Does Group Familiarity Improve Deliberations in Judicial Teams? Evidence from the German Federal Court of Justice

Tilko Swalve*

Collegiality plays a central role in judicial decision-making. However, we still lack empirical evidence about the effects of collegiality on judicial decision-making. In this article, I argue familiarity, an antecedent to collegiality, improves judicial deliberations by encouraging minority dissent and a more extensive debate of different legal viewpoints. Relying on a novel dataset of 21,613 appeals in criminal cases at the German Federal Court of Justice between 1990 and 2016, I exploit quasi-random assignment of cases to decision-making groups to show that judges' pairwise familiarity substantially increases the probability that judges schedule a main hearing after first-stage deliberations. Group familiarity also increases the length of the justification of the ruling. The findings have implications for the way courts organize the assignment of judges to panels.

Keywords: collegiality; deliberation; familiarity; judicial decision-making; minority dissent

I. Introduction

Mounting empirical evidence suggests that the composition of panels matters for judicial decision-making. Judges have been found to adjust their own behavior depending on their colleagues' ideology (Cross & Tiller, 1998; Revesz, 1997; Sunstein et al., 2006), gender (Boyd et al., 2010; Farhang & Wawro, 2004; Peresie, 2005), and race (Kastellec, 2013). Given the significance of a panel's composition, the procedure that selects judges sitting together is important. In many courts, such as the U.S. Federal

Courts of Appeals, judges are assigned to panels for every case anew. In others, such as the German Federal Court of Justice, judges are assigned to *Spruchgruppen*, small groups of judges that sit together on cases for months or even several years with only minor changes to their composition. On the end of the spectrum sit many constitutional courts, where one or two groups of judges often stay together for many years.

A common theme to most assignment procedures is that they aim to promote a collegial atmosphere by increasing judges' familiarity with each other. Even many practices otherwise described as "random" try to compose panels such that each judge on the court sits with every other judge at least a certain number of times (G. Levy & Razin, 2015; M. K. Levy, 2017). Despite being conceptually somewhat elusive, collegiality is often defined as a "deliberatively cultivated attitude" among judges who share a "continuous, open, and intimate relationship" (Coffin, 1994, p. 214). Characteristics of a collegial atmosphere are an emphasis on "collaboration and deliberation" (Kornhauser & Sager, 1993, p. 4), civil and polite exchanges between group members, cohesiveness, and mutual respect (Cohen, 2002; Cross & Tiller, 2008; Wasby, 1987).

Repeatedly, scholars and judges have emphasized the vital role that collegiality and even friendship play when interacting with their peers on the bench (e.g., Baum, 1997, 2006; Edwards, 1998, 2003). Surprisingly, we still lack empirical scholarship analyzing the role of collegiality and interpersonal contact on judicial behavior (for an exception, see Nelson et al., 2022). As Epstein & Knight (2013, p. 19) put it

> [J]udges frequently refer to the importance of collegiality […], and just as frequently, scholars reject it. We should not.

In this article, I argue that familiarity—a necessary antecedent of collegiality (Murphy, 2000)—can improve the judicial decision-making process. Research in social psychology has shown that familiar groups are more productive and efficient, and familiar group members are more comfortable sharing and discussing dissenting opinions and privately held information (Gruenfeld et al., 1996; Reagans et al., 2005; Wittenbaum, 1998; Wittenbaum et al., 2010). Especially in environments where workload and the pressure to conform to a group's majority opinion are high, dissenting opinions are vital for the quality of the group's decision-making, but chronically undersupplied (Asch, 1956; Janis, 1982; Klocke, 2007). Based on these findings, I hypothesize that (a) judges in familiar groups disagree more often with their colleagues during deliberations than judges in less familiar groups and that (b) familiar judges debate a more comprehensive range of perspectives during deliberations than unfamiliar judges.

I estimate the causal effect of familiarity relying on a novel dataset of 21,613 criminal cases between 1990 and 2016 at the German Federal Court of Justice (FCJ). The FCJ's decision-making process offers an ideal setting to test the hypotheses for several reasons. First, the institutional setting produces significant conformity pressure by giving the group a choice between a quick decision-making procedure (court order) and a more laborious process (main hearing). Because the law explicitly demands a unanimous decision for the simple court order, every single group member faces pressure to conform to the group's majority opinion to keep the group's overall workload manageable. Second,

random assignment of cases to decision-making groups allows identifying the causal effect of familiarity on the decision. Given the random assignment, the judges' familiarity is not systematically related to case characteristics that could act as confounders. Finally, the novel data enable us to study judicial behavior on one of the most active and influential European courts. It also sheds light on the judicial deliberation process where it supposedly matters the most: in civil law countries. In the civil law tradition, separate opinion writing is less common, and judges do not enjoy as much public standing and attention as their peers in common law countries (Merryman & Pérez-Perdomo, 2007). Thus, collegial deliberation behind closed doors becomes even more relevant for judges to influence the public and legal discourse.

II. How Group Familiarity Affects Judicial Deliberation

The collegial deliberation among a small group of judges behind closed doors takes a central place in the decision-making process of higher courts. A large empirical scholarship on "panel effects" has provided evidence that panel composition impacts the behavior of judges (e.g., Boyd et al., 2010; Cross & Tiller, 1998; Farhang & Wawro, 2004; Kastellec, 2013).

Several explanations for panel effects have been suggested: acquiescence, deliberation, and strategic considerations with respect to judicial review of higher courts (Hinkle et al., 2020). Acquiescence (or deference) describes the tendency of judges to suppress minority opinions and aim for consensus to be a "good" colleague and alleviate workload (Epstein et al., 2011; Fischman, 2011, 2013). Deliberation refers to the exchange of information, persuasive argument, and the possibility of judges adopting another perspective afterward (Boyd et al., 2010; Edwards, 1998). The third mechanism, avoiding dissent to dodge judicial review by higher courts, is less relevant for decision-making at an apex court like the FCJ. However, we know much less about the factors that make each mechanism more or less likely. What causes judges to engage in effective deliberation, and what makes them avoid dissent and acquiesce?

A possible answer lies in a rich literature in social psychology which has identified professional familiarity as an essential determinant of group performance. Familiarity has been linked to group productivity and higher quality of decision-making in a wide range of working environments, such as coal mining (Goodman & Leyden, 1991), computer software services and development (Espinosa et al., 2007; Huckman & Staats, 2011), cardiac surgery teams (Avgerinos et al., 2019; Avgerinos & Gokpinar, 2017), and Olympic ice hockey teams (Dalal et al., 2017).

Groups may benefit from member familiarity for several reasons. First, team members can use their prior group experience to coordinate activities and allocate tasks (Reagans et al., 2005). Second, familiar teams are more likely to share a mental model about their team. Knowledge about their peers, their attitudes, preferences, skills, and specializations helps team members to interact because they can adjust their actions to the behavior they expect from other members (Mathieu et al., 2000). Third, as teams become increasingly

familiar, the quantity and quality of intra-team communication increases because members establish communication channels and a common language (Gruenfeld et al., 1996). Fourth, familiar team members have fewer conflicts and are better able to separate task from interpersonal conflict (Jehn & Shah, 1997; Shah & Jehn, 1993). Finally, repeated interactions between team members are the foundation to create interpersonal trust, which in turn leads to better performance and productivity (Dirks & Ferrin, 2001).

Interpersonal trust directly relates to team psychological safety, which describes a group climate in which members feel comfortable being themselves and do not have to fear embarrassment or retaliation for taking risk (Edmondson, 1999). In such an environment, individuals share more private, also dissenting, information with other group members (Edmondson, 1999; Siemsen et al., 2009). Information sharing is crucial in the context of collegial deliberations. Groups tend to favor discussing shared information (i.e., information that is available to all group members) over unshared information (i.e., information available uniquely to a single group member) (Stasser & Titus, 1985; Wittenbaum et al., 2010), which can lead to an artificially strong consensus among group members neglecting dissenting information or opinions.

Psychological safety is a necessary condition for the contribution of unshared information and articulation of opinions and knowledge that challenge the group's consensus (Edmondson, 1999). Unfamiliar group members are more uncertain about group norms and are therefore more sensitive to social clues. To avoid social ostracism (Wittenbaum et al., 2010) unfamiliar group members are more likely to copy the behavior of other members and consciously or unconsciously suppress information or opinions that seem dissenting or irrelevant to what other members believe (Gruenfeld et al., 1996). Members with more experience, however, are found to be more comfortable debating unshared information (Wittenbaum, 1998). Information, to be effective, needs not only to be shared but also elaborated. Maynard et al. (2019) finds that more familiar groups are more likely to integrate and build upon shared information. Moreover, group members are more likely to rely on the knowledge of familiar team members compared to information contributed by unfamiliar individuals (Kane et al., 2005).

