2015

Challenging the Randomness of Panel Assignments in the Federal Courts of Appeals

Adam S. Chilton

Marin Levy

CHALLENGING THE RANDOMNESS OF PANEL ASSIGNMENT IN THE FEDERAL COURTS OF APPEALS

Adam S. Chilton† *& Marin K. Levy*‡

A fundamental academic assumption about the federal courts of appeals is that the three-judge panels that hear cases have been randomly configured. Scores of scholarly articles have noted this "fact," and it has been relied on heavily by empirical researchers. Even though there are practical reasons to doubt that judges would always be randomly assigned to panels—such as courts might well want to take into account the scheduling needs of their judges—this assumption has never been tested. This Article is the first to do so, and it calls the assumption into question.

To determine whether the circuit courts utilize random assignment, we have created what we believe to be the largest dataset of panel assignments of those courts constructed to date. Using this dataset, we tested whether panel assignments are, in fact, random by comparing the actual assignments to truly random panels generated by code that we have created to simulate the panel generation process. Our results provide evidence of nonrandomness in the federal courts of appeals.

To be sure, the analysis here is descriptive, not explanatory or normative. We do not ourselves mean to suggest that strict randomness is a desirable goal and indeed note that there are many good reasons for departing from it. Our aim is to test an existing scholarly assumption, and we believe our findings will have implications for the courts, court scholars, and empirical researchers.

---

† Assistant Professor of Law, University of Chicago School of Law.

‡ Associate Professor of Law, Duke University School of Law.

Thanks to Will Baude, Stuart Benjamin, Joseph Blocher, Jamie Boyle, Curt Bradley, Stephen Burbank, Josh Chafetz, Guy Charles, Henry Corrigan-Gibbs, Justin Driver, Frank Easterbrook, Josh Fischman, Jonah Gelbach, Mitu Gulati, William Hubbard, Grayson Lambert, Maggie Lemos, Guangya Liu, Tom Miles, Eric Posner, Richard Posner, Stephen Sachs, Maya Sen, Georg Vanberg, Ben Walther, and Keith Winstein, as well as to the participants of the Duke University School of Law Faculty Workshop and the University of Chicago Law School Faculty Workshop. Special thanks to Jacob Adrian and Sean Chen for their invaluable work on the project. All views expressed here, as well as any errors and solecisms are, of course, our own.

INTRODUCTION

A fundamental academic assumption about the federal courts of appeals is that the three-judge panels that decide cases have been randomly configured. Scores of scholarly articles note this "fact,"[1] with some scholars even going so far as to call the random creation of panels a "hallmark" of the American appellate system.[2] Indeed, an entire quantitative literature on the courts—arguing that the composition of panels has substantive effects on case outcomes—has justified its research

---

1 *See, e.g.*, Michael Abramowicz & Maxwell Stearns, *Defining Dicta*, 57 STAN. L. REV. 953, 1009 (2005) ("In the Courts of Appeals, panels are the product of random draws of three among a larger set of members of the court."); Lee Epstein, William M. Landes & Richard A. Posner, *Why (and When) Judges Dissent: A Theoretical and Empirical Analysis*, 3 J. LEGAL ANALYSIS 101, 110 (2011) ("We assume that members of a panel are chosen randomly from the judges of the court, which is the practice in all circuits." (citation omitted)); Cass R. Sunstein & Thomas J. Miles, *Depoliticizing Administrative Law*, 58 DUKE L.J. 2193, 2197 (2009) (noting that within the federal courts of appeals, "judges are randomly assigned to three-judge panels").

2 Emerson H. Tiller & Frank B. Cross, *A Modest Proposal for Improving American Justice*, 99 COLUM. L. REV. 215, 216 & n.4 (1999) ("[T]he random assignment of federal appellate judges to panels has become a 'hallmark' of the system.").

design and identification strategy for making causal inferences on this basic premise.[3]

And yet, what if the panels in many of the circuit courts were not strictly random? This seems quite possible if not even likely given that courts might want to take into account, among other things, the scheduling needs of the judges and want to space out sittings to keep workloads manageable. And what if taking into account these factors ultimately had an effect on the distribution of a judicial characteristic, such as ideology? While a handful of scholars have considered the possibility of nonrandomness generally,[4] there has been no systematic testing of whether the panels of the twelve circuit courts are in fact nonrandom. This Article is a first step in filling this void, and it calls the academic assumption into question.

To test this randomness assumption, we have built what we believe to be the most comprehensive dataset of federal appellate court panel assignments constructed to date. Our dataset brings together the calendar information for all twelve regional circuits during a five-year period between September 2008 and August 2013.[5] This dataset contains information on

---

3 *See infra* Subpart I.C.

4 *See, e.g.*, Burton M. Atkins & William Zavoina, *Judicial Leadership on the Court of Appeals: A Probability Analysis of Panel Assignment in Race Relations Cases on the Fifth Circuit*, 18 AM. J. POL. SCI. 701, 704 (1974); J. Robert Brown, Jr. & Allison Herren Lee, *Neutral Assignment of Judges at the Court of Appeals*, 78 TEX. L. REV. 1037, 1041–43 (2000); Tom S. Clark, *A Principal-Agent Theory of En Banc Review*, 25 J.L. ECON. & ORG. 55, 57 n.3 (2009).

Note that whether appellate panels are randomly configured is a different question than whether cases are assigned randomly to judges once those panels are in place. On that score, scholars have probed more deeply. *See, e.g.*, Matthew Hall, *Randomness Reconsidered: Modeling Random Judicial Assignment in the U.S. Courts of Appeals*, 7 J. EMPIRICAL LEGAL STUD. 574 (2010) (assessing CASS R. SUNSTEIN ET AL., ARE JUDGES POLITICAL? AN EMPIRICAL ANALYSIS OF THE FEDERAL JUDICIARY (2006) in light of assumed random judicial assignment of cases in the federal courts of appeals).

5 *See infra* Subpart II.C.

over 775[6] judges who have comprised over 10,000 panels during this time.[7]

Using this dataset, we are able to test whether the panels of judges that actually were formed are consistent with random assignment. To do so, we have written code that takes the number of panels that were formed each term in every circuit and the judges who sat in the circuit at that time, and then simulates the random creation of panels a large number of times—in our case, we use simulation to generate over a billion panels.[8] We can then compare the distribution of some observable characteristic in the panels as they were actually formed with the panels that resulted from random assignment, and evaluate the likelihood that panel assignments in a given circuit were random.[9]

To test random assignment in this way, one could select just about any observable characteristic of the judges to evaluate (such as whether or not the distribution of judges who wear eyeglasses is random). We wanted to select a characteristic that was easily observable, objective, had variance, and (unlike whether a judge wears eyeglasses) was relevant to the academic literature. Accordingly, we decided to select whether or

---

6 A follower of the federal courts might be surprised to see such a high figure here, as the number of active appellate judgeships in the country is 179 (and this figure includes the Federal Circuit, which we do not). *See U.S. Courts of Appeals Additional Authorized Judgeships*, ADMIN. OFF. U.S. COURTS, http://www.uscourts.gov/judges-judgeships/authorized-judgeships [http://perma.cc/AGN8-FCUY] (follow the "Additional Authorized Judgeships—Since 1960" hyperlink under the "Courts of Appeals" heading) (listing federal court of appeals judgeships by circuit). However, the 179 figure only includes active appellate judgeships, not senior judgeships, and in some circuits the number of senior judges is close to the number of active judges. Moreover, it should be remembered that many district judges sit by designation at least occasionally on the circuit courts, and those judges are included in our dataset. There are currently 667 active district court judgeships (a figure that also does not include senior judgeships). *U.S. District Courts Additional Authorized Judgeships*, ADMIN. OFF. U.S. COURTS, http://www.uscourts.gov/judges-judgeships/authorized-judgeships [http://perma.cc/AGN8-FCUY] (follow the "Additional Authorized Judgeships—Since 1960" hyperlink under the "District Courts" heading). Finally, it should be remembered that other Article III judges, such as judges from the United States Court of International Trade, occasionally sit by designation, which raises the total number of judges during this time frame as well.

7 *See infra* Subpart II.C. "Judicial Term" here is somewhat artificial. Although some circuits, such as the D.C. Circuit, have formal terms, others do not. We use this phrase to mean the year of sittings between September 1 and August 31 of the following calendar year.

8 For an explanation of why we used simulation to generate one billion possible panel assignments, see *infra* Subpart II.D.

9 For a discussion of the assumptions that were made while generating these predictions, see *infra* Subpart II.C & Subpart II.D. For a discussion of the limitations of our method, see *infra* Subpart III.C.

not a judge was appointed by a Republican president. This characteristic plainly satisfies the first three criteria. On the fourth, the party of the appointing president is a widely used proxy for the ideology of judges,[10] and a great deal of scholarship has focused on testing the influence of ideology on judicial behavior.[11] Since this characteristic of judges (that is, whether they were appointed by a Republican president) is easily observable, objective, has a great deal of variance, and is theoretically important, we believe it is the ideal characteristic to evaluate the distribution of when testing whether panel assignments are random.[12]

Ultimately, our results provide evidence of nonrandomness along this dimension in the federal courts of appeals. Specifically, our primary results suggest that several of the circuit courts have panels that are nonrandom in ways that impact the ideological balance of panels.

There are several key points that should be clearly understood concerning our data and methodology at the outset. First, we have examined deviations from strict randomness in panel composition to report a statistical phenomenon, not to suggest any improper motive by the chief judges, circuit executives, or clerks of court who perform the immensely challenging task of creating a court calendar for sometimes as many as dozens of judges, often six months or a year in advance. As one coauthor has learned from conducting a multiyear qualitative study of nearly half of the courts of appeals,[13] many entirely legitimate factors will inevitably cause deviations from strict randomness. Such factors include spacing judicial assignments (so that no judge has several week-long sittings in a row), accommodating vacation schedules, honoring the scheduling preferences of senior judges (whose additional days of service provide much-needed support to the court as a whole), and honoring recusals for disqualification, health, or other

---

10 *See, e.g.*, Epstein, Landes & Posner, *supra* note 1, at 109; Hall, *supra* note 4, at 581–85; Sunstein & Miles, *supra* note 1, at 2200–01. We of course note, as others have done, that this proxy is not a perfect measure of one's ideology. *See, e.g.*, Hall, *supra* note 4, at 585–86 (suggesting that scholars "be cautious" when "assum[ing] that party identification is simply a proxy for ideology").

11 *See infra* Subpart II.B.

12 *See infra* Subpart II.B.

13 *See* Marin K. Levy, Beyond Strict Randomness: Understanding and Evaluating Panel Assignment in the Federal Courts of Appeals (2015) (unpublished manuscript) (on file with author) (describing, based on qualitative research of five circuit courts, the different factors courts of appeals take into account when creating calendars).

valid reasons. We have no reason to believe that anything beyond a combination of such factors has led to our results.

Second, as the foregoing discussion makes plain, strict randomness will often not be desirable or even feasible in many circuit courts. As one example, in circuits that do not have designated court weeks but instead hear cases on a rolling basis, strict randomness could lead to one judge being assigned to panels two or even three weeks in a row whereas others could have a gap of several months between assignments. Such a schedule would clearly prove unworkable over time, rendering strict randomness inconsistent with sound court administration.

Third, as the preceding paragraphs make plain, we do not purport to—nor could we—replicate the exact processes used by the circuit courts to create panels. For example, a circuit court could have an unstated rule that no more than one senior judge will sit on a panel so that the majority of each panel is comprised of active members of court. This rule, in turn, could have an impact on the ideological balance of panels if, say, the cohort of senior judges had more conservative judges or more liberal judges when compared to the court as a whole—a limitation that we do not use. Again, our position is that rules like this are departures from randomness (and generally are not publicly available). As such, we test whether the panels that were actually formed are what would be expected if the judges were randomly assigned to panels, based on the number of panels that the judges actually sat on in a given term.

Finally, like much empirical scholarship, our results are communicated in terms of statistical significance—that is, what we can ultimately say is whether the chances of a given circuit's panels being randomly configured are quite low or not. It thus may be the case that in any given circuit, the chief judge, circuit executive, or clerk of court utilized a random process and simply produced results that were highly unlikely to occur. Or alternatively, it may be the case that a given circuit used a nonrandom process but produced results that appeared random. That said, although we cannot definitively determine whether a process is nonrandom, we can say that it is unlikely that chance alone can explain the evidence of nonrandomness that we detected overall. Indeed, we can say with roughly 97% confidence that the evidence of nonrandomness that we detected for the ideological balance of panels cannot be

explained by chance alone.[14] Accordingly, our results suggest that the assumption that panels are randomly configured in all of the federal courts of appeals is false.

Although our data and method are not without limitations, we believe our results have several implications: (A) the publicly available information about how federal appellate panels are formed appears to be incomplete; (B) the judicial literature going forward should not assume that panels are in fact random; and (C) the quantitative literature specifically should rethink relying on the randomness assumption when conducting empirical research and the results of some existing studies may need to be reconsidered.

The Article proceeds as follows. In Part I we provide an overview of the long-standing assumption of random panels. Specifically, this Part explains generally how oral argument panels are formed and shows how much of the publicly available information about the process describes it as a random one or at least consistent with a random one. It then considers how broadly the randomness assumption has been held, and its relevance, in the general courts and quantitative literatures. In Part II we detail the methodology of the project, including the construction of a dataset on oral argument panels, the hypothesis we set out to test, and how that hypothesis was ultimately tested. Then, in Part III, we present the findings, which provide evidence of nonrandomness in panel formation. Finally, in Part IV, we briefly consider the implications of these findings for courts, the judicial and quantitative literatures, and our understanding of the federal appellate system more broadly.

I
THE LONG-STANDING ASSUMPTION OF RANDOM PANELS

It has been the practice of the federal courts of appeals for well over a century to hear cases in panels of three judges.[15] Though the creation of panels is technically governed by stat-

---

14 This estimate is based on a number of assumptions. For an explanation of how we calculated these probabilities, see *infra* Subpart III.B.

15 The general form of the current courts of appeals originates with the Act of March 3, 1891 (Evarts Act), ch. 517, 26 Stat. 826. Of course, not all cases receive oral argument; indeed, the majority of cases today that are decided on the merits are decided solely on the briefs. *See* ADMIN. OFFICE OF THE U.S. COURTS, FEDERAL JUDICIAL CASELOAD STATISTICS 2013, tbl.B-1 (2013), http://www.uscourts.gov/Viewer.aspx?doc-/uscourts/Statistics/FederalJudicialCaseloadStatistics/2013/tables/B01Mar13.pdf [http://perma.cc/2CEQ-RL6K] (follow the "Download Data Table" hyperlink) (stating that of the 38,245 cases terminated on the merits in the twelve-month period ending March 31, 2013, 28,593 were submitted on the briefs, meaning decided without oral argument).

ute, the relevant code states only that the panels must consist of three judges and that those panels shall sit at the times and places "as the court directs."[16] Accordingly, it falls upon the circuits not just to create their own panels, but also to determine how to create their own panels. The circuits have responded in different ways with some courts relying upon the chief judge[17] to construct the panels and others relying on the clerk of court[18] or the circuit executive.[19] One commonality is that quite a few circuits state that the process by which they create their panels is a random one or at least would produce results consistent with a random one.[20]

The notion that panels are formed randomly has moved into the various academic literatures. There is a robust general courts literature that has consistently held this assumption. Furthermore, there is a broad, quantitative literature on judicial decision making that has relied on this assumption to reach its results.

In this Part we review the long-standing assumption of panel randomness. First, we consider the publicly available information about panel formation, including a recent monograph by the Federal Judicial Center and the local rules and internal operating procedures of the circuits, in which the formation process is often described as random or consistent with a random process. Second, we discuss the general courts and quantitative literatures respectively, showing how the assumption of randomness has been widely held for decades and has been important to numerous articles along the way.

