Judicial Dark Matter

Nina Varsava,† Michael A. Livermore,†† Keith Carlson,‡ and Daniel N. Rockmore‡‡

Judicial reform efforts aimed at rectifying historical gender and racial inequalities understandably focus on increasing the number of women and people of color on the bench. While this is an important program, this Article sheds light on another aspect of the representation problem, which will not necessarily be resolved through greater diversity in judicial appointments. This problem has to do with the understudied and often opaque practices of judicial administration. Through a large-scale empirical study of federal appellate decisions, we examine the distribution of judges along the lines of gender and race across decision panels and find systematic gender and racial imbalances in representation. We argue that these imbalances are most likely a product of disparities in decision reporting; some decisions, which we call judicial dark matter*, go unreported, resulting in distortions in the representation of judges in reported cases. This is the first study of the representation and distribution of judges by gender and race across decision panels. Ultimately, our findings suggest that assessing the distribution of legal power and influence across gender and racial groups based on the numbers of judges from these groups may be misleading and may create an inflated sense of the influence of judges from historically underrepresented groups. The diversity reform agenda, then, as it is typically cast in the scholarly literature, the political sphere, and the popular media alike, is incomplete. One cannot hope to understand how representation translates into power nor to remedy demographic power imbalances in the judiciary without attending to the features of judicial administration examined here.*

---

† Associate Professor of Law, University of Wisconsin Law School. For helpful comments, we're grateful to Christina Boyd, Anuj Desai, Christopher Drahozal, Sean Farhang, Peter Grajzl, William C. Hubbard, Christine Jolls, Jason Rantanen, and Miriam Seifter, as well as participants of the 2022 Conference on Empirical Legal Studies, the 2022 American Law & Economics Association Conference, the 2022 Midwest Law & Economics Association Conference, and 2022 faculty workshops at NYU School of Law and the Wisconsin Law School. We thank Saloni Bhogale, Jay Chen, Leigha Hildur Vilen, Kelsey Mullins, Yukiko Suzuki, Kou Wang, and Sojung Yun for excellent research assistance. Support for this research was provided by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison with funding from the Wisconsin Alumni Research Foundation.

†† Class of 1957 Research Professor of Law, University of Virginia School of Law.

‡ Tuck School of Business, Dartmouth College.

‡‡ Professor, Department of Computer Science, Dartmouth College; External Professor, Science Steering Committee, Santa Fe Institute.

INTRODUCTION

As of 2022,[1] about 70% of sitting federal judges in the United States were men and about 80% were white.[2] Although several presidents, including President Joe Biden, have made judicial appointments that increase the representation of women and people

[1] The American Bar Association releases a "Profile of the Legal Profession" report annually, but the 2023 iteration did not provide the same breadth of statistics regarding the judiciary's diversity. We therefore provide statistics from the 2022 report.

[2] AM. BAR ASS'N, 2022 ABA PROFILE OF THE LEGAL PROFESSION 6. For a sense of intersectional representation, just 4% of federal judges are Black women. *Id.* at 2. Note that these numbers include only Article III judges.

of color on the federal courts, thereby narrowing the representation gaps, the federal judiciary remains unrepresentative of the general population of the United States.[3]

This problem of underrepresentation has attracted considerable attention, especially in recent years.[4] Judicial reform efforts aimed at rectifying historical gender and racial inequalities understandably focus on increasing the number of women and people of color on the bench.[5] While this is an important program, this Article sheds light on another aspect of the representation problem, which will not necessarily be solved through greater diversity in judicial appointments. This problem has to do with the unseen practices of judicial administration. These practices complicate the assumption that adding diversity to the judiciary will rectify historical power imbalances. Those who wish to transform who wields power from the bench, as well as those who wish to understand how the judiciary allocates power across demographic groups, should attend to the elements of judicial administration that we address here.

Through a series of novel empirical analyses of a large dataset of federal appellate decisions, we study the ways in which judges are distributed across decision panels and find systematic

---

[3] *See New Report on Profession Focuses on Judicial Demographics*, AM. BAR ASS'N (Aug. 1, 2022), https://www.americanbar.org/news/abanews/aba-news-archives/2022/08/new-report-on-profession/ (reporting that, by the summer of 2022, President Biden had appointed sixty-eight judges, 76% of whom are women and 65% of whom are people of color). Forty-nine percent of the general U.S. population is male and 60% is white. 2022 ABA PROFILE OF THE LEGAL PROFESSION, *supra* note 2, at 19.

[4] *See, e.g.*, Amber Fricke & Angela Onwuachi-Willig, *Do Female "Firsts" Still Matter? Why They Do for Female Judges of Color*, 2012 MICH. ST. L. REV. 1529, 1530–31 (2014); Stacy Hawkins, *Trump's Dangerous Judicial Legacy*, 67 UCLA L. REV. DISCOURSE 20, 38 (2019); Mark S. Hurwitz & Drew Noble Lanier, *Diversity in State and Federal Appellate Courts: Change and Continuity Across 20 Years*, 29 JUST. SYS. J. 47, 49 (2008); Alaina Purvis, *Women in the Legal Profession: How Gender Barriers and Attrition Are Keeping Women Out of the Judiciary*, 43 J. LEGAL PRO. 283, 286 (2019).

[5] *See, e.g.*, ALL. FOR JUST., A FAIRER COURT: HOW PRESIDENT BIDEN AND CONGRESS RAISED THE BAR IN 2021, at 8 (2021); Danielle Root, Jake Faleschini & Grace Oyenubi, *Building a More Inclusive Federal Judiciary*, CTR. FOR AM. PROGRESS (Oct. 3, 2019), https://perma.cc/64FX-2W9T; Caroline Fredrickson, *Diversity in Federal Selection During the Biden Administration*, BRENNAN CTR. FOR JUST. (Apr. 5, 2022), https://perma.cc/48QM-JB8L; *Examining the Demographic Compositions of U.S. Circuit and District Courts*, CTR. FOR AM. PROGRESS (Feb. 13, 2020), https://perma.cc/T3H8-ASPS; *January 20, 2021 Snapshot: Diversity of the Federal Bench*, AM. CONST. SOC'Y, https://perma.cc/T5QJ-H9YR; ERINN MARTIN, LAWS.' COMM. FOR C.R. UNDER L., THE FEDERAL BENCH MUST REFLECT THE RACIAL, ETHNIC, AND GENDER DIVERSITY OF THE UNITED STATES 3 (2020) (available at https://perma.cc/YHP3-YU6N).

gender and racial biases.[6] Ultimately, our findings suggest that assessing the distribution of legal power and influence across gender and racial groups based on the numbers of judges from these groups may be misleading and may create an inflated sense of the influence of judges from historically underrepresented groups. The diversity-reform agenda, then, as it is typically cast in the scholarly literature and other legal commentary,[7] is incomplete. One cannot hope to understand how representation translates into power without attending to the features of judicial administration that we study here.

On the federal courts of appeals, judges typically decide cases in panels of three. The courts provide those decisions to litigants. And the legal community and the public can access them, or at least a subset of them, through legal research databases and West's Federal Reporter or Federal Appendix. So-called "published" decisions are reported in the Federal Reporter and are formally binding as precedent, whereas "unpublished" decisions may be reported in the Federal Appendix but do not constitute binding precedent.[8] Nevertheless, even these unpublished decisions can be, and often are, cited by both courts and litigants as persuasive precedent. All accessible decisions, then, can influence judicial decisions in future cases, and thus people's behavior and expectations. And so judges not only resolve disputes when they decide cases and issue decisions; they also clarify, develop, and create law and policy. But, as other recent scholarship has shown, not

---

6 As other scholars have pointed out, "[t]he US Courts of Appeals, in particular, are crucial institutions to examine because they establish the bulk of federal precedent due to the relatively small docket of the Supreme Court, which effectively renders most federal appellate decisions the final word." Laura P. Moyer, John Szmer, Susan Haire & Robert K. Christensen, *"All Eyes Are on You": Gender, Race, and Opinion Writing on the US Courts of Appeals*, 55 L. & SOC. REV. 452, 464 (2021).

7 *See supra* notes 4–5; *see also* Jennifer Bendery, *'Stunning Diversity': How Joe Biden Reshaped the Courts in 2023*, HUFFPOST (Dec. 28, 2023), https://perma.cc/SR24-K688; Tiana Headley & Nicole Sadek, *A Quarter of US Federal Courts Have Never Had a Non-White Judge*, BLOOMBERG (Sept. 5, 2023), https://perma.cc/552Z-THM3.

8 The Federal Appendix was terminated in 2021, and unpublished decisions are no longer included in an official reporter but may still be available on Westlaw and Lexis. *See* Eric Berg, *Pour One Out for the Federal Appendix*, RIPS L. LIBR. BLOG (Oct. 24, 2022), https://perma.cc/ME88-396Q. Even before 2021, not all unpublished decisions were reported in the Federal Appendix. Some that do not appear in the Appendix do appear in legal research databases, and others may not appear anywhere. *See infra* note 180.

all decisions are reported, and so not all decisions can influence law and policy beyond the parties to the dispute.[9]

In this Article, we present results from a large-scale empirical study of reported three-judge decisions issued by the twelve geographic courts of appeals from 2001 through 2017. We examine the relationship between the gender and race of a judge on the one side and, on the other, the number of cases in which the judge appears as a panelist as well as the demographic attributes of co-panelists. We find that the composition of panels in reported cases is skewed along the lines of both gender and race—for example, we see fewer panels with two women than expected based on the number of women on the bench. Further, we find that female judges participate, on average, in fewer reported decisions than male judges. Our findings suggest that certain types of panels, in terms of judge demographics, are more likely to appear in reported decisions than others, which indicates that judges of some demographic groups may be having greater legal influence than others.

Our study shows that existing court operating procedures and rules are not working to prevent judicial administrative decisions from reproducing gender and racial power imbalances. Perhaps shedding light on this reality, which has previously gone unnoticed, will help rectify the problem. But formal reforms to judicial administration may also be necessary to protect against this kind of demographic imbalance. These reforms might include new rules to govern decision reporting, revised protocols for panel construction and case assignment, and procedures for monitoring and tracking case reporting and panel composition.

A further contribution of our study is that it shows that we cannot safely assume that judges are randomly distributed into panels in the case data that is readily available for empirical study. This poses a challenge for studies that attempt to test for causal relationships between judge attributes and case outcomes and generally assume random distribution. Researchers can make use of the empirical tests we develop to check for possible biases in their data that might confound causal claims about the effects of judge attributes such as gender and race on case outcomes.

---

[9] *See* Michael Kagan, Rebecca Gill & Fatma Marouf, *Invisible Adjudication in the U.S. Courts of Appeals*, 106 GEO. L.J. 683, 688 (2018); Merritt E. McAlister, *Missing Decisions*, 169 U. PA. L. REV. 1101, 1118–20 (2021).

Our study accordingly contributes to the empirical and normative literature on judicial decision-making, judicial administration and court reform, and gender and race in the judiciary and legal profession more broadly. Although there is already a substantial body of empirical literature on judicial demographics and decision-making,[10] much less scholarship exists on the relationship between judge attributes and other aspects of the adjudicative process. Ours is the first study, as far as we know, of the representation and distribution of judges by gender and race across decision panels. The study is also a contribution to the continuing awareness across disciplines of the need to investigate and understand "data bias" when drawing statistical inferences from available data.[11]

This Article is organized as follows. In Part I, we explain how achieving gender and racial equality in the judiciary is not just a matter of increasing the numbers of women and people of color among judges; the demographic composition of federal appellate panels, a product of judicial administrative procedures, matters too. And we offer a theory of how demographic imbalances in the composition of decision panels might arise. In Part II, we describe our data and methodology and present results from a series of analyses designed to examine the gender and racial makeup of federal appellate panels; the evidence we uncover suggests that, in multiple ways, judges are not randomly or neutrally distributed across reported cases. In Part III, we discuss the normative implications of our findings, focusing on issues of vote and voice dilution of women and people of color. We then propose possible judicial reforms aimed at increasing the transparency and legitimacy of administrative procedures related to panel creation and decision reporting.

I. MOTIVATION AND THEORY

In this Part, we discuss gender and race representation in the judiciary. We review related literature on the issue, and we explain the relationship between the representation problem and

---

[10] *See infra* notes 148–52 and accompanying text. *See generally* Jeffrey J. Rachlinski & Andrew J. Wistrich, *Judging the Judiciary by the Numbers: Empirical Research on Judges*, 13 ANN. REV. L. & SOC. SCI. 203 (2017) (reviewing literature).

[11] *See, e.g.*, Keith Carlson, Michael A. Livermore & Daniel N. Rockmore, *The Problem of Data Bias in the Pool of Published U.S. Appellate Court Opinions*, 17 J. EMPIRICAL LEGAL STUD. 224, 244–61 (2020).

the demographic composition of decision panels. Further, we explore how judicial administrative decisions could lead to demographic imbalances in panel composition.

A. Gender and Race Representation

The legal profession has historically excluded women and people of color, especially from the highest-status positions.[12] Before 1977, 98% of federal judges were men and 90% were white.[13] Starting with President Jimmy Carter, most presidents took steps to increase the representation of women and people of color. President Donald Trump momentarily reversed that trend—his appointments were 83% white and 76% male—but President Biden has proven committed to increasing the diversity of federal judges.[14] By the summer of 2022, he had appointed sixty-eight judges to the bench, only three of whom are white men. Seventy-six percent of those appointments are women and 65% are people of color.[15] As of 2022, the federal judiciary sat at about 70% male and 80% white.[16]

Increasing the diversity of the judiciary may be a worthwhile goal for multiple reasons. Some scholars focus on the promise of *substantive representation*: the idea here is that when a member of a particular social or demographic group is in a position of influence, the individual will exercise their influence in such a way that furthers the interests of the group.[17] A female judge, for example, may better represent and serve the interests of women

---

12 Other groups are also underrepresented, including the LBGTQ+ population, people with disabilities, and people with lower socioeconomic backgrounds, but for the purposes of this paper, we focus on gender and race.

13 Root et al., *supra* note 5.

14 *Id.*; ALL. FOR JUST., *supra* note 5.

15 *New Report on Profession Focuses on Judicial Demographics*, *supra* note 3.

16 2022 ABA PROFILE OF THE LEGAL PROFESSION, *supra* note 2, at 6.

17 Stephen B. Burbank & Sean Farhang, *Politics, Identity, and Class Certification on the U.S. Courts of Appeals*, 119 MICH. L. REV. 231, 233 (2020) (explaining that the question of substantive representation in the context of government "is concerned with whether governmental actors, in their decision making, actually represent the distinctive preferences or interests of a community that they are associated with") (citing HANNA FENICHEL PITKIN, THE CONCEPT OF REPRESENTATION (1967)); 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, 301 (2004) (describing substantive representation as the idea that "women and minority judges would produce legal policy more advantageous" to their identity groups).

than a male judge.[18] And if men are overrepresented on the bench compared to women, then legal decisions and doctrine could be unjustly biased in favor of men. As Professors Michael Fix and Gbemende Johnson explain, "depending on the content of their jurisprudence, the inclusion of female judges can . . . promote the substantive representation of women's interests and create a court more 'receptive' to the concerns of women."[19] Increasing the representation of women, then, may make for a fairer legal system and, in a society committed to democracy, a more legitimate one.[20]

The idea that women as a group or people of color as a group would decide cases differently from others might seem rooted in a kind of gender or race essentialism that some would want to resist. However, as Professors Sean Farhang and Gregory Wawro point out, we can take seriously the idea that women and people of color might make a difference to adjudicative processes or outcomes without necessarily "indulg[ing] the facile notion that women or racial minority judges are homogeneous in their politics or values," or suggesting that "there is a monolithic 'women's perspective' or 'minority perspective' among judges or anyone else."[21] To pretend that many women or many people of color do not have shared experiences of marginalization and oppression elides the reality that historical injustices do align with identity groups.

---

[18] *See* Purvis, *supra* note 4, at 295 ("Women judges are able to bring their unique perspective to the bench, while also providing symbolic value for women both in the legal community and the community in which they serve.").

[19] Michael P. Fix & Gbemende E. Johnson, *Public Perceptions of Gender Bias in the Decisions of Female State Court Judges*, 70 VAND. L. REV. 1845, 1848 (2017) (quoting Thomas G. Walker & Deborah J. Barrow, *The Diversification of the Federal Bench: Policy and Process Ramifications*, 47 J. POL. 596, 597 (1985)) (citing Nancy Scherer, *Diversifying the Federal Bench: Is Universal Legitimacy for the U.S. Justice System Possible?*, 105 NW. U. L. REV. 587, 627 (2011)); *see* Hurwitz & Lanier, *supra* note 4, at 49 (explaining how "some scholars assert that judicial legitimacy is increased with enhanced levels of nontraditional judges, as their decisions are [informed by nontraditional perspectives]"); Sarah Westergren, *Gender Effects in the Court of Appeals Revisited: The Data Since 1994*, 92 GEO. L.J. 689, 689 (2004) ("[W]omen and minorities are said to bring different political perspectives to the act of judging, which result in decisions that embody substantively different policy outcomes than decisions of their white, male colleagues.").

[20] *See, e.g.*, Fix & Johnson, *supra* note 19, at 1847 (suggesting that "the presence of a judiciary that reflects the composition of the population potentially aids in conferring legitimacy on court decisions and authority"); Fricke & Onwuachi-Willig, *supra* note 4, at 1531 (arguing that "diversifying the federal judiciary with more women and men of color, but particularly with more women of color, is essential to moving forward and strengthening this country's democracy" as it "serve[s] an important symbolic and representative purpose that legitimizes this country's democracy").