These findings apply directly to judicial deliberations. First, because familiar groups are more efficient and productive, familiar judges can elaborate and discuss the same amount of legal issues in less time. A reduction in workload matters, especially in courts with a high number of cases (such as the FCJ). As workload decreases, judges should defer less often and engage in more involved deliberations instead. In line with this argument, Engel and Weinshall (2020) find that a decrease in caseload is associated with a higher likelihood to hear witnesses, a lower probability to issue summary judgments and more elaborate opinions. Second, more increased psychological safety in familiar groups allows judges to voice dissenting views during deliberations without the fear of harming the personal relationships with their colleagues. In familiar groups, judges can expect different viewpoints to be discussed politely and with mutual respect without becoming personal and that their colleagues are open to persuasive arguments instead of insisting on personal convictions. Consistent with this argument, Nelson et al. (2022) present evidence that more frequent interpersonal contact reduces the effect of ideology in appellate review. In a related study, Hinkle et al. (2020) find that counterjudges with longer

co-tenure with the majority judges are more successful, indicating more effective panel deliberations.[1]

However, not all studies find familiarity universally beneficial. Several studies were unable to reproduce a positive effect of familiarity in specific settings (Espinosa et al., 2007; Janssen et al., 2009; Kim, 1997). In the context of judicial decision-making (Richman & Reynolds, 2012, pp. 204–206) maintain that "judicial collegiality [...] appears to be myth" and speculate that judges who know each other well might not be willing to risk personal relationships by disagreeing and engage in vote trading instead.[2] Ultimately, whether team familiarity has benefits for judicial decision-making is still an open empirical question that requires the analysis of data on judges' decision-making under varying degrees of group familiarity. This study provides such an analysis.

III. INSTITUTIONAL BACKGROUND: THE GERMAN FEDERAL COURT OF JUSTICE

The FCJ decides on appeals filed against verdicts of regional (*Landgerichte*) and higher-regional courts (*Oberlandesgerichte*). Regional and higher regional courts almost exclusively rule on cases of serious crime that involve a potential long-term sentence. The workload is split between five permanent benches, called senates[3] (*Senate*), which consist of six or seven associate judges and the senate's chief judge. Each senate is further divided into three subgroups of five judges, called, which share the caseload. Every associate judge serves in two groups, the chief judge in all three. Only occasionally, to substitute for a colleague who is ill or on vacation, judges will be part of judicial panels other than their assigned groups.

Considerations about the collegiality between judges are important for a setting like the FCJ. As in many judicial institutions, judges at the FCJ are appointed for life (until reaching the legal retirement age) and rarely drop out for other reasons than retirement. Thus, judges can generally expect to interact and work with their colleagues on many decisions for several years. The long-time horizon and repeated interactions make collegial relationships especially valuable. One can also expect that after some time at the FCJ, members have become very familiar with their colleagues they share the bench with often.

---

[1]In another related paper, Hinkle (2017) show that longer co-tenure with the author of the majority's opinion decreases the likelihood that a justice on the U.S. Supreme Court writes a dissenting opinion. The finding is equally consistent with (a) the argument that judges avoid dissent to prevent harm to their social relationships and (b) that more familiar judges deliberate more openly and effectively and reach a consensus more often.

[2]Note, however, that Richman and Reynolds (2012) acknowledge a lack of empirical evidence for both positive and negative effects of collegiality.

[3]Hamann (2019, p. 672) remarks that "division" would be a more appropriate translation.

Moreover, all but one criminal law senate reside in the same building in Karlsruhe. Only the fifth senate meets for historical reasons in Leipzig. Geographical proximity and working in the same building has been linked to more frequent personal interaction between judges in and outside the courtroom (Cohen, 2002, pp. 154–160; Nelson et al., 2022; Wasby, 1987), for example, during lunch breaks (Roberts, 2006, p. 376). In the context of the DC circuit, Edwards (2003) also notes that having all the circuit's chambers under the same roof has a positive effect on collegial relationships between judges. Karlsruhe, a city of about 300,000 inhabitants in South-West Germany, is the home to the FCJ and the German Federal Constitutional Court and hence the judicial "capital" of Germany. Given its medium size and the concentration of legal professionals, it is likely that at least some judges also share a social environment. These factors increase the value of personal relationships and the frequency of personal contact.

A. Case Allocation at the FCJ

The allocation of cases to groups of judges proceeds in two steps. First, the lower courts' geographical location determines the responsible senate. Second, within senates, cases are assigned to groups (*Spruchgruppen*). The second step of the assignment procedure, that is, the allocation of cases to groups *within* senates, is essential for the causal identification strategy of this article.

Section 21g (1) and (2) of the German Courts Constitution Acts (Gerichtsverfassungsgesetz)[4] states that "[w]ithin an adjudicating body of several judges, court business shall be allocated among the members by a ruling of all the professional judges belonging to the adjudicating body." It further requires that the ruling specifies how groups are composed and cases are allocated "prior to the beginning of the business year and for the latter's duration."

The exact requirements such principles must satisfy were subject to an extensive jurisprudential debate (for a summary, see Sowada, 2002, pp. 373–404). In 1993, the first civil law senate of the FCJ ruled that the exact composition of the decision-making group must be governed by a mechanism that does not require any action of the senate's president (BGH X ZR 51/92, see also Sangmeister, 1993). The issue was eventually settled in 1994 by the court's Joint Grand Panels (Vereinigte Große Senate).[5] The ruling specifies that abstract case characteristics must determine the allocation of cases to groups of judges. Furthermore, the exact composition of the responsible panel for any specific case arising in the coming business year has to emerge directly from the case allocation plan (*Geschäftsverteilungsplan*, for a discussion of the terminology, see Hamann, 2019, p. 673).

In practice, several options are available that satisfy the requirements of ruling by the Grand Joint Panels. Very common is an allocation scheme based on the last digit of a

---

[4]For an English translation of the original German, see: https://germanlawarchive.iuscomp.org/?p=771#21g

[5]Court Order of the FCJ's Joint Grand Panels, 5.5.1994. Available here: https://www.hrr-strafrecht.de/hrr/1/93/vgs-1-4-93.php

Figure 1: Decision-making process in criminal cases.
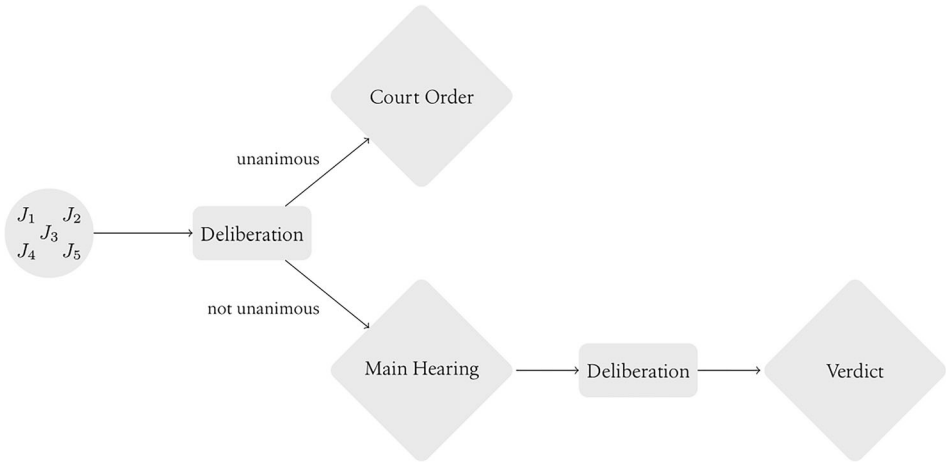
case's docket identifier (*Aktenzeichen*) (Fischer, 2015, p. 713; Sowada, 2002, p. 440).[6] Appendix A shows the first two pages of the case allocation plan of the third criminal law senate for the year 2013 as an example. Other less common options include an assignment based on the first letter of the defendant's name or the date when the case was filed with the court (Sowada, 2002, p. 440). An allocation based on docket identifiers has the advantage of ensuring an even distribution of cases to groups.

B. Decision-Making Process at the FCJ

Figure 1 depicts the decision-making process at the FCJ. Once a case is assigned, the group members (J_1–J_5) discuss the merits of the appeal in a first meeting. On this first deliberation stage, one of the judges serves as a rapporteur, who summarizes the facts of the case for the other judges. After hearing the report and discussion among judges, the group may find that the appeal is obviously unfounded (*offensichtlich unbegründet*) or consider the appeal as obviously well-founded (*offensichtlich begründet*). If all judges unanimously agree on a decision, the group can issue a ruling in the form of a court order (*Beschluss*) immediately. In cases where judges cannot reach a unanimous ruling, the case moves to the second stage: the main hearing. Main hearings involve a thorough document review by all judges and another meeting, including oral argument. The oral argument is followed by a second round of deliberations, after which the group reaches a verdict (*Urteil*) by simple majority rule.

---

[6]Since there are 10 digits to be allocated to three groups, the second to last digit determines the allocation if the last digit is a 0 (Fischer, 2015, p. 713, footnote 24).

The two-stage decision-making process[7] was introduced in 1922 to manage a substantial increase of appeals in criminal cases in the years following WWI to alleviate the increasing workload of the court (Rosenau, 2012).[8] In the past decades, more than 90% of appeals were decided without a main hearing (Rosenau, 2012) which has drawn criticism from the legal community. Initially, groups were supposed to forego the main hearing only if the correct decision outcome was immediately apparent to a trained eye (Rosenau, 2012). However, as observed by legal scholars and practicing judges alike, the criminal senates make extensive use of court orders, not just in "obvious" cases (Fischer, 2015; Rosenau, 2012).

