southern state	−1.1 (2.9)	0.05 (0.17)	−2.3 (2.3)	−0.27 (0.25)	−2.2 (1.8)
	[.810]	[.828]	[.525]	[.501]	[.501]
Career judge	−3.7 (3.5)	−0.07 (0.23)	−2.6 (2.3)	0.50 (0.32)	−2.0 (1.8)
	[.501]	[.828]	[.501]	[.372]	[.501]
Presiding judge	−2.6 (4.9)	3.48 (0.34)	−0.7 (1.8)	0.17 (0.21)	1.6 (1.3)
	[.743]	[<.001]	[.810]	[.553]	[.501]
National judge	31.3 (4.2)	0.42 (0.16)	2.1 (1.7)	−0.06 (0.24)	1.0 (1.3)
	[<.001]	[.053]	[.501]	[.828]	[.578]
Observations	5,181	4,747	370,651	370,651	370,651
Judges	156	113	112	112	112

Note: Each cell comes from a separate regression of the outcome on one characteristic. Estimates in percentage points, standard errors clustered by judge and hearing in parentheses, Benjamini–Hochberg q-values over the 40 cells in brackets. Asked a question: judge-hearings in hearings with at least one attributed question; hearing fixed effects and indicators for the presiding, national, ad hoc, and substitute judge. Camera share: seconds verified on screen as a percentage of the hearing, judges whom the pipeline verified at least once; the same controls plus asking and the place in the order of precedence. Displays: all verified judge-seconds; hearing fixed effects, image quality, and the judge's gender. Observations and judges are those of the regressions on region; age is documented for fewer judges and earlier career for about half of them.

Table A8: Judges who ask the most questions and judges with the most camera time

(a) Most questions

Judge	Elected in respect of	Hearings	With a question	Question turns	Minutes
Robert Spano	Iceland	73	34	35	101
Luis López Guerra	Spain	74	33	35	42
Síofra O'Leary	Ireland	52	28	33	55
Angelika Nußberger	Germany	91	31	32	53
Paulo Pinto de Albuquerque	Portugal	46	30	30	83
Linos-Alexandre Sicilianos	Greece	76	29	29	42
Pere Pastor Vilanova	Andorra	34	21	28	27
Ksenija Turković	Croatia	53	25	27	28
Ineta Ziemele	Latvia	59	25	25	57
Paul Lemmens	Belgium	43	24	24	43

(b) Most camera time

Judge	Elected in respect of	Hearings	Presided	Hours on screen	% of hearing
Guido Raimondi	Italy	106	42	8.2	4.8
Jean-Paul Costa	France	73	65	6.8	6.8
Dean Spielmann	Luxembourg	109	57	5.3	3.2
Angelika Nußberger	Germany	99	4	4.6	2.7
Nicolas Bratza	United Kingdom	85	17	4.6	3.4
Robert Spano	Iceland	80	20	4.4	2.9
Linos-Alexandre Sicilianos	Greece	83	12	4.3	2.9
Christos Rozakis	Greece	50	9	4.2	5.2
Işıl Karakaş	Turkey	92	0	2.7	1.8
András Sajó	Hungary	67	3	2.6	2.4

Note: (a) Hearings in which the judge sat and the speaker record attributes at least one question to a judge; question turns and minutes of questions attributed to the judge. (b) Hearings of judges whom the pipeline verified at least once; hours in which the judge is verified on screen and the mean share of a hearing. Both lists depend on the coverage of the two records: questions that the speaker record leaves unattributed and judges without a usable portrait are missing.

G Judges and Hearings

Table A9 lists the 161 judges who sat in the hearings and Table A10 the 319 hearings. Both tables are built from the hearing metadata, that is, from the composition of the Court and the case details recorded for each webcast. The records spell the names of eight judges in more than one way (172 spellings in all). We count each judge once and print the name as it appears in the Court's list of judges (European Court of Human Rights, 2026b). A judge counts as verified in a hearing when at least one second of the recording passes the bench-verification rule of Table A1. For three hearings the record repeats the bench of another hearing, so the number of judges on the bench is not informative for them.

Table A9: Judges in the dataset

State	Judge	Years	Seats	Subst.	Pres.	Nat.	Verif.
Albania	Kristaq Traja	2007–08	4	1			0
	Ledi Bianku	2008–18	90	15		1	34
	Darian Pavli	2019–26	23	3			18
Andorra	Josep Casadevall	2007–15	125		10	1	15
	Pere Pastor Vilanova	2015–24	45	10	1		24
	Canòlic Mingorance Cairat	2025–26	4	3			0
Armenia	Alvina Gyulumyan	2007–14	54	11		3	19
	Armen Harutyunyan	2015–25	38	14			12
	Vahe Grigoryan	2025–26	3				0
Austria	Elisabeth Steiner	2007–15	58	20		7	3
	Gabriele Kucsko-Stadlmayer	2016–24	46	4		2	0
	András Jakab	2025–26	5	1			0
Azerbaijan	Khanlar Hajiyev	2007–16	69	16		2	23
	Latif Hüseynov	2017–26	34	7			8
Belgium	Françoise Tulkens	2007–12	79		9	2	28
	Paul Lemmens	2012–21	62	13		11	33
	Frédéric Krenc	2022–26	10	2		1	0
Bosnia and Herzegovina	Ljiljana Mijović	2007–11	35	8		2	2
	Faris Vehabović	2013–26	67	14		3	16
Bulgaria	Snejana Botoucharova	2007–08	7	3			1
	Zdravka Kalaydjieva‡	2008–15	51	14		2	0
	Yonko Grozev	2015–23	58	10		1	9
	Diana Kovatcheva	2024–26	6	1			3
Croatia	Nina Vajić	2007–12	54	4	2	1	21
	Ksenija Turković	2013–21	67	13		7	31
	Davor Derenčinović	2022–26	12	5			7
Cyprus	Loukis Loucaides	2007	3				0
	George Nicolaou	2008–16	66	13			0
	Georgios A. Serghides	2016–26	40	7		1	8
Czechia	Karel Jungwiert	2007–12	42	7			15
	Aleš Pejchal	2012–21	63	12		4	18
	Kateřina Šimáčková	2022–26	16	7		1	3

continued

Table A9 (continued)

State	Judge	Years	Seats	Subst.	Pres.	Nat.	Verif.
Denmark	Peer Lorenzen	2007–14	68	2	6		17
	Jon Fridrik Kjølbro	2014–22	67	6	2	4	28
	Anne Louise Bormann	2024–26	10	4			0
Estonia	Rait Maruste	2007–10	26	12			10
	Julia Laffranque	2011–19	50	10		1	13
	Peeter Roosma	2020–25	18	5		1	2
Finland	Matti Pellonpää	2007	1				0
	Päivi Hirvelä	2007–15	74	15	1	1	24
	Pauliine Koskelo	2016–24	39	8		2	16
	Juha Lavapuro	2025–26	5	2			0
France	Jean-Paul Costa	2007–11	74		66	7	70
	André Potocki	2011–20	66	11		8	31
	Mattias Guyomar	2021–26	23		5	5	18
Georgia	Mindia Ugrekhelidze	2007–08	4			1	2
	Nona Tsotsoria	2008–17	70	17		1	20
	Irakli Adeishvili†	2010	1			1	0
	Lado Chanturia	2018–26	44	10	1	1	0
Germany	Georg Ress	2007	1				1
	Renate Jaeger	2007–10	30	4		5	7
	Angelika Nußberger	2011–19	101	8	5	3	69
	Anja Seibert-Fohr	2020–26	23	3		3	12
Greece	Christos Rozakis	2007–11	51		9	2	44
	Spyridon Flogaitis†	2011	1			1	0
	Linos-Alexandre Sicilianos	2011–20	87	6	12	2	60
	Ioannis Ktistakis	2021–26	25	2		1	11
Hungary	András Baka	2007–10	6	1		1	3
	András Sajó	2008–17	67	11	3	5	38
	Péter Paczolay	2017–26	38	14		4	10
Iceland	Davíd Thór Björgvinsson	2007–13	55	12		1	15
	Robert Spano	2013–22	84	6	20	2	64
	Oddný Mjöll Arnardóttir	2024–26	4	1			1
Ireland	Ann Power-Forde	2008–14	53	18		1	8
	Mary Finlay Geoghegan†	2009	1			1	0
	Peter Charleton†	2013	1			1	1
	Síofra O'Leary	2015–24	61	8	14	1	35
	Úna Ní Raifeartaigh	2024–26	6	2			4
Italy	Vladimiro Zagrebelsky	2007–10	26	6		4	0
	Vitaliano Esposito†	2009	1			1	0
	Guido Raimondi	2010–19	111	3	44	10	94
	Ida Caracciolo†	2017	1			1	0
	Raffaele Sabato	2019–26	27	4		4	11
Latvia	Ineta Ziemele	2007–14	72	4	1	4	39
	Alan Vaughan Lowe†	2009	1			1	0
	Jautrīte Briede†	2015	2			2	0
	Mārtiņš Mits	2015–25	41	9		2	9
	Artūrs Kučs	2025–26	4				2

continued

Table A9 (continued)

State	Judge	Years	Seats	Subst.	Pres.	Nat.	Verif.
Liechtenstein	Mark Villiger	2007–15	96	8	1		43
	Carlo Ranzoni	2015–24	36	9			10
	Alain Chablais	2025–26	5	2			1
Lithuania	Danutė Jočienė	2007–13	35	9			8
	Egidijus Kūris	2014–24	50	8		3	15
	Gediminas Sagatys	2024–26	4				2
Luxembourg	Dean Spielmann	2007–15	112	8	60		77
	Georges Ravarani	2016–24	42	3		2	20
	Stéphane Pisani	2024–25	3				2
Malta	Giovanni Bonello	2007–10	23	4		1	2
	Vincent A. De Gaetano‡	2010–19	78	10			0
	Lorraine Schembri Orland	2020–26	23	7		1	8
Moldova	Stanislav Pavlovschi	2007–08	8				0
	Mihai Poalelungi	2008–15	35	8		5	7
	Valeriu Griţco	2013–21	39	9		2	16
	Diana Sârcu	2023–26	12	3			0
Monaco	Isabelle Berro‡	2007–15	93	12	1		0
	Stéphanie Mourou-Vikström	2016–25	46	10			7
	Sébastien Biancheri	2026	1	1			0
Montenegro	Nebojša Vučinić	2009–18	62	14			23
	Ivana Jelić	2018–26	28	5			5
Netherlands	Egbert Myjer	2007–12	46	5		3	20
	Johannes Silvis	2012–16	33	9		3	10
	Jolien Schukking	2017–26	27	6		2	8
	Corinna Wissels	2026	1	1			0
North Macedonia	Margarita Tsatsa-Nikolovska	2007	1				0
	Mirjana Lazarova Trajkovska	2008–17	88	10	2	2	31
	Jovan Ilievski	2017–26	39	15			6
Norway	Sverre Erik Jebens	2007–11	34	5		2	0
	Erik Møse	2011–18	43	8			17
	Dag Bugge Nordén†	2012–18	3			3	0
	Arnfinn Bårdsen	2019–26	43	4		2	19
Poland	Lech Garlicki	2007–12	34	3	4		11
	Krzysztof Wojtyczek	2012–24	64	12		3	29
	Anna Adamska-Gallant	2025–26	6	1		1	4
Portugal	Ireneu Cabral Barreto	2007–11	31	3			0
	Paulo Pinto de Albuquerque	2011–20	64	14		5	20
	Ana Maria Guerra Martins	2021–26	19	6		1	5
	João Manuel da Silva Miguel†	2025	1			1	0
Romania	Corneliu Bîrsan	2007–13	53	10		4	12
	Florin Streteanu	2013–16	3			2	0
	Iulia Antoanella Motoc‡	2014–22	42	10		3	0
	Sebastian Răduleţu	2024–26	9	1		1	0

continued

Table A9 (continued)