---

16 28 U.S.C. § 46(b) (2012). Specifically, subsection (b) reads in full:

> In each circuit the court may authorize the hearing and determination of cases and controversies by separate panels, each consisting of three judges, at least a majority of whom shall be judges of that court, unless such judges cannot sit because recused or disqualified, or unless the chief judge of that court certifies that there is an emergency including, but not limited to, the unavailability of a judge of the court because of illness. Such panels shall sit at the times and places and hear the cases and controversies assigned as the court directs. The United States Court of Appeals for the Federal Circuit shall determine by rule a procedure for the rotation of judges from panel to panel to ensure that all of the judges sit on a representative cross section of the cases heard and, notwithstanding the first sentence of this subsection, may determine by rule the number of judges, not less than three, who constitute a panel.

Id.

17 *E.g.*, LAURAL HOOPER, DEAN MILETICH & ANGELIA LEVY, FED. JUDICIAL CTR., CASE MANAGEMENT PROCEDURES IN THE FEDERAL COURTS OF APPEALS 79 (2d ed. 2011) (Second Circuit).

18 *E.g.*, *id.* at 54 (D.C. Circuit).

19 *E.g.*, *id.* at 67 (First Circuit).

20 *See infra* notes 24–33.

A. Publicly Available Information on Panel Formation

The most recent Annual Report of the Ninth Circuit has a short section on an interesting occurrence: for the first time in Ninth Circuit history, the court had one oral argument panel comprised of all judges from Alaska.[21] The Report then notes that "[a]ppellate panels are randomly assigned" and that therefore this particular panel "resulted from the luck of the draw."[22]

Like the Ninth Circuit, several of the federal appellate courts have stated in various reports that they form their oral argument panels randomly.[23] For example, the most recent monograph on case management from the Federal Judicial Center notes that in the Fourth Circuit, "active judges are randomly assigned to argument panels by a computer program."[24] And the Federal Judicial Center reports that in the Eleventh Circuit, "[t]o ensure complete objectivity in assigning cases, the names of the active judges for the sessions of the court are drawn by lot for the entire court year."[25]

Another set of circuits state that they create their oral argument panels with the goal of approximately equalizing the number of times each judge sits with every other judge. Although the process of panel configuration is therefore not technically a random one, the results should still be consistent with random panel assignment.[26] The Federal Judicial Center states that in the Fifth Circuit, "[a] computer program is used to achieve random assignment of judges to panels," though it includes the caveat that the program also tries to avoid judges sitting together too often in any given court term.[27] The Sixth Circuit states in its Internal Operating Procedures that "[t]he court sits over two-week periods" and "[a]t least six active judges are assigned to one of the two sitting weeks at random"

---

21 JUDICIAL COUNCIL OF THE NINTH CIRCUIT, UNITED STATES COURTS FOR THE NINTH CIRCUIT 2013 ANNUAL REPORT 21 (2014).

22 *Id.*

23 At least one scholar concluded in 1990 that all of the circuit courts claimed to create their oral argument panels randomly. *See* Donald R. Songer, *Criteria for Publication of Opinions in the U.S. Courts of Appeals: Formal Rules Versus Empirical Reality*, 73 JUDICATURE 307, 310 (1990) ("According to circuit rules in each circuit, judges are assigned to panels on a purely random basis."). As this subpart details, we found that the majority of the circuits today state that they utilize either a random process or a process that would produce results consistent with a random process.

24 HOOPER, MILETICH & LEVY, *supra* note 17, at 99.

25 *Id.* at 208.

26 One can think of a series of coin tosses by analogy. Here, it is as if the coin is intentionally being turned to heads 50% of the time and tails 50% of the time as opposed to leaving the matter strictly to chance.

27 HOOPER, MILETICH & LEVY, *supra* note 17, at 115.

with "the balance of the court's active judges . . . assigned to the other sitting week" though goes on to note that "[j]udges are later assigned to panels during the sitting weeks using an automated routine which searches the court's database to determine which active judges have the longest intervals between sitting pairing."[28] Likewise, in the Seventh Circuit, the "[a]ssignment of judges to panels is random except that the circuit executive uses a computer-generated table to ensure that over a two-year period a judge sits approximately the same number of times with every other judge of the court."[29] Differing slightly from the recent Ninth Circuit Annual Report,[30] the Federal Judicial Center states that in the Ninth Circuit, "[t]he random assignment of judges by computer to particular days or weeks on the calendar is intended to equalize the workload among the judges . . . [and] to enable each active judge to sit with every other active and senior judge approximately the same number of times over a two-year period."[31] The Report also notes that a final goal in panel creation is to "assign active judges an equal number of times to each of the locations at which the court holds hearings."[32] And in a similar vein, the Tenth Circuit assignment of judges to argument panels is done "randomly using a software program developed by the court" whereby "[t]he program the court uses to assign judges to panels equalizes the number of times judges sit with one another over a period of one year."[33]

To be sure, not every circuit claims to form its panels randomly or with equalization in mind. For example, some state

---

28 U.S. COURT OF APPEALS FOR THE SIXTH CIRCUIT, SIXTH CIRCUIT INTERNAL OPERATING PROCEDURES, Internal Operating Procedure 34(a)(1) (2014).

29 HOOPER, MILETICH & LEVY, *supra* note 17, at 140. See also *the Seventh Circuit Practitioner's Handbook*, stating:

> Each judge is assigned to sit approximately the same number of times per term with each of his or her colleagues. The calendar of cases to be orally argued in a given week is prepared and circulated to the judges, and the judges advise the chief judge of any disqualifications. The disclosure statements filed pursuant to Circuit Rule 26.1 and Fed. R. App. P. 26.1 are intended to make this process more accurate and, therefore, more helpful. The judges are then randomly assigned by computer to sit in various panels. This separation of the processes of randomly assigning panels and scheduling cases avoids even the remote possibility of the deliberate assignment of an appeal to a particular panel.

U.S. COURT OF APPEALS FOR THE SEVENTH CIRCUIT, PRACTITIONER'S HANDBOOK FOR APPEALS 10 (2014).

30 *See supra* notes 21–22 and accompanying text.

31 HOOPER, MILETICH & LEVY, *supra* note 17, at 174.

32 *Id.*

33 *Id.* at 194.

that certain kinds of panels in the circuit are formed randomly—e.g., panels hearing death penalty cases[34]—but do not purport to randomly configure argument panels. Others are silent on the matter of panel configuration altogether. That said, the majority of the federal appellate courts do state that they form their argument panels randomly or with a goal of largely equalizing co-sittings.

Finally, it is worth noting that at least some of the judges themselves have furthered the claim of randomness. For example, Judge Richard Posner of the Seventh Circuit has stated in his scholarship that argument panels are randomly chosen in all of the circuit courts.[35] Judge Robert Parker of the Fifth Circuit has written that, "[e]xcept in unusual circumstances, an appeal in our court is decided by a panel of three judges" and that "[t]he panels are selected at random."[36] And then-Chief Judge Alex Kozinski of the Ninth Circuit wrote in 2012 that the practice of dissenting from, or concurring in, orders denying rehearings en banc can be beneficial to a judge who was "not assign[ed] . . . to the original three-judge panel" simply because of "the luck of the draw."[37]

In short, quite a few of the federal courts of appeals and several judges state that the process by which they form their argument panels is a random one or that they have deliberately formed panels in a way that should be consistent with a random process.

---

34 *See, e.g.*, U.S. COURT OF APPEALS FOR THE SECOND CIRCUIT, LOCAL RULES AND OPERATING PROCEDURES OF THE COURT OF APPEALS FOR THE SECOND CIRCUIT, Internal Operating Procedure 47.1(b) (2014) [hereinafter SECOND CIRCUIT, LOCAL RULES] ("The clerk assigns judges to death penalty case panels by random drawing from the death penalty case pool"—a pool that "consists of all active judges of the court and those senior judges who have filed with the clerk a statement of willingness to serve on death penalty case panels." The provision further states that "[i]f a judge is unable to serve, that judge's name returns to the pool after the drawing of a replacement. If a random drawing results in the selection of three senior judges, the clerk sets aside the third senior judge's name and continues drawing until the selection of an active judge's name, after which the clerk returns the third senior judge's name, and the names of any senior judges drawn thereafter, to the pool."); HOOPER, MILETICH & LEVY, *supra* note 17, at 87 (describing how, in the Third Circuit, "for each death penalty case, a special panel is constructed, and active judges are randomly assigned to the panel").

35 *See, e.g.*, Richard A. Posner, *A Heartfelt, Albeit Largely Statistical Salute to Judge Richard D. Cudahy*, 29 YALE J. ON REG. 355, 357 (2012) ("[T]he panels that hear cases are randomly selected from the court's judges."); *see also* Epstein, Landes & Posner, *supra* note 1, at 110 ("We assume that members of a panel are chosen randomly from the judges of the court, which is the practice in all circuits." (citation omitted)).

36 Robert M. Parker, *Foreword*, 26 TEX. TECH. L. REV. 265, 266 (1985).

37 Alex Kozinski & James Burnham, *I Say Dissental, You Say Concurral*, 121 YALE L.J. ONLINE 601, 607 (2012).

B. General Courts Literature

The notion that oral argument panels are randomly drawn has been picked up in the academic literature. Numerous articles on the federal courts have noted it.[38] For some of these scholarly works, the randomness assumption serves as a backdrop; it helps to set the context for a particular argument. For others, however, the randomness assumption is more significant, serving, at least in part, as a basis for normative conclusions.

With respect to the first kind of scholarship, quite a few articles include the claim that the federal appellate courts have randomly configured argument panels. For this set of articles, the "fact" that the judges are randomly selected is relevant in some way to the article's analysis. For example, a recent article about the deference courts give to agency interpretation of statutes discusses the random draw of three-judge panels.[39] One of the article's claims is that judges have different views about the degree of deference to afford a particular agency interpretation; because the three judges who hear a case are drawn at random, the article states, the outcome of each case is akin to a lottery.[40] Similarly, an article on the meaning of dicta and the role of stare decisis states that "[i]n the Courts of Appeals, panels are the product of random draws of three among a larger set of members of the court."[41] The article goes on to argue that the randomness of panel composition creates doctrinal instability in the courts of appeals when compared to the Supreme Court by increasing the chance that a panel that does not represent the views of the court "as a whole" will decide a particular matter.[42] Finally, an article on the decentralized nature of federal judicial power assumes random assignment and finds it to be a useful feature of the system in that it helps to avoid various problems associated with group decision making, such as group polarization and herding.[43] In articles such

---

38 For just a few examples, see Joshua A. Douglas, *The Procedure of Election Law in Federal Courts*, 2011 UTAH L. REV. 433, 452 (describing how judges are "randomly selected" to hear cases on review); Samuel P. Jordan, *Early Panel Announcement, Settlement, and Adjudication*, 2007 BYU L. REV. 55, 66–67 (noting "the current practice of randomly assigning judges to appellate panels").

39 *See* Jud Mathews, *Deference Lotteries*, 91 TEX. L. REV. 1349, 1373 (2013).

40 *Id.* at 1372–76.

41 Abramowicz & Stearns, *supra* note 1, at 1009.

42 *Id.* at 1009–10.

43 Ronald J. Krotoszynski, Jr., *The Unitary Executive and the Plural Judiciary: On the Potential Virtues of Decentralized Judicial Power*, 89 NOTRE DAME L. REV. 1021, 1073 (2014) ("Random assignment of judges to panels means that subgroups of the entire court are constituted to hear and decide particular cases.

as these, the assumption of random panels helps to create a context for the underlying analysis.

There is a second set of court scholarship, however, that relies more directly on the claim that panels are randomly created. This scholarship assumes that argument panels are randomly configured and argues that the current system should be changed. Noted federal courts scholar Daniel Meador proposed creating appellate panels designed to specialize in certain subjects—a "method of appellate organization" which would "eliminate" what he took to be the status quo: "random assignments of judges and cases."[44] In a similar vein, Emerson Tiller and Frank Cross noted that random panel assignment has become a "hallmark" of our current federal system,[45] but then advocated abandoning it in favor of a selection process whereby no more than two judges of the same political party could be put on the same panel.[46] Finally, Michael Hasday wrote that "[f]ederal appellate courts employ a random assignment system to select the circuit judges who will serve on any particular three-judge panel" and then argued for allowing the parties to create oral argument panels instead.[47] In short, an important conclusion of these articles is that the assignment of judges to panels should not be random, after it is stated that it is.

In sum, there is a widespread belief in the general courts literature that the oral argument panels of the federal courts are randomly configured. For some scholarship, this "fact" is noted in passing. Elsewhere, however, the randomness assumption is important—it sets the context for, and helps to support, a given argument. For yet other articles, the randomness assumption plays a direct role, as it is the taken status quo that the authors argue against. We turn now to a final set of scholarship in which the randomness assumption plays a crucial role: the empirical literature on courts and judicial decision making.

Different judges will bring different predilections (biases) to the table, but the size of most of the U.S. courts of appeals would seem helpful in avoiding problems associated with choice shift, group polarization, and herding." (footnote omitted)).

44 Daniel J. Meador, *An Appellate Court Dilemma and a Solution Through Subject Matter Organization*, 16 U. MICH. J.L. REFORM 471, 475 (1983).

45 Tiller & Cross, *supra* note 2, at 216 & n.4.

46 *Id.* at 226–32.

47 Michael Hasday, *Ending the Reign of Slot Machine Justice*, 57 N.Y.U. ANN. SURV. AM. L. 291, 291, 298–99 (2000).

C. Empirical Literature

Scholars have long recognized the importance of randomization for causal inference in research generally, as well as in empirical legal scholarship specifically.[48] The reason is simple: when a variable of interest is not randomly assigned, it might be correlated with the outcome scholars are interested in studying.[49] For example, if students self-select into LSAT prep courses, simply showing that students who took the course score better on the LSAT does not demonstrate that the course is effective. This is because students who select to take the class might be more conscientious students than those who do not. In other words, the treatment (taking the course) is systematically related to the outcome variable (LSAT scores). If a large number of students were randomly assigned to take an LSAT prep-course, however, any difference in test results between those students who took the course and those who did not could be attributed to the course.

Given the importance of randomization to causal inference, empirical researchers often rely heavily on random processes to design their studies. To do so, scholars may either conduct experiments where they are able to guarantee the random assignment of treatments,[50] or they find examples where random processes were already used even without intervention from the researcher.[51] One random process that is frequently relied on as the basis for empirical research is the random assignment of judges on the federal courts of appeals to argument panels.[52]

---

48 *See* Daniel E. Ho & Donald B. Rubin, *Credible Causal Inference for Empirical Legal Studies*, 7 ANN. REV. L. SOC. SCI. 17, 18 (2011) (citing R.A. FISHER, STATISTICAL METHODS FOR RESEARCH WORKERS (1925); R.A. FISHER, THE DESIGN OF EXPERIMENTS (1935)).

49 *See, e.g.*, Lee Epstein & Gary King, *The Rules of Inference*, 69 U. CHI. L. REV. 1, 110 (2002) ("[R]andom selection is the only selection mechanism in large-*n* studies that automatically guarantees the absence of selection bias. That is because when we use random sampling we are, by definition, assuring the absence of any association that may exist between selection rules and the variables in our study." (emphasis omitted)).

50 For a discussion of experimental research by legal scholars, see Adam Chilton & Dustin Tingley, *Why the Study of International Law Needs Experiments*, 52 COLUM. J. TRANSNAT'L L. 173, 187–90 (2013).