Anecdotal evidence suggests that the high share of unanimous court orders partly the result of conformity pressure. Thomas Fischer, former chief judge of the second criminal senate, notes that[9]

> There is a costly decision-making procedure [...] and a simple procedure [...]. The simple procedure demands unanimity, and the other one does not. What else is supposed to happen but relentless informal pressure to reach unanimity? (Fischer, 2015, own translation)

Given the high workload of the court and resulting conformity pressures, factors that make group deliberations more efficient and that mediate the pressure to suppress dissenting opinions should be essential for the quality of decision-making. It is important to note that dissent in the first stage of deliberations is different from writing a separate dissenting opinion after deliberations (which is a common practice in the U.S. Supreme Court). The former indicates that judges disagreed on some issues and decided to give a case closer attention and gather more information to resolve differences if possible. Thus, dissenting is only worthwhile for a judge if she can expect that her majority colleagues are open to finding common ground and engaging in a productive second phase of deliberations. On the other hand, a separate dissenting opinion after deliberations signals that judges did not reach a consensus and that differences between judges remained unresolved.

IV. DATA

I analyze 22,163 published and admissible appeals decided by the FCJ's criminal law senates between 1990 and 2016. The court rulings were obtained from the FCJ's website

---

[7]The use of short-cut decision-making procedures is not unique to the FCJ. For example, in the United States, the Federal Rules of Appellate Procedure (2016, p. 40) hold that "[o]ral argument must be allowed in every case unless a panel of three judges who have examined the briefs and record unanimously agrees that oral argument is unnecessary [...]." Like the proceedings at the FCJ, the rules require a unanimous vote. Furthermore, in the vast majority of cases, the U.S. courts of appeals forgo an oral argument (Edwards, 2017).

[8]The FCJ was then called Imperial Court of Justice (*Reichsgericht*).

[9]Original in German: "Es gibt ein aufwendiges Erledigungsverfahren [...] und ein einfaches Verfahren [...]. Das einfache Verfahren setzt Einstimmigkeit voraus, das andere nicht. Was soll da anderes herauskommen als ein gnadenloser informeller Druck zur Einstimmigkeit?" (Fischer, 2015).

(Bundesgerichtshof, 2017) from the year 2000 onwards and obtained directly from the court on request for the cases before 2000.[10,11] Overall, the data contain 25.4% of all (published and unpublished) decisions of the court's criminal panels between 1990 and 2016.[12] From the court's published decisions, I extracted case-specific information automatically and manually where the automatic procedure failed.

A. Dependent Variables

I consider two different dependent variables intended to capture the extent of judicial deliberation in a case. The first is a dummy variable indicating whether a group decided via a verdict (i.e., after main hearing). The variable directly captures whether judges disagreed during the first-stage deliberations to the extent that it made a detailed document review and oral argument necessary. The second dependent variable is the length of the decision's justification measured by a simple word count. Even in cases that are decided via a court order, deliberations can be extensive and controversial. One should expect a more prolonged written justification for a ruling as group members raise more issues.

Figure 2a displays the share of main hearings of all admissible and published appeals in the dataset over time. Main hearings account for 16%–38% of published decisions making them rather the exception than the rule. Therefore, in many cases, judges reach a consensus in the first deliberation stage. Overall, the share of main hearings trends downwards between 1990 and 2010 but has increased since 2011, arguably in response to the critique of the excessive use of court orders (Rosenau, 2012).

Justifications of decisions with a main hearing are, on average, much longer than decisions without (2016: 676 words [court order], 2244 words [verdict]) which comes as no surprise since main hearings indicate issues that cannot quickly be resolved and require further investigation. Moreover, decision length has increased continually since 1990. The average verdict is now more than twice as long as in 1990. The average length of court orders has increased by about 55% between 1990 and 2016.

---

[10]Upon request, the FCJ provided the author with all decisions between 1990 and 1999 that were available in electronic form on CD-ROM. In total, these sum to 6628 observations (30% of the total dataset).

[11]The dataset does not include the small percentage (<1%) of inadmissible appeals usually due to the delayed filing of the appeal or because the panel did not consider the case to be within its responsibilities. The data does not include appeals in which panels did not issue a written justification and subsequently remained without publication.

[12]The coverage of the data is calculated by dividing the number of cases in the dataset to the number of completed cases (*Revisionen*) reported in annual reports of the Federal Statistical Office (Statistisches Bundesamt, 1990). For the years after 1999, the data contain 30.1% of all decisions. These numbers are a bit higher than typical publication rates of U.S. courts (Merrit & Brudney, 2001). While a ruling after a main hearing is (almost) always accompanied by a written and published justification, published decisions for court orders are much less frequent. The data covers 85.9% of all verdicts (98.9% after 1999) (see also Figure B1 in Appendix B).

Figure 2:   Dependent variables over time, 1990–2016, with 95% confidence intervals.

(a)

(b)
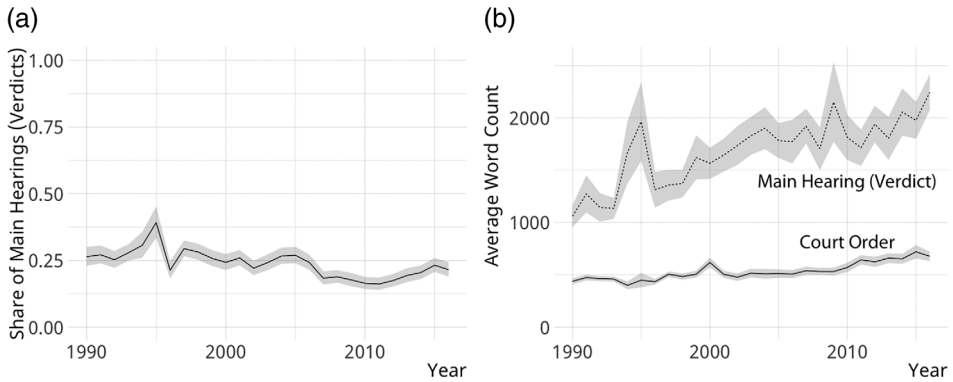
B. Measuring Familiarity

I operationalize familiarity as the number of interactions judges or groups had with each other in the past. I construct two alternative measures of group familiarity: (a) the *median pairwise group familiarity*[13] and (b) the *overall group familiarity*. These two measures correspond to slightly different notions of familiarity. The first measure attributes familiarity to the relationship between two individuals. The measure counts the number of past interactions of the 10 unique pairs of judges within the group. It then aggregates pairwise familiarities to the group level by taking the median.
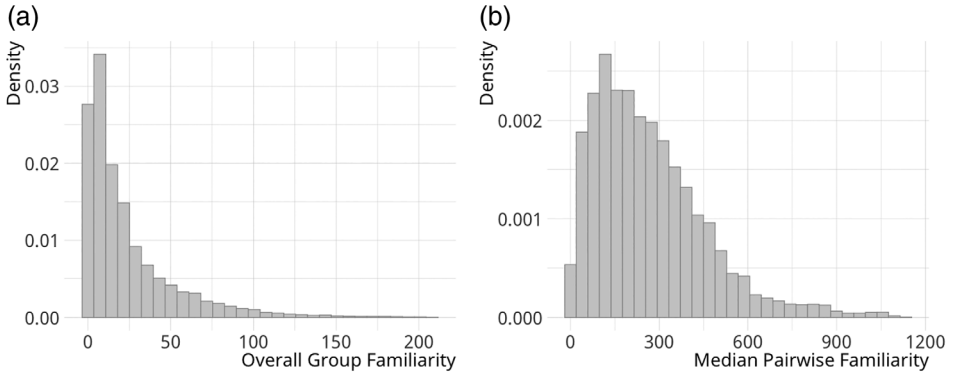
The median pairwise group familiarity captures how well judges know each other on average, but not necessarily how much experience a group has as a whole. By construction, the measure is robust to small changes in group composition, assuming that collegiality and the open exchange of information established via repeated interactions do not entirely break down once the group composition changes slightly.

Overall group familiarity, on the other hand, conceptualizes familiarity exclusively as a group attribute. This measure tallies the number of past cases a particular group has decided together in the past. Because familiarity refers here to unique group compositions instead of pairs of individuals, the measure is much more sensitive to minor group alterations. Even a tiny change in the group's composition resets a group's familiarity to zero. This measure is more appropriate if one assumes that collegial deliberation must be reestablished after a judge leaves and a new colleague joins the panel.

The difference between the two measures is an alternative assumption about the sensitivity of collegial atmosphere to small changes in the group's composition. Consider a group of five judges who have decided many hundred cases together over the past years

---

[13]An alternative measure is the *mean* pairwise group familiarity. Since the median is less vulnerable to the influence of outliers and the distribution of pairwise familiarity counts is highly skewed (see Appendix C, Figure C1), I use the median. All results are robust to using the mean instead (see Appendix D, Tables D1 and D2).

Figure 3: Distribution of familiarity measures.

and therefore have both a high median pairwise group familiarity and a high overall group familiarity. If a single judge leaves the group and is replaced by a new judge, the median pairwise group familiarity would decrease only slightly due to the high familiarity of the other four judges with each other. Overall group familiarity, on the other hand, would drop to zero in this scenario.