State	Judge	Years	Seats	Subst.	Pres.	Nat.	Verif.
Russia	Anatoly Kovler[‡]	2007–12	35	3		9	0
	Andrey Yuryevich Bushev[†]	2010–11	2			2	0
	Olga Alexandrovna Fedorova[†]	2011	1			1	0
	Dmitry Dedov	2013–21	49	9		13	16
	Bakhtiyar Tuzmukhamedov[†]	2019	1			1	0
	Mikhail Lobov	2022	1			1	1
San Marino	Antonella Mularoni	2007–08	10	3			4
	Kristina Pardalos	2010–18	60	13			19
	Gilberto Felici	2018–26	22	3		1	6
Serbia	Dragoljub Popović	2007–15	46	7		2	15
	Branko Lubarda	2015–24	39	8			9
	Mateja Đurović	2025–26	5	1			0
Slovakia	Ján Šikuta	2007–15	67	24		1	26
	Alena Poláčková	2016–25	40	9	1	1	9
Slovenia	Boštjan M. Zupančič	2007–16	71	16	5	3	29
	Marko Bošnjak	2016–25	51	3	8	1	26
	Boštjan Zalar[†]	2018	1			1	0
	Vasilka Sancin	2025–26	4	2			2
Spain	Javier Borrego Borrego	2007	5	1			2
	Luis López Guerra	2008–18	91	12	5	3	54
	Alejandro Saiz Arnaiz[†]	2009–11	2			2	0
	María Elósegui	2018–26	22	4		4	14
Sweden	Elisabet Fura[‡]	2007–12	39	5		1	0
	Helena Jäderblom	2012–18	49	8		3	24
	Erik Wennerström	2019–26	31	5		1	7
Switzerland	Luzius Wildhaber	2007	1		1		0
	Giorgio Malinverni	2008–11	33	5		3	0
	Helen Keller[‡]	2011–20	53	11		6	0
	Andreas Zünd	2021–26	17	2		5	8
Turkey	Rıza Türmen	2007–08	11			6	4
	Gönül Erönen[†]	2008	1			1	0
	Işıl Karakaş	2008–18	96	12		12	53
	Saadet Yüksel	2019–26	28	6		5	9
Ukraine	Volodymyr Butkevych	2007–08	9	1			0
	Ganna Yudkivska	2010–22	77	16		4	34
	Mykola Gnatovskyy	2022–24	8	1			4
United Kingdom	Nicolas Bratza	2007–12	86		17	15	61
	Paul Mahoney	2012–16	31	4		6	11
	Tim Eicke	2016–25	46	4		6	17
	Hugh Mercer	2026	1				1
161 judges, 47 states		2007–26	5,714	913	316	320	2,043

One row per judge, from the composition of the Court recorded for each hearing. State is the state in respect of which the judge was elected or, for ad hoc judges, appointed. Years are the first and last year in which the judge sat in a hearing of the dataset. Seats counts hearings on the bench, of which Subst. as substitute, Pres. presiding, and Nat. in a case against the judge's own state. Verif. counts the hearings in which the judge was verified on screen. [†]Sat only as ad hoc judge. [‡]Left without an official portrait in all or most hearings because the records spell the name in more than one way; never verified. Names follow the Court's list of judges (European Court of Human Rights, 2026b).

Table A10: Hearings in the dataset

Date	Case	Form.	Min.	Bench	Verif.
2007-06-28	Maumousseau and Washington v. France (no. 39388/05)	Ch.	93	9	2
2007-06-29	O'Halloran and Francis v. United Kingdom (nos. 15809/02 and 25624/02)[a]	GC	5	20	1
2007-07-04	Korbely v. Hungary (no. 9174/02)	GC	98	19	4
2007-07-11	Saadi v. Italy (no. 37201/06)	GC	167	20	6
2007-09-04	The Georgian Labour Party v. Georgia (II) (no. 9103/04)	Ch.	90	8	1
2007-09-12	Burden and Burden v. United Kingdom (no. 13378/05)	GC	87	20	3
2007-09-20	Kononov v. Latvia (no. 36376/04) (Part I)	Ch.	–	8	–
2007-09-20	Kononov v. Latvia (no. 36376/04) (Part II)	Ch.	–	8	–
2007-09-26	N. v. United Kingdom (no. 26565/05)	GC	87	19	3
2007-10-18	Tătar and Tătar v. Romania (no. 67021/01)	Ch.	107	8	4
2007-11-14	Kovačič and others v. Slovenia (nos. 44574/98, 45133/98 and 48316/99)	GC	168	20	10
2007-11-21	Yumak and Sadak v. Turkey (no. 10226/03)	GC	93	20	3
2007-11-27	Patriarcat Oecuménique (Fener Rum Patrikligi) v. Turkey (no. 14340/05)	Ch.	102	9	7
2008-01-08	Tahery and Al-Khawaja v. United Kingdom (nos. 22228/06 and 26766/05)	Ch.	111	8	1
2008-01-15	Kart v. Turkey (no. 8917/05)	Ch.	92	9	5
2008-01-16	Demir and Baykara v. Turkey (no. 34503/97)	GC	102	20	9
2008-02-06	Maslov v. Austria (no. 1638/03)	GC	93	20	4
2008-02-27	S. and Michael Marper v. United Kingdom (nos. 30562/04 and 30566/04)	GC	100	17	5
2008-02-28	S.H. and others v. Austria (no. 57813/00)	Ch.	90	9	2
2008-03-19	Salduz v. Turkey (no. 36391/02)	GC	78	20	4
2008-03-26	Sergeï Zolotoukhine v. Russia (no. 14939/03)	GC	84	20	5
2008-04-02	Šilih v. Slovenia (no. 71463/01)	GC	76	20	5
2008-04-22	Turgut and others v. Turkey (no. 1411/03) and Köktepe v. Turkey (no. 35785/03)	Ch.	135	9	4
2008-04-30	Léger v. France (no. 19324/02)	GC	84	20	4
2008-05-13	Medvedyev and others v. France (no. 3394/03)	Ch.	75	10	4
2008-05-14	Paladi v. Moldova (no. 39806/05)	GC	117	20	5
2008-05-21	A. and others v. United Kingdom (no. 3455/05)	GC	110	20	4
2008-06-11	Gorou v. Greece (no. 2) (no. 12686/03)	GC	79	20	4
2008-06-18	Bykov v. Russia (no. 4378/02)	GC	86	20	3
2008-06-19	Association of citizens "Radko" and Paunkovski v. "the former Yugoslav Republic of Macedonia" (no. 74651/01)	Ch.	85	9	2
2008-06-25	Andrejeva v. Latvia (no. 55707/00)	GC	101	20	5
2008-06-26	TV Vest AS & Rogaland Pensjonistparti v. Norway (no. 21132/05)	Ch.	73	9	2
2008-07-01	M. v. Germany (no. 19359/04)	Ch.	64	9	5
2008-07-02	Kozacıoğlu v. Turkey (no. 2334/03)	GC	104	20	7
2008-07-09	Verein Gegen Tierfabriken Schweiz (VGT) v. Switzerland (no. 32772/02)	GC	46	20	3
2008-09-24	Mooren v. Germany (no. 11364/03)	GC	75	20	2
2008-10-07	Opuz v. Turkey (no. 33401/02)	Ch.	149	10	2
2008-10-22	Micallef v. Malta (no. 17056/06)	GC	102	20	5
2008-11-05	Enea v. Italy (no. 74912/01)	GC	102	20	3
2008-11-19	Varnava v. Turkey (no. 16064/90)	GC	127	20	5

continued

Table A10 (continued)

Date	Case	Form.	Min.	Bench	Verif.
2008-12-02	Savino v. Italy (no. 17214/05), Borgo and others v. Italy (no. 42113/04), Persichetti v. Italy (no. 20329/05)	Ch.	124	8	3
2009-01-07	Scoppola v. Italy (no. 10249/03)	GC	77	20	4
2009-02-11	Depalle v. France and Brosset Triboulet and others v. France (nos. 34044/02 and 34078/02)	GC	82	20	2
2009-03-04	Kart v. Turkey (no. 8917/05)	GC	101	20	6
2009-03-10	Suljagić v. Bosnia and Herzegovina (no. 27912/02)	Ch.	99	10	3
2009-03-18	Gäfgen v. Germany (no. 22978/05)	GC	101	20	5
2009-03-24	Ólafsson v. Iceland (no. 20161/06)	Ch.	109	10	4
2009-04-01	Oršuš and others v. Croatia (no. 15766/03)	GC	94	20	11
2009-04-16	Georgia v. Russia (I) (no. 13255/07)	Ch.	95	10	3
2009-05-06	Medvedyev and others v. France (no. 3394/03)	GC	87	20	3
2009-05-12	Gillan and Quinton v. United Kingdom (no. 4158/05)	Ch.	104	10	2
2009-05-20	Kononov v. Latvia (no. 36376/04)	GC	121	20	1
2009-05-26	Muñoz Díaz v. Spain (no. 49151/07)	Ch.	109	9	4
2009-06-03	Sejdić and Finci v. Bosnia and Herzegovina (no. 27996/06)	GC	92	20	3
2009-06-09	Catan and Others v. Republic of Moldova and Russia (nos. 43370/04, 8252/05 and 18454/06)	Ch.	117	11	4
2009-06-17	Guiso-Gallisay v. Italy (no. 58858/00)	GC	91	20	4
2009-07-01	Cudak v. Lithuania (no. 15869/02)	GC	78	20	5
2009-09-02	Carson and others v. United Kingdom (no. 42184/05)	GC	98	20	5
2009-09-16	Tănase v. Moldova (no. 7/08)	GC	54	20	2
2009-09-23	Mangouras v. Spain (no. 12050/04)	GC	75	20	2
2009-10-07	Neulinger and Shuruk v. Switzerland (no. 41615/07)	GC	107	20	7
2009-10-21	Taxquet v. Belgium (no. 926/05)	GC	90	20	3
2009-11-10	Stanev v. Bulgaria (no. 36760/06) and Mitev v. Bulgaria (no. 42758/07)	Ch.	83	8	2
2009-11-18	Demopoulos v. Turkey and 7 other cases (nos. 46113/99, 3843/02, 13751/02, 13466/03, 14163/04, 10200/04, 19993/04 and 21819/04)	GC	140	20	3
2009-12-09	A, B and C v. Ireland (no. 25579/05)	GC	104	20	5
2009-12-16	Şerife Yiğit v. Turkey (no. 3976/05)	GC	90	20	6
2010-01-06	Sanoma Uitgevers B.V. v. the Netherlands (no. 38224/03)	GC	102	20	4
2010-01-20	Sakhnovskiy v. Russia (no. 21272/03)	GC	96	20	4
2010-02-25	Schalk and Kopf v. Austria (no. 30141/04)	Ch.	60	9	2
2010-03-03	McFarlane v. Ireland (no. 31333/06)	GC	96	20	5
2010-03-04	OAO Neftyanaya kompaniya YUKOS v. Russia (no. 14902/04)	Ch.	214	9	2
2010-04-27	Enoukidze and Guirgvliani v. Georgia (no. 25091/07)	Ch.	116	9	4
2010-04-27	Farcaş v. Romania (no. 32596/04)	Ch.	124	9	5
2010-04-28	Paksas v. Lithuania (no. 34932/04)	GC	80	20	2
2010-05-19	Al-Khawaja and Tahery v. United Kingdom (no. 26766/05)	GC	93	20	4
2010-06-08	Atanasiu and Poenaru v. Romania (no. 30767/05) and Solon v. Romania (no. 33800/06)	Ch.	135	9	1
2010-06-09	Al-Skeini and Others v. the United Kingdom (no. 55721/07) and Al-Jedda v. the United Kingdom (no. 27021/08)	GC	184	20	6
2010-06-30	Lautsi v. Italy (no. 30814/06)	GC	124	20	7
2010-09-01	M.S.S. v. Belgium and Greece (no. 30696/09)	GC	181	20	8
2010-09-15	Chiragov and Others v. Armenia (no. 13216/05)	GC	142	20	8
2010-09-15	Sargsyan v. Azerbaijan (no. 40167/06)	GC	103	20	5
2010-09-29	Giuliani and Gaggio v. Italy (no. 23458/02)	GC	107	20	5

continued

Table A10 (continued)