51 For a discussion of these so-called "natural experiments," see Gregory Robinson, John E. McNulty & Jonathan S. Krasno, *Observing the Counterfactual? The Search for Political Experiments in Nature*, 17 POL. ANALYSIS 341 (2009); Jasjeet S. Sekhon & Rocío Titiunik, *When Natural Experiments Are Neither Natural nor Experiments*, 106 AM. POL. SCI. REV. 35 (2012).

52 Claims that panels are randomly assigned are ubiquitous in empirical research on judicial behavior. *See, e.g.*, Andreas Broscheid, *Comparing Circuits: Are Some U.S. Courts of Appeals More Liberal or Conservative Than Others?*, 45 L.

The reason that the randomness assumption is helpful is that it allows researchers to assume that the judges on panels do not correlate with any salient features of the judges or the cases. For example, researchers have relied on the random assignment of judges to test whether judges vote differently when they are assigned to panels with judges who were appointed by a president from the same (or different) political party.[53] Similarly, scholars have studied whether the presence of a female judge on a panel alters the decisions of judges,[54] or whether the presence of a judge that is a racial minority alters

---

& SOC'Y REV. 171, 179 (2011) ("In general, in the U.S. Courts of Appeals three-judge panels are formed by random selection from the pool of sitting circuit judges, and the cases they hear are randomly assigned to panels."); Edward K. Cheng, *The Myth of the Generalist Judge*, 61 STAN. L. REV. 519, 523 (2008) ("[W]ell-established rules and norms within the courts of general jurisdiction require the random assignment of cases to ensure that judges see all case types." (footnote omitted)); Epstein, Landes & Posner, *supra* note 1, at 110 ("We assume that members of a panel are chosen randomly from the judges of the court, which is the practice in all circuits." (citation omitted)); Todd C. Peppers, Katherine Vigilante & Christopher Zorn, *Random Chance or Loaded Dice: The Politics of Judicial Designation*, 10 U. N.H. L. REV. 69, 74 (2012) ("It is especially important to note that each federal circuit follows strict procedures designed to guarantee that appellate judges are randomly selected to each three-judge panel, and that cases are randomly assigned to those panels." (footnote omitted)); Sunstein & Miles, *supra* note 1, at 2197 (noting that within the federal courts of appeals, "judges are randomly assigned to three-judge panels"); Cass R. Sunstein, David Schkade & Lisa Michelle Ellman, *Ideological Voting on Federal Courts of Appeals: A Preliminary Investigation*, 90 VA. L. REV. 301, 303 (2004) (noting, after posing a series of questions related to how ideology affects voting behavior, that "[s]ince judges in a given circuit are assigned to panels (and, therefore, to cases) randomly, the existence of a large data set allows these issues to be investigated empirically").

53 There have been a large number of these papers in this "panel effects" literature. *See, e.g.*, Sunstein & Miles, *supra* note 1; Sunstein, Schkade & Ellman, *supra* note 52.

54 *See* Christina L. Boyd, Lee Epstein & Andrew D. Martin, *Untangling the Causal Effects of Sex on Judging*, 54 AM. J. POL. SCI. 389, 390 (2010) ("[I]n cases implicating sex discrimination . . . we observe sex-based effects: the probability of a judge deciding in favor of the party alleging discrimination decreases . . . when the judge is a male. Likewise, when a woman serves on a panel with men, the men are significantly more likely to rule in favor of the rights litigant."); Sean Farhang & Gregory Wawro, *Institutional Dynamics on the U.S. Court of Appeals: Minority Representation Under Panel Decision Making*, 20 J. L. ECON. & ORG. 299, 299 (2004) ("We find that the norm of unanimity on panels grants women influence over outcomes even when they are outnumbered on a panel."); Donald R. Songer, Sue Davis & Susan Haire, *A Reappraisal of Diversification in the Federal Courts: Gender Effects in the Courts of Appeals*, 56 J. POL. 425, 425 (1994) ("[I]n employment discrimination cases, female judges were significantly more liberal than their male colleagues."); Jennifer L. Peresie, Note, *Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Appellate Courts*, 114 YALE L.J. 1759, 1776 (2005) ("[I]n Title VII sexual harassment and sex discrimination cases, . . . a judge's gender and the gender composition of the panel mattered to a judge's decision.").

the decisions of judges.[55] Additionally, scholars have leveraged the assumption that judges are randomly assigned to study aspects of judicial behavior other than voting on case outcomes, like which judges will write the majority opinion,[56] whether judges will decide to file a dissenting opinion,[57] and even the choice of words used in opinions.[58] In short, the random assignment of judges to panels on the courts of appeals is both frequently assumed and a key feature of empirical research strategies studying judicial behavior.

Although the randomness assumption has been relied upon by empirical researchers in numerous articles, it is important to note that a few scholars have questioned "random assignment" or "judicial assignment" more generally—that is, the overall process by which judges are assigned to panels and those panels are then assigned to cases. For example, in a 2000 article on the assignment of judges to cases in the federal appellate courts, J. Robert Brown, Jr. and Allison Herren Lee describe how there are various ways in which the "randomness" of case assignment is "erode[d]," including how judges are able to change panels even after cases have been assigned.[59] Although these comments are meant to show how the ultimate case assignment is not perfectly random, they also call into question the randomness of panel configuration. In

---

55 *See* Adam B. Cox & Thomas J. Miles, *Judging the Voting Rights Act*, 108 COLUM. L. REV. 1, 1 (2008) ("[T]he so-called 'panel effects' of race are strong, as white judges become substantially more likely to vote in favor of liability when they sit with minority judges."); Adam B. Cox & Thomas J. Miles, *Judicial Ideology and the Transformation of Voting Rights Jurisprudence*, 75 U. CHI. L. REV. 1493, 1536 (2008) ("White judges who sat on panels with at least one African-American judge were considerably more likely to vote in favor of liability, and this effect was evident for both Democratic and Republican appointees."); Jonathan P. Kastellec, *Racial Diversity and Judicial Influence on Appellate Courts*, 57 AM. J. POL. SCI. 167, 167 (2012) ("Randomly assigning a black counterjudge—a black judge sitting with two nonblack judges—to a three-judge panel of the Courts of Appeals nearly ensures that the panel will vote in favor of an affirmative action program.").

56 *See, e.g.*, Sean Farhang, Jonathan P. Kastellec & Gregory J. Wawro, *The Politics of Opinion Assignment and Authorship on the US Court of Appeals: Evidence from Sexual Harassment Cases*, 44 J. LEGAL STUD. S59, S59 (2015) ("[We] find that female and more liberal judges are substantially more likely to write opinions in sexual harassment cases.").

57 *See* Epstein, Landes & Posner, *supra* note 1, at 110.

58 *See* Michael Z. Gill & Andrew B. Hall, How Judicial Identity Changes the Text of Legal Rulings 14 (June 19, 2015) (unpublished manuscript) ("[T]he random assignment of a female or non-white judge systematically causes an overall change in the vocabulary used in published rulings.").

59 *See* Brown & Lee, *supra* note 4, at 1041–42; *see also* Kastellec, *supra* note 55, at 175 n.14 ("The actual procedures employed for panel assignment vary across circuits and allow for some discretion in panel selection, which mitigates against truly random selection.").

describing the general assignment process, other scholars have similarly noted that judges may subsequently trade panel placements with each other or that a judge may be placed on a panel purposely to rehear a case if she was on the original panel that decided the case.[60] Again, while these articles do not directly question the randomness of oral argument panel configuration, their statements indirectly question the assumption.

The scholarly work that comes closest to directly questioning the randomness of panel creation is a 2009 article by political scientist Matthew Hall on the Supreme Court's review of appellate court decisions.[61] Hall notes that while conducting his research a few years prior, he called each federal court of appeals' Clerk's Office to ask whether judicial assignments—meaning judges to panels and then panels to cases—were random.[62] Hall reports that several clerks informed him that assignments were not in fact random in their circuits, and as a consequence he excluded those circuits from his study.[63] It is not clear just what was confirmed to be nonrandom—the panel configuration process, the assignment of panels to cases, or both. In any event, Hall's claims about random assignment have been noted by at least some scholars.[64] That said, we are unaware of any studies that have followed Hall's lead and tried to verify whether panel configuration is actually random. Moreover, Hall's article has actually been cited by prominent scholars for the proposition that "members of a panel are chosen randomly from the judges of the court, which is the practice in all circuits."[65]

In short, scores of quantitative articles have not only assumed the random configuration of panels, but have relied

---

60 *See, e.g.*, Kastellec, *supra* note 55, at 175 n.14 ("[J]udges can trade places on panels in some circuits, and the original judges in a case that requires additional hearings may be selected for such subsequent hearings.").

61 *See* Matthew Hall, *Experimental Justice: Random Judicial Assignment and the Partisan Process of Supreme Court Review*, 37 AM. POL. RES. 195 (2009).

62 *Id.* at 202–03.

63 *Id.*

64 *See, e.g.*, Ryan C. Black & Ryan J. Owens, *Bargaining and Legal Development in the United States Courts of Appeals*, 41 AM. POL. RES. 1071, 1096 n.3 ("Not all circuits randomly assign judges to panels, though most today do." (citing Hall, *supra* note 61)). Interestingly, Black and Owens rely on Hall's research to verify that the D.C. Circuit—the focus of their study—used random assignment. *Id.* Hall bases his claim that the D.C. Circuit uses random assignment on what the Clerk's Office in the D.C. Circuit reported but did not make any independent effort to verify the claim. *See* Hall, *supra* note 61, at 203. Our research, however, found evidence of nonrandomness in the D.C. Circuit. *See infra* Subpart III.A.

65 Epstein, Landes & Posner, *supra* note 1, at 110 (citing Hall, *supra* note 61).

upon this assumption directly in their research design. Although a few researchers have questioned the randomness of how judges are ultimately assigned to cases, we believe it is fair to say that random assignment of judges to panels remains the dominant view in the literature.

II
CHALLENGING THE ASSUMPTION OF RANDOM PANELS

Although the assumption that panels are randomly generated in the courts of appeals is both widely held and critical to scholarship on federal courts and judicial behavior, there are good reasons to believe that it might not be strictly accurate. In conducting qualitative research on the practices of the federal courts of appeals, one of the coauthors learned of several important reasons why a court might want, if not need, to create panels in a non-strictly random fashion.[66] For example, in a circuit that does not have specific court weeks set aside for oral arguments, judges may have scheduling conflicts that need to be accommodated or it be may helpful to ensure that judges have at least some number of weeks between sittings so as not to become overloaded. Taking such considerations into account would require departures from strict randomness, though they are the kinds of departures most would understand if not expect courts to make. Moreover, even if strict randomness were the goal, this same coauthor was informed that there is no standardized process for creating random panels and that some circuits put together the panels by human hand.[67] As discussed more below, research has shown that individuals, as much as they might try, have a difficult time creating results that are consistent with a random process. Accordingly, there is good reason to think that at least some of the calendars will not be random—either because various factors, such as scheduling needs, were taken into account or because the relevant human actors were unable to perfectly create random panels.

Despite the fact that there are plausible reasons to doubt it, the assumption that panels are randomly created has not yet been subject to rigorous empirical scrutiny. One reason, as intimated above, is that many scholars and researchers seem to be unaware of the extent to which there is discretion in calendar creation and the various needs that a court may face;

---

66 *See* Levy, *supra* note 13.
67 *See id.*

as such, scholars have not sufficiently appreciated the reasons to doubt randomness. The other reason is methodological—testing the randomness of panel assignments is quite challenging. First, in order to test random assignment, one must develop a specific testable hypothesis to evaluate a way in which the composition of panels would deviate from randomness. Second, testing whether panels were randomly created requires having data on the panels that actually sat. Although case level data is publicly available in a number of commercial and government databases, the argument calendars for the federal appellate courts are not centrally collected or easily available. Third, of course any individual composition of panels is highly unlikely, and so determining whether panel assignments are random requires determining some baseline distribution of what panels would be expected through a random process and testing the observed panels against this baseline.

In this Part, we first outline several reasons to doubt the assumption of random panel assignment. Second, we develop a hypothesis to test whether the composition of panels might be nonrandom. Third, we describe the original dataset that we have constructed for this project. Fourth, we explain our methodology for empirically testing random assignment of circuit court panels.

A. Motivation

While conducting qualitative research on the federal courts of appeals, one of the coauthors was informed that different circuits have different ways of creating panels.[68] As one example, a sitting circuit court judge noted that in his circuit the calendar had to take into account practical concerns, such as the availability of judges, and that accordingly the panels were not configured by computer but by hand.[69] These comments led to a significant qualitative project, which is still ongoing.[70] But the initial findings from five circuit courts substantiate this judge's statements that there are good reasons to think that the circuits take into account considerations about the schedules of various judges or other circuit interests that would mean creating a calendar in a non-strictly random fashion. Grown out of an effort to balance these various factors, at least several circuits rely on individuals, instead of a computer pro-

68 *Id.* at 17–30.
69 *Id.* at 22.
70 *See generally id.*

gram, to create their oral argument panels.[71] We briefly consider these points here and how they create reasons to doubt the widely held assumption of random panels.

To begin, the initial findings of the coauthor's qualitative study suggest that at least some of the courts of appeals might take into account a variety of factors when creating the oral argument panel.[72] These factors could be specific to one judge—such as Judge A will be away from court in a particular month and so should not be scheduled for that month.[73] These factors could also be more general to the court[74]—such as the circuit believing it important to ensure that each judge have at least a few weeks off between sittings, if possible, and so Judge B will not be scheduled to sit for the second week in October if she has already been scheduled to sit for the first. It is easy to see how a court would take such considerations into account when creating its panels, and therefore there is good reason to think that at least some of the courts would not have a strictly random calendar.

Additionally, at least several of the circuit courts rely on people, and not computers, to create the oral argument panels.[75] As a result, even if the circuits wanted to achieve strictly random panels, it is highly unlikely that they would be able to do so. This is because, as a large body of academic research has shown, "humans have difficulties reproducing . . . random patterns, even when they have incentives to do so."[76] A great deal of this literature has focused on conducting experiments in laboratories showing that human subjects asked to create patterns of random numbers have quite a few nonrandom tendencies—like using certain numbers at higher rates than others, favoring small numbers over large numbers, or not repeating the same number consecutively.[77] In fact,

---

71 *Id.* at 21–30.

72 *See id.*

73 *See, e.g.*, *id.* at 21–23.

74 *See, e.g.*, *id.* at 24–30 (listing, based on interviews with federal circuit judges, a range of possible rationales for creating calendars in a non-strictly random fashion).

75 *Id.* at 22.

76 Bernd Beber & Alexandra Scacco, *What the Numbers Say: A Digit-Based Test for Election Fraud*, 20 POL. ANALYSIS 211, 218–20 (2012) (providing an excellent discussion of research on errors made by humans attempting to replicate random processes).

77 *See* Philip J. Boland & Kevin Hutchinson, *Student Selection of Random Digits*, 49 STATISTICIAN 519, 527 (2000) ("Our results support the claim that humans are unable to generate random sequences even when explicitly asked to do so."); Alphonse Chapanis, *Random-Number Guessing Behavior*, 8 AM. PSYCHOLOGIST 332, 332 (1953) ("With but one exception, [the human subjects] exhibited

even in experiments where subjects are paid more if they are able to effectively randomize, the human subjects are still not able to recreate random processes.[78]

These findings, moreover, are not limited to laboratories. There is also considerable evidence that humans are unable to create random results when they are trying hard to do so and have strong incentives to produce random-looking outcomes. For example, researchers have been caught fabricating data because the data they produced was highly unlikely to have occurred randomly.[79] Additionally, humans' inability to recreate random patterns has been used as a method of detecting election fraud because humans are neither capable of recreating random vote counts[80] nor counts of voter turnout.[81] Given this consistent evidence that humans are incapable of recreating random processes, it seems possible—if not highly probable—that the presence of human discretion over how panels are created in the courts of appeals would mean that the panels would not be random, regardless of how much the courts might want them to be.