[21] Farhang & Wawro, *supra* note 17, at 302.

This is why some critical race scholars resist "both the 'essentialism is per se bad' claim and the 'essentialism can be avoided' claim."[22] Professors Devon Carbado and Cheryl Harris explain:

> There *are* people we might call "Black" (though the content and experiences of blackness are not static but a function of particular social, legal, cultural, and ideological processes). There *is* a social force we might call "racism" (though its content and effects, and the technologies through which it is expressed, are neither transhistorical nor predetermined). And there *is* a phenomenon that we might call whiteness (though its boundaries are never fixed or fully articulated but are constituted and reconstituted in the service of racial power).[23]

Professors Wawro and Farhang further argue that, in our society, with its "long history of race and gender discrimination, including discrimination inscribed into law, the assumption that the race and gender of judges will have no bearing on the policy they make may be [just as] facile" as the idea that all women, or all members of a particular race, think alike.[24] Following this line of thought, we need not make any sweeping generalizations about the kind of differences that women and people of color may bring to the bench to justify inquiring into the question of substantive representation or to take seriously the possibility that, in some contexts, female judges decide cases differently from men, and Black judges differently from white judges.

A large body of empirical literature is aimed at testing the extent to which judges from different demographic groups adjudicate cases differently—both in terms of process and outcomes.[25] This literature presents mixed results: some studies find some differences in voting behavior between male and female judges, while others find no difference, and likewise for race effects.[26] As

---

[22] Devon W. Carbado & Cheryl I. Harris, *Intersectionality at 30: Mapping the Margins of Anti-Essentialism, Intersectionality, and Dominance Theory*, 132 HARV. L. REV. 2193, 2211 (2019) (emphasis in original).

[23] *Id.* at 2214.

[24] Farhang & Wawro, *supra* note 17, at 302.

[25] Studies that test for the effect of judge attributes on other variables of interest, however, may be affected by the data-bias problem that we describe below. *See infra* Part III.A.2.

[26] *See, e.g.*, Kate Malleson, *Justifying Gender Equality on the Bench: Why Difference Won't Do*, FEMINIST LEGAL STUD. 1, 6 (2003) (noting that many general studies, such as those examining differences between male and female judges in criminal sentencing and conviction rates, find minimal discrepancies, but studies of some legal areas, such as sex discrimination, find evidence of meaningful gender differences).

Wawro and Farhang report, "[w]hile studies of the influence of race and gender on judicial behavior have not produced broadly consistent results, a number of the studies have found systematic differences in decision making by judges along racial and gender lines in the area of civil rights," where both "women and racial minority judges appear, on average, to be somewhat more sympathetic than majority group judges [to plaintiffs]."[27] Professor Maya Sen explains how empirical studies testing "whether women and minority judges decide cases differently than their white male counterparts" have generally found that such differences do exist along the lines of both gender and race, but mainly "in the context of substantively salient issues."[28]

Even if increasing the representation of judges from different groups would not make a difference to case outcomes, it could make a difference to other important aspects of the adjudicative process. For example, the quality of the deliberative process may differ between more and less diverse courts, as might qualities of judicial opinions—such as their style and the types of reasons presented. As Sen observes, "descriptive representation" may "be instrumentally important by bringing in viewpoints that might otherwise be unshared."[29] Those different viewpoints may affect the quality of judicial deliberations and the content of judicial opinions even if they do not affect case outcomes. For example, Professors Susan Haire, Laura Moyer, and Shawn Treier find that diversity on federal appellate panels "shape[s] deliberative outputs"; in particular, opinions tend to be more comprehensive, discussing a greater number of issues, when a majority of women or people of color are on the panel.[30] The representation of different groups can matter substantively, then, even if that representation does not affect case outcomes.

Further, minimizing representation gaps may be a worthwhile aim even if members of different groups do not actually adjudicate cases differently. Descriptive representation—

---

27 Farhang & Wawro, *supra* note 17, at 303; *see also* Benjamin Alarie & Andrew Green, *Quantitative Analysis of Judicial Voting* 16 (Dec. 19, 2017) (unpublished manuscript) (available at https://perma.cc/85KY-9C78) (reporting that "in Canada gender has been found to [be] the single most powerful predictor of judicial decisions in equality/non-discrimination cases, with female judges more likely to reach liberal (pro-claimant) decisions").

28 Maya Sen, *Is Justice Really Blind? Race and Reversal in US Courts*, 44 J. LEGAL STUD. S187, S190 (2015).

29 *Id.* at S189.

30 Susan B. Haire, Laura P. Moyer & Shawn Treier, *Diversity, Deliberation, and Judicial Opinion Writing*, 1 J.L. & CTS. 303, 310, 315 (2013).

which just refers to the extent that an institution "mirrors, in salient respects, the composition of the community that it governs"—may matter, for one, because it furthers sociological legitimacy.[31] Especially in a democratic system, citizens may perceive the judiciary as more legitimate if its demographic composition reflects that of the general population.[32] Some have suggested that the overrepresentation of historically privileged groups in positions of power is inherently unjust and that having a more representative judiciary enhances the legitimacy of the judicial system.[33] The idea that descriptive representation is inherently valuable is undertheorized in the legal scholarship, but the idea does seem to reflect a strong intuition that increasing the representation of women and people of color in positions of power matters even if it would not lead to better processes and outcomes for women and people of color in society. Equality—in terms of participation and influence—across salient identity groups may be a good worth promoting for its own sake. Increasing the representation of historically underrepresented groups in positions of power such as judgeships should serve that value.

Increasing the representation of women and people of color on the bench, however, is insufficient—although probably necessary—for ensuring that the purported benefit and value of balanced representation is realized. This is because mere presence on the bench does not translate in a straightforward way to visibility or to the exercise of legal influence. As scholars studying other domains point out, "support for the prediction that increasing numbers [of women in deliberating groups] yield increasing rates of [women's] participation or influence" is "puzzlingly

---

31 Burbank & Farhang, *supra* note 17, at 233.

32 *See, e.g.*, Farhang & Wawro, *supra* note 17, at 301 (explaining that "[a]dvocates of racial and gender diversification of the judiciary have suggested that it will promote the value of descriptive representation by making the judiciary better resemble the public that it governs . . . strengthen[ing] at least the appearance of judicial impartiality [and] the judiciary's legitimacy as a democratic institution"); Hurwitz & Lanier, *supra* note 4, at 49 (suggesting that diversity among judges "enhances the appearance of impartiality for litigants who appear before the court and for the public at large"). Commentators have also referred to the "symbolic" value of descriptive representation. *See, e.g.*, Purvis, *supra* note 4, at 295 (suggesting that female judges provide "symbolic value for women both in the legal community and the community in which they serve"); Westergren, *supra* note 19, at 689 (explaining how some suggest "that greater symbolic representation of women and minorities is important in ensuring fundamental fairness and in redressing past inequalities").

33 *See* Sen, *supra* note 28, at S189 (citing PITKIN, *supra* note 17; Nancy Scherer & Brett Curry, *Does Descriptive Race Representation Enhance Institutional Legitimacy? The Case of the U.S. Courts*, 72 J. POL. 90 (2010)) (describing this normative view).

mixed."[34] This reality has received little attention in scholarship and commentary about gender and racial inequality in the judiciary, however, where the predominant focus has been on increasing the numeric representation of women and people of color.[35]

A notable exception is Professor Rachael Hinkle's work on case publication and judge gender and race. She hypothesized "that judicial diversity may be associated with lower levels of opinion publication and, thus, policy influence," and found evidence to support that idea.[36] Hinkle found that decisions by homogeneous panels are more likely to be published than those by diverse ones (in terms of gender and race).[37] Hinkle concludes that "groups with a larger number of representatives on the federal courts have disproportionate power since they are statistically more likely to be part of a homogeneous decisionmaking group than their less well-represented colleagues."[38] While our research questions differ from Hinkle's, our study is motivated by similar concerns and our results are consistent with the idea that judges from traditionally underrepresented groups may have less opportunity to influence the course of the law.

Our study focuses on the representation of judges in reported cases. First, we examine the gender and racial composition of judge panels in reported cases. That composition might matter for a number of reasons. If judge gender or race affects voting behavior, then panels with two women or people of color might be especially important for the legal influence of those groups (given that

---

[34] CHRISTOPHER F. KARPOWITZ & TALI MENDELBERG, THE SILENT SEX: GENDER, DELIBERATION, AND INSTITUTIONS 16 (2014).

[35] Commentators often seem to suggest that the problem of gender and racial inequality in the judiciary would be resolved if only we increased the numbers of women and people of color on the courts such that their representation was at least proportional to the general population. *See, e.g.*, Westergren, *supra* note 19, at 689 (arguing that "[w]omen must be appointed in numbers proportional to their share of the population," and not as "mere token[s]").

[36] Rachael K. Hinkle, *How Policy Influence Varies with Race and Gender in the US Courts of Appeals*, 8 RSCH. & POL., no. 3, 2021, at 1, 2, 3–5 (2021); *see also* Elizabeth A. Tillman & Rachael K. Hinkle, *Of Whites and Men: How Gender and Race Impact Authorship of Published and Unpublished Opinions in the US Courts of Appeals*, 5 RSCH. & POL., no. 1, 2018, at 1, 6 (studying the relationship between opinion assignment, judge gender and race, and opinion publication). *See generally* Nina Varsava, *Opinion Authorship and Precedential Status*, 101 WASH. U. L. REV. 1593 (2024) (studying the relationship between the gender and race of opinion authors and publication).

[37] Hinkle, *supra* note 36, at 2.

[38] *Id.*

the decision rule is majority vote).[39] Further, panel members influence one another—a phenomenon known in the empirical literature on judicial behavior as "panel effects"[40]—and the ways and extent to which they do so may be related to judge characteristics such as gender and race.[41] Previous empirical research has found that panel composition itself has a substantial effect on case outcomes; in this sense, the whole panel is greater than the sum of its parts, and studies of the relationship between judge attributes and individual votes "will substantially understate the impact of panel composition on case outcomes."[42] A judge's participation and influence may therefore be associated with the demographics of copanelists.

Second, we examine how a judge's gender and race relate to the quantity of reported cases in which the judge appears. It should not be taken for granted that judges will participate in the same numbers of cases or the same numbers of significant cases regardless of their gender and race. And even if assignments are evenly distributed across judges, we might see gender or racial

---

[39] *See supra* notes 25–28 and accompanying text.

[40] Professor Kevin M. Quinn defines panel effects as "the extent to which, and the possible reasons why, attributes of a federal appeals court judge's colleagues on a particular three-judge panel might exert an influence on the judge's decision in a particular case." Kevin M. Quinn, *The Academic Study of Decision Making on Multimember Courts*, 100 CALIF. L. REV. 1493, 1497–98 (2012). *See generally, e.g.*, Stuart Minor Benjamin, ByungKoo Kim & Kevin M. Quinn, *Partisan Panel Composition and Reliance on Earlier Opinions in the Circuit Courts* 1 J.L. & EMPIRICAL ANALYSIS, no. 1, 2024 (finding that the party composition of judge panels affects treatment of precedent). Panel effects could extend beyond effects on votes in cases to other kinds of adjudicative and administrative decisions and behavior as well. *See* Carlos Berdejó, *It's the Journey, Not the Destination: Judicial Preferences and the Decision-Making Process*, 51 U. LOUISVILLE L. REV. 271, 289–317 (2013) (examining the relationship between the ideological composition of panels and "process-related variables"). *See generally, e.g.*, Robert K. Christensen, John Szmer & Justin M. Stritch, *Race and Gender Bias in Three Administrative Contexts: Impact on Work Assignments in State Supreme Courts*, 22 J. PUB. ADMIN. RSCH. & THEORY 625 (2012) (examining panel effects on case disposition time).

[41] Judge Harry T. Edwards of the U.S. Court of Appeals for the D.C. Circuit has made related points in his academic writing. *See, e.g.*, Harry T. Edwards, *Race and the Judiciary* 20 YALE L. & POL'Y REV. 325, 329 (2002) ("A deliberative process enhanced by collegiality and a broad range of perspectives necessarily results in better and more nuanced opinions."); Harry T. Edwards, *The Effects of Collegiality on Judicial Decision Making*, 151 U. PA. L. REV. 1645, 1679, 1656, 1669–70 (2003) (observing that federal appellate judges, who generally decide cases in panels of three, are "willing to listen, persuade, and be persuaded, all in an atmosphere of civility and respect," and suggesting that "[a]ny credible attempt to explain judges' behavior . . . must take account of the collective nature of the enterprise" and that demographic diversity among judges may increase "the richness of deliberation").

[42] Joshua Fischman, *Interpreting Circuit Court Voting Patterns: A Social Interactions Framework*, 31 J.L. ECON. & ORG. 808, 836 (2015).

disparities in representation in reported cases.[43] This is because decisions about whether to report decisions might be associated with the gender and racial composition of decision panels.

As Hinkle observes, "[a] seat at the table matters. But making sure everyone at the table has an equal voice matters too. We must look not only at the representation of women and minorities in policymaking positions, but also at how institutional rules and task distribution shape the power they exert."[44] Although there are various ways in which the distribution of power and voice among sitting judges could be examined, part of our aim with the empirical analyses we present below (Part III) is to explore this distribution from a novel angle. By examining the gender and racial composition of federal appellate panels—the typical decision-making unit at the federal courts of appeals—we shed light on how the presence of judges from different groups translates into participation in reported judicial decisions. As Professors Christopher Karpowitz and Tali Mendelberg observe in a major study of gender and group decision-making, "[d]escriptive representation can build substantive and symbolic representation for women, but only if it grants women authoritative representation, that is, equal status in the group."[45] That is also true, of course, of representation in terms of race and ethnicity. Rather than assuming that individuals from historically underrepresented groups will have voice and influence on the bench equal to individuals from historically dominant groups, we should try to test that ideal, which is what we set out to do here.

B. Judicial Administration and Case Management

We identify two different but not mutually exclusive pathways that would allow for panel compositions that are skewed or biased along demographic lines. By skewed or biased, we mean that the observable panels in judicial decisions do not have the demographic composition that we would expect if, regardless of race and gender, judges participate in (roughly) the same number of cases, and judges and cases are (roughly) randomly assigned to panels. As we discuss below in Part III.A, panels might be biased in this way for a variety of reasons and not necessarily as a result of any nefarious or even intentional manipulation.

---

43 *See infra* Part I.B.2.

44 Hinkle, *supra* note 36, at 6.

45 KARPOWITZ & MENDELBERG, *supra* note 34, at 317.

The two mechanisms we focus on that may lead to skewed panels are (1) nonrandom assignment and (2) unreported cases, or what we call *judicial dark matter* (borrowing the term for unobserved physical matter in cosmology[46]). If the assignment of judges to cases is systematically associated with judge gender and race, then those associations would appear in the makeup of panels. If judges choose not to report some decisions—meaning that these decisions are not publicly available—and if the decision to report is correlated with judge race or gender, that could also produce bias along the lines of race or gender in the panels of reported cases.

---

[46] *See* BARBARA RYDEN, INTRODUCTION TO COSMOLOGY 27 (2d ed. 2016) (defining "dark matter" as "something which is too dim for us to see"). Invisible or inaccessible data has likewise been referred to as "dark data." Zachary D. Clopton & Aziz Z. Huq, *Necessary and Proper Stewardship of Judicial Data*, 76 STAN. L. REV. 893, 906 (2024) (citing DAVID HAND, DARK DATA: WHY WHAT YOU KNOW MATTERS 3–12 (2020)). Professors Zachary Clopton and Aziz Huq discuss the problem of the "'dark data' of the federal courts" more broadly, arguing that lower courts "generate an enormous volume of potentially valuable data" that remains inaccessible to the public and researchers but that could be, and should be, "leveraged . . . for the public good." *Id.* at 898–99. Another mechanism that could create bias in panels is settlement decisions of litigants. For reasons that others have elaborated, though, we doubt that settlement has a substantial effect in this regard because the panel-assignment process leaves minimal time for parties to account for panel assignment in their settlement decisions. *See, e.g.*, Daniel L. Chen & Jasmin K. Sethi, *Insiders, Outsiders, and Involuntary Unemployment: Sexual Harassment Exacerbates Gender Inequality* 25 (Toulouse Sch. of Econ., Working Paper No. 16-687, 2016) (noting that "judges are revealed after litigants file their briefs in Circuit Courts, sometimes only a few days before the hearing, which gives little opportunity and incentive for settlement upon learning the identity of the panel"); Fischman, *supra* note 42, at 812 (explaining that correlations between judge attributes and panels "may arise in subtle ways, for instance, if the announcement of the panel composition leads some parties to settle," but suggesting that "[t]his effect is likely to be small, . . . since most circuits do not announce panel composition until shortly before oral argument"); Samuel P. Jordan, *Early Panel Announcement, Settlement, and Adjudication*, 2007 BYU L. REV. 55, 70–71 (testing whether earlier panel announcement resulted in increased settlement rates and finding only a slight effect). Circuit internal operating procedures specify the timing of panel announcement. The Fifth Circuit's rules, for example, provide that "the court does not release the identity of the panel members until seven days before the beginning of the oral argument session." CLERK'S OFF., U.S. CT. OF APPEALS FOR THE FIFTH CIR., PRACTITIONERS' GUIDE TO THE UNITED STATES COURT OF APPEALS FOR THE FIFTH CIRCUIT 37 (2023).