Date	Case	Form.	Min.	Bench	Verif.
2010-10-13	Springer v. Germany (no. 39954/08), von Hannover v. Germany (no. 40660/08)	GC	–	20	–
2010-11-03	Stummer v. Austria (no. 37452/02)	GC	88	20	6
2010-11-23	Koch v. Germany (no. 497/09)	Ch.	87	9	5
2010-11-24	Bayatyan v. Armenia (no. 23459/03)	GC	75	20	5
2010-11-30	Uzan and Others v. Turkey (no. 18240/03)	Ch.	172	9	5
2010-12-08	Palomo Sánchez and Others v. Spain (nos. 28955/06, 28957/06, 28959/06 and 28964/06)	GC	103	20	7
2010-12-14	Othman v. United Kingdom (no. 8139/09)	Ch.	93	7	2
2011-01-11	Mosley v. the United Kingdom (no. 48009/08)	Ch.	91	10	7
2011-01-12	Kotov v. Russia (no. 54522/00)	GC	87	20	5
2011-02-09	Stanev v. Bulgaria (no. 36760/06)	GC	93	20	3
2011-02-23	S.H. and others v. Austria (no. 57813/00)	GC	81	20	4
2011-03-09	Nejdet Şahin and Perihan Şahin v. Turkey (no. 13279/05)	GC	82	20	4
2011-03-22	V.C. v. Slovakia (no. 18968/07)	Ch.	93	7	6
2011-03-23	Nada v. Switzerland (no. 10593/08)	GC	119	20	5
2011-03-30	Creangă v. Romania (no. 29226/03)	GC	56	20	2
2011-04-12	Gas and Dubois v. France (no. 25951/07)	Ch.	106	10	3
2011-04-13	Aksu v. Turkey (no. 4149/04)	GC	117	20	11
2011-05-04	Sitaropoulos and Others v. Greece (no. 42202/07)	GC	79	20	9
2011-05-17	I.M. v. France (no. 9152/09)	Ch.	105	10	2
2011-05-18	Van der Heijden v. the Netherlands (no. 42857/05)	GC	93	20	6
2011-06-08	Konstantin Markin v. Russia (no. 30078/06)	GC	98	20	7
2011-06-21	Lindheim and Others v. Norway (no. 13221/08), Henriksen v. Norway (no. 2139/10)	Ch.	79	9	4
2011-06-22	Hirsi and Others v. Italy (no. 27765/09)	GC	112	20	5
2011-07-06	Kurić and Others v. Slovenia (no. 26828/06)	GC	111	20	8
2011-08-31	Boulois v. Luxembourg (no. 37575/04)	GC	79	20	1
2011-09-14	Austin and Others v. the United Kingdom (no. 39692/09)	GC	110	20	5
2011-09-22	Georgia v. Russia (II) (no. 38263/08)[c]	Ch.	100	20	0
2011-09-28	Gillberg v. Sweden (no. 41723/06)	GC	93	20	6
2011-10-06	Janowiec and Others v. Russia (nos. 55508/07 and 29520/09)	Ch.	120	8	4
2011-10-12	Centro Europa 7 S.R.L. v. Italy and Di Stefano v. Italy (no. 38433/09)	GC	123	20	6
2011-10-19	Idalov v. Russia (no. 5826/03)	GC	61	20	3
2011-11-02	Scoppola v. Italy (no. 3) (no. 126/05)	GC	94	20	4
2011-11-16	Mouvement raëlien suisse v. Switzerland (no. 16354/06)	GC	97	20	5
2011-11-22	Fernández Martínez v. Spain (no. 56030/07)	Ch.	107	10	7
2011-11-30	Herrmann v. Germany (no. 9300/07)	GC	88	20	6
2011-12-01	X. and Others v. Austria (no. 19010/07)	Ch.	76	9	3
2012-01-11	Vistiņš and Perepjolkins v. Latvia (no. 71243/01) (just satisfaction)	GC	111	20	8
2012-01-25	Catan and Others v. Republic of Moldova and Russia (nos. 43370/04, 8252/05 and 18454/06)	GC	116	20	3
2012-03-07	Animal Defenders International v. United Kingdom (no. 48876/08)	GC	83	20	3
2012-03-21	De Souza Ribeiro v. France (no. 22689/07)	GC	100	20	8
2012-04-04	Fabris v. France (no. 16574/08) (just satisfaction)	GC	74	20	4
2012-04-17	Lutsenko v. Ukraine (no. 6492/11)	Ch.	73	9	4
2012-05-16	El-Masri v. “the former Yugoslav Republic of Macedonia” (no. 39630/09)	GC	94	21	7
2012-06-12	Oleksandr Volkov v. Ukraine (no. 21722/11)	Ch.	94	8	8
2012-06-13	Georgia v. Russia (I) (no. 13255/07) (just satisfaction)	GC	128	21	8

continued

Table A10 (continued)

Date	Case	Form.	Min.	Bench	Verif.
2012-06-19	Telegraaf Media Nederland Landelijke Media B.V. and Others v. the Netherlands (no. 39315/06)	Ch.	107	9	4
2012-08-28	Tymoshenko v. Ukraine (no. 49872/11)	Ch.	94	8	7
2012-09-04	Ladele v. United Kingdom (no. 51671/10), McFarlane v. United Kingdom (no. 36516/10), Eweida v. United Kingdom (no. 48420/10) & Chaplin v. United Kingdom (no. 59842/10)	Ch.	137	9	6
2012-09-18	Vilnes and Others v. Norway (nos. 52806/09 and 22703/10)	Ch.	80	9	3
2012-10-02	Michaud v. France (no. 12323/11)	Ch.	95	8	4
2012-10-03	X. and Others v. Austria (no. 19010/07)	GC	74	21	7
2012-10-10	X. v. Latvia (no. 27853/09)	GC	47	20	7
2012-11-07	Sindicatul "Păstorul cel Bun" v. Romania (no. 2330/09)	GC	114	20	10
2012-11-14	Allen v. United Kingdom (no. 25424/09)	GC	89	20	5
2012-11-28	Vinter and Others v. United Kingdom (nos. 66069/09, 130/10 and 3896/10)	GC	100	20	9
2012-12-12	Maktouf and Damjanović v. Bosnia and Herzegovina (nos. 2312/08 and 34179/08)	GC	82	20	5
2013-01-16	Vallianatos and Others v. Greece (nos. 29381/09 and 32684/09)	GC	99	20	4
2013-01-30	Fernández Martínez v. Spain (no. 56030/07)	GC	113	19	10
2013-02-13	Janowiec and Others v. Russia (nos. 55508/07 and 29520/09)	GC	91	20	2
2013-03-06	O'Keeffe v. Ireland (no. 35810/09)	GC	90	20	6
2013-03-20	Del Río Prada v. Spain (no. 42750/09)	GC	98	20	5
2013-04-03	Söderman v. Sweden (no. 5786/08)	GC	95	20	7
2013-05-15	Vučković and Others v. Serbia (no. 17153/11)	GC	97	20	5
2013-06-26	Marguš v. Croatia (no. 4455/10)	GC	66	20	5
2013-07-10	Ališić and Others v. Bosnia and Herzegovina, Croatia, Serbia, Slovenia and "the Former Yugoslav Republic of Macedonia" (no. 60642/08)	GC	140	20	11
2013-09-04	Center for Legal Resources on behalf of Valentin Câmpeanu v. Romania (no. 47848/08)[a]	GC	112	20	14
2013-09-10	Dubská and Krejzová v. the Czech Republic (nos. 28859/11 and 28473/12)	Ch.	98	9	8
2013-09-18	Svinarenko and Slyadnev v. Russia (nos. 32541/08 and 43441/08)[a]	GC	81	20	2
2013-10-02	Mocanu and Others v. Romania (nos. 10865/09, 45886/07 and 32431/08)	GC	–	20	–
2013-10-15	Natsvlishvili and Togonidze v. Georgia (no. 9043/05)[c]	Ch.	97	20	3
2013-10-16	Hämäläinen v. Finland (no. 37359/09)[c]	GC	98	20	5
2013-11-13	Jeunesse v. the Netherlands (no. 12738/10)	GC	105	20	12
2013-11-27	S.A.S. v. France (no. 43835/11)	GC	100	20	8
2013-12-03	Al Nashiri v. Poland (no. 28761/11) and Husayn (Abu Zubaydah) v. Poland (no. 7511/13)	Ch.	160	10	6
2013-12-11	Hassan v. the United Kingdom (no. 29750/09)	GC	118	20	14
2014-01-22	Chiragov and Others v. Armenia (no. 13216/05)	GC	143	20	14
2014-02-05	Sargsyan v. Azerbaijan (no. 40167/06)	GC	105	20	11
2014-02-12	Tarakhel v. Switzerland (no. 29217/12)	GC	113	20	9
2014-02-19	Jaloud v. the Netherlands (no. 47708/08)	GC	133	20	12
2014-04-16	Mustafa Tunç and Fecire Tunç v. Turkey (no. 24014/05)	GC	99	20	11
2014-05-21	Morice v. France (no. 29369/10)	GC	90	20	8
2014-06-04	Vasiliauskas v. Lithuania (no. 35343/05)	GC	97	20	14
2014-06-18	Parrillo v. Italy (no. 46470/11)	GC	107	20	12
2014-07-09	Delfi AS v. Estonia (no. 64569/09)	GC	101	20	11

continued

Table A10 (continued)