The data contain 99 individual judges, 1515 unique group compositions, and 1143 unique dyads of two judges that sat on at least one case together. On average, two judges have ruled on 316 cases jointly before the present case, suggesting that most judges in a group are often quite familiar with each other. Figure 3 shows the distributions of the two measures of familiarity. Both variables are skewed to the right, with overall group familiarity exhibiting a more significant skew than median pairwise familiarity. The average median pairwise familiarity is 267.2, indicating that judges are relatively well acquainted with each other. Overall group familiarity is much lower. The average group has ruled in its exact composition on 23.9 previous cases.

A drawback of measuring familiarity as previous cases decided together concerns the first years of the data. For the earliest cases in the dataset, every pair or group of judges has the same measure even though there might have been prior interactions. As a result, for the earlier years of the data, calculated familiarity is lower than group members' true familiarity and exhibits less variance. Figure 4 makes this pattern visible. It shows the yearly average of group familiarity (median pairwise, overall) from 1990 to 2016. Due to the construction of the measures, comparisons across time can be difficult. The estimation strategy takes this limitation into account and relies only on the variation within senates and years.

C. Control Variables

Next to the senate-year fixed effects (see the following section), I include two sets of control variables. The first set controls for several case characteristics to ensure that any

Figure 4: Average median and overall group familiarity with 95% confidence intervals, 1990–2016.
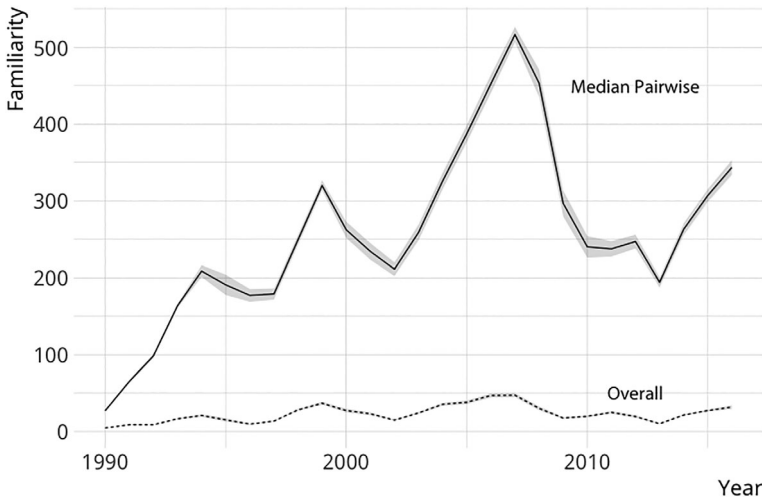

remaining imbalances across decision-making groups do not drive the results. In principle, however, the quasi-random assignment of cases to groups should guarantee that familiarity is independent of case-related variables. The second set of control variables are group characteristics that could potentially be related to both familiarity and the dependent variables.

Appeals can be filed by the defendant, the prosecution, and occasionally by co-plaintiffs. The court's decision text always includes information on the appellant. I created a dummy variable indicating whether the prosecution filed the appeal (coding appeals by co-plaintiffs as "prosecution"). In cases in which there were appeals from both defendant and prosecution, I treated both appeals separately since they can also result in different outcomes (e.g., the appeal of the prosecution is (partially) granted, and the appeal of the defendant rejected). In 86.5% (19,177) of cases, the appeal was filed by the defendant, of which 88% (16,877) resulted in a court order (no main hearing). In contrast, appeals by the prosecution resulted in only 6.3% (188) in a court order. These statistics suggest that the type of appellant is a strong predictor of whether a main hearing occurs (Figure 5).

Moreover, I created eight dummy variables accounting for the different categories of crime at the trial stage. The information on the crime is also available from the published court's decisions. The eight dummies represent the seven most common types of crime and an "other" category that accounts for several not very common crimes at the FCJ (e.g., traffic violations, possession of illegal weapons, human trafficking, etc.).[14] Figure 2 shows the number of appeals from each crime category split by whether they

---

[14]The category "homicide" also includes manslaughter.

Figure 5: Cases by types of crime.
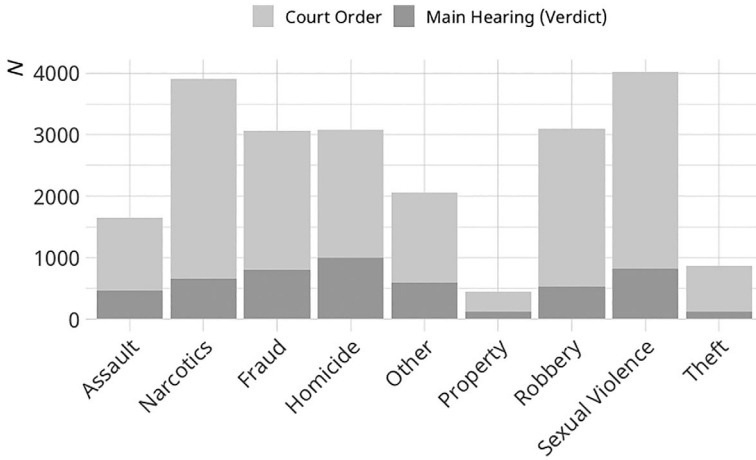
resulted in a main hearing or not. The most common types of crime are concerned with sexual violence or narcotics. Overall, differences between the crimes concerning the frequency of main hearings are minor.

Several studies have documented so-called freshman effects. Freshman or newcomer justices display more ideological instability in their voting behavior (Hagle, 1993; Wood et al., 1998), write fewer majority opinions (Bowen & Scheb, 1993), fewer separate opinions (Bowen, 1995; Brenner & Hagle, 1996; Hettinger et al., 2003) and adhere to precedent at higher rates (Hurwitz & Stefko, 2004). Since new judges are naturally also less familiar with the other judges, group familiarity and the presence of a freshman judge are likely to be correlated. To isolate the effect of familiarity from freshman effects, I include a dummy variable indicating whether the group includes a judge with less than 12 months of experience on the court.[15] The start dates of judges were gathered from the court's official press releases. In total, 6103 (27.5%) cases included a freshman judge, of which 4810 were court orders (78.8%, cases without freshman judges: 76.3%).

I also control for the average age of the group of judges at the time of the decision. Older judges are presumably more familiar with their (older) colleagues. If more senior judges dissent more (or less) often or are more (or less) prolix, then age would be an omitted variable. Information on the birth years of the judges comes from Wikipedia. In the data, the average age of panels was 56.15 years (SD = 2.34).

One of the more prominent debates in German jurisprudence in recent years has centered around the role of the rapporteur judge (Fischer et al., 2013; Katholnigg, 1992). At the FCJ, each case is assigned to a group member who presents a

---

[15]I explored alternative operationalizations (6, 9, and 24 months) with similar results.

Table 1:   Wald Statistics of Logistic Regression

| | DV: Mean hearing | | | | | |
| | (1) | | | (2) | | |
	χ^2	df	p	χ^2	df	p
Median pairwise familiarity	14.18	4	0.0067			
Nonlinear	4.33	3	0.2276			
Overall group familiarity				5.05	4	0.2825
Nonlinear				4.86	3	0.1826
Freshman	0.00	1	0.9676	0.32	1	0.5687
Average age	0.43	1	0.5128	0.20	1	0.6565
Appellant (prosecution)	3443.23	1	<0.0001	3444.46	1	<0.0001
Crime dummies	171.85	8	<0.0001	173.63	8	<0.0001
Senate-year dummies	481.14	134	<0.0001	472.54	134	<0.0001
Total	3754.66	149	<0.0001	3756.67	149	<0.0001

case summary to her colleagues. The fact that the other judges of the group may rely to a large extent on the presentation could give the rapporteur considerable influence over the case's outcome (Sowada, 2002, pp. 449–451). Unfortunately, the rapporteur's identity can neither be inferred from the decision text nor from (most of) the senates' internal case allocation plans.[16] Luckily for this study, the assignment of the rapporteur occurs *after* the assignment to a decision-making group (Sowada, 2002, p. 447). Thus, even if the required data were available, controlling for the rapporteur judge could introduce post-treatment bias and should be avoided (e.g., King & Zeng, 2006, pp. 147–148).

V. ESTIMATION STRATEGY

To identify the causal effect of group familiarity, I exploit the quasi-random allocation of cases to groups of judges *within* senates. Note that the identification strategy differs from previous research on panel effects, which relies on the assumption of random panel compositions (Boyd et al., 2010; Kastellec, 2013; Sunstein et al., 2004). In the FCJ, panels are not composed randomly for each case. Instead, panel composition remains largely fixed, and cases are assigned to the different groups via a quasi-random procedure. This difference is vital since recent studies have questioned the assumption of random panel composition in the context of the U.S. Courts of Appeals (M. K. Levy & Chilton, 2015; M. K. Levy, 2017).

There is no reason to assume a linear effect of familiarity on the probability of a main hearing and the decision length. Familiarity may even have a marginally decreasing impact on the outcome of group deliberations: a slight increase in familiarity when

---

[16]In practice, the rapporteur assignment process varies across senates. In some, the senate's president determines the rapporteur, while others use a procedure based on abstract case characteristics similar to the case-to-group allocation scheme described in the following section (Fischer, 2015, p. 713, footnote 24).