Another possible mechanism of nonrandom panel compositions is judge recusals. We doubt that recusals would lead to systematic imbalances in panels along the lines of race or gender, but it is a possibility we do not attempt to test here and cannot rule out. Another potential source of imbalances that we do not explore here comes from motions panels, which make decisions about whether to proceed on the merits. Motions panels sometimes issue orders that appear in our data, and those panels are not necessarily assigned by the same procedure as merits panels. *See* Will Baude, *How Seventh Circuit Motions Panels Work*, WASH. POST (Dec. 3, 2014), https://perma.cc/ERQ3-X9GG (discussing motions-panel procedures); *see also* Ed Whelan, *Judge Easterbrook Responds*, NAT'L REV. (Dec. 2, 2014) https://perma.cc/79HQ-RWMT (discussing motions-panel procedures).

1. Case assignment.

The assignment of judges to panels and of panels to cases on the federal courts of appeals is generally understood to be random or roughly random. The journalist's guide produced by the United States Courts states that "[a]ppeals normally are decided by randomly assigned three-judge panels" and "[j]udges play no role in panel assignments."[47] In a major empirical study of judicial administration, Professors Jennifer Barnes Bowie, Donald Songer, and John Szmer interviewed judges across the federal courts of appeals, inquiring about the assignment process; the authors reported that, "[w]hile the specific procedures used to assign judges and cases to panels vary across the circuits, all the judges . . . indicated they were confident that the process was essentially random and contained no political bias."[48] As Professors Adam Chilton and Marin Levy observe, "[t]he notion that panels are formed randomly has moved into the various academic literatures"; "[t]here is a robust general courts literature that has consistently held this assumption" and "a broad, quantitative literature on judicial decision making that has relied on this assumption to reach its results."[49] For example, in their empirical study of the relationship between judge gender and judicial opinion content, Haire, Moyer, and Treier assert that "[c]ase assignment to panels is, by all accounts, an essentially randomized process, with exceptions made in some situations when cases are related to each other" and some "adjustments . . . to prevent imbalances across panels in terms of case difficulty."[50]

Although there is no statutory requirement of random assignment, the courts of appeals have internal operating procedures

---

47 *Appellate Courts and Cases—Journalist's Guide*, U.S. CTS., https://perma.cc/X23J-N2LM; *see also* Allison Orr Larsen & Neal Devins, *Circuit Personalities*, 108 VA. L. REV. 1315, 1349 (2022) (asserting that "federal appellate judges hear cases in randomly-assigned panels of three that continuously shuffle").

48 JENNIFER BARNES BOWIE, DONALD R. SONGER & JOHN SZMER, VIEW FROM THE BENCH AND CHAMBERS 49 (2014); *see also* Adam Cox & Thomas J. Miles, *Judging the Voting Rights Act*, 108 COLUM. L. REV. 1, 17 (2008) ("[W]ithin circuits, appellate judges are randomly assigned to panels."); Tracey E. George & Albert H. Yoon, *Chief Judges: The Limits of Attitudinal Theory and Possible Paradox of Managerial Judging*, 61 VAND. L. REV. 1, 2–4 (2008) ("[M]ost courts have instituted procedures that result in roughly random assignment of judges to cases."); Richard A. Posner, *A Heartfelt, Albeit Largely Statistical, Salute to Judge Richard D. Cudahy*, 29 YALE J. ON REGUL. 355, 357 (2012) ("[T]he panels that hear cases are randomly selected from the court's judges.").

49 Adam S. Chilton & Marin Levy, *Challenging the Randomness of Panel Assignments in the Federal Courts of Appeals*, 101 CORNELL L. REV. 1, 8 (2015).

50 Haire et al., *supra* note 30, at 306.

mandating it.[51] Nevertheless, previous studies have uncovered some, although quite limited, evidence of nonrandom assignment.[52] Chilton and Levy studied panels on the twelve geographical courts of appeals from 2008 to 2013. They looked at the distribution of judges by appointing party using a similar methodology to ours,[53] and they found some evidence of nonrandom assignment in four circuits.[54] Their study examined only oral argument panels, however,[55] and as the authors observe, most cases are decided without oral argument.[56] Further, the decision to grant oral argument is not random and may correlate with judge attributes.[57] Chilton and Levy's evidence of nonrandomness

---

51 The U.S. Code provides that the courts of appeals "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," but includes no guidance on how judges are to be assigned to panels, nor panels to cases. 28 U.S.C § 46. Chilton and Levy report that most "of the federal appellate courts . . . state that they form their argument panels randomly or with a goal of largely equalizing co-sittings"; in the latter case, "[a]lthough the process of panel configuration is . . . not technically a random one, the results should still be consistent with random panel assignment." Chilton & Levy, *supra* note 49, at 9, 11; *see also* J. Robert Brown & Allison H. Lee, *Neutral Assignment of Judges at the Court of Appeals*, 78 TEX. L. REV. 1037, 1069 (2000) (surveying circuit rules on assignment and reporting that "[a]ll circuits purport to use a system of random assignment of judges and cases").

52 Hinkle, *supra* note 36, at app. A & n.1 (noting that "[t]here is evidence that assignment of judges to panels is not always strictly random because of scheduling and logistical constraints" but "once panels are formed the cases assigned to each panel are selected randomly with only rare exceptions"); *see also* Alma Cohen, *The Pervasive Influence of Political Composition on Circuit Court Decisions* 11–12 (Harvard Public Law Working Paper No. 1109, 2024) (observing that "several empirical studies have examined [the] random-assignment assumption and found it to be empirically valid" and that, while two recent studies have found evidence of nonrandomness, "even these two studies conclude that any such deviations are small").

53 For details, see *infra* Part II.B.1.

54 Chilton & Levy, *supra* note 49, at 39–40.

55 *Id.* at 24.

56 *Id.* at 26.

57 *See* Stefanie A. Lindquist, *Bureaucratization and Balkanization: The Origins and Effects of Decision-Making Norms in the Federal Appellate Courts*, 41 U. RICH. L. REV. 659, 663 (2007) (explaining that the courts of appeals have "guidelines for when opinions should be published or oral argument granted," "[b]ut the application and interpretation of these guidelines are shaped profoundly by the participating judges' views and behavior regarding appropriate or adequate appellate process"); *see also* Penelope Pether, *Sorcerers, Not Apprentices: How Judicial Clerks and Staff Attorneys Impoverish U.S. Law*, 39 ARIZ. ST. L.J. 1, 11 n.56 (2007) (describing the process for disposing of cases without oral argument at the Ninth Circuit).

is thus likely a product, at least in part, of oral argument decisions rather than panel assignment itself, and may overestimate nonrandomness in panels construction overall.[58]

In another study, Levy investigated panel assignment at several courts of appeals qualitatively, through interviews with judges and court administrators, and found no evidence to suggest that those responsible for assigning judges to panels (chief judges or court administrators) intentionally manipulated panels to affect the composition of judge attributes represented on panels.[59] She did find that courts considered "logistical or efficiency-based factors" in panel assignment, taking into account, for example, "the personal schedules" and "calendar preferences" of judges, in particular senior-status judges.[60]

Given evidence that panel assignment is not strictly random, we might expect some gender or race bias in panel compositions as a result. For example, if female judges are more likely to take vacation during the same time—say, when schools are on spring break—that might affect panel composition on the margins. However, given the rules providing for random assignment of judges to panels and panels to cases, and the limited evidence of nonrandomness that previous studies have uncovered, it is unlikely that deviations from random assignment are creating any notable gender and race biases in panel composition.

2. Case reporting.

This brings us to the second pathway that may lead to skewed panels in reported decisions: decision reporting. A large body of literature, both empirical and normative, focuses on *decision*

---

[58] Note that, while we uncover more widespread evidence of nonrandomness in panels than Chilton and Levy, our results are consistent with little or no randomness in initial panel construction. That is because we examine the distribution of judges as they appear in reported cases, and not the distribution that appears earlier in the adjudicative process (which is what Levy and Chilton examine). Indeed, we leverage results of previous studies finding limited evidence of nonrandom assignment to argue that our findings are more likely a product of nonrandomness in decision reporting (the "missing decisions" pathway). Chilton and Levy do not examine this aspect of judicial administration.

[59] Marin Levy, *Panel Assignment in the Federal Courts of Appeals*, 103 CORNELL L. REV. 65, 68 (2017); *see also* Matthew Hall, *Randomness Reconsidered: Modeling Random Judicial Assignment in the U.S. Court of Appeals*, 7 J. EMPIRICAL LEGAL STUD. 574, 578 (2010) (finding, through phone interviews with the clerks' offices of the twelve geographic courts of appeals, that "judges [are] randomly assigned to cases through the use of a computer program or simply by drawing names out of a hat in most circuits," but that "in the Second, Third, and Sixth Circuits, judges were not randomly assigned to panels").

[60] Levy, *supra* note 59, at 68.

publication as a central aspect of judicial administration in the federal courts of appeals. Today, each court of appeals designates a large proportion of its decisions as "not for publication"—a practice that has been widely criticized[61] but that many experts believe is necessary for reasons of case management and efficiency.[62] "Unpublished" decisions are available through legal research databases and were generally reported in the Federal Appendix from 2001 to 2021, but they are not technically precedential.[63] They nevertheless do have persuasive value, and litigants and courts may and do cite them.[64] These two types of decisions—published and unpublished—have been characterized as two "tracks" or "tiers" of appellate cases.[65]

In recent years, some scholars have drawn attention to a "third tier of federal appellate decisions."[66] These decisions are unpublished and are also unreported even in the Federal Appendix.[67] These decisions represent an additional branch of the

---

61 *See* K.K. DuVivier, *Are Some Words Better Left Unpublished?: Precedent and the Role of Unpublished Decisions*, 3 J. APP. PRAC. & PROCESS 397, 405–06 (2001) ("Some commentators have criticized unpublished opinions for creating a body of 'secret' law, and other commentators have stated that permitting decisions to go unpublished invites sloppy decisions, less judicial accountability, and a lack of uniformity."); Kristen Marie Hansen, *The U.S. Legal System: Common Values, Uncommon Procedures*, 69 BROOK. L. REV. 689, 726 (2004) (arguing that "it is only through the complete eradication of the use of unpublished, non-precedential decisions that the common law can function in a fair and just manner").

62 *See* Dean A. Morande, *Publication Plans in the United States Courts of Appeals: The Unattainable Paradigm*, 31 FLA. ST. U. L. REV. 751, 755 (2004) ("Selective publication plans favoring unpublished opinions were promoted in response to the exponentially expanding volume of cases before the courts."); William L. Reynolds & William M. Richman, *An Evaluation of Limited Publication in the United States Courts of Appeals: The Price of Reform*, 48 U. CHI. L. REV. 573, 593 (1981) ("The major impetus for the limited publication movement has been the dramatically increasing caseload of the circuit courts.").

63 Hinkle, *supra* note 36, at 2 ("Legal doctrine, often formalized in court rules, mandates that only opinions that a panel designates as 'published' are binding precedent, that is they are required to be applied throughout the circuit"); McAlister, *supra* note 9, at 1114 n.65 ("In every circuit, decisions that are not designated for publication are not binding precedent."); Berg, *supra* note 8.

64 Before 2007, some circuits prohibited litigants from citing unpublished decisions, but since 2007 the Federal Rules of Appellate Procedure have required courts to permit citation to all types of decisions. FED. R. APP. P. 32.1(a); FED. R. APP. P. 32.1 advisory committee's note.

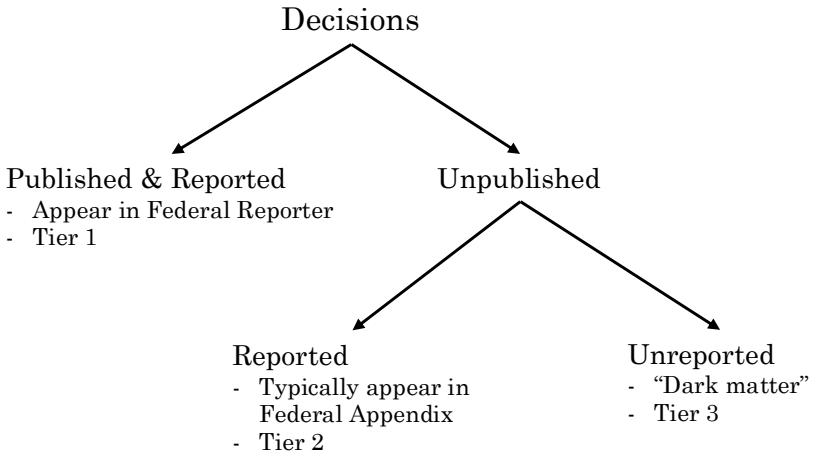
65 McAlister, *supra* note 9, at 1106 & n.29.

66 *Id.* at 1106 (studying unreported or "missing" decisions at the twelve geographical federal courts of appeals). *See generally* Jason Rantanen, *Missing Decisions and the United States Court of Appeals for the Federal Circuit*, 170 U. PA. L. REV. ONLINE 73 (2022) (studying missing decisions at the Federal Circuit).

67 McAlister, *supra* note 9, at 1106. Scholars have also recently drawn attention to, and criticized, the phenomenon of inaccessible federal district court decisions. *See, e.g.*,

category of unpublished decisions that has only started receiving scholarly attention recently. These three tiers of federal appellate decisions are represented in Figure 1.

FIGURE 1: DECISION TIERS AT THE FEDERAL COURTS OF APPEALS



Many commentators have apparently assumed that all federal appellate decisions appear as officially published decisions in the Federal Reporter or else as unpublished ones in the Federal Appendix, and likewise that all decisions regardless of publication status are publicly available through court websites and legal research databases. Former Ninth Circuit Court of Appeals Chief Judge Alex Kozinski,[68] for example, claims that "of course, all dispositive rulings, whether designated for inclusion in an official reporter or not, are [now] widely available online through Westlaw and Lexis, as well as in hard copy in West's Federal Appendix."[69]

Peter W. Martin, *District Court Opinions that Remain Hidden Despite a Long-Standing Congressional Mandate of Transparency—The Result of Judicial Autonomy and Systemic Indifference*, 110 L. LIBR. J. 305, 313 (2018); Elizabeth Y. McCuskey, *Submerged Precedent*, 16 NEV. L.J. 515, 516 (2016).

[68] Judge Kozinski resigned from the federal judiciary in 2017 after several women claimed he had sexually harassed them. *See* Matt Zapotosky, *Judge Who Quit over Harassment Allegations Reemerges, Dismaying Those Who Accused Him*, WASH. POST (July 24, 2018), https://perma.cc/CZ26-BCHK.

[69] *Unpublished Judicial Opinions: Hearing Before the Subcomm. on Courts, the Internet & Intellectual Property of the H. Comm. on the Judiciary*, 107th Cong. 30 (2002) (statement of Judge Alex Kozinski); *see also* Richard B. Cappalli, *The Common Law's Case Against Non-Precedential Opinions*, S. CAL. L. REV. 755, 756 (2003) ("These [nonprecedential] opinions were once called 'unpublished' and were distributed only to the parties to

But scholars have started to question that claim and some have uncovered evidence of a substantial body of decisions that are unreported and may be inaccessible. These are the decisions we refer to as *judicial dark matter*. Focusing on the First Circuit and comparing docket records from PACER to decisions available on the Lexis online database, Professor Merritt McAlister presents evidence of a substantial number of unreported or "missing" decisions.[70] As Professors Zachary Clopton and Aziz Huq explain, "PACER is a government-run database that charges per page for access to court records," technically enabling "access to almost any specific item of judicial data."[71] However, they observe, PACER's "kludge-cluttered and time-consuming interface mean that, in practice, public access is tightly constrained and relatively costly."[72] Professors Michael Kagan, Rebecca Gill, and Fatma Marouf, focusing on immigration cases, likewise find evidence that surprisingly many decisions are unavailable on Lexis and Westlaw; they refer to these decisions as "invisible," observing that "nominally unpublished merits cases actually come in two varieties: those that are actually available on standard research databases, and those that are not."[73] No prior studies have, as far as we are aware, explored the relationship between judge attributes and missing decisions. But the existing research sets the stage for this kind of study.

Circuit rules provide that panels may issue a decision without an opinion or even memorandum, although the rules do not explicitly state that such decisions may be unreported and made available only to the parties.[74] The U.S. Courts official website

---

the appeal, but they are now widely available through online databases and through the Federal Appendix."); Morgan Hazelton, Rachael K. Hinkle & Jee Seon Jeon, *Sound the Alarm? Judicial Decisions Regarding Publication and Dissent*, 44 AM. POL. RSCH. 649, 652 (2016) [hereinafter Hazelton et al., *Judicial Decisions*] ("In the modern era, electronic publishing has allowed nearly all decisions to be accessible, but the practice of marking decisions as unpublished is still widespread."); Andrew T. Solomon, *Making Unpublished Opinions Precedential: A Recipe for Ethical Problems & Legal Malpractice?*, 26 MISS. COLL. L. REV. 185, 188–89 (2007) ("By 2005, . . . the full-text of nearly every federal appellate opinion was published in either the Federal Appendix or the Federal Reporter.").

70 *See* McAlister, *supra* note 9, at 1126–32.

71 Clopton & Huq, *supra* note 46, at 7.

72 *Id.*

73 Kagan, Gill & Marouf, *supra* note 9, at 688.

74 *See, e.g.*, U.S. CT. OF APPEALS FOR THE FIFTH CIR., RULES AND INTERNAL OPERATING PROCEDURES 43 (2024); U.S. CT. OF APPEALS FOR THE D.C. CIR., HANDBOOK OF PRACTICE AND INTERNAL PROCEDURES 54 (2021) (listing "[f]our possible forms for disposing of cases that have been considered by a merits panel," including "a judgment or order without memorandum").

provides statistics on the number of cases terminated in the courts of appeals per year, and these numbers are far greater than the number of decisions reported. For example, in 2009 and 2010, 61,024 and 58,319 cases were terminated, respectively.[75] However, our data indicates that 25,806 and 24,037 decisions were reported in those years. Previous empirical studies, then, together with circuit rules and statistics on case terminations in the federal appellate courts, suggest that the possible scope for our dark matter theory is considerable. Judicial discretion over whether to report a decision could result in systematic imbalances in panel composition in terms of gender or race in the population of reported decisions. A substantial proportion of cases are likely disposed of without a reported decision. If the gender or race of judges affects the likelihood that a panel's decision will be reported, then we should expect to see biased panel compositions.