Date	Case	Form.	Min.	Bench	Verif.
2014-09-03	Khoroshenko v. Russia (no. 41418/04)	GC	78	20	11
2014-09-24	Roman Zakharov v. Russia (no. 47143/06)	GC	100	20	11
2014-10-08	Bouyid v. Belgium (no. 23380/09)	GC	82	20	6
2014-10-14	Tagayeva and Others v. Russia (no. 26562/07)	Ch.	118	10	6
2014-11-12	Gherghina v. Romania (no. 42219/07)	GC	114	20	11
2014-12-03	F.G. v. Sweden (no. 43611/11)	GC	94	20	10
2014-12-10	Al-Dulimi and Montana Management Inc. v. Switzerland (no. 5809/08)	GC	158	20	8
2014-12-17	Pentikäinen v. Finland (no. 11882/10)	GC	104	20	14
2015-01-07	Lambert and Others v. France (no. 46043/14)	GC	124	20	5
2015-01-14	Murray v. the Netherlands (no. 10511/10)	GC	108	20	9
2015-01-21	Dvorski v. Croatia (no. 25703/11)	GC	92	20	11
2015-01-28	Perinçek v. Switzerland (no. 27510/08)	GC	134	20	6
2015-02-04	Mozer v. the Republic of Moldova and Russia (no. 11138/10)	GC	116	20	7
2015-03-04	Schatschaschwili v. Germany (no. 9154/10)	GC	93	20	7
2015-04-01	Biao v. Denmark (no. 38590/10)	GC	94	20	9
2015-04-08	Avotiņš v. Latvia (no. 17502/07)	GC	114	21	11
2015-04-15	Couderc and Hachette Filipacchi Associés v. France (no. 40454/07)	GC	91	21	3
2015-05-13	Bédat v. Switzerland (no. 56925/08)	GC	106	21	7
2015-05-19	Gouarré Patte v. Andorra (no. 33427/10)	Ch.	63	7	5
2015-05-20	de Tommaso v. Italy (no. 43395/09)	GC	119	21	9
2015-06-03	İzzettin Doğan and Others v. Turkey (no. 62649/10)	GC	106	21	4
2015-06-10	Armani Da Silva v. the United Kingdom (no. 5878/08)	GC	100	21	12
2015-06-17	Baka v. Hungary (no. 20261/12)	GC	88	21	4
2015-06-23	Nasr and Ghali v. Italy (no. 44883/09)	Ch.	89	9	5
2015-07-01	Jeronovičs v. Latvia (no. 44898/10)	GC	66	21	3
2015-07-08	Karácsony and Others v. Hungary (no. 42461/13) and Szél and Others v. Hungary (no. 44357/13)	GC	84	21	8
2015-09-02	G.I.E.M s.r.l v. Italy (no. 1828/06), Hotel Promotion Bureau s.r.l. and Rita Sarda s.r.l. v. Italy (no. 34163/07) and Falgest s.r.l. and Gironda v. Italy (no. 19029/11)	GC	156	21	10
2015-09-16	Paposhvili v. Belgium (no. 41738/10)	GC	102	21	7
2015-10-07	Buzadji v. Republic of Moldova (no. 23755/07)	GC	85	21	14
2015-10-21	Hutchinson v. the United Kingdom (no. 57592/08)	GC	87	21	8
2015-11-04	Magyar Helsinki Bizottság v. Hungary (no. 18030/11)	GC	100	21	12
2015-11-25	Ibrahim and Others v. the United Kingdom (nos. 50541/08, 50571/08, 50573/08 and 40351/09)	GC	130	21	11
2015-12-02	Dubská and Krejzová v. the Czech Republic (nos. 28859/11 and 28473/12)	GC	100	21	8
2015-12-09	Paradiso and Campanelli v. Italy (no. 25358/12)	GC	110	21	15
2015-12-16	Béláné Nagy v. Hungary (no. 53080/13)	GC	86	21	11
2016-01-06	Muršić v. Croatia (no. 7334/13)	GC	73	21	8
2016-01-13	A and B v. Norway (nos. 24130/11 and 29758/11)	GC	83	21	9
2016-01-27	Lhermitte v. Belgium (no. 34238/09)	GC	84	21	8
2016-02-24	J.K. and Others v. Sweden (no. 59166/12)	GC	69	21	7
2016-03-02	Lupeni Greek Catholic Parish and Others v. Romania (no. 76943/11)	GC	110	21	11
2016-04-20	Khamtokhu and Aksenchik v. Russia (nos. 60367/08 and 961/11)	GC	86	21	11
2016-05-25	V.M. and Others v. Belgium (no. 60125/11)	GC	94	21	9
2016-06-01	Moreira Ferreira v. Portugal (no. 2) (no. 19867/12)	GC	79	21	11
2016-06-22	Khlaifia and Others v. Italy (no. 16483/12)	GC	127	21	12

continued

Table A10 (continued)

Date	Case	Form.	Min.	Bench	Verif.
2016-06-29	Al Nashiri v. Romania (no. 33234/12)	Ch.	–	9	–
2016-06-29	Abu Zubaydah v. Lithuania (no. 46454/11)	Ch.	100	9	6
2016-07-06	Simeonovi v. Bulgaria (no. 21980/04)	GC	101	20	16
2016-08-31	Medžlis Islamske zajednice Brčko and Others v. Bosnia and Herzegovina (no. 17224/11)	GC	75	20	13
2016-09-14	Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland (no. 931/13)	GC	92	21	10
2016-10-12	Károly Nagy v. Hungary (no. 56665/09)	GC	108	20	11
2016-10-19	Regner v. the Czech Republic (no. 35289/11)	GC	108	20	15
2016-11-09	Fábián v. Hungary (no. 78117/13)	GC	118	20	14
2016-11-16	Lopes de Sousa Fernandes v. Portugal (no. 56080/13)	GC	118	18	13
2016-11-30	Bărbulescu v. Romania (no. 61496/08)	GC	118	20	15
2017-01-11	Harkins v. the United Kingdom (no. 71537/14)	GC	128	21	11
2017-01-25	Garib v. the Netherlands (no. 43494/09)	GC	96	19	6
2017-02-08	Correia de Matos v. Portugal (no. 56402/12)	GC	102	20	12
2017-03-08	Merabishvili v. Georgia (no. 72508/13)	GC	99	20	5
2017-03-22	Ramos Nunes De Carvalho e Sá v. Portugal (no. 55391/13)	GC	87	20	7
2017-05-23	Sharxhi and Others v. Albania (no. 10613/16)	Ch.	133	8	7
2017-06-14	Naït-Liman v. Switzerland (no. 51357/07)	GC	104	20	11
2017-07-12	Zubac v. Croatia (no. 40160/12)	GC	71	20	5
2017-10-18	Denisov v. Ukraine (no. 76639/11)	GC	88	20	8
2017-11-07	Big Brother Watch and Others v. the United Kingdom (no. 58170/13), Bureau of Investigative Journalism and Alice Ross v. the United Kingdom (no. 62322/14) and 10 Human Rights Organisations and Others v. the United Kingdom (no. 24960/15)	Ch.	174	9	8
2017-11-15	Nicolae Virgiliu Tănase v. Romania (no. 41720/13)	GC	68	20	10
2017-11-22	Berlusconi v. Italy (no. 58428/13)	GC	114	20	9
2017-11-29	Ilnseher v. Germany (nos. 10211/12 and 27505/14)	GC	96	20	10
2017-12-06	Molla Sali v. Greece (no. 20452/14)	GC	104	20	11
2017-12-20	Beuze v. Belgium (no. 71409/10)	GC	86	20	14
2018-01-17	S., V. and A. v. Denmark (no. 35553/12)	GC	95	20	7
2018-01-24	Navalnyy v. Russia (nos. 29580/12, 36847/12, 11252/13, 12317/13 and 43746/14)	GC	98	20	8
2018-02-14	Murtazaliyeva v. Russia (no. 36658/05)	GC	89	20	14
2018-03-07	Fernandes de Oliveira v. Portugal (no. 78103/14)	GC	95	20	9
2018-03-14	Lekić v. Slovenia (no. 36480/07)	GC	117	20	11
2018-03-28	Güzelyurtlu and Others v. Cyprus and Turkey (no. 36925/07)	GC	156	20	5
2018-04-18	Ilias and Ahmed v. Hungary (no. 47287/15)	GC	97	20	9
2018-04-18	Z.A. and Others v. Russia (nos. 61411/15, 61420/15, 61427/15 and 3028/16)	GC	–	20	–
2018-05-23	Georgia v. Russia (II) (no. 38263/08)	GC	183	20	5
2018-06-06	Rooman v. Belgium (no. 18052/11)	GC	101	20	8
2018-09-26	N.D. and N.T. v. Spain (no. 8675/15)	GC	173	20	13
2018-10-03	Mihalache v. Romania (no. 54012/10)	GC	55	20	9
2018-10-17	Strand Lobben and Others v. Norway (no. 37283/13)	GC	108	20	9
2018-11-13	Elçi v. Turkey (no. 63129/15), Ahmet Tunç and Others v. Turkey (no. 4133/16) and Tunç and Yerbasan v. Turkey (no. 31542/16)	Ch.	179	9	5
2018-11-21	Magyar Kétfarkú Kutya Párt v. Hungary (no. 201/17)	GC	91	20	8
2018-11-28	López Ribalda and Others v. Spain (no. 1874/13)	GC	121	20	8
2019-01-22	A.M. v. France (no. 12148/18)	Ch.	118	8	4

continued

Table A10 (continued)

Date	Case	Form.	Min.	Bench	Verif.
2019-04-24	M.N. and Others v. Belgium (no. 3599/18)	GC	146	20	6
2019-05-15	S.M. v. Croatia (no. 60561/14)	GC	74	20	8
2019-06-12	Slovenia v. Croatia (no. 54155/16)	GC	90	20	13
2019-07-10	Big Brother Watch and Others v. the United Kingdom (nos. 58170/13, 62322/14 and 24960/15)	GC	168	20	6
2019-07-10	Centrum för rättvisa v. Sweden (no. 35252/08)	GC	–	20	–
2019-09-11	Ukraine v. Russia (re Crimea) (no. 20958/14)	GC	177	20	5
2019-09-18	Selahattin Demirtaş v. Turkey (no. 2) (no. 14305/17)	GC	163	20	11
2019-09-25	Muhammad and Muhammad v. Romania (no. 80982/12)	GC	118	20	7
2019-10-09	Gestur Jónsson and Ragnar Halldór Hall v. Iceland (nos. 68273/14 and 68271/14)	GC	108	20	8
2019-12-04	Mugemangango v. Belgium (no. 310/15)	GC	89	20	4
2019-12-18	Albert and Others v. Hungary (no. 5294/14)	GC	101	20	6
2020-01-15	X and Others v. Bulgaria (no. 22457/16)	GC	111	20	9
2020-02-05	Guðmundur Andri Ástráðsson v. Iceland (no. 26374/18)	GC	125	20	8
2020-02-26	Hanan v. Germany (no. 4871/16)	GC	158	20	10
2020-06-10	M.A. v. Denmark (no. 6697/18)[b]	GC	131	20	3
2020-06-17	Kurt v. Austria (no. 62903/15)[b]	GC	118	20	7
2020-06-24	Savran v. Denmark (no. 57467/15)[b]	GC	110	20	7
2020-07-01	Vavřička v. the Czech Republic (no. 47621/13) and 5 other applications (nos. 3867/14, 73094/14, 19306/15, 19298/15 and 43883/15)	GC	112	20	7
2020-10-14	NIT S.R.L. v. the Republic of Moldova (no. 28470/12)[b]	GC	88	20	7
2020-10-21	Denis and Irvine v. Belgium (nos. 62819/17 and 63921/17)[b]	GC	86	20	5
2021-01-20	Khasanov and Rakhmanov v. Russia (nos. 28492/15 and 49975/15)[b]	GC	96	20	6
2021-01-27	Abdi Ibrahim v. Norway (no. 15379/16)[b]	GC	108	20	6
2021-05-19	Grzęda v. Poland (no. 43572/18)[b]	GC	130	20	6
2021-05-26	Savickis and Others v. Latvia (no. 49270/11)[b]	GC	107	13	4
2021-06-16	Beeler v. Switzerland (no. 78630/12)	GC	106	20	6
2021-07-07	Vegotex International S.A. v. Belgium (no. 49812/09)[b]	GC	105	20	4
2021-09-29	H.F. and Others v. France (nos. 24384/19 and 44234/20)	GC	161	20	6
2021-11-03	L.B. v. Hungary (no. 36345/16)	GC	99	20	8
2022-01-26	Ukraine and the Netherlands v. Russia (nos. 8019/16, 43800/14 and 28525/20)	GC	237	20	9
2022-02-02	Halet v. Luxembourg (no. 21884/18)	GC	116	19	8
2022-02-23	McCallum v. Italy (no. 20863/21)	GC	87	20	4
2022-02-23	Sanchez-Sanchez v. the United Kingdom (no. 22854/20)	GC	138	20	6
2022-03-09	Hurbain v. Belgium (no. 57292/16)	GC	102	20	6
2022-03-23	Macatė v. Lithuania (no. 61435/19)	GC	108	20	8
2022-06-29	Sanchez v. France (no. 45581/15)	GC	134	20	9
2022-09-27	Pietrzak v. Poland and Bychawska-Siniarska and Others v. Poland (nos. 72038/17 and 25237/18)	Ch.	186	8	7
2022-10-18	Pagerie v. France (no. 24203/16)	Ch.	101	10	4
2023-01-18	Yüksel Yalçınkaya v. Türkiye (no. 15669/20)	GC	92	20	0
2023-03-01	Humpert and Others v. Germany (nos. 59433/18, 59477/18, 59481/18 and 59494/18)	GC	116	20	0
2023-03-29	Carême v. France (no. 7189/21)	GC	166	22	9
2023-03-29	Verein KlimaSeniorinnen Schweiz and Others v. Switzerland (no. 53600/20)	GC	96	22	4

continued

Table A10 (continued)