Figure 6: The effect of familiarity on the probability of a main hearing.
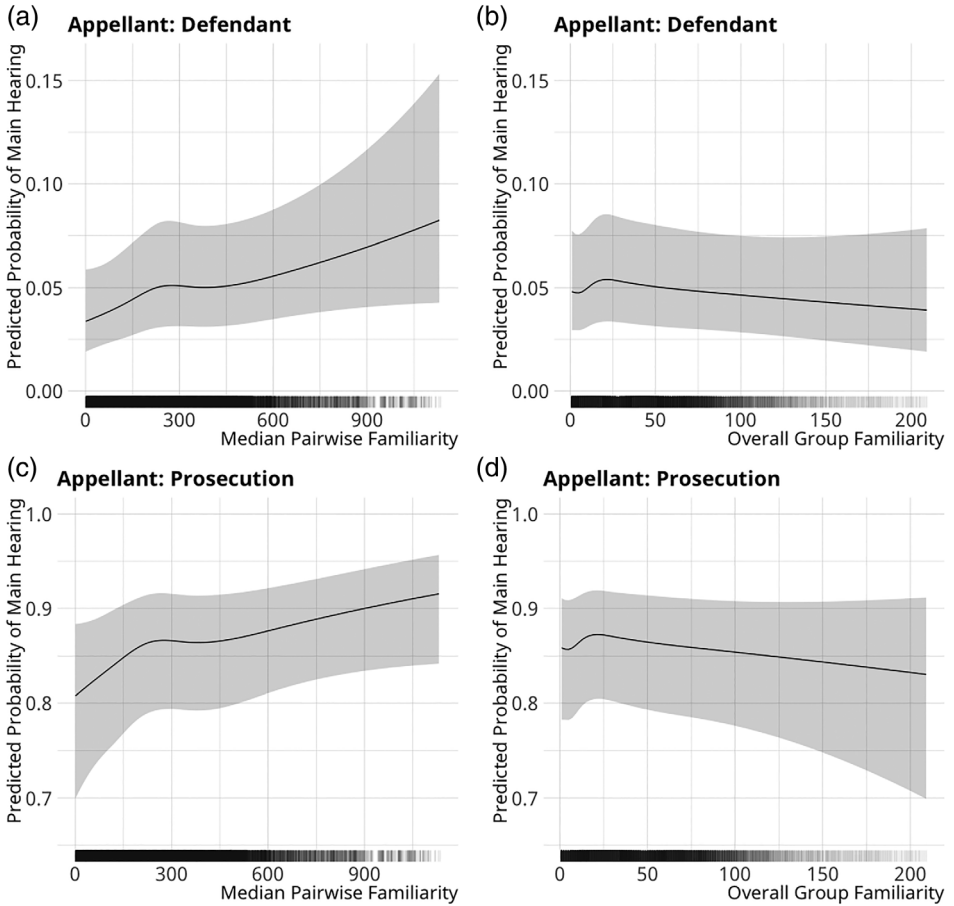
judges are relatively unfamiliar with each other will have a stronger effect on collegiality than a slight increase in familiarity when the group of judges has had years of prior experience with one another. A popular choice to capture potential nonlinearities is regression modeling with a restricted cubic spline (Harrell Jr, 2015). Cubic splines divide the x-axis into intervals by a specified number k of thresholds (knots) and then estimate the relationship between the dependent and independent variable piecewise for each interval. Restricted cubic splines assume a linear relationship in the tails which means that the regression estimates the linear variable plus $k-2$ piecewise cubic variables.

Whether a group scheduled a main hearing is a binary indicator calling for logistic regression. The continuous variable decision length can be appropriately modeled with ordinary least squares. The causal identification strategy as described above requires

Table 2:   ANOVA of OLS Regression

	DV: Decision length (word count)					
	(1)			*(2)*		
	df	F	p	df	F	p
Median pairwise familiarity	4	3.05	0.0159			
Nonlinear	3	3.52	0.0143			
Overall group familiarity				4	14.04	<0.0001
Nonlinear				3	9.27	<0.0001
Freshman	1	0.53	0.4661	1	3.86	0.0493
Average age	1	4.67	0.0306	1	7.69	0.0056
Main hearing	1	5550.05	<0.0001	1	5566.52	<0.0001
Appellant dummy	1	151.10	<0.0001	1	153.34	<0.0001
Crime dummies	8	101.04	<0.0001	8	100.80	<0.0001
Senate-year dummies	134	8.35	<0.0001	134	8.62	<0.0001
Regression	150	71.29	<0.0001	150	71.72	<0.0001

including senate-year fixed effects. Senate-year fixed effects ensure that the estimation only relies on the variance in familiarity created by the random assignment procedure to groups within senates in a given year.
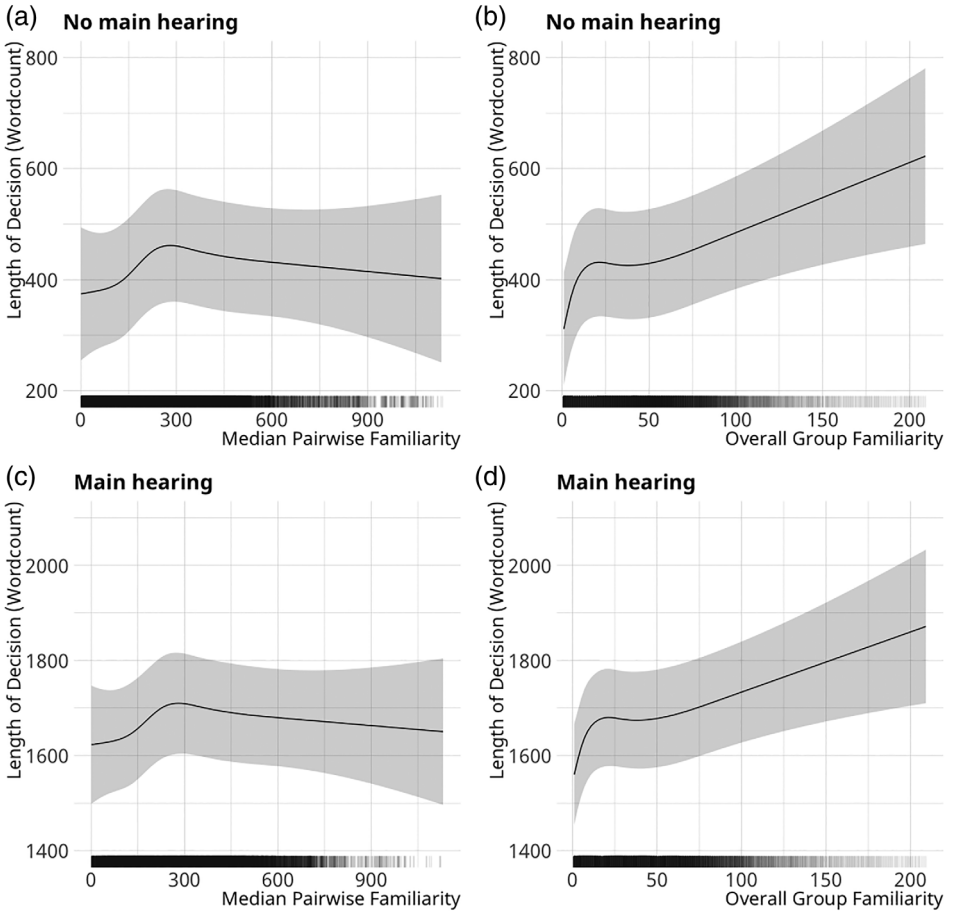
VI. RESULTS

According to the theory, one should expect that familiarity increases the probability of a main hearing. The results presented in the analysis of variance (ANOVA) Table 1 support this prediction for median pairwise familiarity ($p < 0.01$).[17] The Wald test for joint significance of the nonlinear terms is not statistically significant at conventional levels ($p \approx 0.23$). On the other hand, overall group familiarity does not show any significant effect on the likelihood of a main hearing ($p \approx 0.28$). Furthermore, judges are more likely to grant main hearings to appeals filed by the prosecution ($p < 0.01$). The variables "Freshman" and "Average Age" do not have any statistically significant impact.

Figure 6 visualizes the predicted probability that a main hearing is scheduled for appeals by the defendant over the empirical range of the familiarity measures.[18] The causal effect of median pairwise familiarity is nearly linear and substantial in size. For appeals by defendants (Figure 6a), the chances for a main hearing increase by 58.3% (3.6%–5.7%) between the 5th and the 95th percentile. A similar dynamic holds for appeals brought before the court by the prosecution (Figure 6c). Chances for a main hearing increase by

---

[17]The results of replicating the analysis with a sample restricted to the years 1995-2016 are presented in Appendix E, Tables E1 and E2.

[18]The other variables are adjusted to their means or modes: senate = 5, year = 2008, appellant = prosecution, crime = sexual violence, freshman = 0, average age = 56.15.

Figure 7: The effect of familiarity on decision length.

about 6.2 percentage points from 81.7% to 87.9%. For overall group familiarity, Figure 6b,d shows no pronounced dynamic in either direction.

The results reveal substantial differences between deliberations where judges know each other very little or not at all and groups with a high level of familiarity. However, only pairwise familiarity matters for minority dissent. For the overall group familiarity, which is much less robust to small changes in the group's composition, the results did not establish an equivalent effect.

If familiarity increases group productivity and private information sharing and encourages dissenting perspectives, one should see more controversial deliberations. As theorized, a more elaborate discussion should be reflected in a longer justification of a ruling. The results presented in Table 2 lend support to the theoretical prediction.