If, on the other hand, the reporting decision is just a matter of how important or difficult the case is, or some other case variables, then reporting decisions would not create imbalances in the panels represented in reported decisions. While there are rules governing publication decisions (even if these rules are vague and seem to leave considerable room for judicial discretion), as far as we can tell no such rules or procedures exist to guide decision-making about whether to report a case.[76] It should perhaps not be surprising, then, if the likelihood that a decision is reported depends on attributes of the judges assigned to the case.[77]

Given that decisions designated as "not for publication" but reported in the Federal Appendix are often accompanied by reasoned opinions and often cited, we might question if there remains a real distinction between officially published cases on the one hand and unpublished cases on the other. The more critical distinction might be between reported, available decisions on the one side and unavailable decisions—the judicial dark matter—on the other. This critical body of federal appellate decisions has

---

[75] *Table B-1. U.S. Courts of Appeals—Appeals Filed, Terminated, and Pending, by Circuit, During the 12-Month Period Ending December 31, 2009*, U.S. CTS., https://perma.cc/HWG6-SHBL; *Table B-1. U.S. Courts of Appeals—Appeals Filed, Terminated, and Pending, by Circuit, During the 12-Month Period Ending December 31, 2010*, U.S. CTS., https://perma.cc/TE4L-HFGV.

[76] *See infra* Part III.B.1.

[77] If administrators or staff members make reporting decisions before cases are assigned to panels, however, then we should not expect decision reporting to be associated with judge attributes (assuming more or less random assignment of cases to panels).

flown largely under the radar and warrants greater scholarly attention. McAlister emphasizes the pressing "need to examine what the missing decisions themselves say about how the appellate system administers justice."[78]

Judicial dark matter poses serious challenges for empirical study. Others have sought to study unreported decisions by comparing numbers of dispositions reported in the statistics released by the Administrative Office of the U.S. Courts and docket information on PACER to decisions available on legal research databases.[79] Here we take a different and novel approach, exploring judicial dark matter indirectly. By examining the compositions of panels in reported cases, we can make inferences about the types of panels and cases that are represented in the population of unreported cases.

To summarize this Part, the gender and racial composition of the panels of reported cases has a critical relationship to gender and racial equality on courts. If we care about the participation and influence of women and people of color in the judiciary, we should be concerned not only about increasing the number of judges from these groups, but also about the number and types of cases they appear in and how judges are distributed, in terms of race and gender, across reported decisions. Because there may be some nonrandomness in the assignment of cases to judges and, more critically, judges may exercise considerable discretion over whether to issue a reportable decision in a case, it cannot be assumed that judicial panels in the body of reported decisions will be neutral in terms of gender and race composition.

II. EMPIRICAL STUDY

In this Part, we present a series of empirical analyses designed to test, from multiple angles, the extent to which the panels of reported cases are gender- and race-neutral.

A. Data and Empirical Strategy

We rely on a curated dataset that has been made publicly available by Harvard Law School's Caselaw Access Project.[80] Our study period extends from 2001 through 2017, includes both published and unpublished decisions, and covers all the geographic

---

78 McAlister, *supra* note 9, at 1154–55.

79 *See supra* notes 70–73 and accompanying text.

80 Our data is described in more detail in Appendix Section A.

circuit courts. We excluded the Federal Circuit due to its unusual subject-specific jurisdiction. We downloaded over 350,000 decisions, and this constituted our primary dataset, which we augmented with demographic information that is available from the Federal Judicial Center (FJC). We focus on three-judge panels, dropping en banc panels as well as the rare panels that are made up of fewer than three judges. There are 355 court of appeals judges, 994 district court judges, and 10 Supreme Court Justices represented in our data. District court judges are included because they sometimes sit "by designation" on appellate panels as visiting judges; Supreme Court Justices, both active and retired, sometimes also sit on appellate panels.

As mentioned in the previous Part, empirically studying a phenomenon that is not directly observed presents challenges. But by identifying patterns in phenomena that are directly observed, it is possible to draw inferences about phenomena that are not. This is the methodology used by cosmologists to infer the existence of dark matter in the universe: the distribution of observable matter, along with Einstein's equations of general relativity, lead to the prediction that there is a large amount of additional (unseen) matter of unknown composition.[81] For our study of panel composition, the body of reported (published and unpublished) decisions represent the observable phenomenon, and from patterns in those observations we are able to draw inferences about the data-generating processes that lead to both the observable and unobservable (i.e., both reported and unreported) cases.

Our analyses expand on those used in prior work to study unpublished decisions in the U.S. courts of appeals.[82] The basic strategy we use to study panel composition is to assume a null model, meaning that judge gender and race do not affect the construction of panels or the likelihood that panels will report their decisions. Based on this null model, we can expect our data to have certain statistical characteristics. If the data do not have these characteristics, we can infer that reality does not match the null model, and that gender and race are associated with either panel construction or case reporting.

---

[81] *See* RYDEN, *supra* note 46, at 23 (explaining that "[t]he standard method of detecting dark matter is by measuring its gravitational effect on luminous matter").

[82] *See* Carlson et al., *supra* note 11, at 234–36. Our methods and the description of them here borrow from this earlier work, but our data and broader research questions differ.

A simple analogy can help illustrate the idea. Under a null hypothesis of a fair coin and a random sample of observations, heads and tails are equally likely outcomes of coin tosses. With a sufficiently large number of observations, if the reported results suggest that one outcome is more likely than the other, then one can infer that either the coin is not fair or there is some bias in how the results are reported. Our empirical approach for investigating panel composition is based on this model.

To study the representation of individual judges by gender and race in reported cases, we can take a more basic approach: we test whether gender and race are associated with the number of cases in which a judge appears as a panelist per year. We conduct this test on the data overall, controlling for circuit and year (interacted). We also examine individual-judge representation at a disaggregated level, testing for gender- and race-based disparities within circuits, years, and circuit-years.

B. Empirical Results

In this Section, we summarize the results from a series of tests to determine whether the demographic composition of panels in the pool of reported cases is consistent with a null model in which a judge is equally likely to appear on a case panel regardless of their gender or race.[83] Overall, we find strong evidence that the pool of reported cases is inconsistent with that model, which suggests either nonrandomness in panel construction or case assignment, or imbalances in case reporting that track judge demographics. For the reasons discussed in Part I, we believe that disparities in case reporting are most likely responsible for our results, given the limited evidence of nonrandomness in the prior literature and the evidence to suggest that large numbers of decisions go unreported.[84]

1. Panel composition tests.

Here we test the distribution of panels in reported cases according to gender and race composition. For a binary judicial characteristic (such as the binary male/female gender in the FJC), there are four potential panel types. For judges that are typed as *A* or *B*, the panel types are all-*A* panels (*AAA*), two-*A* panels (*AAB*), two-*B* panels (*ABB*), and all-*B* panels (*BBB*). For a hypothetical circuit

---

[83] Interested readers can find more detail in Appendix Section B.1–2.

[84] *See supra* Part I.B.

with twelve judges, eight of whom are type *A*, and all of whom have an equally likely chance of being drawn for a panel, the expected distribution of panel types would be as follows[85]:

TABLE 1: ILLUSTRATION OF EXPECTED PANEL COMPOSITIONS

AAA	*AAB*	*ABB*	*BBB*
25%	51%	22%	2%

It bears noting that the use of three-judge panels can exacerbate the underrepresentation of underrepresented groups.[86] In this hypothetical circuit, type *A* judges make up two-thirds of the total, but panels in which type *A* judges are the majority (i.e., *AAA* and *AAB*) make up over three-quarters of the total.

In our primary composition test, for each circuit-year pair in our data (e.g., the Seventh Circuit in 2005) we construct an expected distribution of the four panel types, and then compare the observed distribution to the expected distribution. We include judges of all types (i.e., regular active judges, as well as chief, senior, and visiting judges). We first determined, based on the case data, the number of reported cases in which each judge appears. Based on those numbers, we then estimated how many cases with each gender and racial composition (e.g., two women) we should expect to see if panels were neutral with respect to judge gender and race.[87] There may be some variance from the expected distribution even if the panel construction process is completely random. We used a chi-square test to estimate the statistical probability that the observed distribution was created through a random process.

The results of this analysis indicate that panel composition with respect to both race and gender is skewed in a statistically

---

85 The probability of drawing an *AAA* panel is eight out of twelve for an *A* judge in the first slot, seven out of eleven for the second slot, and six out of ten for the third slot, collectively 8/12 x 7/11 x 6/10 ≈ 0.25. For *BBB* panels, the probability is 4/12 x 3/11 x 2/10 ≈ 0.02. For *AAB* panels, the probability is 3(8/12 x 7/11 x 4/10) ≈ 0.51, where the initial multiplier, 3, accounts for the various permutations of panels comprised of two *A* judges and one *B* judge (i.e., *AAB*, *ABA*, *BAA*). Likewise, for *ABB* panels, the probability is 3(8/12 x 4/11 x 3/10) ≈ 0.22.

86 *See* Farhang & Wawro, *supra* note 17, at 304–05 (observing that, even if we assume that panels are randomly constructed, "[t]he probability of drawing two minority judges on a three-judge panel is substantially lower than the proportion of minority judges in the pool" and that "compared with the federal district courts where each case is heard by a single judge, . . . the institution of the federal appellate panel has the potential to considerably dilute the translation of minority representation into doctrinal output representative of minority views where they differ from majority views").

87 For further detail, see Appendix Section B.1.

significant way.[88] In the aggregate, we find that there are about seventeen hundred fewer cases with two-woman panels than expected and three hundred fewer with two judges of color than expected.[89] At a disaggregated level, for gender we find panel composition bias in all twelve circuits and in all seventeen years, and for race in almost all circuits and years.[90]

The first panel composition test takes as given, based on the observed data, the number of reported cases in which each judge appears. One might wonder, however, how the panel compositions of observed cases compare to what we would expect to see if the judges in a given circuit and year participated in equal numbers of cases. To address this question, we conducted an additional panel composition test using a different approach. It would be reasonable to expect regular, active judges to serve on roughly the same number of cases per year, but not so for special-status judges, and so in this test we included only regular, active judges (dropping, for example, visiting and senior-status judges).[91]

We then compared the actual distribution of panel types (i.e., the number of reported cases with each panel type in our data) to the distribution we would expect to see if active judges in a given circuit and year participate in the same number of cases. The results of this test, like the first one, indicate that the panels of reported cases are neither gender- nor race-neutral. As in the first panel composition test, the results indicate that panels composed of two women are underrepresented among reported decisions. But, unlike the first test, this one indicates that panels with two judges of color are overrepresented.

A possible source of nonrandomness in the compositions of panels in reported cases is the "batching" of cases to particular panels. Although there seems to be no standard procedure across circuits and no set procedure codified in circuit operating rules, it is common knowledge that separate panels are not necessarily created for every case.[92] Instead, panels are created and then cases are assigned to them in batches. This process could create nonrandomness in the panels of reported cases, since the initial panel assignment in effect "sticks" across multiple cases. If, for

---

88 *See* Appendix Section B.1.

89 *See* Appendix Section B.1.

90 *See* Appendix Section B.1.

91 For further detail, see Appendix Section B.1.

92 *See* Chilton & Levy, *supra* note 49, at 41–42 (discussing in detail this process and the difficulty of identifying the procedures that courts use).

example, a two-woman panel is created to begin with, the probability of that two-woman panel reappearing would be set according to the number of cases each initial panel receives; judges would not be reshuffled between each case, which is what would have to happen to expect a random distribution of judges across all cases. To test the extent to which this kind of batching process might be responsible for our findings of nonrandomness, we conducted a simulation in which we randomly constructed panels of three judges within each circuit-year based on the numbers of cases in our data with three active judges and the attributes of those judges, and then assigned cases to panels in batches ranging from one to twenty cases.[93] We found that even the largest batch sizes only rarely generated nonrandomness to the same extent as observed in our data. While not definitive, this exercise indicates that even if every panel in every circuit is assigned twenty cases at a time, that practice would be highly unlikely to produce the degree of nonrandomness that we observe for some of the panel composition types.

As discussed above, there are different potential pathways that could account for the effects we identify.[94] It is possible that there are biases in how panels are constructed, such that (for example) panels with two female judges and one male judge are less likely to be created than other panel types. It is also possible that this panel type is just as likely to be created and hear cases, but less likely to issue reported decisions, and therefore less likely to show up in our data. Our data and empirical strategy do not allow us to identify which of these two causal mechanisms is primarily responsible for the effects we observe, and they are not mutually exclusive. However, as discussed above,[95] existing evidence indicates both that panel creation is a largely random, or quasi-random process, and that large numbers of decisions go unreported. Accordingly, we believe that disparities in decision reporting likely have a greater explanatory role here than nonrandomness in panel construction.

2. Representation tests.

To further test for disparities in representation in reported cases overall, we examine whether gender or race predicts the

---

93 For technical details, see Appendix Section B.1.

94 *See supra* Part I.B.

95 *See supra* Part I.B.

number of cases in which a judge appears per year, conditioning on circuit and year. For this purpose, we ran a regression analysis to predict the number of cases each judge hears in each year, based on demographic information about the judge. In one model, we include regular, active judges only; in a second model, we include all judges (except visiting ones), but control for special statuses (e.g., chief and senior).[96] We find a positive and statistically significant relationship between judge gender (male) and number of cases in both models, while judge race does not have a significant effect in either.[97] Our preferred model (in which we are able to include more of the data) indicates that male judges appear in about eighteen more cases per year than female ones.

We also examine our data at a disaggregated level to test whether race and gender are associated with over- or underrepresentation within each circuit-year pair (e.g., the Ninth Circuit in 2005).[98] In a substantial number of circuit-years (over 20% for both gender and race at $p < 0.05$), the numbers of reported cases in which male judges appear and in which white judges appear depart from the numbers that would be expected if representation were balanced—that is, not systematically associated with judge gender and race.[99]

Our empirical study examines the demographic makeup of the panels producing reported decisions from multiple angles. Each of our analyses indicates that judges are not randomly distributed across reported cases. Instead, judge race and gender are systematically associated with the race and gender of the rest of the panel. Judge gender and race are also associated with the number of cases a judge appears in; while the within-circuit-year variation evens out with respect to race, female gender overall is associated with underrepresentation. Our findings suggest that the demographic composition of judges represented in reported decisions does not reliably reflect the composition of judges on courts.

III. IMPLICATIONS AND PRESCRIPTIONS

In this Part we first discuss explanations for and normative implications of our findings. We then propose reforms to judicial

---

96 For details, see Appendix Section B.2.

97 *See* Appendix Section B.2.

98 *See* Appendix Section B.3.

99 *See* Appendix Section B.3.

administration, and refinements to the empirical study of judicial decision-making, responsive to the concerns our findings raise. Given the descriptive and exploratory nature of our study, the explanations and prescriptions we discuss are necessarily tentative. But our results do point to a pressing need to start a discussion of these issues.

A. Explanations and Implications

In this Section, we explore how decisions regarding case assignment and reporting could lead to demographic imbalances in panel composition, and we discuss why such imbalances might be cause for concern.

1. Distortions in representation.

a) Panel composition. Our analysis indicates that the demographic makeup of the judges on the panels that appear in available decisions is systematically biased. We do not see the kind of heterogeneity along the lines of gender and race that we would expect if one judge was equally likely to appear with another judge regardless of the gender and race of the judges. And our results show that the demographic heterogeneity of the panels we see in reported cases cannot be inferred from the demographic representation of judges serving on the federal courts.

In our primary panel composition test (which includes judges of all types, including senior and visiting),[100] we find that cases with two-woman panels are systematically underrepresented, as are cases with panels including two people of color (POC). The effect sizes are meaningful, especially for gender. The analysis indicates that, if the gender composition of panels were unbiased, we would expect to see 46,552 cases in our data with panels composed of two women. We instead see 44,879 such cases, 1,673 fewer than expected. Similarly, in the absence of bias, we would expect to see 25,877 cases with panels made up of two people of color, whereas we observed 25,564 such cases, 313 fewer than expected. On this analysis, cases with a two-judge majority of women or people of color appear less often than expected even after accounting for the descriptive underrepresentation of those groups on the federal courts. In our secondary analysis, which excludes special-status judges and compares the actual distribution

---

[100] *See* Appendix Section B.1.

of panels to the one we would expect to see if all judges participated in a roughly equal number of cases, we likewise find that two-woman panels are underrepresented, but here we find that panels with two judges of color are overrepresented. This suggests that the underrepresentation of that panel type identified in the first test may be a result of special-status judges rather than regular, active judges. In contrast, the underrepresentation of panels with two women appears to be more systemic. There are various possible explanations for this finding, and our data and empirical methods were not designed to identify the causal pathway. Future research using different methods, including qualitative ones, could help with that identification. Nevertheless, we explore some possible explanations for the finding here.