Date	Case	Form.	Min.	Bench	Verif.
2023-04-12	Communauté genevoise d'action syndicale (CGAS) v. Switzerland (no. 21881/20)	GC	124	20	4
2023-07-05	Nealon and Hallam v. the United Kingdom (nos. 32483/19 and 35049/19)	GC	102	20	5
2023-07-12	Fabbri and Others v. San Marino (nos. 6319/21, 6321/21 and 9227/21)	GC	157	20	7
2023-09-27	Duarte Agostinho and Others v. Portugal and 32 Others (no. 39371/20)	GC	251	22	18
2023-11-28	Karsai v. Hungary (no. 32312/23)	Ch.	109	8	7
2023-12-13	Ukraine v. Russia (re Crimea) (nos. 20958/14 and 38334/18)	GC	4	20	1
2024-01-10	Pindo Mulla v. Spain (no. 15541/20)	GC	200	20	10
2024-03-06	Ships Waste Oil Collector B.V. and Others v. the Netherlands (nos. 2799/16, 2800/16, 3124/16 and 3205/16)	GC	135	21	7
2024-05-15	Semenya v. Switzerland (no. 10934/21)	GC	163	22	7
2024-06-04	G.R.J. v. Greece and A.E. v. Greece (nos. 15067/21 and 15783/21)	Ch.	148	9	5
2024-06-12	Ukraine and the Netherlands v. Russia (nos. 8019/16, 43800/14, 28525/20 and 11055/22)	GC	265	21	7
2024-09-18	Mansouri v. Italy (no. 63386/16)	GC	120	20	8
2024-11-20	Kovačević v. Bosnia and Herzegovina (no. 43651/22)	GC	230	20	9
2024-12-18	Danileţ v. Romania (no. 16915/21)	GC	152	20	11
2025-02-12	C.O.C.G. and Others v. Lithuania (no. 17764/22)	GC	151	22	7
2025-02-12	H.M.M. and Others v. Latvia (no. 42165/21)	GC	145	22	7
2025-02-12	R.A. and Others v. Poland (no. 42120/21)	GC	122	13	5
2025-02-26	Tsaava and Others v. Georgia (nos. 13186/20, 16757/20, 20129/21, 20175/21 and 39382/21)	GC	111	22	7
2025-05-07	Yasak v. Türkiye (no. 17389/20)	GC	148	18	6
2025-09-17	Jesus Pinhal v. Portugal (nos. 48047/15 and 2276/20)	GC	154	20	3
2025-11-19	Grande Oriente d'Italia v. Italy (no. 29550/17)	GC	124	20	6
2026-01-21	Kuijt v. the Netherlands (no. 19365/19)	GC	165	20	10
2026-03-04	Tergek v. Türkiye (no. 39631/20)	GC	140	20	6
2026-03-25	Kavala v. Türkiye (no. 2) (no. 2170/24)	GC	164	20	5
2026-04-01	S.O. v. Spain (no. 5742/22)	GC	164	20	8
2026-09-22	Ukrainian Helsinki Human Rights Union on behalf of ten Ukrainian children v. Russia (no. 6719/23)	Ch.	–	9	–
319 hearings			560 h	5,714	2,043

Hearings in order of date; case names and application numbers follow the Court's webcast listing. Form. is the formation, Grand Chamber (GC) or Chamber (Ch.). Min. is the length of the recording in minutes. Bench is the number of judges recorded for the hearing, substitutes included, and Verif. the number of them verified on screen; a dash marks a hearing without a processed recording. [a]Delivery of a judgment. [b]Held by videoconference. [c]The record lists the bench of another hearing: for Georgia v. Russia (II), 22 September 2011, that of the Grand Chamber hearing of 23 May 2018; for Natsvlishvili and Togonidze v. Georgia, 15 October 2013, that of Svinarenko and Slyadnev v. Russia, 18 September 2013; for Hämäläinen v. Finland, 16 October 2013, that of Mocanu and Others v. Romania, 2 October 2013.

H Additional Figures



Figure A2: Judges and hearings.

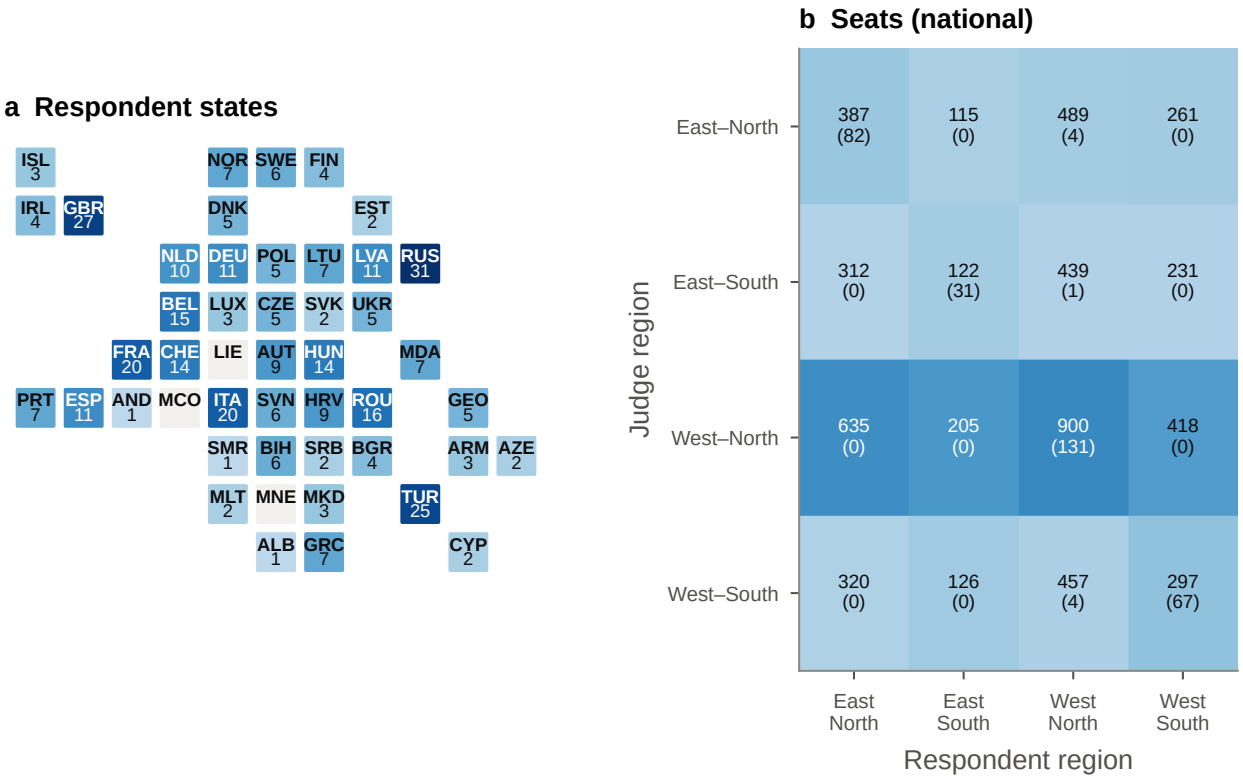

Figure A3: Respondent states and region pairings. Note: National-judge seats in parentheses.

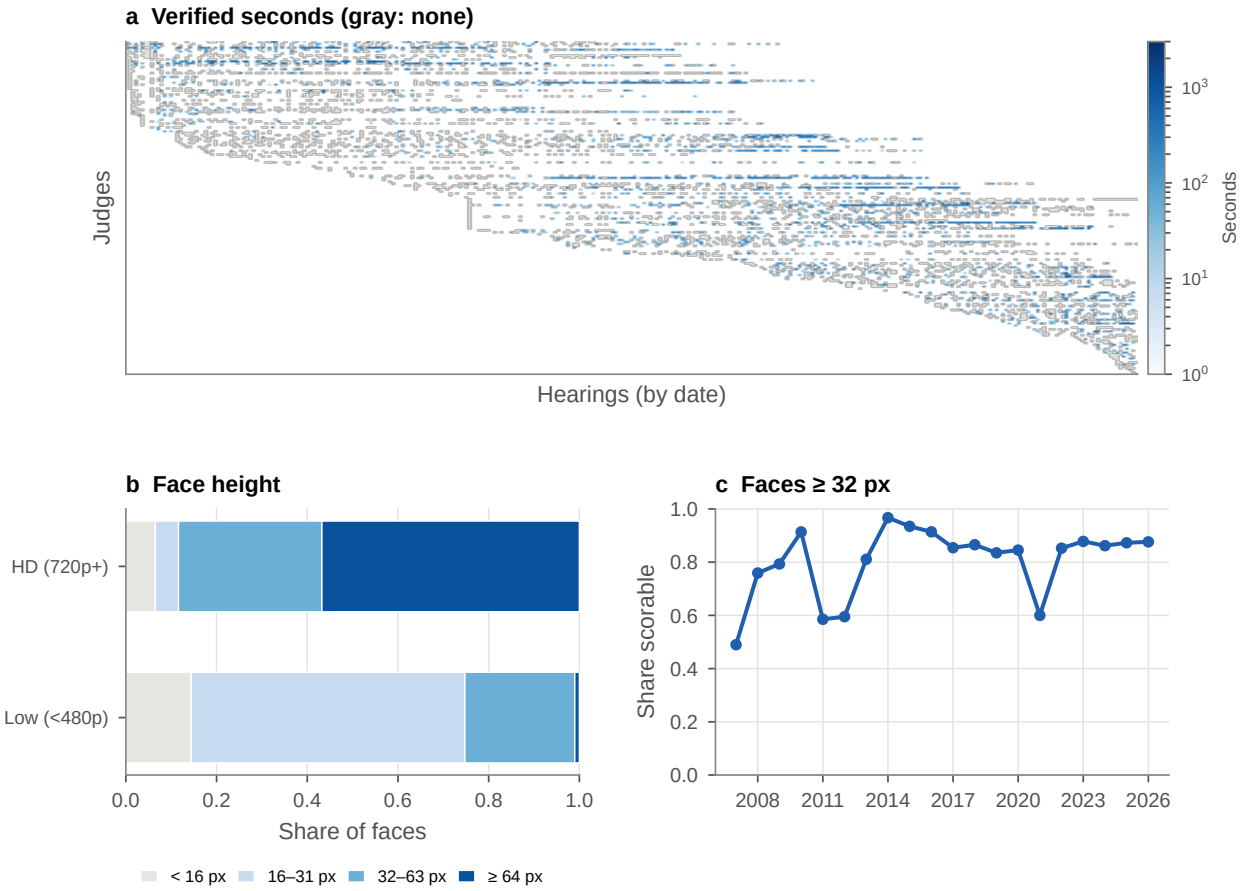

Figure A4: Facial Properties.