Taken together, the initial results of the qualitative research provide reasons to question the randomness assump-

---

marked preferences for certain digits . . . ."); Gustave J. Rath, *Randomization by Humans*, 79 AM. J. PSYCHOL. 97, 102 (1996) ("[Human subjects] are poor randomizers, as has been found in other studies."); *see also* Beber & Scacco, *supra* note 76, at 218–20.

78 Beber & Scacco, *supra* note 76, at 219; *see* David V. Budescu, *A Markov Model for Generation of Random Binary Sequences*, 13 J. EXPERIMENTAL PSYCH. 25, 37–38 (1987); James E. Mosimann, Claire V. Wiseman & Ruth E. Edelman, *Data Fabrication: Can People Generate Random Digits?*, 4 ACCOUNTABILITY RES. 31, 42–44 (1995); Amnon Rapoport & David V. Budescu, *Randomization in Individual Choice Behavior*, 104 PSYCHOL. REV. 603, 612, 614 (1997).

79 *See, e.g.*, Sanaa Al-Marzouki, Stephen Evans, Tom Marshall & Ian Roberts, *Are These Data Real? Statistical Methods for the Detection of Data Fabrication in Clinical Trials*, 331 BRIT. MED. J. 267, 270 (2005) (concluding data in a diet trial that supposedly relied upon randomized intervention "were either fabricated or falsified"); James E. Mosimann, John E. Dahlberg, Nancy M. Davidian & John W. Krueger, *Terminal Digits and the Examination of Questioned Data*, 9 ACCOUNTABILITY RES. 75, 79, 82–83, 87, 92 (2002) (identifying four cases of falsified data that others had portrayed as random).

80 Beber & Scacco, *supra* note 76, at 219; Walter R. Mebane, Jr., *Election Forensics: The Second-Digit Benford's Law Test and Recent American Presidential Elections*, *in* ELECTION FRAUD: DETECTING AND DETERRING ELECTORAL MANIPULATION 162, 165–68 (R. Michael Alvarez, Thad E. Hall & Susan D. Hyde eds., 2008) (developing a test for voter fraud with "counts generated using some simple random process").

81 Mikhail Myagkov, Peter C. Ordeshook & Dimitry Shakin, *Fraud or Fairytales: Russia and Ukraine's Electoral Experience*, 21 POST-SOVIET AFF. 91, 126–30 (2005); Misha Myagkov, Peter C. Ordeshook & Dmitry Shakin, *The Disappearance of Fraud: The Forensics of Ukraine's 2006 Parliamentary Elections*, 23 POST-SOVIET AFF. 218, 226–27 (2007).

tion. There are numerous factors a court might understandably want to take into account when creating panels that would cause departures from randomness. Moreover, even if that were not the case, simply because of human involvement in at least some of the courts, it seems unlikely that all of the panels in the circuit courts would be strictly random. As such, we concluded that it was important to quantitatively test the randomness of the panels of the courts of appeals.

B. Hypothesis

Before testing whether panel assignments are random, there is a threshold question of what significant characteristic should, in theory, be randomly distributed if the assignment process is a random one. For example, in experimental research, it is common practice to make sure that treatments are randomly assigned across subjects in the sample being studied.[82] This is done by testing whether subjects in the control group have the same relevant characteristics as subjects in the treatment groups. This may include checking to make sure that groups have the same ratio of men to women, the same education levels, or the same racial breakdown.[83] Although the specific characteristics evaluated may change based on the subject being studied, the key is that relevant variables must be identified to evaluate random assignment in an experimental setting.

In testing random assignment in the federal appellate courts, there were several criteria we kept in mind in selecting the relevant judicial characteristic. First, we wanted to select a characteristic that was easily observable. It may be that testing the randomness of panel assignment based on the judges' internal preferences—say, about supporting the business community—would be interesting, but it would be impossible without access to this kind of information. Second, it was important to select a characteristic that must be objectively defined. For example, if we were to test the distribution of "new" judges, we could decide to define judges as being "new" in their

---

82 *See* Ben B. Hansen & Jake Bowers, *Covariate Balance in Simple, Stratified and Clustered Comparative Studies*, 23 STAT. SCI. 219, 219 (2008) ("In a controlled, randomized experiment, treatment and control groups should be roughly the same—balanced—in their distribution of pretreatment variables.").

83 *See, e.g.*, Stephen Chaudoin, *Promises or Policies? An Experimental Analysis of International Agreements and Audience Reactions*, 68 INT'L ORG. 235, 242 (2014) (checking to make sure that the treatment and control groups are balanced based on the respondents age, sex, race, marital status, and education level).

first year or in their first five years on the bench. This injects a subjective decision into the process that, depending on how it was made, could influence the results. Third, because it is difficult to evaluate whether a characteristic with little variance is randomly distributed, the characteristic must have a great deal of variance. For example, one could test the distribution of judges who have last names ending in "Z," but as this would be such a small number of judges,[84] assessing whether the panels appeared random along this dimension would be quite challenging. Finally, we wanted to select a characteristic that is relevant to the academic literature. For example, one could test the distribution of judges that wear eyeglasses—a characteristic that satisfies all of the other criteria listed above—and yet, the results of such a study would be of limited value. That is, if it turned out that panels appeared nonrandom based on this measure, the finding of nonrandomness would be important but it is not clear what more could be said; a more valuable study would then have a characteristic that is theoretically important.

A characteristic that meets all four criteria—and the one we elected to test—is the ideology of judges assigned to panels.[85] More specifically, we elected to test whether the distribution of judges based on the political party of the president that appointed them was random. The party of the president that appointed a judge is easily observable, objectively defined, has a great deal of variance, and is theoretically important.

---

84 Out of the over 775 judges in our study, only six have last names beginning with the letter "Z": Judges James Block Zagel, Lawrence Paul Zatkoff, Thomas Samuel Zilly, Jennifer Guerin Zipps, Rya Weickert Zobel, and Jack Zouhary.

85 An earlier draft of this project also tested the gender balance of panels and the balance of minority judges on panels. We no longer test the balance of these characteristics because there are relatively few women and minorities serving as federal judges. *See* Barry J. McMillion, *U.S. Circuit and District Court Judges: Profile of Select Characteristics*, CONG. RES. SERV., 12 (Mar. 19, 2014), https://www.fas.org/sgp/crs/misc/R43426.pdf [http://perma.cc/J9TN-4VKV] ("[W]omen comprise 32.1% of active U.S. circuit court judges and 10.5% of senior status circuit judges . . . ."); *id.* at 14 ("[C]ircuit court judges who are white account for 76.5% of all active judges and 90.3% of senior judges."); *id.* at 21 ("31.8% of active U.S. district court judges are women . . . . 12.8% of senior district court judges currently serving are women."); *id.* at 22 ("74.4% of active district court judges are white . . . . Of senior district court judges, 89.5% are white . . . ."). This poses a problem because our method for testing for nonrandomness—using a chi-squared test—requires that all of the "bins" have counts higher than five. *See* ROBERT M. LAWLESS, JENNIFER K. ROBBENNOLT & THOMAS S. ULEN, EMPIRICAL METHODS IN LAW 258–260 (2010). Many of the circuits had fewer than five total panels with three women or three minorities on the panel, thus making it inappropriate to use our method to test whether panel assignments are random along these dimensions.

On the last point in particular, as noted already, there is a large body of empirical research demonstrating that the ideology of judges—most often defined by the party of the appointing president—has an influence on the decisions of judges and panels.[86] Much of this research has relied on the assumption that the ideological composition of panels is random.

For these reasons, the hypothesis we test is whether the distribution of judges appointed by Republican presidents is consistent with the distribution produced by a strictly random process.

C. Data

In order to test whether the ideological composition of panels is random, we created a dataset based on the oral argument panels of all of the twelve regional circuit courts[87] during the five-year span between September 1, 2008, and August 31, 2013. Over the course of a year, we were fortunate enough to

---

86 *See, e.g.*, CASS R. SUNSTEIN ET AL., ARE JUDGES POLITICAL? AN EMPIRICAL ANALYSIS OF THE FEDERAL JUDICIARY (2006); Deborah Beim & Jonathan P. Kastellec, *The Interplay of Ideological Diversity, Dissents, and Discretionary Review in the Judicial Hierarchy: Evidence from Death Penalty Cases*, 76 J. POL. 1074, 1074, 1081–82 (2014) ("[T]he existence of ideological diversity on a panel—and the potential for dissent—plays a significant role in judicial decision making."); Jonathan P. Kastellec, *Hierarchical and Collegial Politics on the U.S. Courts of Appeals*, 73 J. POL. 345, 356–59 (2011) (finding that having one judge from a political party on a panel has the greatest effect when that judge is aligned with the political ideology of the Supreme Court); Jonathan P. Kastellec, *Panel Composition and Voting on the U.S. Courts of Appeals Over Time*, 62 POL. RES. Q. 377, 381–85 (2011) (studying the ideological compositions of panels over time and finding that the relationship between panel composition and judicial behavior is a relatively recent phenomenon); Thomas J. Miles & Cass R. Sunstein, *Do Judges Make Regulatory Policy? An Empirical Investigation of* Chevron, 73 U. CHI. L. REV. 823, 823, 827 (2007) ("Republican appointees demonstrated a greater willingness to invalidate liberal agency decisions and those of Democratic administrations. These differences are greatly amplified when Republican appointees sit with two Republican appointees and when Democratic appointees sit with two Democratic appointees."); Thomas J. Miles & Cass R. Sunstein, *The Real World of Arbitrariness Review*, 75 U. CHI. L. REV. 761, 767, 785 (2008) (finding "panel effects are substantial" when examining the invalidation of agency rulemaking: "Democratic appointees typically show increasingly liberal voting patterns as the number of Democratic appointees [on a panel] increases, and Republican appointees typically show increasingly conservative voting patterns as the number of Republican appointees increases."); Sunstein, Schkade & Ellman, *supra* note 52, at 304–05 ("A judge's ideological tendency . . . is likely to be amplified if she is sitting with two judges from the same political party."); Tiller & Cross, *supra* note 2, at 226–32 (examining "the frequency of partisan dominance . . . on federal circuit court panels and the benefits of requiring split partisan assignments").

87 We define the "regional circuit courts" as consisting of the First through Eleventh Circuits and the District of Columbia Circuit Court of Appeals.

be given the calendar information by each individual court.[88] To our knowledge, no dataset like this has ever been amassed.

Regarding the substance of the data, for the most part what we received were calendar pages from each court.[89] A typical page would list the sitting date, the location of the sitting, the three judges who heard the case, and sometimes the case name.[90] Less typically, we received a page that listed all of the panels as they existed for a week, without any case information.[91] Although most of the pages were fairly straightforward to interpret, we corresponded with nearly every Clerk's Office to be sure that our understanding of the relevant dates and panels was accurate.

Regarding the form of the data, we typically received a series of PDF documents of the calendar pages (though on one occasion we received a hard copy of the circuit's calendar pages). In total, the calendar pages for all twelve courts for this five-year time period translated into several thousand pages of information. In order to extract the data accurately, code was written to parse the individual calendar pages and pull the relevant information (such as the names of the judges and the sitting date).[92] In two of the circuit courts—the Third and the Fifth—code extraction proved impossible, due to the formatting of the calendar pages, and so hand-coding and entry was used.[93] From these sources, we were able to create our dataset, which contains approximately ten thousand panels across all circuits in this time frame. One coauthor then performed a verification process on the data, spot-checking between 10% and 20% of all panels in the code-extracted

---

88 Some circuits provide past calendar information online, but this information is typically for a short time frame (say, the past six months, year, or two years at most). *See, e.g.*, U.S. COURT OF APPEALS FOR THE SECOND CIRCUIT, *Court Calendar*, http://ww2.ca2.uscourts.gov/calendar/ [http://perma.cc/U9QH-DSV2] (providing court calendars only as far in the past as August 2013). The only exception to this general rule at the time of this Article's writing is the Ninth Circuit, which provides several years worth of calendar information on its court website. *See* U.S. COURTS FOR THE NINTH CIRCUIT, *Oral Argument Dates & Locations*, http://www.ca9.uscourts.gov/calendar/ [http://perma.cc/U9QH-DSV2]. But in order to obtain the information for our full time frame, we were in touch with the Clerk's Office directly for each of the circuit courts.

89 These calendar pages are on file with the authors.

90 See, for example, a calendar page of the D.C. Circuit, the First Circuit, or the Second Circuit Court of Appeals. These calendar pages are on file with the authors.

91 See, for example, the calendar pages of the Fifth Circuit Court of Appeals. These calendar pages are on file with the authors.

92 We owe a debt of gratitude to Sean Chen for this work.

93 With many thanks to Jacob Adrian for this work.

circuits,[94] and 100% of all panels in the hand-extracted circuits. Due to these processes, we are confident that we accurately captured the panels as we received them during this time period.

A few qualifications about the data should be noted. First, the data we obtained speaks only to oral argument panels. As such, it does not speak to cases that were decided by nonargument panels, which have different configuration rules.[95] As noted earlier, the majority of cases that are decided on the merits are not decided with oral argument.[96] That said, the literatures—both the general courts literature and the empirical literature—have focused upon the cases decided by oral argument, presumably because such cases are generally understood to be the cases that the courts deem to be important, particularly complex, or novel.[97] Moreover, the courts of appeals have focused on these panels as well (the courts say virtually nothing in their internal operating procedures or local rules about how their nonargument panels are formed).[98] Given that the focus of the academic literatures and the courts themselves has been on the oral argument panels, and not simply panels generally, we think it is reasonable to focus on such panels.

Second, we relied upon the data that we received from each circuit and the circuits occasionally gave us different kinds of information. For example, the Eleventh Circuit provided the general argument panel data exclusive of capital cases, whereas other circuits, such as the Ninth Circuit, did include capital cases in their data. Some of these inconsistencies we were able to cure by reaching out to the relevant Clerk's Office and asking for additional data, but this was not possible in

---

94 If anything, the ten-to-twenty percent verification figure underestimates the spot-checking that we performed. One of the coauthors went through the data by hand, counting out every ten panels and checking that panel against the panel that had been extracted. As such, the process not only directly checked every tenth panel but also indirectly checked the panels in-between (as, if there was an extra panel in the original data, it was then caught). Accordingly, it would not be inaccurate to say that our verification process indirectly checked all of the panels in our time frame.

95 *See* Levy, *supra* note 13.

96 *See supra* note 15.

97 *See* JUDITH A. MCKENNA ET AL., FED. JUDICIAL CTR., CASE MANAGEMENT PROCEDURES IN THE FEDERAL COURTS OF APPEALS 10 (2000) (noting that certain case characteristics are "likely to trigger oral argument," including "novel issues, complex issues, extensive records, and numerous parties").

98 *See, e.g.*, SECOND CIRCUIT, LOCAL RULES, supra note 34, at Local Rule 34.2 (listing the local rules governing the Non-Argument Calendar without identifying the procedure for how nonargument panels are formed).

every instance. The remaining kinds of discrepancies between circuits are minor and we believe only amount to a small number of cases.