As other scholars have suggested, judges might engage in strategic publication decisions for policy or reputational reasons, whereby they favor nonpublication of decisions that do not align with their policy preferences or that they would rather not draw attention to.[101] Professor David Law explains how "some judges are prepared to acquiesce to decisions that run contrary to their ideological preferences if the case remains [unpublished], but may be driven to dissent if the majority insists on publication."[102] Judges might approach decision reporting in the same kind of strategic way, preferring that a decision go unreported if the judge disfavors that decision and wishes to limit its influence.[103]

Other research suggests that, in various contexts including the legal profession, women's competence and expertise are more likely to be questioned than men's.[104] This kind of disparity, and

---

[101] *See, e.g.*, Ryan W. Copus, *Statistical Precedent: Allocating Judicial Attention*, 73 VAND. L. REV. 605, 649–50 (2020); *see also* Ben Grunwald, *Strategic Publication*, 92 TUL. L. REV. 745, 766 (2018) (discussing this kind of strategic maneuvering in the context of decision publication); Hazelton et al., *Judicial Decisions*, *supra* note 69, at 653 (same); Hinkle, *supra* note 36, at 2; William L. Reynolds & William M. Richman, *The Non-Precedential Precedent—Limited Publication and No-Citation Rules in the United States Courts of Appeals*, 78 COLUM. L. REV. 1167, 1201 n.168 (1978) (same).

[102] David S. Law, *Strategic Judicial Lawmaking: Ideology, Publication, and Asylum Law in the Ninth Circuit*, 73 U. CIN. L. REV. 817, 820 (2004); *see also* Hinkle, *supra* note 36, at 2 ("A key way to compromise is to leave the opinion unpublished so that the judge unhappy with the result can be assured it will not shape the law of the circuit.").

[103] In a recent work, one of us studies the relationship between judge demographics and opinion publication and discusses explanations for gender and racial differences in publication that overlap with the explanations for differences in decision reporting that we explore here. *See generally* Varsava, *supra* note 36.

[104] *See* Deborah L. Rhode, *Diversity and Gender Equity in Legal Practice*, 82 U. CIN. L. REV. 871, 878 (2014) (arguing that "women, like minorities, often fail to receive the presumption of competence enjoyed by white men"); Vicki C. Jackson, *What Judges Can*

the effect it could have on self-confidence, might mean that a lone male judge is more inclined to push for nonreporting when they dislike the majority's decision or to threaten to write separately if a decision is reported. A lone male judge might also be more likely to persuade the group to dispose of such a case without reporting a decision. That would result in an underrepresentation of panels with two women.[105]

Further, if women are less comfortable with conflict or feel more of a sense of responsibility to uphold norms and appearances of collegiality, that could lead to an overrepresentation of observed panels with two male judges. Previous studies suggest that women are more conflict avoidant than men and have a greater "desire to be part of a unified whole."[106] A lone woman on a panel might be less likely to propose nonreporting or to threaten to write a separate opinion if a decision is reported. In contrast, a lone male judge might be more likely to offer to go along with the outcome preferred by his female copanelists in exchange for nonreporting, giving those judges a way to avoid conflict and an incentive not to report the decision.[107]

Learn from Gender Bias Task Force Studies, 81 JUDICATURE, no. 1, 1997, at 15, 20 (summarizing—as co-chair of the Special Committee on Gender of the D.C. Circuit Task Force on Gender, Race, and Ethnic Bias—the findings of task forces in various circuits that "suggest that colleagues and attorneys may evaluate female judges more harshly than male judges"); KARPOWITZ & MENDELBERG, *supra* note 34, at 53, 70–71 (observing that women "are much more likely than men to underrate their competence, qualifications, and achievement," and noting that "[e]xperimental studies show that men enjoy a higher status than women in discussions, unless the subject is commonly perceived to be a feminine one"); Moyer et al., *supra* note 6, at 454 (explaining how, "[w]hen tasks are masculine-typed, women are more likely than men to display inaccurately low self-perceptions of their performance . . . and to underestimate their intelligence relative to men" (citation omitted)); Afra Afsharipour & Matthew Jennejohn, *Gender and the Social Structure of Exclusion in U.S. Corporate Law*, 90 U. CHI. L. REV. 1819, 1843–44 (2023) (studying gender inequality among corporate attorneys and observing that "[w]omen face biases about their commitment and availability to work, as well as about their competence and ability to successfully develop business, thus hindering their path to professional advancement").

105 Women receive lower scores in judicial performance evaluations, even after controlling for indicators of judicial competence like education, disciplinary history, experience, and reversal rates, suggesting systemic biases against women in the judiciary. Rebecca D. Gill, Sylvia R. Lazos & Mallory M. Waters, *Are Judicial Performance Evaluations Fair to Women and Minorities? A Cautionary Tale from Clark County, Nevada*, 45 L. & SOC'Y REV. 731, 749–50 (2011); Susan Brodie Haire, *Rating the Ratings of the American Bar Association Standing Committee on Federal Judiciary*, 22 JUST. SYS. J. 1, 8 (2001); Rebecca D. Gill, *Implicit Bias in Judicial Performance Evaluations: We Must Do Better Than This*, 35 JUST. SYS. J. 271, 278 (2014); Maya Sen, *How Judicial Qualification Ratings May Disadvantage Minority and Female Candidates*, 2 J.L. & CTS. 33, 34 (2014).

106 KARPOWITZ & MENDELBERG, *supra* note 34, at 65, 68.

107 On the tendency and value of suppressing judicial disagreement, see MORGAN HAZELTON, RACHAEL HINKLE & MICHAEL NELSON, THE ELEVATOR EFFECT: CONTACT AND

As other commentators point out, norms of consensus and collegiality shape decision-making at the federal appellate courts.[108] In a recent book, Professors Morgan Hazelton, Rachael Hinkle, and Michael Nelson argue that "collegiality affects nearly every aspect of judicial behavior," where collegiality is defined "as behavior by individuals that is intended to maintain relationships with colleagues" and "to make interpersonal relationships better."[109] Public expressions of disagreement in the form of concurring or dissenting opinions may harm the judiciary's sociological legitimacy and may also damage the judges' relationships with one another.[110] Women might be more inclined to avoid causing that kind of disruption, and they might also face greater adverse consequences if they introduce friction and disrupt collegial decision-making.[111] These kinds of gender differences could result

---

COLLEGIALITY IN THE AMERICAN JUDICIARY 81–82 (2023) [hereinafter HAZELTON ET AL., ELEVATOR EFFECT].

[108] As Hazelton and her coauthors note, "[m]any scholars have observed that there is a strong norm of dissent avoidance in the circuit courts." *Id.* at 10 (citing Lee Epstein, William M. Landes & Richard A. Posner, *Why (and When) Judges Dissent: A Theoretical and Empirical Analysis*, 3 J. LEGAL ANALYSIS 101 (2011); RICHARD A. POSNER, HOW JUDGES THINK (2008); and Fischman, *supra* note 42). Hazelton and her coauthors argue that, given this norm, "when considering one's vote, it is strategic for a circuit judge to consider that a dissent could rock the boat. Dissents not only have the potential to undermine the legitimacy of the court, they also often cause more work for the majority." *Id.* (citing BARNES BOWIE ET AL., *supra* note 48); *see also* Farhang & Wawro, *supra* note 17, at 306 (observing that "extremely high rates of consensus on federal appellate panels prevail even within particularly contentious issue areas, where measures of individual judges' voting and measures of panel outcomes show wide ideological variation"); Haire et al., *supra* note 30, at 307 (explaining how decision-making at the federal appellate courts "has been characterized as decision 'by committee' as opposed to the more 'autocratic' decision-making role of the district court judge" (citation omitted)); Larsen & Devins, *supra* note 47, at 1321 (arguing that "[f]ederal appeals judges should try to preserve the consensus-driven decision-making model that is the hallmark of their courts"); Edwards, *supra* note 41, at 1656.

[109] HAZELTON ET AL., ELEVATOR EFFECT, *supra* note 107, at 5, 18 (citing Helen V. Collier, *Collegiality Among Judges: No More High Noons*, 31 JUDGES J. 4 (1992)).

[110] *Id.* at 51–52; *see also* Larsen & Devins, *supra* note 47, at 1349 (pointing out that "[d]issents and concurrences take time and, as such, present challenges to court administration as well as collegial decision-making").

[111] *See* COMM'N ON WOMEN IN THE PRO. & MINORITY CORP. COUNS. ASS'N, YOU CAN'T CHANGE WHAT YOU CAN'T SEE: INTERRUPTING RACIAL AND GENDER BIAS IN THE LEGAL PROFESSION, EXECUTIVE SUMMARY 8 (2018) ("Women of all races reported pressure to behave in feminine ways, including backlash for masculine behaviors."); Joan C. Williams & Veta T. Richardson, *New Millennium, Same Glass Ceiling? The Impact of Law Firm Compensation Systems on Women*, 62 HASTINGS L.J. 597, 637 (2011) ("Studies show that women are often faulted for lacking in collegiality or for having personality problems for behavior that, in a man, is seen [in a positive light]."). As Hazelton and her coauthors observe in their recent book on collegiality and judicial behavior, "[s]tudies suggest that demographic characteristics combine with institutional structures and interpersonal norms in ways that differentially affect the ability of women and nonwhite participants to

in disproportionately high numbers of two-man panels in the population of reported decisions.

Although our findings are somewhat mixed, they raise a number of concerns, especially our finding that two-woman panels are underrepresented (which we observed in both of our analytical approaches). First, if two female judges are more likely to hold similar views about a case than a male and female judge, then the apparent underrepresentation of two-woman panels in reported cases indicates that the voting power of women in reported decisions is diluted overall.

Second, as Karpowitz and Mendelberg report, "[s]cholarship on group composition suggests that gender composition has a variety of powerful effects on the group decision and on individual attitudes."[112] Judicial deliberations in cases are for the most part private: we cannot observe the conferences during which judges discuss how cases should be decided, nor the exchanges among judges in the process of producing judicial opinions and issuing final decisions.[113] But group deliberations are observable in other decision-making contexts that may be analogous to the judicial one in key ways. Karpowitz and Mendelberg find that when women are in the numerical majority in a group, women's average participation in group discussion is almost as high as men's, but women's participation decreases substantially when they are in the numerical minority.[114] Men, however, are not affected by the

---

have equal footing in deliberative environments"; they note that, "[a]pplied to our findings, this research would suggest that the effects of collegiality vary according to the sociodemographic characteristics of judges." HAZELTON ET AL., ELEVATOR EFFECT, *supra* note 107, at 242. But they did not study the relationship between these demographic attributes and collegiality in any detail, and the authors encourage future research on the topic. *See id.* Other recent research on federal appellate decisions does not find evidence to suggest that women are less likely to write separate opinions than men. *See generally* Varsava, *supra* note 36. Dissent and concurrences rates, however, might be associated with the gender of copanelists. We are not aware of existing research on this topic.

112 KARPOWITZ & MENDELBERG, *supra* note 34, at 70.

113 We can observe oral arguments, however. Studying oral arguments at the U.S. Supreme Court, Professor Tonja Jacobi and now-attorney Dylan Schweers find that female Justices speak less than male Justices, that female Justices are more likely to be interrupted by attorneys as well as fellow Justices, and that female Justices are less likely to interrupt other Justices. *See* Tonja Jacobi & Dylan Schweers, *Justice, Interrupted: The Effect of Gender, Ideology, and Seniority at Supreme Court Oral Arguments*, 103 VA. L. REV. 1379, 1483 (2017) (concluding that "it follows from this pattern of interruptions that there is a marked difference in the relative degree of influence of the women and the men on the Court"); *id.* at 1398.

114 KARPOWITZ & MENDELBERG, *supra* note 34, at 73, 87.

gender of other group members in this way.[115] Numerical-minority status, then, seems to dilute the voices of women but not of men. The authors suggest that this is because "when men are women's predominant conversation partners, . . . [women are] expected to behave in less assertive and more accommodating ways than they would if surrounded by many women."[116] A female judge might have less influence in group decision-making when she is the lone woman on a panel than she would if she was on a panel with another woman.

A new study by Professor Alma Cohen of the relationship between the political composition of appellate panels and decision outcomes finds that adding one Republican-appointed judge to a panel of Democrat-appointed judges makes more of a difference to the outcome than does adding one Democrat-appointed judge to a panel of Republican-appointed judges.[117] Cohen argues that this finding calls into question proposals to mandate that there be at least one Democrat appointee and one Republican appointee on every panel, since the effect will be to skew outcomes more "toward those associated with Republican judges."[118] While it is unclear whether the same kind of pattern holds for gender and race (and Cohen appears not to have tested this), given the history of gender and racial power imbalances, we have reason to expect a similar pattern—and perhaps even a more pronounced one than for political affiliation. So, although an overrepresentation of one-woman panels does increase panel diversity in one sense, such that women are dispersed across more panels, two-woman panels might be especially important for substantive representation purposes.

Further, Karpowitz and Mendelberg find that women are especially disadvantaged in terms of group influence when they are

---

115 *Id.*; *see also* Haire et al., *supra* note 30, at 306 (reporting that "research on diversity and group dynamics emphasizes inequality between group members and the importance of relative status," and explaining that "[g]iven the dominance of white males in positions of power in political and legal institutions, . . . the influence of a single female or minority may be limited because norms of interaction within the group are defined by the [dominant] majority").

116 KARPOWITZ & MENDELBERG, *supra* note 34, at 73.

117 Cohen, *supra* note 52, at 51–52.

118 *Id.* at 52. The political diversity idea was first proposed by Professors Emerson Tiller and Frank Cross. *See generally* Emerson H. Tiller & Frank B. Cross, *A Modest Proposal for Improving American Justice*, 99 COLUM. L. REV. 215 (1999). Professors Cass Sunstein and Thomas Miles discuss the proposal and some problems with it. *See* Cass Sunstein & Thomas Miles, *Depoliticizing Administrative Law*, 58 DUKE L.J. 2193, 2227–28 (2009).

"in the gender minority *under majority rule*."[119] In that situation, "deliberation is a negative experience [for women] in which their speech is interrupted in a dismissive manner and their words rarely affirmed."[120] "The key," Karpowitz and Mendelberg emphasize, "is that the factors of numbers and institutional norms or rules jointly shape the status of deliberators within the group."[121] The authors find that highly educated women are not immune from this effect of group representation, but to the contrary are actually most affected by it.[122] This finding suggests that we might reasonably expect gender power imbalances to persist among judges, despite their high levels of education.[123]

Given that the federal courts of appeals decide cases by majority rule, female judges might be more likely to participate in deliberations and influence decision-making on par with men when women are in the panel majority. Panels with two women, then, might be especially important for reducing gender disparities in legal impact.[124] Karpowitz and Mendelberg recommend that deliberating bodies with a minority of women adopt a unanimity decision rule, suggesting that this would reduce gender disparities in participation and improve deliberation.[125] Majority vote is well established as the decision rule for the courts of appeals, however, and that rule is unlikely to change. And besides, a majority-vote decision rule may be necessary in the adjudicative context, given the limited amount of time and resources to dispose of cases. However, as we discuss in Part III.B below, more modest reforms could help enhance the voices of women and other groups on appellate panels.

---

119 KARPOWITZ & MENDELBERG, *supra* note 34, at 231 (emphasis added); *see also id.* at 251.

120 *Id.* This idea is supported by Jacobi and Schweer's study of interruptions of Supreme Court Justices. *See generally* Jacobi & Schweers, *supra* note 113.

121 KARPOWITZ & MENDELBERG, *supra* note 34, at 18.

122 *See id.* at 328.

123 Others have suggested that we should not expect to see significant gender differences in judicial behavior because of the similar backgrounds of judges. *See, e.g.*, Rebecca D. Gill, Michael Kagan & Fatma Marouf, *The Impact of Maleness on Judicial Decision Making: Masculinity, Chivalry, and Immigration Appeals*, 7 POL. GRPS. & IDENTITIES 509, 511 (2017) (describing this point of view as "the organizational account," but not endorsing it).

124 Exploring another way in which judges from traditionally marginalized groups may have less opportunity to exert influence on the law than other judges, Professor Maya Sen finds that Black district court judges are more likely to be reversed on appeal than their white counterparts, even after controlling for various factors that we might expect to affect reversal rates. *See* Sen, *supra* note 28, at S221.

125 KARPOWITZ & MENDELBERG, *supra* note 34, at 250–51, 270–71.

Third, even if we set aside the matter of gender and racial equality on the bench, we have further reason to care about the composition of appellate panels. This is because diversity in decision-making bodies is associated with improved deliberation and better decisions.[126] As Professors Matthew Spitzer and Eric Talley suggest, people "may become more extreme when interacting with like minded counterparts."[127] Haire, Moyer, and Treier similarly explain that "[w]hen a group is composed of like-minded individuals, members tend to focus on shared information and overlook issues, easily reaching consensus on a position that potentially fails to identify errors and reflects a more extreme position than suggested by members' individual thinking."[128] In contrast, diversity in terms of background and experience can enhance information sharing and improve deliberation.[129]

Focusing on ideological diversity on federal appellate panels, proxied by political party, Spitzer and Talley use a game theoretic model to show that "ideologically heterogeneous panels are more likely to incentivize broad information production than are homogenous ones": "mixed panels produce more information, which—through deliberation—brings about more informed decisions."[130] They conclude that "mixed panels produce not only different results but also better results than their homogenous counterparts."[131] Assuming that some degree of substantive representation holds for women and people of color on courts, gender and racial diversity on panels might, like ideological diversity, lead to better deliberations and decisions. For this reason, our

---

126 *See* Matthew L. Spitzer & Eric L. Talley, *Left, Right, and Center: Strategic Information Acquisition and Diversity in Judicial Panels*, 29 J.L. ECON. & ORG. 638, 639–44 (2013). *See generally* Haire et al., *supra* note 30.