Figure A5: Judge crops by predicted class. Note: Labels are the classifier's output with probability. Frames © ECHR-CEDH.

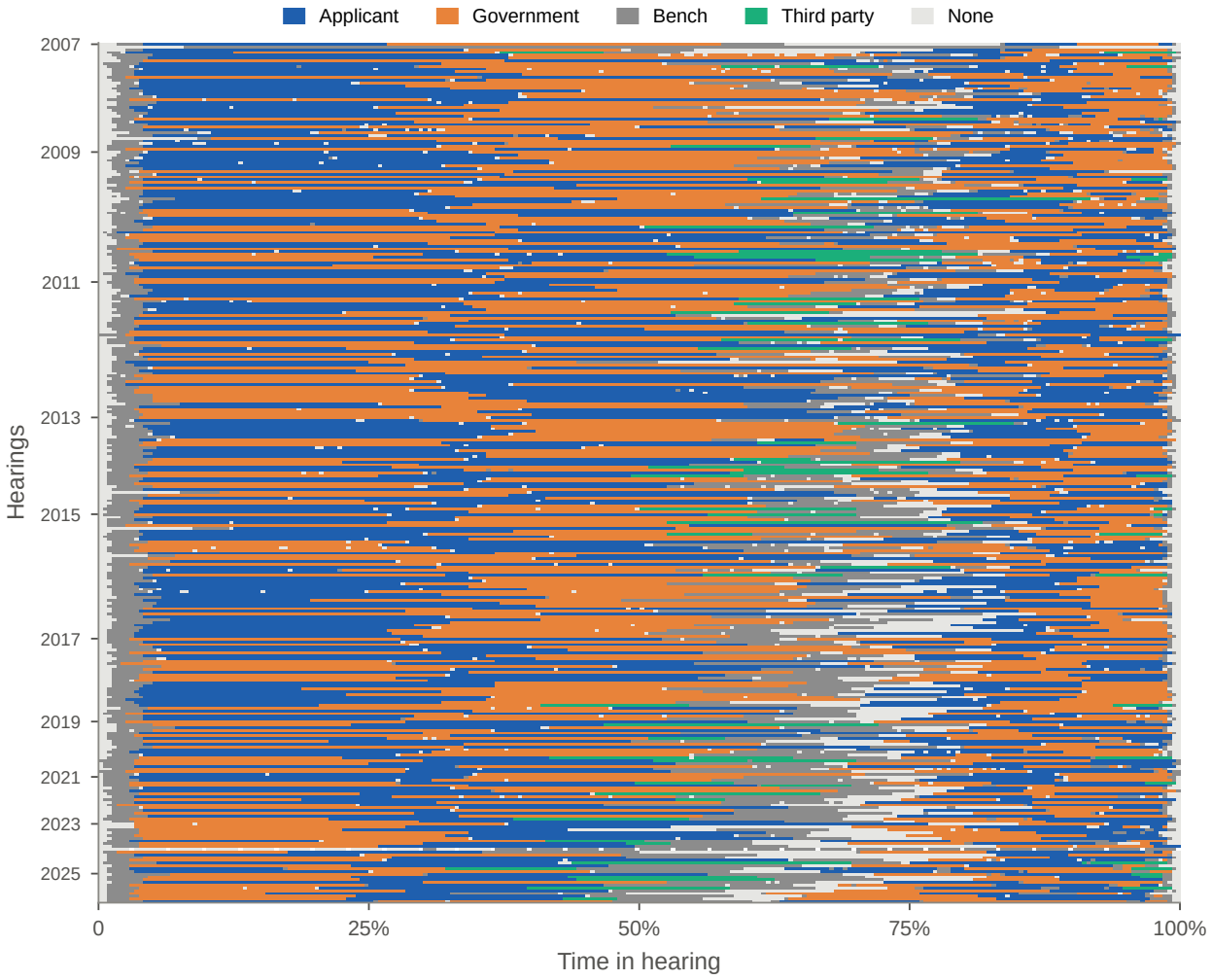

Figure A6: Speaker by second, all hearings.

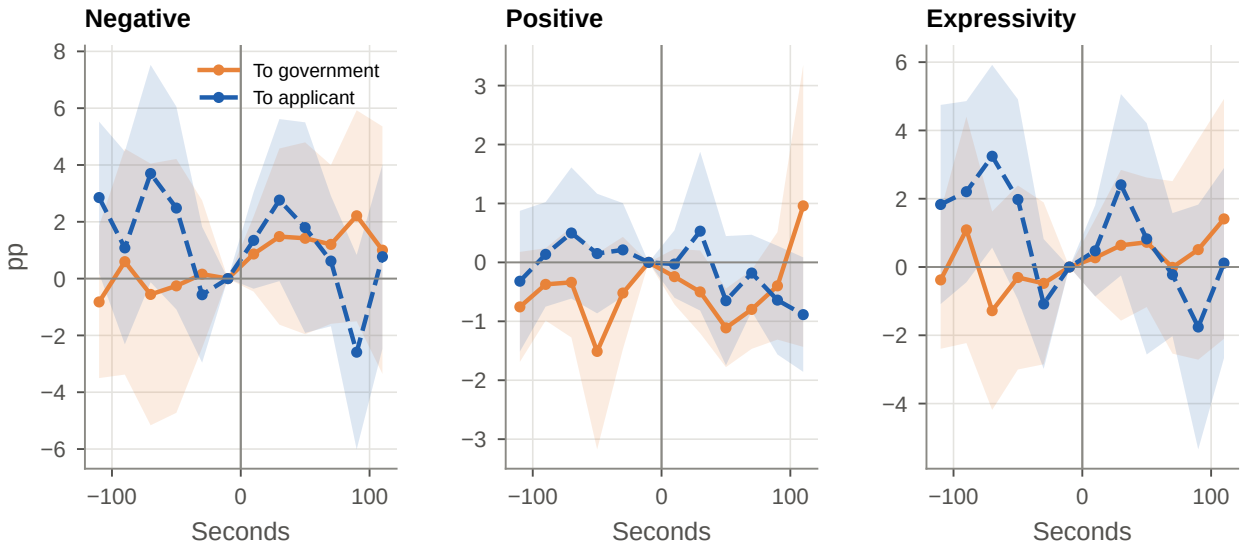

Figure A7: Changes of speaker between parties. Note: Lines: change to the government and to the applicant. Reference bin: −20 to 0 s.

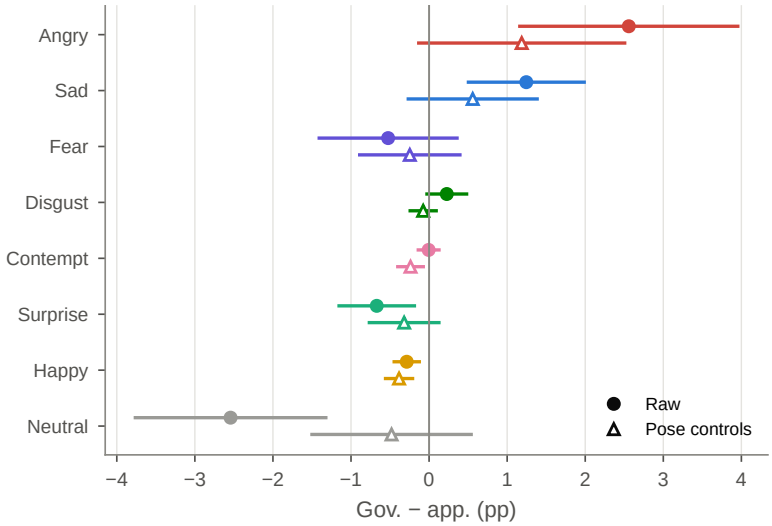

Figure A8: Difference between parties by class. Note: Raw and with frontal-score controls; 95% CIs.

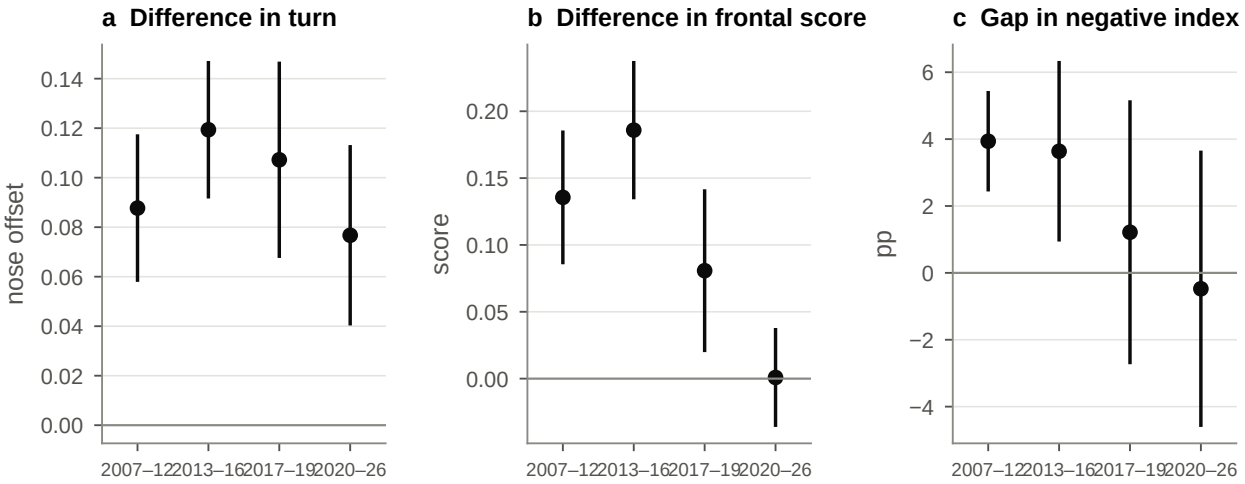

Figure A9: By period. Note: Government's turn minus applicant's, within judge-hearing; 95% CIs.