Third, based on limitations with our data, we were not able to consistently identify what constituted a "draw" of a panel for each circuit. For example, some circuits may draw a panel and then have those judges sit together for one morning, but other circuits may draw a panel and then have those judges sit together for three days. Ideally, we would be able to consistently identify these draws from the data itself. The reality, however, was dramatically more complicated. For example, in some cases two judges would sit together for multiple days, but have a different third judge each day. Or the same set of three judges would hear cases several days apart, with those judges sitting with different judges during the days in-between. Ultimately, we defined a panel as a group of three judges who sat to hear cases during a particular session of court on a particular day. Accordingly, if three judges were listed as hearing a set of cases at 9:00 a.m. and then later at 1:00 p.m., even on the same day, we would count this grouping as two separate panels. This approach seemed consistent with the interpretation of most circuits. However, some circuits had different measurement units. As a result, the fact that we were unable to identify what constituted a "draw" for each circuit, and instead relied upon this proxy, certainly has the potential to bias our results. To guard against this concern, we also perform our analysis using an alternative definition of panels as a robustness check in subpart III.B.

Finally, our data is based on panels as they actually sat and not how they were originally scheduled. Within each circuit, though practices vary, typically the calendar will be set for some time period—say, the following six months or even term—well in advance of the actual sittings.[99] Inevitably, there will be some last-minute adjustments to the calendar, as judges have scheduling conflicts or other reasons for not being able to sit on a particular date. The data that we received from each circuit was of the sittings as they actually occurred, and we decided that this was the relevant data on the ground that scholarship on judicial behavior relies on assuming that the judges who actually heard a case were randomly chosen. Accordingly, even if it was possible to tell in one circuit that Judge D was

99 *See, e.g.*, HOOPER, MILETICH & LEVY, *supra* note 17, at 79 (noting that in the Second Circuit "[t]he chief judge selects the panels for the year before the start of the term").

actually scheduled for a day and was replaced at the last moment by Judge E—as it occasionally was—we recorded the panel as a sitting with Judge E. In a related vein, some circuits occasionally had panels switch one or two members partway through the day. That is, every so often one would see the following on a calendar page: Judges F, G, and H sat to hear several cases on a given day and then Judge I replaced Judge H for only one case, presumably because of a case-specific recusal. Rather than try to determine what panels were scheduled originally, which would have been impossible across all circuits, we decided to code both panels as unique panels. In short, we coded only the panels that sat and all of the panels that sat.

Using these approaches, we amassed the data from the oral argument panels from all twelve circuits in this time frame. The number of panels by term for each circuit is presented in Table 1. As Table 1 shows, although the number of panels formed in each circuit is fairly stable from term to term, there is considerable variation across circuits.[100] Most notably, only 298 oral argument panels were formed in the First Circuit over this five-year period, but 2301 panels were formed in the Ninth Circuit in the same time frame.[101] This variation is consistent with the fact that the courts of appeals generally hear different numbers of appeals each year based on the size of their docket and the circuit norms about what kinds of cases should receive oral argument.[102]

100 *See infra* Table 1.
101 *See infra* Table 1.
102 *See* Marin K. Levy, *The Mechanics of Federal Appeals: Uniformity and Case Management in the Circuit Courts*, 61 DUKE L.J. 315, 355–60, 366–75 (2011).

TABLE 1. TOTAL NUMBER OF PANELS BY TERM

		Term					
		2008	2009	2010	2011	2012	Total
D.C.	Circuit	115	99	102	96	98	510
1st	Circuit	65	59	59	54	61	298
2nd	Circuit	294	248	254	266	270	1,332
3rd	Circuit	134	140	132	135	128	669
4th	Circuit	115	112	104	132	134	597
5th	Circuit	197	174	173	154	179	877
6th	Circuit	252	233	236	260	225	1,206
7th	Circuit	137	123	125	114	110	609
8th	Circuit	174	193	168	162	144	841
9th	Circuit	450	502	469	423	457	2,301
10th	Circuit	102	89	83	80	75	429
11th	Circuit	132	139	157	139	128	695
Total		2,167	2,111	2,062	2,015	2,009	10,364

In addition to collecting information on the panels formed in each circuit, we also compiled data on every judge that participated in at least one panel during this time frame. This includes information on every active circuit court judge, senior judge, and visiting judge (including active and senior district judges and other Article III judges).

Table 2 reports the total number of judges who sat on at least one panel during each term. Just as with the number of panels reported in Table 1, Table 2 shows that the number of judges who sat on at least one panel is fairly stable from term to term, but there is considerable variation across circuits.[103] For example, only 13 judges sat on at least one panel in the D.C. Circuit in each term, but over 100 sat on at least one panel in the Ninth Circuit in each term[104]—a discrepancy due to circuit size and also the use of visiting judges.[105]

103 *See infra* Table 2.

104 *See infra* Table 2.

105 *Compare* HOOPER, MILETICH & LEVY, *supra* note 17, at 49 ("The [D.C. Circuit] has not used the services of visiting judges for several years."), *and id.* ("The [D.C. Circuit] has 11 [active] judgeships."), *with id.* at 165 (describing the Ninth Circuit's procedure for orienting visiting judges), *and id.* ("The [Ninth Circuit] has 29 authorized [active] judgeships.").

TABLE 2. TOTAL NUMBER OF JUDGES SITTING ON AT LEAST ONE PANEL

		Term				
		2008	2009	2010	2011	2012
D.C.	Circuit	13	13	13	13	13
1st	Circuit	24	18	18	13	17
2nd	Circuit	63	58	47	48	58
3rd	Circuit	45	45	29	36	27
4th	Circuit	35	35	28	30	34
5th	Circuit	28	31	25	25	27
6th	Circuit	67	62	62	80	70
7th	Circuit	20	25	25	23	23
8th	Circuit	27	26	34	22	23
9th	Circuit	101	133	137	130	120
10th	Circuit	24	27	24	22	22
11th	Circuit	44	43	53	47	58

In order to test the hypothesis laid out in Subpart II.B, as a proxy for ideology we collected information on whether each judge was appointed by a Republican president.[106] This data was largely compiled from a database maintained by the Federal Judicial Center.[107] The exception is that complete biographical data was not available through the Federal Judicial Center for judges of the United States Court of International Trade,[108] so for these judges we instead gathered missing biographical information by directly contacting the clerk of that court. Table 3 provides summary statistics for the biographical information for judges included in our sample.[109]

106 We readily acknowledge that the party of the appointing president is an imperfect proxy for whether a judge is liberal or conservative. *See supra* note 10. There have been conservative judges appointed by Democrats, and liberal judges appointed by Republicans. That said, using the party of the appointing president is consistently used as a measure of ideology in the judicial behavior literature. *See supra* note 10.

107 *See* FED. JUDICIAL CTR., *Biographical Directory of Federal Judges, 1789-present*, http://www.fjc.gov/history/home.nsf/page/export.html [http://perma.cc/T3YG-J4LC] (follow the "Database Export" hyperlink).

108 *See id.*

109 It is worth noting that Table 3 provides summary statistics using the individual judge as the unit of observation. In other words, 9 of the 13 judges (69%) on the D.C. Circuit that heard at least one case were appointed by a Republican president. Since the judges did not all sit on the same number of panels, it does not necessarily follow that 69% of panel spots were filled by Republican-appointed judges. As we discuss in Subpart II.D, *infra*, however, our code takes the percentage of panels that individual judges actually sat on into account when creating artificial panels.

TABLE 3. PERCENTAGE OF JUDGES APPOINTED BY A REPUBLICAN PRESIDENT

		Term				
		2008	2009	2010	2011	2012
D.C.	Circuit	0.69	0.69	0.69	0.69	0.69
1st	Circuit	0.54	0.72	0.67	0.62	0.53
2nd	Circuit	0.54	0.53	0.47	0.44	0.45
3rd	Circuit	0.69	0.69	0.62	0.58	0.52
4th	Circuit	0.86	0.57	0.54	0.53	0.53
5th	Circuit	0.82	0.81	0.72	0.72	0.59
6th	Circuit	0.63	0.68	0.61	0.60	0.54
7th	Circuit	0.85	0.72	0.44	0.48	0.74
8th	Circuit	0.89	0.69	0.76	0.64	0.61
9th	Circuit	0.57	0.56	0.55	0.45	0.44
10th	Circuit	0.71	0.56	0.67	0.59	0.55
11th	Circuit	0.68	0.65	0.57	0.57	0.60

D. Methodology

In addition to creating an original dataset for this project, we have also written code to test for random panel assignment. The reason that we did so is that testing for random assignment of panels is not a straightforward proposition. After all, any combination of panel assignments is theoretically possible when a random process is used. For example, even if a coin is fair, it is possible to flip heads ten times in a row—it is just highly unlikely. In the case of a coin flip, calculating the probability of different combinations of flips is a fairly easy math problem. When there are multiple probabilistic events occurring simultaneously—such as 13 judges being selected to form 510 different three-judge panels in the D.C. Circuit (or 239 judges being selected to form 2301 different three-judge panels in the Ninth Circuit!)—the math required to calculate the probability of an event occurring becomes all but impossible fairly quickly.

In these complicated situations, one widely used approach to calculate the probability of an event occurring randomly is simulation.[110] The basic intuition is that if you are interested

[110] This approach is also commonly referred to as Monte Carlo simulation. For an overview of using simulation to study probability, see THOMAS M. CARSEY & JEFFREY J. HARDEN, MONTE CARLO SIMULATION AND RESAMPLING METHODS FOR SOCIAL SCIENCE (2014). Monte Carlo simulation has been used in a range of legal and social science research. *See, e.g.*, David S. Abrams, Marianne Bertrand & Sendhil Mullainathan, *Do Judges Vary in Their Treatment of Race?*, 41 J. LEGAL STUD. 347,

in calculating how likely an event is to occur through a random process, you can program a computer to complete that process a large number of times, and then count the number of times the event you are interested in occurred.[111] For example, if you wanted to know how likely it is to flip heads in 10 consecutive coin tosses—and you did not feel like calculating (1/2)^10—you could program a computer to randomly draw heads or tails 10 times, add up the number of heads, and then repeat the process 100,000 times. Any common laptop can complete this process in a matter of seconds and would reveal that 10 heads in a row should occur roughly 100 out of every 100,000 times.

This is exactly the kind of function that we have written code to perform,[112] only ours is slightly more complicated because the process we are interested in is more complicated than coin flips. Our code was designed to account for three key inputs for every circuit for each term. The first is the number of panels that were formed in a given circuit that term. For example, in the 2008 term there were 115 oral argument panels

---

349–50, 357–64 (2012) (using Monte Carlo simulation to "construct [a] counterfactual" for the authors' study and noting "this technique could benefit a large array of empirical studies facing similar constraints"); Eric D. Chason, *Naked and Covered in Monte Carlo: A Reappraisal of Option Taxation*, 27 VA. TAX REV. 135, 172–99 (2007) (using Monte Carlo simulation to test two taxing methods); Robert J. Franzese Jr. et al., *Modeling History Dependence in Network-Behavior Coevolution*, 20 POL. ANALYSIS 175, 177–78, 187–89 (2012) (using Monte Carlo simulation to evaluate and compare two coevolutionary strategies); Daniel E. Ho, *Foreword: Conference Bias*, 10 J. EMPIRICAL LEGAL STUD. 603, 603 n.1, 607 n.10 & fig.2 (2013) (using Monte Carlo simulation to measure "publication bias"); Kate Litvak, *Monte Carlo Simulation of Contractual Provisions: An Application to Default Provisions in Venture Capital Limited Partnership Agreements*, 98 CORNELL L. REV. 1495, 1507–12 (2013) (using Monte Carlo simulation to, in part, "develop a simulation methodology for coding the relative economic importance of contract terms"); Michael Peress, *Small Chamber Ideal Point Estimation*, 17 POL. ANALYSIS 276, 276, 280–84 (2009) (using Monte Carlo simulation to prove the author's proposed ideal point estimator is "an improvement over conventional estimators").

111 As Carsey and Harden explain, "the typical Monte Carlo simulation involves drawing multiple random samples of data from an assumed [Data Generation Process (DGP)] that describes the unobserved process in the larger population of how a phenomenon of interest is produced. It is the true or real DGP that scholars are ultimately interested in evaluating. Of course, we rarely know what the true DGP is in the real world—we just see the sample data it produces. Most of our research is about trying to uncover the underlying DGP or test predictions that emerge from different theories about what the DGP looks like." CARSEY & HARDEN, *supra* note 110, at 6. In our case, the Data Generation Process is how panels are formed, and the theory we are trying to test is whether those panels are formed randomly.

112 All of the code we have written for our project was written in the statistical programming environment "R." R is an open-source, freely available program that is widely used in the social sciences for empirical research. More information on R can be found at the website for the Comprehensive R Archive Network (CRAN), http://cran.r-project.org/ [http://perma.cc/N7GU-QGMQ].

formed on the D.C. Circuit.[113] The second input is all of the judges who sat on at least one panel on that circuit during the given term. To continue with our example, there were 13 judges who sat on at least one oral argument panel in the D.C. Circuit in the 2008 term.[114] Finally, we input the percentage of the total panels that each judge sat on during a given term. That is, just like we would have to tell the computer that 50% of coin flips should result in heads, we tell the computer that Judge X sat on 25.21% of the oral argument panels in the D.C. Circuit in 2008. It is worth noting that because we draw panels based on the percentage of panels that judges sat on, the drawing of each panel is an independent event. In other words, for every simulation there is a 25.21% chance Judge X will be selected for the first panel we create for the 2008 term, and even if Judge X is selected for that panel, there is still a 25.21% chance that he will be selected for the second panel.[115]

Using this information about the actual number of panels formed and the judges who comprised them, our code then creates one random set of panels for a given term. Figure 1 illustrates how the code we have written completes this process. The basic idea is that after we have set the number of times that a judge's name should be included in a pool of judges and the number of panels that need to be formed, the computer randomly picks names from that pool to complete the number of panels that should be formed for a given term (of course, while including the caveat that the same judge cannot be selected for the same panel more than once). After the com-

---

113 For a complete breakdown of the number of oral argument panels formed in each term by circuit, see *supra* Table 1.

114 For a complete breakdown of the number of judges who sat on at least one oral argument panel in each term by circuit, see *supra* Table 2.

115 Although this may result in simulations where judges are used more times than they actually were (e.g., Judge X appearing in 30% of panels in one simulation) or fewer times than they actually were (e.g., Judge X appearing in 20% of panels in the next simulation), over a large number of simulations the average number of times that judges were included in panels should approach the actual probability. There are three reasons for conducting our simulations in this way. First, we believe that treating each panel formation as an independent event reflects the actual panel formation process more closely than treating the number of panels that a given judge sat on as fixed. Second, treating each panel formation as an independent event allows us to evaluate whether the actual results deviate from nonrandomness by using a chi-squared test. This is valuable because a chi-squared test is an intuitive and widely used method of evaluating whether an actual distribution is different from an expected distribution in a statistically significant way. *See infra* notes 120–23 and accompanying text. Third, it is less computationally intense to run code that treats the formation of each panel as an independent event where all judges are available based on their actual availability for each simulation.

puter has created one random set of panels for a given term (e.g., 115 panels for the D.C. Circuit in 2008),[116] we have it count our quantity of interest: the number of judges appointed by a Republican president on each panel.