The table presents the ANOVA statistics of ordinary least squares regressions of decision length on median pairwise familiarity and overall group familiarity. I control for main hearing, the appellant, the presence of a freshman judge, the judges' average age, and crime and senate-year fixed effects. For both measures of familiarity, the coefficients for the main effect and the nonlinear terms are statistically significant ($p < 0.05$). Figure 7 shows the predicted length of decisions with and without a main hearing for both familiarity measures. For both measures, there is a sharp initial rise in the decision length. An increase of median pairwise familiarity from 1 to 300 extends the justification court orders from approximately 374 to 461 words (Figure 7a). Justifications after a main hearing expand from 1623 to 1709 words (Figure 7d). These constitute increases by 23.2% and 5.3%, respectively. Beyond a median pairwise familiarity there is no further increasing effect. The decision length even decreases slightly. A strong initial increase can also be observed regarding overall group familiarity. Between a group without previous decisions in the exact composition together and a group with an experience of 20 cases, the decision length after a main hearing increases by 8.5% (124 words) (Figure 7d) and by 59% for court orders (Figure 7b). Over the entire range of the overall group familiarity, the differences are even more substantial. Increasing overall group familiarity from 1 to 209 increases prolongs justifications by 254 words (court orders: 121%, verdicts: 17.5%).

Also the group's average age and the presence of a freshman judge have a statistically significant influence on decision length. Older judges are less prolix than their younger colleagues. In model (2), increasing the average age from its 5th to its 95th percentile shortens justifications by 129 words. In contrast, the presence of a freshman judge adds 32 words to the decision.

Overall, the results strongly support the hypothesis that group familiarity leads to longer justifications of decisions. As far as decision length serves as a proxy for the number of different arguments raised during deliberations, the results imply that familiarity—especially overall group familiarity—contributes substantially to collegiality, leading to more civil, involved, diverse, and eventually better deliberations.

VII. Conclusion

Does group familiarity improve the deliberations of judges? The findings provide causal evidence that group familiarity substantially increases both the probability of minority dissent and the length of the court's justification. It appears that familiarity cultivates a collegial atmosphere that promotes information sharing, more discussion of different opinions, and lowers barriers for individual judges to speak out against a majority opinion within the group.

These findings echo the voices of scholars and practicing judges who have emphasized the influence of collegiality and interpersonal contact on the behavior of judges (Baum, 1997, 2006; Edwards, 1998, 2003; Nelson et al., 2022). If one accepts the conclusion that familiarity positively impacts group deliberation, the question arises of how courts should ideally organize panel composition to enhance it. Should the court create stable groups of judges that sit together for years, or should a panel be formed by a random draw

of judges for each case? In other words, how much familiarity is necessary to create a collegial atmosphere during deliberations, and how is this level of familiarity best achieved? The results of this study give us at least two crucial answers to these questions.

First, group familiarity often has decreasing marginal benefits: even modest increases in familiarity can significantly influence the group's deliberations when judges do not know each other very well. In contrast, increasing familiarity matters less when judges have shared the bench many times in the past already. This finding speaks against practices that create stable groups of judges who sit together for several years. Over time, the benefits of collegiality decrease, and other factors, such as the lack of new input from outside the group and evolving strong group identity, may become detrimental to group deliberation.

The second implication of the findings is that it can make a difference whether familiarity is thought of as a group attribute or something created between a pair of judges. The empirical analysis reveals that both types of familiarity matter, but in different ways. Overall group familiarity had no statistically significant effect on the likelihood of a main hearing but proved to increase the length of justifications strongly. In contrast, median pairwise familiarity substantially impacted the probability of a main hearing but had a much weaker influence on the decision length. The conceptual difference between the two types of familiarity also matters in practice. Creating groups by random draws of judges for each case may ensure that judges within a court all get to know one another over time but barely increases overall group familiarity. In sum, the findings advocate for the composition of panels as fluid teams. These groups stay together for some time but reshuffle in regular intervals. This practice generates both pairwise and overall group familiarity. By composing new groups once the marginal effect of familiarity becomes too small, the court would create collegiality within panels while simultaneously allowing judges to build a network of collegial relationships with many colleagues over time.

Finally, while judicial deliberation is a central aspect of judicial decision-making in all legal traditions, it is particularly relevant for civil law countries. Empirical legal scholarship is still much focused on courts in common law countries where judges can make their opinions heard by writing separately and are more prominent intellectual public figures in general (Merryman & Pérez-Perdomo, 2007). In the civil law tradition, separate opinion writing is discouraged, and decisions are announced as the group's consensus, even without a vote tally. This fact only further elevates the importance of well-structured, well-organized judicial deliberations.

References

Asch, S. E. (1956). Studies of independence and conformity: I. a minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, 70(9), 1–70.

Avgerinos, E., Fragkos, I., & Huang, Y. (2019). Team familiarity in cardiac surgery operations: The effects of hierarchy and failure on team productivity. *Human Relations*, 73(9), 1278–1307.

Avgerinos, E., & Gokpinar, B. (2017). Team familiarity and productivity in cardiac surgery operations: The effect of dispersion, bottlenecks, and task complexity. *Manufacturing & Service Operations Management*, 19(1), 19–35.

Baum, L. (1997). *The puzzle of judicial behavior*. University of Michigan Press.

Baum, L. (2006). *Judges and their audiences: A perspective on judicial behavior*. Princeton University Press.

Bowen, T. (1995). Consensual norms and the freshman effect of the United States supreme court. *Social Science Quarterly*, 76(1), 222–231.

Bowen, T., & Scheb, J. M., II. (1993). Freshman opinion writing on the U.S. Supreme Court, 1921–1991. *Judicature*, 76(5), 239–243.

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

Brenner, S., & Hagle, T. M. (1996). Opinion writing and acclimation effect. *Political Behavior*, 18(3), 235–261.

Bundesgerichtshof. (2017). *Entscheidungsdatenbank des Bundesgerichtshofs*. https://juris.bundesgerichtshof. de/cgi-bin/rechtsprechung/list.py?Gericht=bgh\&Art=en\&Datum=Aktuell\&Sort=12288

Coffin, F. M. (1994). *On appeal: Courts, lawyering, and judging*. W.W. Norton & Company.

Cohen, J. (2002). *Inside appellate courts. The impact of court organization on judicial decision making in the United States Courts of Appeals*. University of Michigan Press.

Cross, F. B., & Tiller, E. H. (1998). Judicial partisanship and obedience to legal doctrine: Whistleblowing on the federal courts of appeals. *Yale Law Journal*, 107(7), 2155–2176.

Cross, F. B., & Tiller, E. H. (2008). Understanding collegiality on the court. *University of Pennsylvania Journal of Constitutional Law*, 10, 257–271.

Dalal, D. K., Nolan, K. P., & Gannon, L. E. (2017). Are pre-assembly shared work experiences useful for temporary-team assembly decisions? A study of olympic ice hockey team composition. *Journal of Business and Psychology*, 32(5), 561–574.

Dirks, K. T., & Ferrin, D. L. (2001). The role of trust in organizational settings. *Organization Science*, 12(4), 450–467.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350–383.

Edwards, H. T. (1998). Collegiality and decision making on the D.C. circuit. *Virginia Law Review*, 84(7), 1335–1370.

Edwards, H. T. (2003). The effects of collegiality on judicial decision making. *University of Pennsylvania Law Review*, 151(5), 1639.

Edwards, H. T. (2017). *Collegial decision making in the US courts of appeals*. NYU School of Law, Public Law Research Paper, 17–47.

Engel, C., & Weinshall, K. (2020). Manna from heaven for judges: Judges' reaction to a quasi-random reduction in caseload. *Journal of Empirical Legal Studies*, 17(4), 722–751.

Epstein, L., & Knight, J. (2013). Reconsidering judicial preferences. *Annual Review of Political Science*, 16, 11–31.

Epstein, L., Landes, W. M., & Posner, R. A. (2011). Why (and when) judges dissent: A theoretical and empirical analysis. *The Journal of Legal Studies*, 3(1), 101–137.

Espinosa, J. A., Slaughter, S. A., Kraut, R. E., & Herbsleb, J. D. (2007). Familiarity, complexity, and team performance in geographically distributed software development. *Organization Science*, 18(4), 613–630.

Farhang, S., & Wawro, G. (2004). Institutional dynamics on the U.S. court of appeals: Minority representation under panel decision making. *Journal of Law, Economics, and Organization*, 20(2), 299–330.

Fischer, T. (2015). *Die Augen Des Revisionsgerichts*. http://www.zeit.de/gesellschaft/zeitgeschehen/ 2015-06/bundesgerichtshof-justiz-fischer-im-recht/komplettansicht

Fischer, T., Eschelbach, R., & Krehl, C. (2013). Das Zehn-Augen-Prinzip—Zur Revisionsgerichtlichen Beschlusspraxis in Strafsachen. *StV Strafverteidiger*, 6, 395–402.

Fischman, J. (2011). Estimating preferences of circuit judges: A model of consensus voting. *Journal of Law and Economics*, 54(4), 781–809.

Fischman, J. (2013). Interpreting circuit court voting patterns: A social interactions framework. *Journal of Law, Economics, and Organization*, 31(4), 808–842.

Goodman, P. S., & Leyden, D. P. (1991). Familiarity and group productivity. *Journal of Applied Psychology*, 76(4), 578–586.