I Regression Results

Table A11: Coefficients (probability units)

Hyp.	Term	Outcome	Specification	Est.	SE	p	q	N
C1	gov	Negative	cell jhFE w	0.035	0.007	$< .001$	$< .001$	1,174
C1	gov	Positive	cell jhFE w	−0.003	0.001	0.001	0.004	1,174
C1	gov	Expressivity	cell jhFE w	0.025	0.006	$< .001$	0.001	1,174
C2	national judge	Negative	hearingFE	0.011	0.023	0.632	0.931	418
C2	national judge	Negative	judgeFE	0.013	0.007	0.060	–	428
C2	national judge	Negative	bothFE	−0.020	0.019	0.297	–	390
C2	national judge	Positive	hearingFE	−0.001	0.003	0.748	0.931	418
C2	national judge	Positive	judgeFE	−0.001	0.002	0.737	–	428
C2	national judge	Positive	bothFE	−0.003	0.003	0.454	–	390
C2	national judge	Expressivity	hearingFE	0.002	0.017	0.915	0.931	418
C2	national judge	Expressivity	judgeFE	0.000	0.009	0.980	–	428
C2	national judge	Expressivity	bothFE	−0.023	0.015	0.137	–	390
C4	both eu	Negative	hearingFE	−0.023	0.018	0.211	0.872	418
C4	both eu	Negative	judgeFE	0.006	0.013	0.641	–	428
C4	both eu	Negative	bothFE	−0.062	0.046	0.193	–	390
C4	both eu	Positive	hearingFE	−0.002	0.012	0.837	0.931	418
C4	both eu	Positive	judgeFE	−0.001	0.001	0.572	–	428
C4	both eu	Positive	bothFE	−0.002	0.009	0.844	–	390
C4	both eu	Expressivity	hearingFE	−0.013	0.020	0.516	0.929	418
C4	both eu	Expressivity	judgeFE	0.004	0.009	0.654	–	428
C4	both eu	Expressivity	bothFE	−0.043	0.048	0.376	–	390
C5a	same eastwest	Negative	hearingFE	0.020	0.030	0.514	0.929	418
C5a	same eastwest	Negative	judgeFE	0.014	0.010	0.171	–	428
C5a	same eastwest	Negative	bothFE	−0.025	0.034	0.470	–	390
C5a	same eastwest	Positive	hearingFE	−0.003	0.003	0.296	0.872	418
C5a	same eastwest	Positive	judgeFE	−0.002	0.002	0.153	–	428
C5a	same eastwest	Positive	bothFE	0.005	0.004	0.267	–	390
C5a	same eastwest	Expressivity	hearingFE	0.002	0.021	0.931	0.931	418
C5a	same eastwest	Expressivity	judgeFE	0.002	0.009	0.803	–	428
C5a	same eastwest	Expressivity	bothFE	−0.023	0.030	0.446	–	390
C5b	same northsouth	Negative	hearingFE	−0.007	0.014	0.624	0.931	418
C5b	same northsouth	Negative	judgeFE	−0.010	0.012	0.411	–	428
C5b	same northsouth	Negative	bothFE	−0.024	0.017	0.164	–	390
C5b	same northsouth	Positive	hearingFE	−0.000	0.001	0.860	0.931	418
C5b	same northsouth	Positive	judgeFE	−0.001	0.001	0.405	–	428
C5b	same northsouth	Positive	bothFE	−0.001	0.002	0.717	–	390
C5b	same northsouth	Expressivity	hearingFE	−0.010	0.011	0.339	0.872	418
C5b	same northsouth	Expressivity	judgeFE	−0.018	0.009	0.057	–	428
C5b	same northsouth	Expressivity	bothFE	−0.025	0.013	0.065	–	390
C6	dem distance	Negative	hearingFE	0.009	0.067	0.891	0.931	382
C6	dem distance	Negative	judgeFE	−0.019	0.013	0.144	–	400
C6	dem distance	Negative	bothFE	0.174	0.161	0.290	–	356
C6	dem distance	Positive	hearingFE	0.025	0.025	0.323	0.872	382
C6	dem distance	Positive	judgeFE	0.005	0.003	0.061	–	400
C6	dem distance	Positive	bothFE	−0.023	0.026	0.386	–	356
C6	dem distance	Expressivity	hearingFE	0.042	0.055	0.447	0.929	382
C6	dem distance	Expressivity	judgeFE	−0.011	0.008	0.187	–	400

Hyp.	Term	Outcome	Specification	Est.	SE	p	q	N
C6	dem distance	Expressivity	bothFE	0.196	0.139	0.168	–	356
C3	national judge × renewable term	Negative	hearingFE	−0.005	0.039	0.887	–	418
C3	national judge × renewable term	Positive	hearingFE	0.002	0.006	0.799	–	418
C3	national judge × renewable term	Expressivity	hearingFE	−0.000	0.031	0.997	–	418
C8	female × disc	Negative	hearingFE	0.091	0.091	0.321	–	418
C8	female × disc	Positive	hearingFE	0.002	0.010	0.839	–	418
C8	female × disc	Expressivity	hearingFE	0.086	0.055	0.120	–	418
B1	core integrity	Negative	judgeFE ctrl	0.004	0.008	0.613	0.788	1,788
B1	core integrity	Positive	judgeFE ctrl	0.000	0.001	0.745	0.788	1,788
B1	core integrity	Expressivity	judgeFE ctrl	0.002	0.006	0.788	0.788	1,788
D1	d neg	Negative	hearingFE	−0.082	0.134	0.542	0.542	344
D1	d pos	Positive	hearingFE	−0.861	0.646	0.189	0.38	344
D1	d expr	Expressivity	hearingFE	−0.406	0.138	0.005	–	344
D2	neg	dissent any on neg	hearingFE	0.001	0.089	0.996	–	1,518
D2	pos	dissent any on pos	hearingFE	0.013	0.387	0.972	–	1,518
D2	expr	dissent any on expr	hearingFE	−0.010	0.109	0.924	–	1,518
D2	neg	sep any on neg	hearingFE	0.091	0.091	0.316	–	1,518
D2	pos	sep any on pos	hearingFE	−0.089	0.399	0.824	–	1,518
D2	expr	sep any on expr	hearingFE	0.050	0.106	0.642	–	1,518
E2	female	Negative	hearingFE quality	−0.112	0.029	$< .001$	–	211,944
E2	female	Positive	hearingFE quality	0.008	0.005	0.070	–	211,944
E2	female	Expressivity	hearingFE quality	−0.067	0.021	0.002	–	211,944

Standard errors clustered by judge and hearing. q: Benjamini–Hochberg within family across primary outcomes and primary hypotheses (primary specifications only). D1 q-values cover the tests on negative and positive display; expressivity is exploratory. Specifications: jhFE = judge×hearing fixed effects; hearingFE/judgeFE/bothFE = fixed effects; ctrl = with controls.

Table A12: Additional analyses

Hyp.	Term	Outcome	Specification	Est.	SE	p
Pose by party						
	government turn	Frontal score	seconds, judge×hearing FE	0.140	0.018	$< .001$
	government turn	Face height (px)	seconds, judge×hearing FE	0.076	0.274	0.782
	government turn	|Yaw| (degrees)	seconds, faces $\geq$64 px	−7.168	1.009	$< .001$
Pose by party, by period						
	government turn	Frontal score	2007–2019	0.147	0.018	$< .001$
	government turn	Frontal score	2020–2026	0.001	0.018	0.961
Difference between parties by period						
C1	government turn	Negative	2007–2019	0.037	0.007	$< .001$
C1	government turn	Positive	2007–2019	−0.003	0.001	0.002
C1	government turn	Expressivity	2007–2019	0.027	0.006	$< .001$
C1	government turn	Negative	2020–2026	−0.005	0.020	0.816
C1	government turn	Positive	2020–2026	−0.009	0.005	0.094
C1	government turn	Expressivity	2020–2026	0.003	0.012	0.777
Difference between parties with image controls						
C1	government turn	Negative	no controls	0.035	0.007	$< .001$
C1	government turn	Negative	frontal score (quadratic)	0.013	0.007	0.054
C1	government turn	Negative	frontal, face size, faces in frame, shot	0.012	0.006	0.059
C1	government turn	Negative	frontal, turn, nose, roll	0.005	0.005	0.273
C1	government turn	Negative	all landmarks, position, light, size, shot	0.006	0.005	0.211
C1	government turn	Negative	pose-by-size cells	0.014	0.006	0.025
C1	government turn	Positive	no controls	−0.003	0.001	0.001

Hyp.	Term	Outcome	Specification	Est.	SE	p
C1	government turn	Positive	frontal score (quadratic)	−0.004	0.001	$< .001$
C1	government turn	Positive	frontal, face size, faces in frame, shot	−0.004	0.001	$< .001$
C1	government turn	Positive	frontal, turn, nose, roll	−0.002	0.001	0.078
C1	government turn	Positive	all landmarks, position, light, size, shot	−0.002	0.001	0.049
C1	government turn	Positive	pose-by-size cells	−0.004	0.001	$< .001$
C1	government turn	Expressivity	no controls	0.025	0.006	$< .001$
C1	government turn	Expressivity	frontal score (quadratic)	0.006	0.006	0.305
C1	government turn	Expressivity	frontal, face size, faces in frame, shot	0.005	0.005	0.352
C1	government turn	Expressivity	frontal, turn, nose, roll	0.002	0.005	0.670
C1	government turn	Expressivity	all landmarks, position, light, size, shot	0.002	0.005	0.611
C1	government turn	Expressivity	pose-by-size cells	0.006	0.005	0.241
C1	government turn	Negative	frontal score ≥ 0.8	−0.003	0.009	0.745
C1	government turn	Negative	frontal score < 0.5	0.024	0.012	0.040
C1	government turn	Positive	frontal score ≥ 0.8	−0.002	0.002	0.161
C1	government turn	Positive	frontal score < 0.5	−0.000	0.001	0.657
C1	government turn	Expressivity	frontal score ≥ 0.8	−0.005	0.009	0.532
C1	government turn	Expressivity	frontal score < 0.5	0.016	0.010	0.132
Difference between parties, adjusted cells						
C1	government turn	Negative (adj.)	cells, judge×hearing FE	0.009	0.006	0.176
C1	government turn	Positive (adj.)	cells, judge×hearing FE	−0.002	0.001	0.050
C1	government turn	Expressivity (adj.)	cells, judge×hearing FE	0.004	0.006	0.516
Affinity, adjusted						
C2	national judge	Negative (adj.)	hearing FE	0.014	0.020	0.508
C2	national judge	Positive (adj.)	hearing FE	−0.002	0.003	0.482
C2	national judge	Expressivity (adj.)	hearing FE	0.002	0.015	0.884
C4	both eu	Negative (adj.)	hearing FE	−0.026	0.016	0.114
C4	both eu	Positive (adj.)	hearing FE	−0.003	0.012	0.793
C4	both eu	Expressivity (adj.)	hearing FE	−0.015	0.016	0.356
C5a	same eastwest	Negative (adj.)	hearing FE	0.016	0.027	0.555
C5a	same eastwest	Positive (adj.)	hearing FE	−0.003	0.003	0.393
C5a	same eastwest	Expressivity (adj.)	hearing FE	−0.001	0.018	0.948
C5b	same northsouth	Negative (adj.)	hearing FE	−0.002	0.014	0.876
C5b	same northsouth	Positive (adj.)	hearing FE	−0.001	0.001	0.636
C5b	same northsouth	Expressivity (adj.)	hearing FE	−0.007	0.010	0.507
C6	dem distance	Negative (adj.)	hearing FE	0.019	0.063	0.767
C6	dem distance	Positive (adj.)	hearing FE	0.022	0.025	0.379
C6	dem distance	Expressivity (adj.)	hearing FE	0.048	0.048	0.325
National judge, other checks						
C2	national judge	Negative	hearing FE, one respondent	0.004	0.024	0.884
C2	national judge	Positive	hearing FE, one respondent	−0.001	0.003	0.880
C2	national judge	Expressivity	hearing FE, one respondent	−0.002	0.017	0.889
Votes, adjusted						
D1	differential	Negative (adj.)	hearing FE	−0.058	0.139	0.679
D1	differential	Positive (adj.)	hearing FE	−0.848	0.657	0.203
D1	differential	Expressivity (adj.)	hearing FE	−0.431	0.150	0.006
D1	differential	Expressivity	hearing FE, no national judges	−0.412	0.148	0.008
D1	differential	Expressivity	hearing FE, weighted	−0.510	0.165	0.003
Families A and B, adjusted						
A3	president	Negative (adj.)	hearing FE	0.031	0.019	0.100
B1	core integrity	Negative (adj.)	judge FE, controls	0.001	0.007	0.872
B2	interstate	Negative (adj.)	judge FE, controls	0.007	0.011	0.540
A3	president	Positive (adj.)	hearing FE	−0.004	0.002	0.102