FIGURE 1. SIMULATING RANDOM ASSIGNMENT

	Panels to be Formed			Count
1.	Rogers	Henderson	Garland	1
2.	Garland	Griffith	Williams	2
3.	Griffith	Rogers	Sentelle	2
4.	Randolph	Tatel	Rogers	1
5.	Garland	Sentelle	Henderson	2
.				
.				
.				
k.	Griffith	Garland	Tatel	1

This process is then repeated for each term in the circuit and the results are added up. For example, there were a total of 510 oral argument panels formed in the D.C. Circuit during the five terms in our sample,[117] and our first random draw of 510 panels resulted in: 11 panels with 0 judges appointed by a Republican; 102 panels with 1 judge appointed by a Republican; 271 panels with 2 judges appointed by a Republican; and 126 panels with 3 judges appointed by a Republican. We then

116 *See supra* Table 1.
117 *See supra* Table 1.

simulate this process 100,000 times.[118] This means that, in total, we created over a billion artificial panels.[119]

After doing so, we can calculate the probability that the actual panels were randomly formed by comparing the actual panels to the simulated panels. To do so, we use a common approach to test whether an actual distribution is statistically different from an expected distribution: calculating a chi-squared test statistic.[120] We specifically calculate the Pearson's chi-squared test statistic to compare the actual number of panels in each circuit with 0 Republican appointees, 1 Republican appointee, 2 Republican appointees, and 3 Republican appointees to the mean values produced by our simulations.[121] The test statistic can then be compared to the

---

118 There is no set rule on how many times simulations should be run to produce reliable estimates, and the number 100,000 is admittedly somewhat arbitrary. In their book on using Monte Carlo Simulation methods, Carsey and Harden advise:

> [Determining] how many times [running a simulation] is enough is not certain, but most simulation studies include at least 500 to 1,000 repetitions, and some include many more than that. When determining the number of repetitions, you have to balance the increased precision that emerges from using a larger number of repetitions against the amount of time it takes for the simulation to run.

CARSEY & HARDEN, *supra* note 110, at 72 (footnote omitted). As Carsey and Harden suggest, running 1000 simulations is a very common convention in the legal and political science literature. *See, e.g.*, Abrams, Bertrand & Mullainathan, *supra* note 110, at 361 ("The process is then repeated 1,000 times . . . ."); Beber & Scacco, *supra* note 76, at 215 n.10 ("We simulate three runs of 1000 random draws from each distribution."); Stefano M. Iacus, Gary King & Giuseppe Porro, *Multivariate Matching Methods That Are Monotonic Imbalance Bounding*, 106 J. AM. STAT. ASS'N. 345, 357 (2011) ("We generate 1000 random datasets . . . ."). The reason that we elected to run 100,000 simulations, however, is to be overly cautious. By running 100,000 simulations, we are able to all but eliminate the probability that any of our results would be driven by "lucky" (or unlucky) simulations.

119 For each simulation, we create an artificial panel for each of the 10,000 panels that actually occurred between September 2008 and August 2013. So we created over 10,000 panels 100,000 times (10,000 x 100,000 = 1,000,000,000) for a total of over 1 billion panels simulated for this project.

120 *See* LAWLESS, ROBBENNOLT & ULEN, *supra* note 85, at 248–54 (explaining the meaning of the chi-squared statistic).

121 The chi-squared test statistic is calculated by: $\chi^2 = \sum_{i=1}^{n} [\frac{(Observed\ value\ i - Expected\ value\ i)^2}{Expected\ value\ i}]$. In our case, the observed values are the actual number of panels with 0, 1, 2, or 3 Republican appointees for each circuit, and the expected values for each circuit are the simulated mean values. For example, assume that a given circuit has 75 panels with 0 Republican appointees, 125 panels with 1 Republican appointee, 125 panels with 2 Republican appointees, and 75 panels with 3 Republican appointees. Assume further that our simulations found that the simulated mean values were 100 panels with 0 Republican appointees, 100 panels with 1 Republican appointee, 100 panels with 2 Republican appointees, and 100 panels with 3 Republican appointees. The test statistic would then be calculated as:

critical values of the χ^2 distribution with three degrees of freedom.[122] In our case, we define circuits as displaying statistically significant evidence of nonrandomness when this test produces a p-value lower than 0.1.[123]

Having discussed our reasons for doubting the randomness assumption, the specific hypothesis we set out to test, and the data and methodology we used to test that hypothesis, we now turn to our results.

III
TESTING THE ASSUMPTION OF RANDOM PANELS

In this Part, we present the results of using our simulation method to test whether circuit court oral argument panels are randomly created. First, we present our primary analysis of whether the ideological balance of panels is consistent with random assignment. This analysis reveals evidence of nonrandom assignment within the circuit courts of appeals. Second, we examine whether our findings are robust using an alternative approach to define panels. Specifically, to guard against the possibility that our initial results are due to definitional decisions, we adopt an alternative definition of panels that considers sets of judges that hear cases together on multiple days to be a single "panel." Although, as we explain below, we believe that using the alternative approach to defining panels is inappropriate because it actually masks over critical sources of nonrandomness, there is still evidence of

---

$\chi^2 = \frac{(75-100)^2}{100} + \frac{(125-100)^2}{100} + \frac{(125-100)^2}{100} + \frac{(75-100)^2}{100}$. This would produce a test statistic of 25. This can then be compared to a chosen critical value of the $\chi^2(3)$ distribution (that is, the chi-squared distribution with three degrees of freedom). This would reveal that a test statistic of 25 has a p-value of less than 0.001.

122 Although we are examining the number of panels with 0, 1, 2, or 3 Republican appointees, there are three degrees of freedom—instead of four—because of the constraint that the number of panels for the four combinations of panels must sum to the total number of panels. For example, if we know that there are 400 panels in a given circuit over a set time frame, and there are 75 panels with 0 Republican appointees, 125 panels with 1 Republican appointee, and 125 panels with 2 Republican appointees, we would be able to know that the number of panels with 3 Republican appointees is 75.

123 As an alternative approach to calculating statistical significance, for each of our simulations for a given circuit we calculated the chi-squared test statistic. That is, for each *simulation i*, we calculated the test statistic by using the number of panels with 0, 1, 2, and 3 Republican appointees produced by *simulation i* as the actual values and the overall simulated mean number as the expected values. We then saved those test statistics for all 100,000 simulations for each circuit. We then compared the test statistic produced by comparing the actual results to the simulated means to the distribution of test statistics produced by our 100,000 simulations. These values were nearly identical to the p-values produced by comparing our overall test statistics to the $\chi^2(3)$ distribution.

nonrandom assignment when using this approach. Third, we summarize our results and discuss some caveats that should be kept in mind when evaluating our evidence of nonrandomness, before turning to the implications of our findings.

A. Primary Results

Our primary results test whether the ideological balance of actual oral argument panels was likely to have occurred if the panels were strictly randomly formed. To do so, we compared the actual number of judges appointed by Republican presidents on each panel to the expected number of judges appointed by Republican presidents based on the simulated panels we created using the process outlined in subpart II.D.

The results of this analysis are presented graphically in Figure 2.[124] Each row in Figure 2 presents the results from a different circuit. For each circuit, the four graphs are density plots of the number of times that each of the four outcomes occurred in our 100,000 simulations. While the density curves are the simulated results, the vertical lines are the actual result. When the actual ideological composition of panels differs from the expected ideological composition of panels in a statistically significant way, then the density curve for the circuit is bold and solid.[125] When the actual ideological composition of panels does not differ from the expected ideological composition of panels, then the density curve for the circuit is dotted.

It is important to note that a statistically significant result for a given circuit does not necessarily mean that there are large differences between the actual ideological composition of

---

124 The results in Figure 2 are also presented in a table in Appendix A. For a discussion of the merits of using simulation techniques to present results graphically, see generally Gary King, Michael Tomz & Jason Wittenberg, *Making the Most of Statistical Analyses: Improving Interpretation and Presentation*, 44 AM. J. POL. SCI. 341 (2000); *see also* Jonathan P. Kastellec & Eduardo L. Leoni, *Using Graphs Instead of Tables in Political Science*, 5 PERSP. ON POL. 755, 756 (2007) ("We show that graphs [as opposed to tables] better communicate relevant information from both data summaries and regression models, including comparing values across variables or models and the sign and significance of predictors."). Although statistical results have been presented using graphs in the social sciences for over a decade, in the last few years it has become increasingly common in the empirical legal studies movement as well. *See, e.g.*, Adam S. Chilton & Christopher A. Whytock, *Foreign Sovereign Immunity and Comparative Institutional Competence*, 163 U. PA. L. REV. 411, 445–46, 450, 453–70 (2015); Daniel E. Ho, *Fudging the Nudge: Information Disclosure and Restaurant Grading*, 122 YALE L.J. 574, 597, 602–04, 610–11, 615–16, 621, 628–38 (2012); Bert I. Huang, *Lightened Scrutiny*, 124 HARV. L. REV. 1109, 1124–25, 1131–35 (2011).

125 For a discussion of how we define statistical significance, see *supra* text accompanying notes 121–23.

panels and the expected ideological composition of panels. Instead, there can be a statistically significant difference even if only a handful of panels are different between the actual and expected results.

As Figure 2 shows, there is evidence that the panels are nonrandom in four circuits: the D.C. Circuit, the Second Circuit, the Eighth Circuit, and the Ninth Circuit.

FIGURE 2. SIMULATED DISTRIBUTION OF JUDGES APPOINTED BY REPUBLICANS OVER A FIVE-YEAR PERIOD (PART 1)

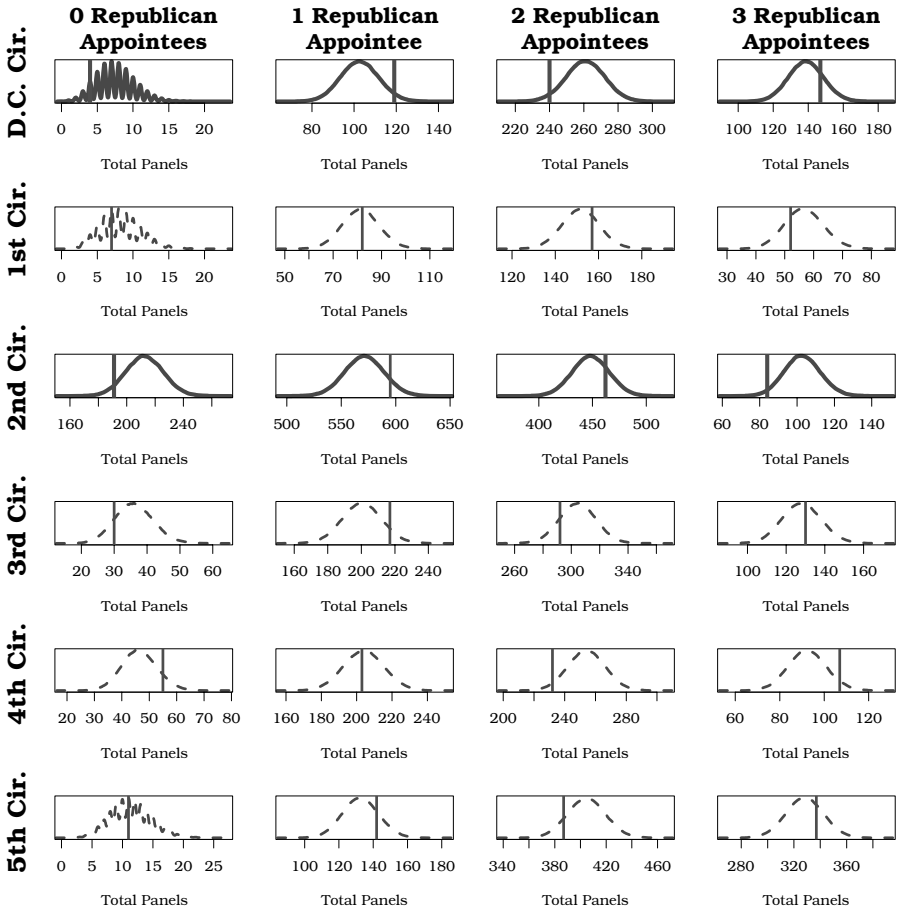
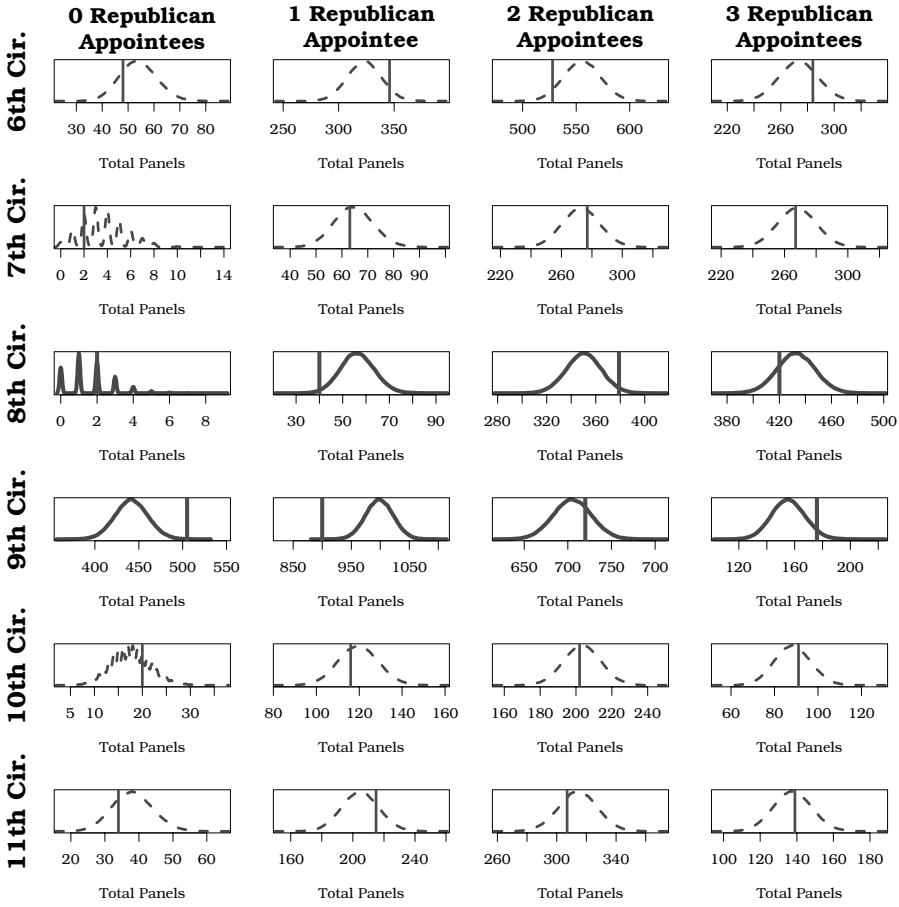


FIGURE 2. SIMULATED DISTRIBUTION OF JUDGES APPOINTED BY REPUBLICANS OVER A FIVE-YEAR PERIOD (PART 2)



For the D.C. Circuit, the Pearson's chi-squared test statistic is 6.42, which has a p-value of 0.09. The size of this test statistic is driven by two differences between the actual composition of panels and the expected composition of panels. First, there were more actual panels with 1 Republican appointee (119) than the expected number produced by our simulations (103).[126] Second, there were fewer actual panels with 2 Republican appointees (240) than the expected number produced by our simulations (261).[127]

For the Second Circuit, the Pearson's chi-squared test statistic is 7.05, which has a p-value of 0.07. The size of this test

126 *See supra* Figure 2, Part 1.
127 *See supra* Figure 2, Part 1.

statistic is also driven by two differences between the actual composition of panels and the expected composition of panels. First, there were fewer actual panels with 0 Republican appointees (191) than the expected number produced by our simulations (213).[128] Second, there were fewer actual panels with 3 Republican appointees (84) than the expected number produced by our simulations (103).[129]

For the Eighth Circuit, the Pearson's chi-squared test statistic is 7.63, which has a p-value of 0.06. The size of this test statistic is also driven by two differences between the actual composition of panels and the expected composition of panels. First, there were fewer actual panels with 1 Republican appointee (40) than the number produced by our simulations (56).[130] Second, there were more actual panels with 2 Republican appointees (379) than the number produced by our simulations (350).[131]

For the Ninth Circuit, the Pearson's chi-squared test statistic is 21.84, which has a p-value of < 0.01. The size of this test statistic is due to across-the-board deviations between the actual composition of panels and the expected composition of panels. The two biggest deviations, however, are the number of panels with 0 Republican appointees and the number of panels with 1 Republican appointee. There were more actual panels with 0 Republican appointees (505) than the number produced by our simulations (442).[132] Additionally, there were fewer actual panels with 1 Republican appointee (900) than the number produced by our simulations (999).[133]

For the other eight circuits—the First, Third, Fourth, Fifth, Sixth, Seventh, Tenth, and Eleventh—we did not find any statistically significant evidence of nonrandomness.[134] Of course, it is important to note that the panel assignments in these circuits may not be truly random. It may be the case that the panels were created in a nonrandom way, but happened to fall within the band of what would be produced by a random process. Relatedly, it is worth noting that it may not be the case that all of the circuits that we found statistically significant

---

128 *See supra* Figure 2, Part 1.
129 *See supra* Figure 2, Part 1.
130 *See supra* Figure 2, Part 2.
131 *See supra* Figure 2, Part 2.
132 *See supra* Figure 2, Part 2.
133 *See supra* Figure 2, Part 2.
134 It is worth noting that the results for the Fourth Circuit were close to statistical significance. *See supra* Figure 2, Part 1. For the Fourth Circuit, the Pearson's chi-squared test statistic is 5.77, which has a p-value of 0.11.

results for used a nonrandom process. It may instead be the case that some (or all) of the evidence of nonrandomness was simply due to chance—after all, results will be statistically significant at the 10% level just by chance 10% of the time.[135] As we discuss in subpart III.B., however, we are able to calculate that the probability of all our results being due to chance is less than 3%.