127 Spitzer & Talley, *supra* note 126, at 640.

128 Haire et al., *supra* note 30, at 305.

129 *Id.* at 306.

130 Spitzer & Talley, *supra* note 126, at 639, 670.

131 *Id.* at 672; *see also* Cass R. Sunstein, David Schkade & Lisa Michelle Ellman, *Ideological Voting on Federal Courts of Appeals: A Preliminary Investigation*, 90 VA. L. REV. 301, 348 (2004) (presenting empirical evidence to suggest that "[t]he existence of [political] diversity on a panel is likely to . . . move the panel's decision in the direction of what the law requires" and "the existence of politically diverse judges and a potential dissent increases the probability that the law will be followed"); Frank B. Cross & Emerson H. Tiller, *Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing on the Federal Courts of Appeals*, 107 YALE L.J. 2155, 2176 (1998) (finding evidence to suggest that "[w]hile a partisan split panel does not negate all partisan influences on *Chevron* review, it clearly moderates such influences and makes doctrine more likely to be followed").

finding that panels with three men are underrepresented in reported cases, and those with one woman overrepresented, may be seen as salutary.[132]

If a woman or a person of color is likely to contribute more to group deliberation when the judge is not in the gender or racial minority, however, then panels with two women and panels with two people of color might be especially important for realizing the advantages of heterogeneity in decision panels. Along the same lines, Karpowitz and Mendelson suggest that "conditions that most empower women—majority rule with many women—also produce a good deliberative exchange": "a fuller, more robust engagement of deliberators with each other," which makes for "more informed choices."[133] For the sake of information sharing, it may be better if judges were distributed such that we have more panels with two women or two people of color.

The findings of Karpowitz and Mendelson might help explain why scholars have found mixed support for the idea that diversity in judicial panels, in terms of gender and race, improves the quality of judicial opinions.[134] Haire, Moyer, and Treier examined the relationship between the number of "nontraditional" judges (women or racial minorities) on a panel and the quality of deliberative output, measured by the number of legal issues discussed in the majority opinion for the case. The number of issues did not change with the presence of a single woman or person of color on the panel; "however, when the demographic majority shifted [to a majority of nontraditional judges], issue coverage increased."[135] The authors note that their findings are consistent with other research showing that individual women have less influence in minority-woman groups than majority-woman ones.[136] And the authors conclude that, "[t]o the extent that diversity on the panel shaped deliberative outputs, it appeared to do so only when the

---

[132] Both composition tests reveal this pattern, with the one-woman overrepresentation statistically significant in both tests and the three-man underrepresentation significant in Test I only. *See* Appendix Section B.1.

[133] KARPOWITZ & MENDELBERG, *supra* note 34, at 272. This idea is consistent with research on gender and corporate boardroom dynamics finding that having three or more women directors on the board made for a "culture change" that "improves the board's overall performance"; the researchers suggest that women directors become more active when more women directors join the board. Alison M. Konrad & Vicki W. Kramer, *How Many Women Do Boards Need?*, HARV. BUS. REV. (Dec. 2006), https://perma.cc/S3YD-WE2Q.

[134] Haire et al., *supra* note 30, at 304.

[135] *Id.* at 310–15.

[136] *Id.* at 308.

majority of judges are women or minorities."[137] Panels with two women or two people of color, then, might be important not only for reducing gender- and race-based disparities in legal influence but also for the quality of adjudication.

b) Case representation. Our representation analysis indicates that female judges appear in fewer reported cases per year than male judges. The analysis does not reveal statistically significant racial differences in this regard. Our results suggest that we should expect a female judge to appear in about eleven or eighteen fewer cases than a male judge in a given year (depending on the model specification).[138] This does not mean that female judges necessarily participate in fewer cases than male judges, but it does indicate that they participate in fewer reported cases, the type most likely to make a legal difference beyond the parties to the case. Whether this effect is a result of imbalances in case assignment or in reporting, an implication is that female judges are less visible and may shape the law and understandings of it less than male judges.[139]

There are a few (not mutually exclusive) possible explanations for this finding. First, less important cases might be more likely to be assigned to women, such that their cases are less likely to warrant reporting. Relatedly, judges may be less likely to perceive decisions produced by women as worthy of reporting, because the work of female judges may be unfairly undervalued in relation to the work of male judges.[140]

---

137 *Id.* at 315.

138 *See* Appendix Section B.2.

139 It is of course possible that the additional unreported decisions issued by men are trivial, which would mean that the gender disparity may not actually be substantively interesting or concerning. Our study design does not allow us to rule out this possibility.

140 *See* Christensen et al., *supra* note 41, at 627 (pointing out that "status stereotypes often result in diminished expectations of competence for minority-classed groups," which in the judicial profession includes women and people of color); MEERA DEO, UNEQUAL PROFESSION: RACE AND GENDER IN LEGAL ACADEMIA 111 (2019) (finding, through a major qualitative study of law professors, that both women and people of color, and especially women of color, confront presumptions of incompetence in their professional roles); Jordana Goodman, *Ms. Attribution: How Authorship Credit Contributes to the Gender Gap*, 25 YALE J.L. & TECH. 309, 328 (2023) (suggesting that, in the context of patent law practice, "[w]omen—and especially women of color—are generally required to provide more evidence of competence than their male peers, which means they may need to have better work product for a longer period of time to be recognized as an author on the final document"); Monica C. Schneider & Angela L. Bos, *Measuring Stereotypes of Female Politicians*, 35 POL. PSYCH. 245, 259 (2013) (finding that "female politicians score significantly lower than [males] on leadership and competence, two characteristics central to being a successful politician"); *see also supra* note 105 and accompanying text.

Second, female judges may be subjected to heightened standards and scrutiny, and they might be more inclined to issue unreported decisions as a result.[141] Others have pointed out that unpublished decisions allow judges "to hide outside the public glare."[142] That may be an overstatement, because even unpublished decisions are broadly accessible. They are also citable by litigants.[143] Unreported decisions, however, do provide a safe space to avoid scrutiny. And women might be more likely to favor nonreporting for that reason.[144]

Third, women might be compelled to issue higher numbers of unreported decisions because they are under pressure to put more

---

[141] *See* COMM'N ON WOMEN IN THE PRO. & MINORITY CORP. COUNS. ASS'N, *supra* note 111, at 7 ("Women of color, white women, and men of color reported that they have to go 'above and beyond' to get the same recognition and respect as their colleagues."); Moyer et al., *supra* note 6, at 453 (reporting that, "in fields dominated by a particular group, individuals who do not fit into the profession's stereotype may adopt perfectionistic tendencies and set exceptionally high standards for their work, in order to demonstrate they have legitimately earned their position"); Robert L. Nelson, Ioana Sendroiu, Ronit Dinovitzer & Meghan Dawe, *Perceiving Discrimination: Race, Gender, and Sexual Orientation in the Legal Workplace*, 44 L. & SOC. INQUIRY 1051, 1052 (2019) ("Gender and racial stereotypes afford individual members of privileged gender or ethno-racial groups the presumption of competence while women and racial minorities are held to a higher standard than their white male counterparts."). Historically underrepresented groups in other professions face the same kind of challenge. *See, e.g.*, Sylvia Maxfield, Mary Shapiro, Vipin Gupta & Susan Hass, *Gender and Risk: Women Risk Taking and Risk Aversion*, 25 GENDER MGMT. 586, 594–95 (2010) (noting that women CEOs may "want to manage their hyper-visibility, the phenomena of incurring intense scrutiny as a result of being in the minority," and explaining how "any risky decision undertaken by one of those few women is subject to increased analyses and opinion"). If women are more risk averse than men, that is another reason why they might be more likely to favor nonreporting, since an unreported decision is much less likely to have far-reaching negative consequences than a reported one that affects future cases. *Cf.* Paola Sapienza, Luigi Zingalesb & Dario Maestripieric, *Gender Differences in Financial Risk Aversion and Career Choices Are Affected by Testosterone*, 106 PNAS 15268, 15268 (2009) (finding that "[w]omen are, on average, more risk averse than men in financial decision-making").

[142] Boyce F. Martin, Jr., *In Defense of Unpublished Opinions*, 60 OHIO ST. L.J. 177, 180 (1999).

[143] *See supra* note 64 and accompanying text.

[144] A related possibility is that decisions not to report are sometimes a subtle means of exercising power, since judges might hide controversial decisions through nonreporting. Others have suggested that judges might use nonpublication in this way. *See* Elizabeth Earle Beske, *Rethinking the Nonprecedential Opinion*, 65 UCLA L. REV. 808, 822 (2018) (observing that "it is beyond question that there are [unpublished] opinions that state new legal principles or seemingly draw controversial conclusions" and so should be published given publication rules); Rachael K. Hinkle, *Publication and Strategy in the U.S. Court of Appeals*, 179 J. INST. & THEORETICAL ECON. 121, 137 (2023) (arguing that "the lack of accountability [in unpublished decisions] creates an opportunity for panels to use nonpublication to shield a potentially vulnerable ruling from review and reversal").

work into and spend more time on reported ones than men.[145] Professor Laura Moyer and her coauthors argue that "high-achieving women and people of color in the legal profession have adopted strategies for dealing with pressure and societal expectations and that they carry these strategies with them onto the bench."[146] And they find evidence to suggest that women and people of color put greater effort into their written opinions than other judges.[147] Women and people of color might also put greater effort into decisions that are reported; some judges might make up for that extra work by issuing more unreported decisions.

2. Data bias in empirical studies.

There is a substantial body of empirical literature, spanning decades, that attempts to study the effects of judge characteristics on judicial decision-making.[148] Much of this work examines the relationship between judge attributes and decision outcomes.[149] The standard empirical approach hypothesizes a relationship between a judicial attribute (such as liberal ideology) and case outcomes (such as prodefendant decisions in criminal cases). Correlations between the attribute and the outcome are interpreted to imply a causal relationship, such that the assignment of a judge with some attribute to a case affects the

---

[145] *See* Moyer et al., *supra* note 6, at 453 (suggesting that prior scholarship "raises the possibility that, whether consciously or subconsciously, judges who are people of color and/or women are likely to feel pressure to work harder than white men to explain and to justify their decisions to relevant audiences like co-panelists, litigants, and even the Supreme Court . . . to be persuasive"). It is also possible that some judges aim at (and may succeed in) maximizing the impact of their reported decisions by issuing higher proportions of unreported ones into which they can put less effort, thus saving more time and energy for the reported ones.

[146] *Id.* at 464.

[147] *Id.* ("[T]he evidence suggests that female and minority appellate judges are more likely to overprepare in achievement-related tasks like writing majority opinions, spending more effort than their male colleagues to justify to key audiences (litigants, other judges, and the Supreme Court) that their decision is legally correct.").

[148] *See* Mark S. Hurwitz & Ashlyn Kuersten, *Changes in the Circuits: Exploring the Courts of Appeals Databases and the Federal Appellate Courts*, 96 JUDICATURE, no. 1, 2012, at 23 (providing a brief overview with a focus on Professor Donald Songer's dataset). *See generally* Rachlinski & Wistrich, *supra* note 10 (reviewing literature). For an early example, see generally Sheldon Goldman, *Voting Behavior on the United States Courts of Appeals, 1961–1964*, 60 AM. POL. SCI. REV. 374 (1966). For a recent example, see generally Robert S. Erikson, *Appellate Court Assignments as a Natural Experiment: Gender Panel Effects in Sex Discrimination Cases*, 19 J. EMPIRICAL L. STUD. 423 (2022).

[149] *See* LEE EPSTEIN, WILLIAM M. LANDES & RICHARD A. POSNER, THE BEHAVIOR OF FEDERAL JUDGES: A THEORETICAL AND EMPIRICAL STUDY OF RATIONAL CHOICE 153–206 (2013).

chances of a particular outcome. The most-studied attribute is probably judicial ideology, typically understood in terms of partisan affiliation.[150] But there is also considerable work that examines other factors, such as prior work experience[151] and demographic variables like race and gender.[152]

This literature widely rests on two assumptions. First, that cases are assigned to judges as if at random, meaning that judicial characteristics are uncorrelated with case characteristics in ways that affect the outcome of a decision. Second, that the mechanisms through which cases are selected to be observed by researchers are not confounded with the phenomenon being studied. Together, we call these the *attribute-to-outcome inference assumptions*. If these assumptions are not met, then studying observable cases can lead to biased estimates of relationships between judge characteristics and case outcomes.

Others have pointed out that the attribute-to-outcome inference assumptions are unwarranted when it comes to published cases, since judges exercise discretion over whether to issue a published or unpublished decision.[153] Since a judge's

---

[150] *See id.* at 65–100. *See generally* FRANK B. CROSS, DECISION MAKING IN THE U.S. COURTS OF APPEALS (2007); CASS R. SUNSTEIN, DAVID SCHKADE, LISA M. ELLMAN & ANDRES SAWICKI, ARE JUDGES POLITICAL?: AN EMPIRICAL ANALYSIS OF THE FEDERAL JUDICIARY (2006).

[151] *See generally* Stuart S. Nagel, *Judicial Backgrounds and Criminal Cases*, 53 J. CRIM. L., CRIMINOLOGY & POL. SCI. 333 (1962) (examining the effect of prior experience as a prosecutor on judicial decision-making); Rob Robinson, *Does Prosecutorial Experience "Balance Out" a Judge's Liberal Tendencies?*, 32 JUST. SYST. J. 143 (2011) (same).

[152] *See e.g.*, Christina L. Boyd, Lee Epstein & Andrew D. Martin, *Untangling the Causal Effects of Sex on Judging,* 54 AM. J. POL. SCI. 389, 392 (2010) (observing that the literature, as of 2010, already included "over thirty systematic, multivariate analyses of the extent to which female judges make decisions distinct from their male colleagues . . . or cause male judges to behave differently [in terms of outcomes voted for] than they otherwise would"). *See generally* Richard Fox & Robert Van Sickel, *Gender Dynamics and Judicial Behavior in Criminal Trial Courts: An Exploratory Study*, 21 JUST. SYS. J. 261 (2000) (examining gender); Jonathan P. Kastellec, *Racial Diversity and Judicial Influence on Appellate Courts*, 57 AM. J. POL. SCI. 167 (2013) (examining race).

[153] *See* Grunwald, *supra* note 101, at 775 (arguing "that scholars should not ignore unpublished decisions any longer"); Denise M. Keele, Robert W. Malmsheimer, Donald W. Floyd & Lianjun Zhang, *An Analysis of Ideological Effects in Published Versus Unpublished Judicial Opinions*, 6 J. EMPIRICAL LEGAL STUD. 213, 236 (2009) (arguing that "future scholars should not continue to ignore the rich source of information available in unpublished opinions when drawing conclusions regarding judicial behavior"); Deborah Jones Merritt & James J. Brudney, *Stalking Secret Law: What Predicts Publication in the United States Courts of Appeals*, 54 VAND. L. REV. 71, 116 (2001) ("Today, a scholar who studies only published opinions from the United States Courts of Appeals does so at his or her peril."); Law, *supra* note 102, at 5 (calling "into question the validity of existing empirical research on the courts of appeals, given the frequent tendency of such work to omit unpublished decisions from analysis"). *See generally* Harry T. Edwards & Michael A.

decision concerning publication may be influenced by the same characteristics that are hypothesized to affect case outcomes, correlations between outcomes and judicial characteristics in published decisions may be the result of judge effects on publication decisions, rather than on outcomes.

In recent work, three of us demonstrated that the pool of published decisions that most studies have relied on is biased along the axis of judge ideology.[154] These results call into question the validity of studies that rely on published decisions to make causal inferences about the relationship between judge attributes and case outcomes, which much of the prior work on judicial decision-making does.[155] The two most important datasets in the field—one constructed by political scientist Donald Songer and the other by legal scholar Cass Sunstein—both contain published decisions only.[156] For these datasets, which have been used in a substantial number of studies, there is no solid foundation of support for the attribute-to-outcome assumptions.

Empiricists sensitive to the problem of bias in published decisions have included both published and unpublished decisions in their analyses in an effort to avoid the issue.[157] This represents an improvement over prior approaches. The analyses reported in this Article, however, show that even this more inclusive strategy does not resolve the problem of data bias. Accordingly, our findings raise some doubts about causal claims based on analyses of the pool of available decisions appearing in the Federal Reporter and Federal Appendix, extending the insights of previous work on the data bias problem in published decisions.

Because judge attributes such as race and gender might affect whether a panel reports a decision at all, studies that look at the relationship between these judge attributes and case

---

Livermore, *Pitfalls of Empirical Studies that Attempt to Understand the Factors Affecting Appellate Decisionmaking*, 58 DUKE L.J. 1895 (2009); Fischman, *supra* note 42.

154 *See* Carlson et al., *supra* note 11, at 228.

155 *See id.*; Fischman, *supra* note 42, at 823 tbl.1 (listing fourteen prominent studies, nine of which rely on published opinions alone). Examples of studies that rely on published decisions exclusively include CROSS, *supra* note 150, and Kastellec, *supra* note 152.

156 *See generally* Ashlyn Kuersten & Donald Songer, *Presidential Success Through Appointments to the United States Courts of Appeals*, 24 JUST. SYS. J. 283 (describing and relying on the Songer data); Sunstein et al., *supra* note 131 (describing and relying on the Sunstein data).

157 *See generally* Fischman, *supra* note 42; Jennifer L. Peresie, *Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Appellate Courts*, 114 YALE L.J. 1759 (2005).

outcomes—based on available decisions—may miss an important part of the story. For example, it may appear that female judges are more likely to side with plaintiffs in sexual harassment cases,[158] when in fact they may just be more likely to report their decisions when they side with plaintiffs. In other words, decisions about whether to issue a reportable decision for a case might explain some of the relationships that others have observed between judicial attributes and case outcomes.