Hyp.	Term	Outcome	Specification	Est.	SE	p
B1	core integrity	Positive (adj.)	judge FE, controls	0.001	0.001	0.283
B2	interstate	Positive (adj.)	judge FE, controls	−0.006	0.003	0.023
A3	president	Expressivity (adj.)	hearing FE	0.038	0.012	0.003
B1	core integrity	Expressivity (adj.)	judge FE, controls	0.001	0.005	0.909
B2	interstate	Expressivity (adj.)	judge FE, controls	0.010	0.012	0.425
Smile width						
	government turn	Smile width	no controls	0.000	0.001	0.791
	government turn	Smile width	frontal score (quadratic)	0.002	0.002	0.452
	government turn	Positive	no controls	−0.002	0.002	0.246
	government turn	Positive	frontal score (quadratic)	−0.004	0.002	0.068
Pose: further checks						
	government turn	Frontal score	government spoke first	0.133	0.019	$< .001$
	government turn	Frontal score	applicant spoke first	0.145	0.020	$< .001$
C1	government turn	Negative	frontal (quadratic); government first	0.020	0.010	0.052
C1	government turn	Negative	frontal (quadratic); applicant first	0.008	0.006	0.144
C1	government turn	Negative	judge×hearing×frontal-decile FE	0.017	0.006	0.010
C1	government turn	Negative	same, 20 bins	0.018	0.007	0.014
C1	government turn	Positive	judge×hearing×frontal-decile FE	−0.004	0.001	$< .001$
C1	government turn	Expressivity	judge×hearing×frontal-decile FE	0.006	0.005	0.279
C1	government turn	Negative	frontal ≥ 0.6	0.006	0.008	0.460
C1	government turn	Negative	frontal ≥ 0.7	0.002	0.009	0.864
C1	government turn	Negative	frontal ≥ 0.9	0.005	0.010	0.631
C1	government turn	Negative	frontal < 0.5	0.024	0.012	0.040
C1	government turn	Negative	faces $\geq$64 px, no controls	0.035	0.014	0.017
C1	government turn	Negative	faces $\geq$64 px; |yaw|, |pitch|, frontal	0.006	0.010	0.562
How the two turns differ in the picture (landmarks, all faces)						
	government turn	Sideways turn (nose offset)	seconds, judge×hearing FE	0.097	0.010	$< .001$
	government turn	Absolute turn	seconds, judge×hearing FE	−0.087	0.011	$< .001$
	government turn	Vertical nose position	seconds, judge×hearing FE	−0.005	0.003	0.073
	government turn	Roll (degrees)	seconds, judge×hearing FE	−4.349	0.449	$< .001$
	government turn	Face position, horizontal	seconds, judge×hearing FE	−0.001	0.003	0.647
	government turn	Brightness	seconds, judge×hearing FE	−0.260	0.458	0.571
Difference between parties with landmark controls						
C1	government turn	Negative	signed turn (cubic)	0.014	0.007	0.063
C1	government turn	Negative	frontal, turn, nose, roll	0.005	0.005	0.273
C1	government turn	Negative	all landmarks, position, light, size, shot	0.006	0.005	0.211
C1	government turn	Negative	judge×hearing×turn-bin FE	0.019	0.006	0.003
C1	government turn	Negative	same, plus nose and roll	0.009	0.005	0.061
C1	government turn	Negative	adjusted cells, slopes from party turns	0.006	0.006	0.366
C1	government turn	Positive	signed turn (cubic)	−0.004	0.001	$< .001$
C1	government turn	Positive	frontal, turn, nose, roll	−0.002	0.001	0.078
C1	government turn	Positive	all landmarks, position, light, size, shot	−0.002	0.001	0.049
C1	government turn	Positive	judge×hearing×turn-bin FE	−0.004	0.001	$< .001$
C1	government turn	Positive	same, plus nose and roll	−0.001	0.001	0.223
C1	government turn	Positive	adjusted cells, slopes from party turns	−0.002	0.001	0.013
C1	government turn	Expressivity	signed turn (cubic)	0.004	0.006	0.456
C1	government turn	Expressivity	frontal, turn, nose, roll	0.002	0.005	0.670
C1	government turn	Expressivity	all landmarks, position, light, size, shot	0.002	0.005	0.611
C1	government turn	Expressivity	judge×hearing×turn-bin FE	0.008	0.005	0.159
C1	government turn	Expressivity	same, plus nose and roll	0.003	0.005	0.535
C1	government turn	Expressivity	adjusted cells, slopes from party turns	0.002	0.006	0.676
Presiding judges and others						

Hyp.	Term	Outcome	Specification	Est.	SE	p
	government turn	Sideways turn	presiding	0.103	0.013	$< .001$
C1	government turn	Negative	presiding; no controls	0.050	0.010	$< .001$
C1	government turn	Negative	presiding; frontal score	0.026	0.009	0.007
C1	government turn	Negative	presiding; frontal, turn, nose, roll	0.003	0.008	0.670
C1	government turn	Negative	presiding; adjusted cells	0.021	0.009	0.035
	government turn	Sideways turn	not presiding	0.092	0.011	$< .001$
C1	government turn	Negative	not presiding; no controls	0.024	0.006	$< .001$
C1	government turn	Negative	not presiding; frontal score	0.004	0.007	0.583
C1	government turn	Negative	not presiding; frontal, turn, nose, roll	0.002	0.006	0.734
C1	government turn	Negative	not presiding; adjusted cells	−0.000	0.006	0.997
By period						
	government turn	Sideways turn	2007–2012	0.088	0.014	$< .001$
	government turn	Frontal score	2007–2012	0.136	0.024	$< .001$
C1	government turn	Negative	2007–2012	0.039	0.007	$< .001$
	government turn	Sideways turn	2013–2016	0.119	0.013	$< .001$
	government turn	Frontal score	2013–2016	0.186	0.025	$< .001$
C1	government turn	Negative	2013–2016	0.036	0.013	0.009
	government turn	Sideways turn	2017–2019	0.107	0.019	$< .001$
	government turn	Frontal score	2017–2019	0.081	0.028	0.011
C1	government turn	Negative	2017–2019	0.012	0.019	0.522
	government turn	Sideways turn	2020–2026	0.077	0.018	$< .001$
	government turn	Frontal score	2020–2026	0.001	0.018	0.961
C1	government turn	Negative	2020–2026	−0.005	0.020	0.816
Gap against difference in frontal score, across judge-hearings (WLS, two-way clustered)						
C1	intercept	Negative differential		0.017	0.008	0.046
C1	slope	Negative differential		0.140	–	$< .001$
Geometric proxies, faces ≥64 px (in SD units)						
	government turn	Brow raise	no controls	−0.334	0.059	$< .001$
	government turn	Brow raise	landmark controls	−0.050	0.038	0.188
	government turn	Eye closure	no controls	0.427	0.101	$< .001$
	government turn	Eye closure	landmark controls	0.194	0.055	0.001
	government turn	Smile width	no controls	0.011	0.042	0.791
	government turn	Smile width	landmark controls	0.032	0.061	0.600
Votes: benches whose members voted differently (55 votes, 22 benches)						
D1	differential	Negative	hearing FE, clustered by hearing	−0.773	1.305	0.560
D1	differential	Positive	hearing FE, clustered by hearing	−15.204	4.772	0.004
D1	differential	Expressivity	hearing FE, clustered by hearing	−4.354	0.841	$< .001$

Standard errors clustered by judge and hearing. "adj." outcomes are adjusted for head pose, face size, faces in frame and shot type (Section 6.5). Estimates in probability units except pose, face height and smile width. – indicates a standard error not reported for that estimate.

Table A13 reports the exploratory tests of Hypotheses 1 to 6 that condition on the ruling. The sample consists of the 1,400 judge-hearings with at least ten verified seconds and a coded vote (266 hearings, 105 judges). Votes differ among the observed judges in 95 of these hearings. National judges, of whom there are 46 in 92 observations, voted for their own state in 39 instances. Among the judges who voted for the respondent state, national judges have a positive display that is 0.99 points higher than that of their colleagues in the same hearing ($p = 0.042$; 1.00 after the image adjustment, $p = 0.071$). The interaction for national judges ranges from 0.57 to 0.96 points when one national judge at a time is left out, and it is 0.46 points without weights. With 20 tests of Hypotheses 1 to 5 in the table, the smallest q-value is 0.27.

Table A13: Hypotheses 1 to 6 conditional on the judge's ruling

		Positive display			Negative display			
Tie × ruling for the state		Est.	(SE)	p	Est.	(SE)	p	N
Classifier output								
H1	National judge	0.79	(0.40)	0.051	−4.41	(2.57)	0.090	1,378
H2	Both from EU states	−0.42	(0.34)	0.22	2.86	(1.28)	0.028	1,378
H3	Same East/West group	−0.02	(0.27)	0.94	0.77	(1.48)	0.60	1,378
H4	Same North/South group	0.36	(0.19)	0.069	1.41	(1.14)	0.22	1,378
H5	Democracy scores 0.1 closer	−0.02	(0.08)	0.79	0.33	(0.32)	0.32	1,283
H6	Woman judge, discrimination case[a]	−0.41	(0.46)	0.37	1.72	(5.18)	0.74	1,378
Adjusted for head orientation and image properties								
H1	National judge	0.51	(0.38)	0.18	−3.96	(2.35)	0.094	1,378
H2	Both from EU states	−0.40	(0.35)	0.26	2.54	(1.43)	0.078	1,378
H3	Same East/West group	−0.17	(0.26)	0.52	1.04	(1.31)	0.43	1,378
H4	Same North/South group	0.38	(0.19)	0.045	0.19	(1.09)	0.86	1,378
H5	Democracy scores 0.1 closer	−0.02	(0.08)	0.81	0.11	(0.31)	0.74	1,283
H6	Woman judge, discrimination case[a]	−0.42	(0.48)	0.38	−2.08	(4.07)	0.61	1,378

Note: Judge-hearing means over all seconds in which the judge is on screen and not speaking, for judges with a coded vote. Each row is the interaction between the tie named and an indicator that the judge's vote favored the respondent state, in percentage points, from a model with judge and hearing fixed effects, controls for face size and the share of bench shots, and weights equal to the number of seconds. Standard errors clustered by judge and hearing. [a]Interaction of woman judge, discrimination case and a vote for the applicant; the gender of applicants is not coded.

J Speaker and Phase Coding

The record of speakers and phases, which our files call the context layer, was produced by ChatGPT 6.1 Sol in an agentic workflow. It read the transcripts, the Court's session information and, for some hearings, video frames, and wrote structured tables for each hearing: speech turns, speaker segments, phases, judges' questions, answers, arguments, legal references, shots, a per-second frame table, and status and quality tables. Its own status tables describe most hearings as having had targeted review only. A chronological visual pass was completed for 36 hearings. We use the per-second side and phase fields.