B. Robustness Checking

In addition to our primary results, we also evaluated whether our analysis was robust to alternative ways of defining panels. As we discussed in subpart II.C, we believe that the most supportable way to define panels is as three judges who sit together for the same court session on the same day. We made the decision to define panels using that definition for two reasons. First, our qualitative research suggests that this definition of panels corresponds to how many of the circuits conceive of panels.[136] Second, this is also an objective definition of panels that does not require making any arbitrary judgments on where to draw lines. This latter point both minimizes "researcher degrees of freedom" (that is, the choices that we are able to make that may influence statistical findings)[137] and makes it easier for other researchers to replicate our results.

That said, it is certainly the case that some circuits "draw" panels and have the judges sit together on more than one day. For example, a circuit may draw a panel of three judges and have those judges sit together on three consecutive days, or have the three judges sit together on Monday and then again on Thursday. It is not obvious, however, how to appropriately account for the fact that some circuits may have judges sit together on multiple days.

One reason that it is not clear how to appropriately account for the fact that some circuits elect to have judges sit together on multiple days is that the decision to do so is itself a potential source of nonrandomness. If a circuit randomly draws Judges A, B, and C to sit together and Judges X, Y, and Z to sit together, but then has Judges A, B, and C sit together for three days while Judges X, Y, and Z only sit together for one

---

135 For a longer discussion of this issue, see *infra* Subpart III.C.

136 *See* Levy, *supra* note 13.

137 *See, e.g.*, Joseph P. Simmons, Leif D. Nelson & Uri Simonsohn, *False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant*, 22 PSYCHOL. SCI. 1359, 1360–62 (2011) (discussing and defining "researcher degrees of freedom").

day, this may be a way that nonrandomness is introduced into the process.[138] As a result, counting Judges A, B, and C as one panel and also counting Judges X, Y, and Z as one panel would have the effect of masking nonrandomness.

Ultimately, although we believe the way we defined panels in our primary results is the most appropriate way to do so, we also have experimented with trying to account for the fact that when a circuit has three judges sit together on multiple days, it may be a single "draw." To do so, we removed panels from our data whenever the same three judges had sat together within the proceeding seven days. So if Judges A, B, and C sat together for court sessions on Monday, Tuesday, and Wednesday of the same week, the second two days were removed from our dataset. Table 4 presents the number of panels for each circuit using this alternative definition of panels. This reduces the number of panels in our sample from 10,364 to 6675 (leaving roughly 64% of the original total).[139]

TABLE 4. PANELS BY TERM USING ALTERNATIVE DEFINITION OF PANELS

		Term					
		2008	2009	2010	2011	2012	Total
D.C.	Circuit	90	84	89	91	86	440
1st	Circuit	63	56	56	53	59	287
2nd	Circuit	206	183	207	204	191	991
3rd	Circuit	61	69	57	63	49	299
4th	Circuit	113	110	104	132	134	593
5th	Circuit	59	52	49	48	51	259
6th	Circuit	157	142	143	164	146	752
7th	Circuit	137	121	123	114	110	605
8th	Circuit	69	82	72	73	67	363
9th	Circuit	248	288	288	242	265	1,331
10th	Circuit	101	88	82	79	74	424
11th	Circuit	55	60	65	71	80	331
Total		1,359	1,335	1,335	1,334	1,312	6,675

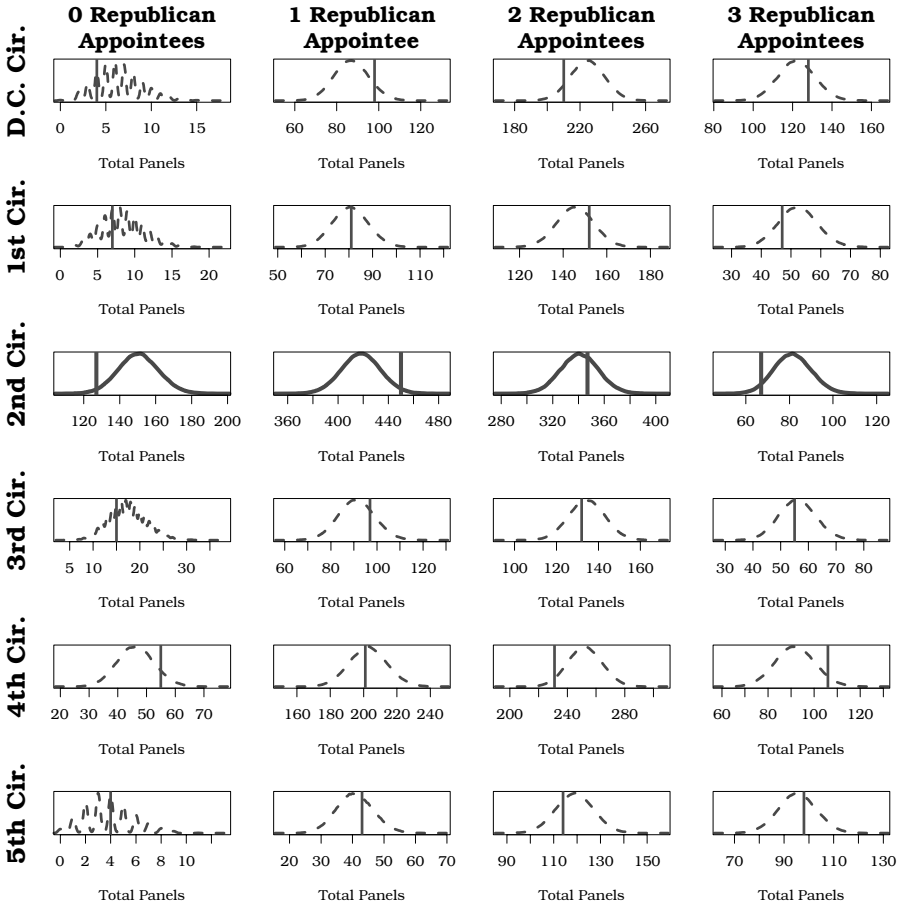
In addition to recalculating the number of panels in this way, we also recalculated the percentage of panels that each

[138] To be clear, the decision to have Judges A, B, and C sit together for three days while Judges X, Y, and Z only sit together for one day may be completely explainable by benign reasons. For example, one of the judges—say Judge Z—may have limited availability due to her senior status or a scheduling conflict.

[139] *Compare supra* Table 1, *with infra* Table 4.

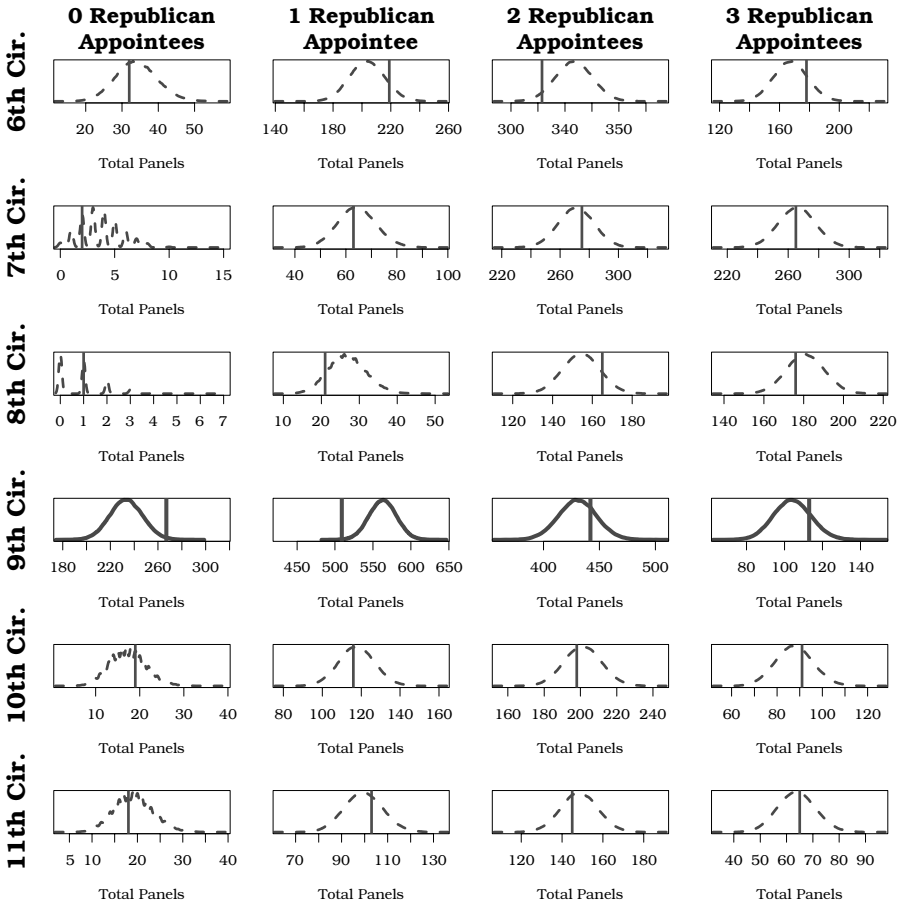
judge sat on in every term using this new definition of panels. Using this alternative definition of panels and the related new data on judge availability, we resimulated our results using the same process outlined in subpart II.D. The results of this analysis are presented in Figure 3.[140] As Figure 3 shows, there is now evidence that the ideological balance of panels is nonrandom in two circuits: the Second Circuit and the Ninth Circuit.

FIGURE 3. SIMULATED DISTRIBUTION OF JUDGES APPOINTED BY REPUBLICANS OVER A FIVE-YEAR PERIOD AFTER REMOVING "DUPLICATE PANELS" (PART 1)



140 The results in Figure 3 are also presented in a table in Appendix B.

FIGURE 3. SIMULATED DISTRIBUTION OF JUDGES APPOINTED BY REPUBLICANS OVER A FIVE-YEAR PERIOD AFTER REMOVING "DUPLICATE PANELS" (PART 2)



For the Second Circuit, the Pearson's chi-squared test statistic is 8.76, which has a p-value of 0.03. The size of this test statistic is primarily driven by three differences between the actual composition of panels and the expected composition of panels. First, there were fewer actual panels with 0 Republican appointees (127) than the expected number produced by our simulations (150).[141] Second, there were more actual panels with 1 Republican appointee (450) than the expected number produced by our simulations (418).[142] Third, there were fewer actual panels with 3 Republican appointees (67) than the expected number produced by our simulations (81).[143]

141 *See supra* Figure 3, Part 1.
142 *See supra* Figure 3, Part 1.
143 *See supra* Figure 3, Part 1.

For the Ninth Circuit, the Pearson's chi-squared test statistic is 11.07, which has a p-value of 0.01. The size of this test statistic is driven by two differences between the actual composition of panels and the expected composition of panels. First, there were more actual panels with 0 Republican appointees (267) than the number produced by our simulations (233).[144] Second, there were fewer actual panels with 1 Republican appointee (509) than the number produced by our simulations (563).[145]

For the other ten circuits, we did not find any statistically significant evidence of nonrandomness when using this approach to define panels.[146]

C. Discussion

We believe that our findings have a number of implications for both the judiciary and scholars.[147] Before discussing the implications of our primary findings, however, it is important to acknowledge that no empirical study is without limitations, and ours is no exception. We believe that it is critical—and consistent with best practices—to openly discuss the strengths and weaknesses of any empirical research strategy.[148]

First, although we have worked with all twelve regional circuit courts to assemble what we believe to be the most comprehensive dataset of panel assignments constructed to date, that does not mean that our data does not have limitations. As we discussed in subpart II.C, our analysis is based on data of oral argument panels, and the data is only on panels as they actually sat and not on how they were initially formed. Additionally, we define a draw of a panel as judges who sat together on a particular day for a particular session, and in some cases circuits drew panels for different periods. Although we try to account for these facts as best as possible—by, for example, conducting the robustness check presented in subpart III.B—these limitations have the potential to influence our findings.

---

144 *See supra* Figure 3, Part 2.

145 *See supra* Figure 3, Part 2.

146 *See supra* Figure 3. Although it is worth noting that the results for the Fourth Circuit were once again close to statistical significance: the Pearson's chi-squared test statistic is 5.64, which has a p-value of 0.12.

147 For more on the implications of our findings, *see infra* Part IV.

148 *See* Epstein & King, *supra* note 49, at 49–54. As Epstein and King note, "a basic premise of all empirical research—and indeed of every serious theory of inference—is that all conclusions are uncertain to a degree." *Id.* at 50. Given that reality, as Epstein and King persuasively argue, it is perhaps especially important that legal scholars conducting empirical research directly acknowledge assumptions and sources of uncertainty. *Id.*

Second, our method cannot definitively identify whether the courts used a strictly random process during this five-year period; it instead can reveal only if the actual panels formed were unlikely based on the distribution of results that a strictly random process would have produced. As we have noted, there of course may be some instances where circuits used a nonrandom process that our method does not identify, and other instances where circuits used a random process that produced unlikely distributions—after all, panels that are unlikely to happen randomly should randomly happen sometimes.[149] As a result, even in circuit courts where the actual distribution of judges appointed by a Republican president was near identical to the distribution produced by our simulation, it still may be the case that a nonrandom process was used. Correspondingly, even in circuits where we found evidence that a circuit's distribution of judges appointed by a Republican president was an outlier, it may mean nothing. What we can say is whether the overall distributions—within circuits or across all the circuits—are unlikely to have happened randomly. This is not a defect of our research design but simply an inherent feature of using observational data to test for randomness.[150]

All that said, it is possible to calculate the probability that the evidence of nonrandomness that we found can be attributed to chance—and that probability is incredibly low. We have undertaken this test for our primary results reported in subpart III.A in two ways. First, we have calculated the probability that we would have found 4 results that are statistically significant at the 0.1 level by chance in 12 opportunities to do so.[151] Since the outcome for each circuit is independent, results that are statistically significant at the 0.1 level—which is the standard we use—should occur by chance 10% of the time. By simulating 12 flips of a weighted coin—that is, one

---

149 *See supra* Subpart III.A.