Even if judge race and gender have no effect on case outcomes, these factors might affect whether a panel issues a reportable decision in a case. While there is a growing awareness about this problem of data bias in the context of studies using published cases only, there has been relatively little recognition of the possibility that this problem persists in a different form in studies that use more comprehensive data. Our analysis suggests that the pool of available court of appeals decisions is biased according to judge race and gender. Because judges with certain attributes are more likely to appear in reported cases than others, and because panels of certain combinations of judge race and gender are more likely to occur than others, observable relationships between judge attributes and case outcomes may or may not be reflective of the actual relationships between the two. This is not to say, however, that studies testing relationships between judge attributes and case outcomes (or other case variables of interest) in reported cases are unimportant or unilluminating. Reported cases, after all, are the ones that most influence the law (even though they do not represent all cases that influence the lives of litigants). And so understanding the relationship between judge attributes and case outcomes in the universe of reported cases is still of interest, even if those cases do not represent the whole story.

B. Prescriptions

In this Section, we propose some possible reforms to judicial administration and refinements to the empirical study of judicial decision-making, aimed at addressing the concerns our study raises.

---

158 *See generally* Fischman, *supra* note 42; Peresie, *supra* note 157.

1. Reporting rules.

To the extent that judges are randomly assigned to panels and panels randomly assigned to cases, this randomization is not doing the kind of work that it is designed, and generally believed, to do. Either there is some kind of systematic interference in existing random-assignment practices along the lines of gender and race, or decisions about case reporting introduce gender- and race-based biases into the panel compositions of reported cases. More robust and transparent randomization procedures as well as randomization checks and accountability systems may help ensure that randomized assignment is operating as it should. Other scholars have proposed reforms to random-assignment rules to counteract biased panels. Professors Robert Brown and Alison Lee, for example, argue that "[n]eutrality requires a well-defined and automatic system for determining panel composition and case assignments" and suggest that the Federal Rules of Appellate Procedure could be revised "to require random assignment."[159] Even if panels are perfectly random, however, decisions regarding case reporting can lead to biases in the panels of reported cases. And so, improvements targeting panel construction and case assignment alone are unlikely to solve the problem of biases in panel composition. Indeed, others have tested random assignment in the courts of appeals and have found only modest evidence of incomplete randomization.[160]

Formal rules to govern administrative decisions concerning decision reporting might help mitigate gender and racial biases in the panels of reported decisions. Currently, there is no public accountability for these decisions, they are utterly nontransparent, and judges apparently have broad discretion over them. A new rule requiring panels to report all decisions, even if there is only a very short written opinion or none at all, might help counteract gender- and race-based disparities in reported decision panels. This reform should be relatively low cost, especially since the Federal Appendix has been discontinued and unpublished decisions appear only in online databases and on court websites.

Currently, the federal circuits have internal rules governing decision publication but no such rules for decision reporting. As far as we can tell, current circuit rules do not directly address unreported decisions. Nor do the Federal Rules of Appellate

---

159 Brown & Lee, *supra* note 51, at 1104, 1108.

160 *See supra* Part I.B.1.

Procedure. So, another possibility would be for the circuits to issue such rules—or perhaps, better yet, the Federal Rules of Appellate Procedure could be revised to include them—which might lead to less judicial discretion and more consistency in case reporting. The rules concerning decision publication, however, are vague, and would seem to allow for considerable discretion and inconsistent application across judges.[161] Efforts to revise appellate rules, then, should not model reporting rules on existing publication rules, but should rather develop concrete criteria to govern case reporting and publication.

2. Transparency.

In the current system, decisions regarding reporting are opaque, and unreported decisions themselves lack transparency. A requirement that all decisions are to be reported and included in legal research databases would increase transparency. Better yet would be a requirement that a panel not only report its decision but also include an explanation, however brief, for its disposition of the case. This kind of requirement would serve a monitoring and accountability function, and would serve values of procedural justice.[162]

Even if circuits maintain the current three-tier system of published, unpublished but reported, and unreported cases, they could increase transparency by keeping and presenting statistics on the numbers of cases every judge disposed of by decision type in each term or year.[163] This way, disparities among judges or groups of judges that might raise concerns and warrant interven-

---

[161] *See* Rachel Brown, Jade Ford, Sahrula Kubie, Katrin Marquez, Bennett Ostdiek & Abbe R. Gluck, *Is Unpublished Unequal? An Empirical Examination of the 87% Nonpublication Rate in Federal Appeals*, 107 CORNELL L. REV. 1, 23 (2021) (reporting that the publication rules in each of the circuits "give courts significant freedom when making publication decisions, effectively asking judges some version of the question, 'Do you think that this opinion is the type of opinion that should be published?'"); Copus, *supra* note 101, at 649 (noting that there are "no uniformly enforced or practiced guidelines for making the publication decision" and that "hence judges exercise considerable discretion in deciding when an opinion should be published, i.e., when an opinion will become law"). *See generally* Varsava, *supra* note 36 (finding that attributes of opinion authors are associated with publication rates).

[162] *See* Merritt E. McAlister, *"Downright Indifference": Examining Unpublished Decisions in the Federal Courts of Appeals*, 118 MICH. L. REV. 533, 582–93 (2020) (observing that even some reported, but unpublished, decisions contain little to no reasoning and arguing in favor of a minimum presentation of reasoning, for the sake of procedural justice).

[163] *See supra* Part I.B.2 (discussing the tiers of decision).

tions could more readily be detected. But it is also up to researchers to shed light on judicial practices and patterns that might otherwise fly under the radar, which is what we have aimed to do in this Article.

3. Alternative panels.

Another possible kind of reform, a more radical one, would involve a move away from three-judge panels.[164] As we explained above, numeric underrepresentation on courts is exacerbated when judges decide cases in panels of three.[165] On the hypothetical court we described above, where we have twelve judges in total—eight with one attribute (for example, male) and four with another (for example, female)—we should expect 76% of those panels to have a minority of women (i.e., zero or one woman), even if judges were neutrally distributed across three-judge panels.[166] If, in contrast, judges were to sit in panels of two, then we should expect minority-woman panels (meaning panels with no women) to occur in 42% of cases.[167]

Now, this system would raise some considerable difficulties—most obviously, a panel might not be able to reach a decision in a case because it is evenly divided. In that event, the district court decision could be affirmed by default, just as decisions of the courts of appeals are affirmed when the Supreme Court is equally divided over the outcome. In some cases, en banc review might then be appropriate.

Given persistent disparities in representation on courts, and the apparent exacerbation of these disparities through the three-judge panel system and case assignment or reporting practices, it might be worth considering the possibility of differently structured decision panels.[168]

---

164 Note, however, that this might require statutory reform, since the U.S. Code currently provides that the courts of appeals "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." 28 U.S.C. § 46.

165 *See supra* Part II.B.1.

166 *See supra* Part II.B.1.

167 Forty-two percent represents the probability of drawing an *AA* panel, which is eight out of twelve for an *A* judge in the first slot and seven out of eleven for the second slot—collectively, $8/12 \times 7/11 \approx 0.42$.

168 To counteract similar kinds of disparities along the lines of ideology, others have proposed nonrandom assignment procedures to make panels more ideologically balanced and to avoid extreme panels (for example, those with three Democratic appointees). *See* Tiller & Cross, *supra* note 118, at 232–34.

4. Interpretation of empirical studies.

As discussed in the previous Section, our work shows that even in studies that include both published and unpublished decisions, the attribute-to-outcome inference assumptions may be violated. Accordingly, correlations in these datasets between judicial attributes and outcomes are not necessarily indicative of causal relationships. Reporting decisions, like publication decisions, are endogenous to the variables under study. Judicial attributes appear to interact with case characteristics in ways that predict whether a case leads to a published or reported decision. This introduces selection bias that must be addressed before we can be confident about causal claims regarding the effects of judicial attributes on case outcomes and other case variables of interest.

That said, it is important to note that the findings in this Article do not invalidate the statistical correlations that empirical legal scholars have found—the issue is purely one of interpretation. So long as the assignment of cases to panels is functionally random, then prior analyses do indicate a causal effect of some kind. What is not known is whether correlations are due to the influence of judicial attributes on case outcomes or on case reporting. It is possible that the race, gender, party affiliation, and other attributes do in fact affect case outcomes. But it is also possible that selection effects fully account for the correlations that have led scholars to draw that conclusion. At this stage in its development, the field of empirical legal studies has not adequately teased out these two different causal mechanisms—or even taken seriously the possibility that the latter one is driving results.

This is a distinction with normative consequences. Courts can be understood as performing two basic social functions. The first is an *arbitration* or *dispute resolution* function that involves settling conflicts between individual parties. Conflicts are inevitable in society, and courts are called on to resolve them. The existence of this forum provides a normalized, nonviolent means to address a wide range of disagreements, from child custody disputes between former spouses to the legitimacy of criminal sanctions.

The second function of courts is to *determine and announce rules of conduct*. In the course of settling disputes, courts are sometimes called upon to clarify the meaning of ambiguous legal language or fill in gaps in legal rules, and in doing so, they distinguish between lawful and unlawful conduct. When these decisions are reported, they provide notice to the broader community about how legal norms will be interpreted in particular circumstances.

When judges make decisions concerning the outcomes of particular cases, those outcomes have direct consequences for the parties involved. A decision settles the dispute between the parties. But in justifying outcomes reached, courts also issue opinions that set more general rules and standards. Law is clarified and created when opinions are published and serve as precedent in future decisions. But even unpublished, reported opinions have informational value, for example predictive value concerning how future courts are likely to decide similar cases. They might play a normative role as well, as a means of justifying similar resolutions of like cases.

The choice of whether to publish or report decisions raises an important set of normative concerns that can be usefully informed by empirical analysis.[169] If judges publish or report decisions with strategic factors in mind, or differentially publish or report decisions under different circumstances, that can affect the rule-announcing function of courts and have major consequences for the state and direction of the law. But such decisions do not bear on courts' dispute-resolution function. If, however, different judges facing similar cases reach different outcomes, that raises questions about the fairness and predictability of the legal system.[170] To the extent that prior research of judicial decision-making fails to tease apart publication and reporting decisions on the one hand and outcome decisions on the other, its normative implications are unclear.

A claim that gender, race, or partisan affiliation affects case outcomes, which is purported to be supported by empirical data, may influence how the judiciary is perceived by the public and other institutions in ways that affect the credibility of the courts. If these claims are correct, then updated perceptions are appropriate. But if they are incorrect, these claims will lead to potentially costly misperceptions about the operation of the legal system.

An example may help illustrate the stakes. A relatively robust finding from prior studies of judicial behavior is that party affiliation affects outcomes in criminal law cases, with Democrat-appointed judges favoring defendants and Republican-appointed

---

169 *See, e.g.*, Hazelton et al., *Judicial Decisions*, *supra* note 69, at 649 (discussing publication). Our discussion of normative concerns here builds on Carlson et al., *supra* note 11, at 245–57.

170 *See generally* Joshua B. Fischman, *Measuring Inconsistency, Indeterminacy, and Error in Adjudication*, 16 AM. L. & ECON. REV. 40 (2014).

judges favoring the government.[171] If this finding is accurate, then defendants would be justified in questioning the legitimacy of outcomes in their cases because they appear to reflect, at least to some degree, simple partisan disagreement rather than a neutral application of the law. Similarly, interested citizens with preferences over how the law is applied in individual cases would be justified in orienting their voting behavior to the potential effects of party control of the White House and Senate on the partisan makeup of the bench.

By contrast, if prior findings concerning the relationship between judicial party affiliation and outcomes in criminal cases actually resulted from differences in publication or reporting decisions, rather than differential treatment of individual cases, the normative and political consequences are entirely different. An individual criminal defendant may not have any interest in whether a decision is published or reported. Partisan differences in publication or reporting do not go to the fairness, neutrality, or legitimacy of the judicial system from the perspective of the individual litigant (unless a reporting decision affects the outcome itself).[172] However, such decisions do matter for the broader political community because they affect the stock of law that is relevant to future proceedings. For example, if Republican-appointed judges are more likely to publish decisions in criminal cases in which the government wins, and Democrat-appointed judges are more likely to publish decisions in criminal cases in which defendants win, the partisan composition of the bench will affect the shape of the law over time. Citizens and politicians would be justified in taking this effect into account when making political choices, although for different reasons than in the alternative where partisan affiliation affects outcomes in individual cases.

Given rule-of-law norms, litigants have a legitimate expectation of neutrality and impartiality when they come before courts. But the same considerations may be less applicable to the determination of legal rules that apply to the broader community. Political influence over lawmaking is expected, and indeed desirable, in democratic societies. When courts act in their rule-announcing role, some degree of partisan influence may be more acceptable. Because decisions about whether to publish or report cases are more distant from a court's dispute-resolution function,

---

171 *See* Daniel R. Pinello, *Linking Party to Judicial Ideology in American Courts: A Meta-Analysis*, 20 JUST. SYS. J. 219, 240–41 (1999).

172 *See supra* notes 101–03 and accompanying text.

and more closely related to its rule-announcing function, the influence of party and other factors (such as race and gender) on these decisions do not raise the same kinds of rule-of-law concerns. Untangling the true causal mechanisms at play in judicial decision-making, and disaggregating influence over outcomes from influence over publication and reporting, should be a priority for future empirical legal research.

For the time being, researchers and those who engage with and report on empirical legal findings should be mindful of the different plausible causal mechanisms that lead to statistical correlations found in analyses of published or reported cases. Responsible research will call attention to these different mechanisms and remain agnostic between them unless study design enables credible distinctions to be made. In particular, we should avoid attributing correlations between judicial attributes and case outcomes to a causal influence on outcomes absent a specified basis for drawing that inference. Doing otherwise invites potential misunderstandings with real social and political consequences.

CONCLUSION

The existing empirical literature on judge demographics and decision-making focuses largely on voting behavior and case outcomes. But federal appellate judges make many critical administrative decisions in their judicial role, including which cases will be granted oral argument, which cases will be adjudicated by which judges, which cases will result in published decisions, and whether a decision will be made available beyond the parties to the case. Comparatively little empirical attention has been directed at judge demographics and judicial administration.[173] This Article helps fill that gap.

More research is needed to better understand the causal mechanisms behind our findings of skewed representation on the panels of reported decisions. Future research could also explore possible intersectional effects of race and gender on panel composition and investigate in more detail possible differences across

---

[173] *See* Moyer et al., *supra* note 6, at 463:

> While much of the extant work on gender and race in the judiciary focuses on the voting behavior of judges, another profitable avenue for inquiry focuses on the micro-foundations of racialized and gendered institutions, identifying ways in which actors and institutional rules or norms interact to shape expectations and how such interactions shape the distribution of power within an institution.

racial groups. The relatively low number of judges in the race categories aside from white presents challenges for statistical study of judge race as well as gender-race interactions, and an empirical approach that differs from the one used here might be necessary to get traction on these questions.

After a long history of severe underrepresentation, both women and people of color now make up substantial proportions of federal judges.[174] The normative literature concerned with gender and racial equality in the judiciary has focused on increasing the numeric representation of women and people of color on courts. And commentators concerned with gender and racial inequities in the legal system widely celebrate the growing numbers of women and people of color on the bench.[175] A focus on these numbers alone, however, elides possible distortions in how judges are represented in reported judicial decisions. And participation in such decisions is the primary way in which judges influence law and policy.

We find that the demographic composition of the federal courts of appeals does not translate in a straightforward or neutral way to representation on the panels of reported decisions. Our analyses indicate that female judges are underrepresented on these panels and that the gender and racial composition of panels is skewed—in particular, two-woman panels appear to be systematically underrepresented, raising concerns about the voice and the voting power of female judges. Overall, our study shows how power and voice might be unequally distributed across judges on the federal bench in ways that have gone, and could easily continue to go, unnoticed.

APPENDIX

A. Data

Our case data source is Harvard Law School's Caselaw Access Project (CAP). CAP constructed its collection by creating digital versions of the approximately forty thousand bound volumes

---

[174] *See supra* Part I.A.

[175] *See generally* Al Sharpton & Martin Luther King III, *Biden Has Been Revolutionary on Judicial Diversity—States Should Learn from Him*, THE HILL (Feb. 20, 2022), https://perma.cc/8F7K-Z85U; Stacy Hawkins, *Trump's Dangerous Judicial Legacy*, 67 UCLA L. REV. DISCOURSE 20 (2019); John P. Collins, Jr., *Judging Biden*, 75 SMU L. REV. F. 150 (2022).

owned by the Harvard Law School Library.[176] CAP describes the digitization process as follows:

> The Harvard Law School Collection was digitized on site at Langdell Hall. Members of our team created metadata for each [reporter] volume, including a unique barcode, reporter name, title, jurisdiction, publication date and other volume-level information. We then used a high-speed scanner to produce JP2 and TIF images of every page. A vendor then used OCR [optical character recognition] to extract the text of every case, creating case-level XML files. Key metadata fields, like case name, citation, court and decision date, were corrected for accuracy, while the text of each case was left as raw OCR output. In addition, for cases from volumes not yet in the public domain, our vendor redacted any headnotes.[177]

Our analyses focus on the federal appellate courts, which issue both published and unpublished decisions. Published decisions are those that are designated by the courts for publication in West's Federal Reporter. The West Reporter has long been the case law reporter for U.S. courts of appeals decisions.[178] As discussed above, so-called unpublished decisions have actually been publicly available for many years.[179] Nonetheless, they are designated by courts as "not for publication." Decisions with this designation used to appear in West's Federal Appendix (until it was terminated in 2021).[180] West began producing the Federal

---

[176] For details on case coverage, see *About*, CASELAW ACCESS PROJECT, https://perma.cc/P4WM-DUFA. The full CAP dataset is freely available for others to access with a researcher account. *See Documentation*, CASELAW ACCESS PROJECT, https://perma.cc/56AF-HEQ9.