Gruenfeld, D. H., Mannix, E. A., Williams, K. Y., & Neale, M. A. (1996). Group composition and decision making: How member familiarity and information distribution affect process and performance. *Organizational Behavior and Human Decision Processes*, 67(1), 1–15.

Hagle, T. M. (1993). "Freshman effects" for the Supreme Court. *American Journal of Political Science*, 37(4), 1142–1157.

Hamann, H. (2019). The German federal courts dataset 1950: From paper archives to linked open data. *Journal of Empirical Legal Studies*, 16(3), 671–688.

Harrell, F. E., Jr. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis.* Springer Series in Statistics. Springer International Publishing.

Hettinger, V. A., Lindquist, S. A., & Martinek, W. L. (2003). The role and impact of chief judges on the United States courts of appeals. *The Justice System Journal*, 24(1), 91–117.

Hinkle, R. K. (2017). Panel effects and opinion crafting in the US courts of appeals. *Journal of Law and Courts*, 5(2), 313–336.

Hinkle, R. K., Nelson, M. J., & Hazelton, M. L. W. (2020). Deferring, deliberating, or dodging review: Explaining counterjudge success in the US courts of appeals. *Journal of Law and Courts*, 8(2), 277–300.

Huckman, R. S., & Staats, B. R. (2011). Fluid tasks and fluid teams: The impact of diversity in experience and team familiarity on team performance. *Manufacturing & Service Operations Management*, 13(3), 310–328.

Hurwitz, M. S., & Stefko, J. V. (2004). Acclimation and attitudes: "Newcomer" justices and precedent conformance on the Supreme Court. *Political Research Quarterly*, 57(1), 121–129.

Janis, I. L. (1982). *Groupthink: Psychological studies of policy decisions and fiascoes.* Houghton Mifflin Boston.

Janssen, J., Erkens, G., Kirschner, P. A., & Kanselaar, G. (2009). Influence of group member familiarity on online collaborative learning. *Computers in Human Behavior*, 25, 161–170.

Jehn, K. A., & Shah, P. P. (1997). Interpersonal relationships and task performance: An examination of mediation processes in friendship and acquaintance groups. *Journal of Personality and Social Psychology*, 72(4), 775–790.

Kane, A. A., Argote, L., & Levine, J. M. (2005). Knowledge transfer between groups via personnel rotation: Effects of social identity and knowledge quality. *Organizational Behavior and Human Decision Processes*, 96(1), 56–71.

Kastellec, J. (2013). Racial diversity and judicial influence on appellate courts. *American Journal of Political Science*, 57(1), 167–183.

Katholnigg, O. (1992). Zur Geschäftsverteilung bei obersten Gerichtshöfen des Bundes und innerhalb ihrer Senate. *Neue Juristische Wochenschrift*, 36, 2256–2260.

Kim, P. H. (1997). When what you know can hurt you: A study of experiential effects on group discussion and performance. *Organizational Behavior and Human Decision Processes*, 69(2), 165–177.

King, G., & Zeng, L. (2006). The dangers of extreme counterfactuals. *Political Analysis*, 14(2), 131–159.

Klocke, U. (2007). How to improve decision making in small groups: Effects of dissent and training interventions. *Small Group Research*, 38(3), 437–468.

Kornhauser, L. A., & Sager, L. G. (1993). The one and the many: Adjudication in collegial courts. *California Law Review*, 81(1), 1–59.

Levy, G., & Razin, R. (2015). Correlation neglect, voting behavior, and information aggregation. *American Economic Review*, 105(4), 1634–1645.

Levy, M. K. (2017). Panel assignment in the federal courts of appeals. *Cornell Law Review*, 103(1), 65–116.

Levy, M. K., & Chilton, A. S. (2015). Challenging the randomness of panel assignment in the federal courts of appeals. *Cornell Law Review*, 101(1), 1–56.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, 85(2), 273–283.

Maynard, M. T., Mathieu, J. E., Gilson, L. L., Sanchez, D. R., & Dean, M. D. (2019). Do I really know you and does it matter? Unpacking the relationship between familiarity and information elaboration in global virtual teams. *Group & Organization Management*, 44(1), 3–37.

Merrit, D. J., & Brudney, J. J. (2001). Stalking secret law: What predicts publication in the United States courts of appeal. *Vanderbilt Law Review*, 54, 71–121.

Merryman, J. H., & Pérez-Perdomo, R. (2007). *The civil law tradition: An introduction to the legal systems of Europe and Latin America* (3rd ed.). Stanford University Press.

Murphy, M. R. (2000). Collegiality and technology. *Journal of Appellate Practice and Process*, 2(2), 455–461.

Nelson, M. J., Hazelton, M. L. W., & Hinkle, R. (2022). How interpersonal contact affects appellate review. *Journal of Politics*, 84(1). https://doi.org/10.1086/714783

Peresie, J. (2005). Female judges matter: Gender and collegial decisionmaking in the federal appellate courts. *Yale Law Journal*, 114(7), 1759–1790.

Reagans, R., Argote, L., & Brooks, D. (2005). Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. *Management Science*, 51(6), 869–881.

Revesz, R. L. (1997). Environmental regulation, ideology, and the D.C. circuit. *Virginia Law Review*, 83(8), 1717–1772.

Richman, W. M., & Reynolds, W. L. (2012). *Injustice on appeal: The United States courts of appeals in crisis*. Oxford University Press.

Roberts, J. G. (2006). What makes the D.C. circuit different?: A historical view. *Virginia Law Review*, 92(3), 375–389.

Rosenau, H. (2012). Die offensichtliche Ungesetzlichkeit der "Ou"-Verwerfung nach §349 Abs. 2 StPO in der Spruchpraxis des BGH. *Zeitschrift Für Internationale Strafrechtsdogmatik*, 5, 195–205.

Sangmeister, B. (1993). BGH, 30. 3. 1993 X ZR 51/92. Zur senatsinternen Geschäftsverteilung beim BGH. *JuristenZeitung*, 48(14), 733–740.

Shah, P. P., & Jehn, K. A. (1993). Do friends perform better than acquaintances? The interaction of friendship, conflict, and task. *Group Decision and Negotiation*, 2(2), 149–165.

Siemsen, E., Roth, A. V., Balasubramanian, S., & Anand, G. (2009). The influence of psychological safety and confidence in knowledge on employee knowledge sharing. *Manufacturing & Service Operations Management*, 11(3), 429–447.

Sowada, C. (2002). *Der Gesetzliche Richter Im Strafverfahren*. Walter de Gruyter.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology*, 48(6), 1467–1478.

Statistisches Bundesamt (1990). Rechtspflege—Strafgerichte. Fachserie 10/Reihe 2.3.

Sunstein, C. R., Ellmann, L. M., & Schkade, D. (2004). Ideological voting on federal courts of appeals: A preliminary investigation. *Virginia Law Review*, 90, 301–354.

Sunstein, C. R., Schkade, D., Ellman, L. M., & Sawicki, A. (2006). *Are judges political?: An empirical analysis of the federal judiciary*. Brookings Institution Press.

Wasby, S. L. (1987). Communication in the ninth circuit: A concern for collegiality. *University of Puget Sound Law Review*, 11, 73–138.

Wittenbaum, G. M. (1998). Information sampling in decision-making groups: The impact of members' task-relevant status. *Small Group Research*, 29(1), 57–84.

Wittenbaum, G. M., Shulman, H. C., & Braz, M. E. (2010). Social ostracism in task groups: The effects of group composition. *Small Group Research*, 41(3), 330–353.

Wood, S. L., Keith, L. C., Lanier, D. N., & Ogundele, A. (1998). "Acclimation effects" for the Supreme Court justices: A cross-validation, 1888–1940. *American Journal of Political Science*, 42(2), 690–697.

Appendix A: Example of a Case Allocation Plan

BUNDESGERICHTSHOF
3. Strafsenat

BESCHLUSS

vom 20. Dezember 2012

Grundsätze, nach welchen die Senatsmitglieder an den
beim Senat anfallenden Sachen im Jahre 2013 mitwirken

I.   Revisionsverfahren und andere Verfahren, in denen in der Fünferbesetzung zu entscheiden ist

1.   Der mit 1 : 7 Richtern besetzte Senat entscheidet in drei Sitzgruppen; diese sind - außer dem Vorsitzenden - wie folgt besetzt:

Sitzgruppe I: RiBGH Pfister
                 Ri'inBGH Sost-Scheible
                 RiBGH Dr. Schäfer
                 RiBGH Mayer
                 RiBGH Gericke

Nimmt Ri'in BGH Sost-Scheible an der Beratung teil, so scheidet das dienstjüngste Mitglied der Spruchgruppe, das ansonsten zur Mitwirkung berufen wäre, aus der Spruchgruppe aus; in Sachen, in denen dem dienstjüngsten Mitglied der Spruchgruppe die Berichterstattung obliegt, scheidet das nächste dienstältere Mitglied der Spruchgruppe aus.