150 The same principle has become a major talking point of the analytic sports community, which has often driven home that evidence of outliers alone does not "prove" anything. *See, e.g.*, Aaron Schatz, *Super Bowl XLIX's Insane Penalty Stats*, FOOTBALL OUTSIDERS (Jan. 27, 2015), http://www.footballoutsiders.com/stat-analysis/2015/super-bowl-xlixs-insane-penalty-stats [http://perma.cc/VM44-C8S8] (arguing that one year of unusual data regarding penalties called against the Seattle Seahawks does not prove that officials are biased against that team). Instead, the existence of an outlier is merely evidence that should be evaluated in context. In our case, our quantitative evidence is supported by qualitative evidence collected through extensive interviews and discussions with judges and clerks. *See generally* Levy, *supra* note 13.

151 That is, the chance of finding evidence of nonrandomness in four out of twelve circuits.

that produces a "1" 10% of the time and a "0" 90% of the time—we can then determine the distribution of how many statistically significant results should have occurred by chance.[152] Simulation suggests that 4 or more statistically significant results should occur by chance less than 3% of the time. That is, we can say with roughly 97% confidence that our results are not attributable to chance alone.

Second, it is also possible to test the overall significance of our primary results by taking the sum of the Pearson's chi-squared test statistics for all twelve circuits. The sum of these test statistics is 60.65. The p-value of a test statistic of 60.65 with 36 degrees of freedom (3 degrees of freedom x 12 circuits = 36 total) is 0.006. In other words, using this approach we can say with roughly 99% confidence that the deviations from randomness across twelve circuits would not have occurred by chance.[153]

Of course, the alternative analysis reported in subpart III.B only found statistically significant results for two circuits. Just as with our primary results, it is also possible to calculate the probability that our results using this approach are attributable to chance. First, using simulation to calculate the probability of finding 2 or more statistically significant results at the 0.1 level in 12 tries suggests that we cannot reject the null hypothesis that our results are attributable to chance. That is, 34% of simulations resulted in 2 or more statistically significant results randomly occurring. Second, the sum of the Pearson's chi-squared test statistics for all twelve circuits is 38.42. The p-value of a test statistic of 38.42 with 36 degrees of freedom is 0.36. In other words, using this approach we cannot say with confidence that our subpart III.B results are not attributable to chance.

Although this may lead some readers to conclude that our analysis has not produced convincing evidence of nonrandom assignment in the circuit courts of appeals, we believe that interpretation of our results would be a mistake for several

---

152 Just as with our primary analysis, we performed these simulations 100,000 times. *See supra* notes 119–21 and accompanying text.

153 The second approach to evaluating the probability of nonrandomness in our overall results is likely preferable to the first approach because it accounts for the degree of deviation from expected outcomes in every circuit instead of simply treating nonrandomness as a binary outcome. In other words, the first approach treats the Fourth Circuit as "nonrandom" even though it has a p-value of 0.11 and the D.C. Circuit as "random" because it has a p-value of 0.09. The second approach, however, accounts for the degree of nonrandomness by taking the sum of the test statistics. We report the first approach primarily, however, to be conservative in our estimate.

reasons. First, we believe that removing panels of judges that sat together multiple times in a set time period removes one critical way that nonrandomness is introduced into the circuit courts.[154] Second, when using an alternative definition of defining panels, we still found two circuits—the Second and the Ninth—with a p-value of 0.03 or lower. There is less than a 5% probability of two results out of 12 with a p-value of 0.03 or lower occurring randomly. Finally, our qualitative research—reported elsewhere[155]—found evidence that the circuit courts use nonrandom processes that are consistent with our quantitative findings reported in this paper. In sum, we believe our findings produce strong evidence that strictly random processes are not always used to create oral argument panels in the circuit courts of appeals.

IV
IMPLICATIONS

We believe that our results challenging the randomness of panel assignment in the courts of appeals have several implications. Fully appreciating their significance may ultimately require additional research and scrutiny of existing scholarship. For now, however, we briefly consider the implications for the three groupings identified at the outset of the Article: courts, general court scholars, and empirical researchers.

A. Courts

First, our findings suggest that there is a much more nuanced story about how the courts configure their panels than that they rely on a random process. As previously noted, it may be that there are a host of considerations at work—from the schedules of the judges, to recusals, to the return of cases on remand.[156] Accounting for these factors makes it impossible to achieve strict randomness, and scholars and practitioners should be aware of that fact.

Second, these findings could be relevant to the courts themselves. As previously stated, the qualitative evidence of one of the coauthors strongly suggests that the causal mechanism is an attempt to balance a series of otherwise arguably benign factors.[157] The cumulative effect, though, would be im-

---

154 For more on why we believe that our primary approach to defining panels is more supportable, see *supra* Subpart III.B.

155 *See* Levy, *supra* note 13, at 20–30.

156 *See supra* Subpart II.A.

157 *See supra* Subpart II.A.

portant to know. For example, as noted earlier, it could be that a court with more panels of a particular ideological makeup than would occur randomly relies on an ancillary circuit rule—such as no panel may have more than one senior judge. If that circuit had quite a few senior judges who were Republican appointees, say, then this rule would affect the ideological balance of panels.[158] It is important to know the effect of such a rule so that the court could then determine if it would be worth continuing with such a practice.

Finally, the discussion of the potential reliance on these other factors in constituting panels makes it all the more clear that it is important to know how the circuits are, in fact, creating their argument panels. Specifically, it would be useful to better understand what factors the courts are taking into account and how those factors might differ from circuit to circuit. Rather than focusing on *whether* panels are nonrandom, then, future discussions would do well to focus on *how* and *why* they are nonrandom.[159]

B. Court Scholars

Turning now to scholars who generally write on the federal courts, the implications of our findings are fairly straightforward. As a general matter, such scholars would do well to no longer assume that all courts of appeals randomly create oral argument panels. For some scholarship—those articles that simply mention random panels as a passing fact—this change in assumption will likely not be significant. For other kinds of scholarship, however, the implications will be more meaningful.

As noted in Part I,[160] there is a fair amount of scholarship that assumes panels are randomly configured and for which the assumption is relevant. Returning to a few earlier examples, one article made the argument that because oral argument panels are randomly drawn, the outcomes in agency cases are akin to lottery results.[161] Another article claimed that random panels were important because they helped to ensure that the courts of appeals did not fall prey to typical group problems, such as polarization.[162] Results suggesting

158 *See* McMillion, *supra* note 85, at 8 fig.3 (reporting that, as of March 7, 2014, 64% of all senior circuit judges were appointed by a Republican president).
159 *See* Levy, *supra* note 13, at 32–35.
160 *See supra* Subpart I.B.
161 *See supra* notes 39–40 and accompanying text.
162 *See supra* note 43 and accompanying text.

that panels are not randomly configured, and moreover, that the ideological balance of panels is affected in some circuits, call these kinds of arguments into question.

Furthermore, Part I also notes several articles that assume panels are randomly configured and then argue against that assumed state of affairs. One prominent example is the contribution by Tiller and Cross, which calls for courts to create panels with no more than two judges of either political party.[163] Our findings show that in at least some of the circuits, there are fewer panels with either all Democrat appointees or all Republican appointees than would be expected.[164] Thus, those scholars who have taken the panels to be randomly configured and then argued for a change to the status quo might now want to question the premise and this, in turn, could lead to different prescriptive conclusions.

C. Empirical Researchers

Finally, our results have important implications for researchers that use empirical methods to study judicial behavior. As we discussed in Part I,[165] random processes are incredibly important for empirical research because they make it possible to move beyond correlations and towards causation. In part because of the importance of randomization to empirical research, researchers have relied heavily on the assumed random assignment of judges to panels on the federal courts of appeals to study judicial behavior. Our results, however, provide evidence that the fundamental assumption that panel assignments are random may not be valid.

The main import of this finding is that empirical researchers should recognize and address the fact that panels may not be fully random. Unfortunately, there is no easy solution to this problem. As we previously discussed,[166] in one article the political scientist Matthew Hall excluded from his study circuits that he had reason to believe did not use random judicial assignment.[167] Although it may seem on first glance that the easiest response to our findings would be to simply exclude from future studies the four circuits that we identified as dis-

---

163 *See* Tiller & Cross, *supra* note 2, at 216–18, 232–34.

164 The Second Circuit, for example, appears to have fewer panels with no judges appointed by a Republican president and fewer panels with all three judges appointed by a Republican president than expected by our model. *See supra* Figure 2, Part 1.

165 *See supra* Subpart I.C.

166 *See supra* text accompanying notes 61–65.

167 *See* Hall, *supra* note 4, at 580.

playing evidence of nonrandomness,[168] we do not believe that this fix would be sufficient for two reasons. First, there is no reason to believe that these results are static. Instead, chief judges change, Clerks' Offices change personnel, and the processes used by the circuits to create panels change over time.[169] As a result, simply excluding these circuits may not be an adequate solution for scholars studying other time periods than those in our study. Second, several of the circuits we identified as having nonrandom assignments—like the Second Circuit and Ninth Circuit[170]—are among the largest and highest-profile circuits. Simply excluding these circuits and others from empirical research would likely be an unsatisfying response to evidence of nonrandom panel assignment.

Ultimately, our findings could affect the findings of numerous articles. For some studies, our results may strengthen their core findings; for other studies, our results may in some ways weaken them. Evaluating the full scope of the consequences of nonrandomness will both require a case-by-case evaluation of the research used in other studies, as well as more research on the ways that circuit courts deviate from random assignment. But the primary takeaway is that researchers should be cautious when making the fundamental assumption that judges are randomly assigned to panels in the federal courts of appeals.

CONCLUSION

How panels are formed in the federal appellate courts is an important question. To date, it has been assumed by many scholars that the answer to this question is "randomly." Challenging that assumption has implications for our ability to measure other key aspects of the courts, such as to what extent case outcomes are affected by the membership of the panel. It further shifts the way we study and understand the processes of the courts more generally.

Ultimately, more work is needed to understand the true assignment of judges to panels and its effects. There is a need for qualitative work to pinpoint how, precisely, panels are formed. There is also a need for future quantitative work to test whether our results carry over to different time frames. This Article has provided a key step in better understanding the

168 Those circuits are the D.C. Circuit, Second Circuit, Eighth Circuit, and Ninth Circuit. *See supra* Subpart III.A.

169 *See* Levy, *supra* note 13, at 20–30.

170 *See supra* Part III.

practices of the federal judiciary; we look forward to the gains to come.

APPENDIX A.
SIMULATED DISTRIBUTION OF JUDGES APPOINTED BY REPUBLICANS PANELS DEFINED BY JUDGES SITTING TOGETHER FOR A COURT SESSION FROM 2008 TO 2012

	Actual Panels	Simulated Mean	90% Confidence Interval		Test Statistic	P-Value
D.C. Circuit					**6.42**	**0.09**
0 Republicans	4	8	3	12		
1 Republican	119	103	88	118		
2 Republicans	240	261	242	279		
3 Republicans	147	139	122	155		
1st Circuit					**0.76**	**0.86**
0 Republicans	7	8	4	13		
1 Republican	82	81	69	94		
2 Republicans	157	152	138	166		
3 Republicans	52	57	46	68		
2nd Circuit					**7.05**	**0.07**
0 Republicans	191	213	191	235		
1 Republican	595	571	542	601		
2 Republicans	462	448	420	477		
3 Republicans	84	103	87	119		
3rd Circuit					**2.93**	**0.39**
0 Republicans	30	36	27	46		
1 Republican	217	200	181	220		
2 Republicans	292	304	283	326		
3 Republicans	130	128	112	145		
4th Circuit					**5.77**	**0.11**
0 Republicans	55	46	36	57		
1 Republican	203	204	185	223		
2 Republicans	232	254	234	274		
3 Republicans	107	92	79	107		
5th Circuit					**1.68**	**0.64**
0 Republicans	11	11	6	17		
1 Republican	142	133	116	150		
2 Republicans	387	405	381	429		
3 Republicans	337	328	305	352		
6th Circuit					**4.04**	**0.25**
0 Republicans	48	54	42	66		
1 Republican	346	323	298	348		
2 Republicans	528	556	527	584		
3 Republicans	284	273	250	297		

	Actual Panels	Simulated Mean	90% Confidence Interval		Test Statistic	P-Value
7th Circuit					**0.88**	**0.83**
0 Republicans	2	4	1	7		
1 Republican	63	65	52	77		
2 Republicans	277	273	253	293		
3 Republicans	267	268	248	288		
8th Circuit					**7.63**	**0.06**
0 Republicans	2	2	0	4		
1 Republican	40	56	45	69		
2 Republicans	379	350	327	374		
3 Republicans	420	433	409	457		
9th Circuit					**21.84**	**<0.01**
0 Republicans	505	442	411	473		
1 Republican	900	999	960	1038		
2 Republicans	720	705	669	742		
3 Republicans	176	155	136	175		
10th Circuit					**0.52**	**0.91**
0 Republicans	20	18	11	25		
1 Republican	116	120	104	135		
2 Republicans	202	203	186	220		
3 Republicans	91	88	75	102		
11th Circuit					**1.14**	**0.77**
0 Republicans	34	38	29	48		
1 Republican	215	205	185	225		
2 Republicans	307	314	293	336		
3 Republicans	139	138	121	155		

APPENDIX B.
SIMULATED DISTRIBUTION OF JUDGES APPOINTED BY REPUBLICANS PANELS DEFINED BY JUDGES SITTING TOGETHER WITHIN 7-DAY PERIOD FROM 2008 TO 2012

	Actual Panels	Simulated Mean	90% Confidence Interval		Test Statistic	P-Value
D.C. Circuit					**3.48**	**0.32**
0 Republicans	4	6	3	11		
1 Republican	98	87	74	101		
2 Republicans	210	225	207	242		
3 Republicans	128	122	107	138		
1st Circuit					**1.07**	**0.78**
0 Republicans	7	8	4	13		
1 Republican	81	81	68	93		
2 Republicans	152	146	132	159		
3 Republicans	47	52	42	63		
2nd Circuit					**8.76**	**0.03**
0 Republicans	127	150	132	169		
1 Republican	450	418	393	444		
2 Republicans	347	341	317	365		
3 Republicans	67	81	68	96		
3rd Circuit					**0.75**	**0.86**
0 Republicans	15	17	11	24		
1 Republican	97	91	78	104		
2 Republicans	132	134	120	149		
3 Republicans	55	56	45	67		
4th Circuit					**5.64**	**0.12**
0 Republicans	55	46	36	57		
1 Republican	201	203	184	221		
2 Republicans	231	252	233	272		
3 Republicans	106	92	78	106		
5th Circuit					**0.41**	**0.94**
0 Republicans	4	4	1	7		
1 Republican	43	41	31	51		
2 Republicans	114	119	106	132		
3 Republicans	98	95	83	108		
6th Circuit					**3.53**	**0.32**
0 Republicans	32	34	25	44		
1 Republican	219	204	184	224		
2 Republicans	323	347	324	369		
3 Republicans	178	168	149	186		

	Actual Panels	Simulated Mean	90% Confidence Interval		Test Statistic	P-Value
7th Circuit					**0.86**	**0.84**
0 Republicans	2	4	1	7		
1 Republican	63	64	52	77		
2 Republicans	275	271	251	291		
3 Republicans	265	266	246	286		
8th Circuit					**2.17**	**0.51**
0 Republicans	1	1	0	3		
1 Republican	21	27	19	35		
2 Republicans	165	154	139	170		
3 Republicans	176	181	165	197		
9th Circuit					**11.07**	**0.01**
0 Republicans	267	233	211	256		
1 Republican	509	563	533	592		
2 Republicans	442	430	402	459		
3 Republicans	113	104	88	121		
10th Circuit					**0.37**	**0.95**
0 Republicans	19	17	11	24		
1 Republican	116	118	103	133		
2 Republicans	198	201	184	218		
3 Republicans	91	88	74	102		
11th Circuit					**0.30**	**0.96**
0 Republicans	18	19	12	26		
1 Republican	103	99	86	113		
2 Republicans	145	148	134	163		
3 Republicans	65	64	52	76		