Sitzgruppe II: RiBGH Hubert
                  RiBGH Dr. Schäfer
                  RiBGH Gericke
                  Ri'inBGH Dr. Spaniol

Sitzgruppe III: RiBGH Pfister

RiBGH Hubert

RiBGH Mayer

Ri'inBGH Dr. Spaniol

2. a)   Die Zugehörigkeit einer Sache zu einer Sitzgruppe bestimmt sich nach der Endziffer des vom Generalbundesanwalt nach Maßgabe seiner Verfügung vom 16.5.1997 vergebenen Aktenzeichens, wobei die "0" außer Betracht bleibt:

Sitzgruppe I   Endziffer 1, 4, 7

Sitzgruppe II   Endziffer 2, 5, 8

Sitzgruppe III   Endziffer 3, 6, 9.

b)   Für Anfrageverfahren nach § 132 Abs. 2 GVG, für Entscheidungen nach dem Gesetz über parlamentarische Untersuchungsausschüsse sowie für Entscheidungen des Senats als gemeinschaftliches oberes Gericht in Verfahren nach dem Außenwirtschaftsgesetz gilt die Regelung zu a) mit der Maßgabe, dass es auf die Endziffer des ARs-Aktenzeichens des Senats ankommt.

3. a)   Rechtlich oder tatsächlich eng zusammenhängende Sachen werden einer anderen Sitzgruppe zugewiesen, wenn dies sachgerecht ist, um unterschiedliche Beurteilungen oder Doppelarbeit zu vermeiden. Besteht ein solcher Zusammenhang zwischen einer im Senat bereits anhängig gewesenen Beschwerde-, Haftprüfungs- oder Revisionssache und ist der dortige Berichterstatter nicht Mitglied der zuständigen Sitzgruppe, so tritt er anstelle des dienstjüngsten Mitgliedes in die Sitzgruppe ein.

b)   Rückläufer werden in der Sitzgruppe des ersten Durchgangs beraten. 3 a) Satz 2 gilt entsprechend.

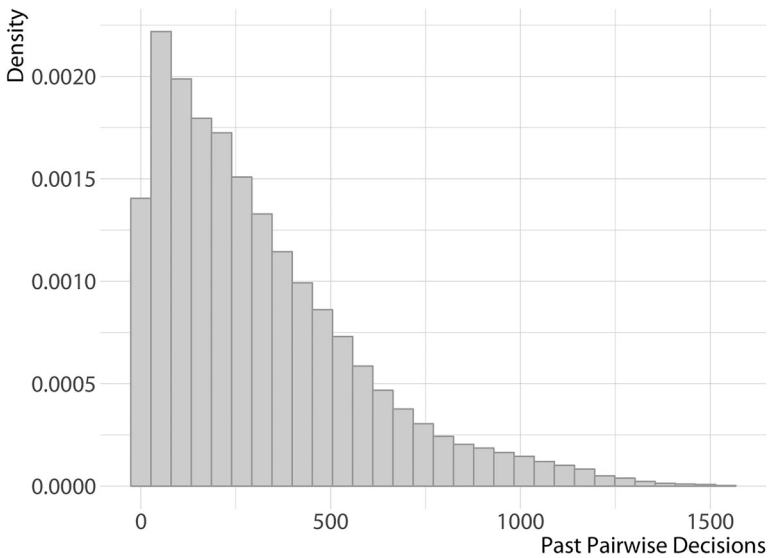# Appendix B: Coverage of the Dataset

Figure B1: Share of decisions included in the dataset.

Share of Published Decisions
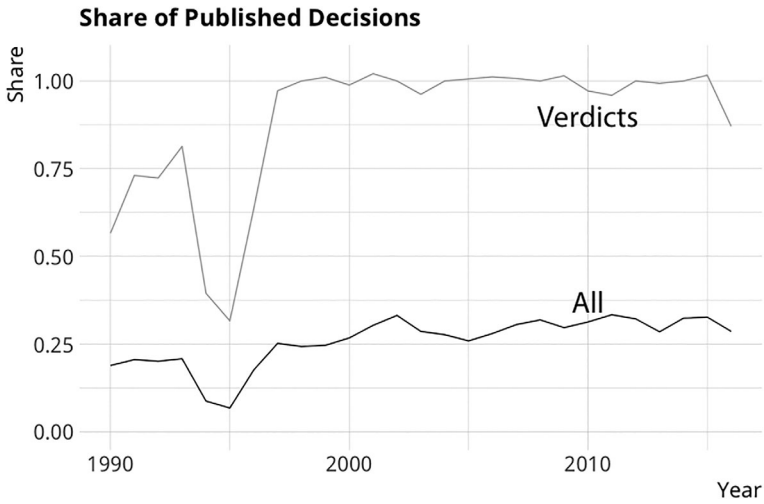
Appendix C: Distribution of Past Pairwise Decisions

Figure C1: Distribution of past decisions by pairs of judges.

APPENDIX D: ROBUSTNESS: MEAN PAIRWISE FAMILIARITY

D.1. DV: Main Hearing

Table D1:   Wald Test of Logistic Regression

	DV: Main hearing					
	(1)			(2)		
	χ^2	df	p	χ^2	df	p
Mean pairwise familiarity	11.08	4	0.0256	9.66	4	0.0466
Nonlinear	8.33	3	0.0396	8.20	3	0.0421
Freshman				0.00	1	0.9517
Average age				0.08	1	0.7783
Appellant (prosecution)	3444.21	1	<0.0001	3443.92	1	<0.0001
Crime dummies	172.15	8	<0.0001	172.05	8	<0.0001
Senate-year dummies	466.68	134	<0.0001	446.94	134	<0.0001
Total	3759.17	147	<0.0001	3758.99	149	<0.0001

In model (1), mean pairwise familiarity is statistically significant on the 95% confidence level. The terms accounting for nonlinearities in the effect of mean pairwise familiarity on the probability of a main hearing are also significant on the 95% confidence level. Model (2) includes the variables Freshman and Average Age as controls. Note that both Freshman and Average Age are statistically insignificant. The effect of mean pairwise familiarity and its terms accounting for nonlinearity remain both statistically significant on the 95% confidence level.

Both the Akaike Information Criterion (model (1): 15080.55, model (2): 15084.46) and the log-likelihood ratio test (χ^2(df = 2) = 0.0902, $p = 0.9559$) indicate that model (1) is preferable.

D.2. DV: Word Count

Table D2:   ANOVA of OLS Regression

	DV: Decision length (word count)					
	(1)			(2)		
	df	F	p	df	F	p
Mean pairwise familiarity	4	4.69	0.0009	4	9.40	<0.0001
Nonlinear	3	4.21	0.0055	3	5.91	0.0005
Freshman				1	3.68	0.0550
Average age				1	16.34	0.0001
Main hearing	1	5548.03	<0.0001	1	5551.94	<0.0001
Appellant dummy	1	151.60	<0.0001	1	150.90	<0.0001
Crime dummies	8	101.01	<0.0001	8	100.36	<0.0001
Senate-year dummies	134	8.34	<0.0001	134	8.17	<0.0001
Regression	148	72.29	<0.0001	150	71.54	<0.0001

In both models mean pairwise familiarity and its nonlinear terms are statistically significant on the 99% confidence level. Average Age in model (2) is statistically significant on the 99% level of confidence, while Freshman is significant on the 90% level of confidence.

Both the Akaike Information Criterion (model (1): 357864.5, model (2): 357846.0) and the log-likelihood ratio test ($\chi^2(\mathrm{df} = 2) = 22.523$, $p < 0.0001$) indicate that model (2) is preferable.

Overall, the results show that the findings of the main analysis are robust to using the mean pairwise familiarity instead of the median pairwise familiarity.

APPENDIX E: ROBUSTNESS: REANALYZING DATA 1995–2016

E.1. DV: Main Hearing

Table E1: Wald Statistics of Logistic Regression, 1995–2016

	DV: Main hearing					
	(1)			(2)		
	χ^2	df	p	χ^2	df	p
Median pairwise familiarity	15.21	4	0.0043			
Nonlinear	4.33	3	0.2276			
Overall group familiarity				7.23	4	0.1244
Nonlinear				7.18	3	0.0663
Freshman	0.02	1	0.8987	0.32	1	0.5725
Average age	0.39	1	0.5323	0.23	1	0.6296
Appellant (prosecution)	3090.54	1	<0.0001	3090.83	1	<0.0001
Crime dummies	146.42	8	<0.0001	148.06	8	<0.0001
Senate-year dummies	364.72	109	<0.0001	366.23	109	<0.0001
Total	3332.41	124	<0.0001	3334.64	124	<0.0001

E.2. DV: Word Count

Table E2: ANOVA of OLS Regression, 1995–2016

	DV: Decision length (word count)					
	(1)			(2)		
	df	F	p	df	F	p
Median pairwise familiarity	4	2.69	0.0295			
Nonlinear	3	3.05	0.0275			
Overall group familiarity				4	14.12	<0.0001
Nonlinear				3	9.77	<0.0001
Freshman	1	136	0.2428	1	5.89	0.0152
Average age	1	4.41	0.0357	1	7.59	0.0059
Main hearing	1	4933.28	<0.0001	1	4945.48	<0.0001
Appellant dummy	1	156.71	<0.0001	1	159.19	<0.0001
Crime dummies	8	90.65	<0.0001	8	90.22	<0.0001
Senate-year dummies	109	7.37	<0.0001	109	7.48	<0.0001
Regression	125	75.59	<0.0001	125	76.14	<0.0001

The results show that the findings of the main analysis are robust to limiting the dataset to the years 1995–2016.