[177] *About*, *supra* note 176.

[178] *See* E. ALLAN FARNSWORTH, AN INTRODUCTION TO THE LEGAL SYSTEM OF THE UNITED STATES 55–58 (4th ed. 2010).

[179] *See supra* Part I.B.2.

[180] Some unpublished decisions appear on Lexis or Westlaw but not in the Federal Appendix. These are not included in our analysis, and for our purposes here we consider them to be part of the category of "unreported" decisions. Based on correspondence with Thomson Reuters West representatives, our understanding is that courts themselves decide what to send to West for inclusion in the Federal Reporter and Federal Appendix. In contrast, correspondence with Westlaw and Lexis representatives indicated that these electronic databases use their own (apparently proprietary and confidential) selection mechanisms to determine which decisions to include. Accordingly, the reporting of decisions in the Federal Reporter and Federal Appendix would seem to reflect more the judgments of the courts themselves, and the inclusion in Lexis and Westlaw more the judgments of those databases. A possible question for future research is how the demographic representation of the panels in cases that appear in Lexis or Westlaw compares to the representation in West's official reporters.

Appendix in 2001.[181] After obtaining a researcher account to access the CAP data, we downloaded all decisions that appear in the Federal Reporter, 3d series and the Federal Appendix from 2001 through 2017 for all circuits.[182] We then filtered out cases from the Court of Appeals for the Federal Circuit, given its unique jurisdictional mandate.[183] There are 367,943 raw observations in our data for analysis.

We took advantage of the extensive demographic information about Article III judges made publicly available by the Federal Judicial Center (FJC).[184] We merged the FJC data with the CAP data using judge names, which, given the limited number of federal judges, amounts to a unique identifier in most cases. To identify judges for each observation, we constructed Python scripts using case metadata (circuit and year), case textual information, and FJC data to extract the judges listed on the panel that issued each decision. Misspellings and OCR errors were uncommon but not entirely absent. We iterated this script over multiple rounds of testing and verification until the correct matching rate for judge identification, cross-referenced with an external source (the online Westlaw database), reached over 99% accuracy on a set of 404 randomly selected cases.

Because our analyses take three-judge panels (which represent the vast majority of cases) as the unit of analysis, we dropped all decisions in which a three-judge panel could not be identified

---

[181] The first volume appeared in September 2001. *See* ROBERT TIMOTHY REAGAN, MEGHAN DUNN, DAVID GUTH, SEAN HARDING, ANDREA HENSON-ARMSTRONG, LAURAL HOOPER, MARIE LEARY, ANGELIA LEVY, JENNIFER MARSH & ROBERT NIEMIC, CITING UNPUBLISHED OPINIONS IN FEDERAL APPEALS 23 n.38 (2005). The Federal Appendix was discontinued in 2021. Berg, *supra* note 8.

[182] At the time we collected the data, CAP's case coverage ended partway through 2018. Since the data for that year was incomplete, we dropped it from our analyses.

[183] The Federal Circuit's jurisdiction is unique in that it is nationwide and is specialized and exclusive in terms of subject matter (including matters of international trade, intellectual property, government contracts, and monetary claims against the U.S. government). *See* 28 U.S.C. § 1295. All the other federal appellate courts have jurisdiction based on geographic area (and the same comparatively broad subject matter jurisdiction). Various other studies of federal appellate decisions have likewise excluded the Federal Circuit. *See, e.g.*, BARNES BOWIE ET AL., *supra* note 48, at 22; Carlson et al., *supra* note 11, at 236; Moyer et al., *supra* note 6, at 458 n.6; Corey R. Yung, *Judged by the Company You Keep: An Empirical Study of the Ideologies of Judges on the United States Courts of Appeals*, 51 B.C. L. REV. 1133, 1162 n.218 (2010); *see also* Beth Zeitlin Shaw, *Please Ignore This Case: An Empirical Study of Nonprecedential Opinions in the Federal Circuit*, 12 GEO. MASON L. REV. 1013, 1030 (explaining "differences in the nature of the Federal Circuit compared to other federal appellate courts").

[184] *See Biographical Directory of Article III Federal Judges, 1789–Present*, FED. JUD. CTR., https://perma.cc/Q8NC-FPSE.

(see Table 2). Primarily, these were en banc panels consisting of more than three judges. We excluded en banc panels because they are relatively uncommon and the process of constructing them and assigning them cases is drastically different than for typical three-judge panels. En banc review is discretionary and can be expected to relate to the characteristics of the case; further, these panels are considerably larger, and judge characteristics may affect assignment to them.[185] The remainder of non-three-judge cases consisted of ones for which we identified fewer than three judges. Such cases make up a small proportion of all cases in the data and are roughly evenly distributed across circuits and years.

TABLE 2: CASES BY CIRCUIT

	Total Accessed	Three-Judge Panel
First Circuit	7,352	7,253
Second Circuit	30,491	29,533
Third Circuit	24,508	24,202
Fourth Circuit	55,151	54,819
Fifth Circuit	47,863	47,406
Sixth Circuit	24,461	24,169
Seventh Circuit	18,931	18,658
Eighth Circuit	21,443	21,243
Ninth Circuit	79,639	78,306
Tenth Circuit	22,921	22,589
Eleventh Circuit	29,113	28,769
D.C. Circuit	6,070	5,970
Total	367,943	362,917

Note: Data drawn from the Harvard Law School Caselaw database. Both published and unpublished decisions issued from 2001 through 2017 by all U.S. courts of appeals, except the Federal Circuit, are included.

[185] For example, senior-status judges generally do not participate in en banc decisions. *See, e.g.*, 4TH CIR. RULE 35(c):

> A court en banc shall consist of all eligible, active and participating judges of the Court, except that any senior judge of the Court may (1) participate in en banc rehearing of a decision of a panel of which the judge was a member or (2) continue to participate in the decision of a case or controversy that was heard or reheard by the en banc court at a time when the judge was in regular active service.

A total of 355 court of appeals judges, 994 district court judges, and 10 Supreme Court Justices appear in our data.[186] District judges appear because they sometimes serve on appellate panels as visiting judges "sitting by designation." We rely on the FJC's data on judge race/ethnicity and gender to generate those variables in our analysis.[187] Table 3 provides details on the demographic characteristics of the judges in our data.[188]

TABLE 3: JUDGE DEMOGRAPHICS

	Female	Male	White	African American	Asian American	Hispanic	Other
District Courts	238	756	809	95	23	61	6
Courts of Appeals	82	273	301	29	5	20	0
Supreme Court	2	8	8	1	0	1	0

Note: Gender and racial/ethnic classifications are drawn from the FJC data. The category "Other" includes judges who are identified in the FJC data as African American/Hispanic, American Indian, Asian American/Pacific Islander, Hispanic/white, and Pacific Islander/white.

B. Results

1. Panel composition tests.

For our primary panel composition test, we construct the expected distribution based on the frequency with which each judge

---

186 There are ten judges from the U.S. Court of International Trade in our data. We treat them as district court judges for purposes of our analysis. Judges from the Court of Appeals for the Federal Circuit also appear in our data because they sometimes serve as visiting judges on the other circuits.

187 The FJC data places judges into one of two gender categories: female and male. Judges are further identified according to eight racial or ethnic groups (African American, American Indian, Hispanic, white, Afro-Latino, Asian American, Pacific Islander, Chaldean), and the FJC allows for mixed categorizations (e.g., African American/Hispanic). The majority of judges are identified as white (exclusively). Judges identified as African American, Hispanic, and Asian American are the next three largest categories (respectively). We place the remaining judges, who fall into FJC categories with only one or two members, in a category of "Other." This group contains no court of appeals judges and six district court judges.

188 Throughout this paper we describe judges in the terms that the FJC uses for gender and race/ethnicity categories. Although the latter categories represent not only race but also ethnicity, we generally refer to them as racial groups.

appears in our data in that circuit-year. This judge-frequency approach accounts for the fact that different judges may contribute differently at different stages of their career or for exogenous reasons such as illness. Senior judges, in particular, have differential workloads from active judges and can decide to participate in fewer or more cases based on their preferences.[189] We also include visiting judges sitting by designation in this analysis; they can be expected to participate in far fewer cases than "home" judges. For each circuit-year, we assign each judge a frequency corresponding to the number of cases in which we observe that judge. For example, if a circuit reports two hundred cases in a given year, and for twenty of those cases Judge Smith is on the panel, then Judge Smith's frequency for that year is 10%. The sum of all of the frequencies will be 300%, by construction (100% x 3 panel slots).

We find that panel composition with respect to both race and gender is skewed. In the aggregate, we find that there are roughly seventeen hundred fewer two-woman cases than expected and three hundred fewer two-POC cases than expected. Table 4 presents these aggregate results.

TABLE 4: PANEL COMPOSITION TEST I

	♂♂♂	♂♀♂	♀♂♀	♀♀♀
Observed	151,085***	163,015***	44,879***	3,938*
Expected	[153,163]	[159,386]	[46,552]	[3,816]
	○○○	○□○	□○□	□□□
Observed	199,767	135,719	25,564*	1,867***
Expected	[200,078]	[135,364]	[25,877]	[1,597]

Note: ♂ = male judge; ♀ = female judge; ○ = white judge; □ = POC judge. Observed number of cases with each panel type are given, with the expected number below in brackets. Significance levels were generated using a chi-squared test. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.

Looking in a more disaggregated fashion, a chi-square test comparing the set of expected frequencies for cases of each panel type to the observed frequencies for gender showed statistically significant panel composition bias for all twelve circuits[190] and for

[189] *See* Levy, *supra* note 59, at 69 n.12 ("Senior judges elect how much to sit—they can choose to hear a caseload that is 25% of that of an active judge or, say, 75%.").

[190] All circuits showed significant effects at the $p < 0.1$ level; eleven circuits showed significant effects at the $p < 0.05$ level; ten circuits showed significant effects at the $p < 0.01$ level; and five circuits showed significant effects at the $p < 0.001$ level.

all seventeen years.[191] At the circuit-year level, 81 of the 122 circuit-years tested showed significant gender bias.[192] With respect to race, eleven of twelve circuits showed significant panel composition bias,[193] as did sixteen of seventeen years.[194] At the circuit-year level, thirty-nine of the sixty-two circuit-years tested showed significant racial bias.[195]

We construct an additional panel composition test based on an alternative approach. Here, rather than taking judge frequency as a given (as in the prior analysis), we calculate an expected judge frequency for the subset of judges that should have the most consistent participation. We construct this expected distribution of the appearances of active (i.e., nonsenior, nonvisiting) judges in each circuit-year based on an assumption that all nonvisiting active judges in a circuit have an equal likelihood of being selected for a panel and all panels receive the same number of cases.[196] If this is true, then each active judge on the circuit will appear, on average, in an equal number of decisions, and appearances in a given year will not be correlated with judge race or gender. Table 5 presents the results.

---

191 All years showed significant effects at the $p < 0.05$ level; fifteen years showed significant effects at the $p < 0.01$ level; and fourteen years showed significant effects at the $p < 0.001$ level.

192 Eighty-one circuit-years showed significant effects at the $p < 0.05$ level; seventy circuit-years showed significant effects at the $p < 0.01$ level; and sixty circuit-years showed significant effects at the $p < 0.001$ level. Note that the number of circuit-years tested is not equal to the number of circuit-years in our data because the chi-square test requires at least five expected and observed counts in each category to be valid. Accordingly, if a circuit-year has fewer than five observations in any category, we exclude it from the test.

193 Eleven circuits showed significant effects at the $p < 0.05$ level; seven circuits showed significant effects at the $p < 0.01$ level; and six circuits showed significant effects at the $p < 0.001$ level.

194 Sixteen years showed significant effects at the $p < 0.05$ level; fourteen years showed significant effects at the $p < 0.01$ level; twelve years showed significant effects at the $p < 0.001$ level.

195 Thirty-nine circuit-years showed significant effects at the $p < 0.05$ level; thirty-four showed significant effects at the $p < 0.01$ level; and thirty-three showed significant effects at the $p < 0.001$ level.

196 In addition to dropping visiting judges and senior judges from this analysis, we drop chief judges and judges in their first two years or final year on the bench since we should not expect duration of service in those years to be consistent across judges. Some judges joined a court late in one year and did not start appearing in cases regularly until part way through the next year, which is why we excluded the first two years of service.

TABLE 5: PANEL COMPOSITION TEST II[197]

	♂♂♂	♂♀♂	♀♂♀	♀♀♀
Observed	34,023	45,193***	14,214***	1,161***
Expected	[34,192]	[43,970]	[15,020]	[1,409]
	○○○	○□○	□○□	□□□
Observed	46,014*	38,968†	9,167***	442***
Expected	[46,321]	[39,233]	[8,510]	[527]

Note: ♂ = male judge; ♀ = female judge; ○ = white judge; □ = POC judge. Observed number of cases with each panel type are given, with the expected number below in brackets. Significance levels were generated using a chi-squared test. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; †$p < 0.10$.

We again find that panel composition with respect to both race and gender is skewed. In the aggregate, we find that there are about eight hundred fewer two-woman cases than expected. We observe fewer cases with three-man or zero-man panels than expected, but more with two-man panels. Turning to race, in contrast to the first composition test, here we see more cases with two people of color than expected. We observe fewer with three, two, or no white judges.

Each of the courts of appeals has its own practices regarding how many cases are assigned to each panel of judges. There appears to be substantial variation across circuits in terms of how many cases are assigned to each panel that is constructed; moreover, it is difficult to ascertain the exact parameters within circuits and the practices seem to be messy and inconsistent.[198] If multiple cases are assigned to each panel, that might result in a skewed distribution of panel types in cases. To explore this possible mechanism, we conducted a simulation in which we randomly constructed panels of three judges within each circuit-year based on the numbers of cases in our data with three active judges and

[197] We also ran this test on cases with panels of two active judges and one active judge. We found statistically significant differences for gender and race in these tests as well. For cases with two active judges, we observe more with zero women than expected and fewer with one and with two women. We likewise observe more cases with zero people of color than expected and fewer with one and with two people of color. For cases with one active judge, we observed fewer with zero women than expected and more with one woman. We likewise observed fewer with zero people of color and more with one person of color.

[198] *See* Chilton & Levy, *supra* note 49, at 27 (describing the considerable heterogeneity across circuits in terms of number of cases assigned to panels and the lack of clarity around these practices).

the attributes of those judges.[199] (This is the same set of judges and cases considered in the analysis reported in Table 5 above.) We ran this simulation one thousand times to test the likelihood that we would see the kind of nonrandomness in panel composition that we observe if judges were randomly distributed to observed panels and panels received different numbers of cases (i.e., differently sized "batches" of cases), between one and twenty. At the corpus-aggregated level, we then count the observed frequencies of each panel type in our data. Table 6 reports the number of simulations that had fewer than the observed case count for each panel type. For example, with a batch size of one, one hundred out of the one thousand simulations generated fewer cases with all-male panels than the actual, observed number; all one thousand simulations generated fewer cases with two-man panels, and none of the simulations generated fewer cases with two-woman panels or all-woman panels.

TABLE 6: SIMULATION BASED ON SIZE OF CASE BATCHES

Batch Size	♂♂♂	♂♀♂	♀♂♀	♀♀♀	○○○	○□○	□○□	□□□
1	100	1,000	0	0	17	42	1,000	1
2	228	1,000	0	0	52	111	1,000	1
3	261	1,000	0	0	110	133	1,000	12
4	268	1,000	0	2	117	196	1,000	34
5	266	1,000	0	3	164	212	1,000	48
10	345	990	11	13	252	286	990	115
15	398	982	21	33	272	313	979	177
20	412	961	40	65	308	340	959	234

Note: ♂ = male judge; ♀ = female judge; ○ = white judge; □ = POC judge. Greyed cells represent runs where only a small percentage (<5%) of simulations were as extreme as the observed distributions—e.g., for cases with two-man panels, which we found to be overrepresented in the data, less than 5% of the simulations generated as many or more such cases than observed.

We find that although simulations with larger batch sizes sometimes generate extreme panel distributions (i.e., distributions that are very different from the expected distribution under the null hypothesis of gender- and race-neutrality), such simulations are quite rare. For example, with batch sizes of ten cases, only ten out of one thousand (1%) generated a larger number of

[199] We consider only judges that are not in their first, second, or final year of service, not chief judges, not senior-status judges, and not visiting judges. *See supra* note 196.

cases with two-man panels than we observed in the data. Even with batch sizes of twenty, fewer than forty out of one thousand (<4%) generated a larger number of cases with two-man panels than observed. The inverse is the case for two-woman panels: only forty out of one thousand simulations had fewer cases with two-woman panels than what we observed. With respect to race, cases with one-white-judge panels were observed in the data at higher-than-expected rates; even with a batch size of twenty, less than 4% of simulations generated more such cases than observed. This exercise suggests that even if every panel in every circuit were assigned twenty cases at a time, that practice would be very unlikely to explain the full extent of imbalances in composition types that we observe.

2. Representation test: aggregate level.

Second, we estimate relationships between judge features and frequency of appearance in our data overall through an ordinary least squares regression in which we have observations for every judge for each year the judge appeared on panels. The dependent variable of interest is the number of cases in which the judge appears. The independent variables of interest are the demographic information that we collected, specifically gender and the four racial/ethnic categories described above.[200] We specify two models, each accounting in a different way for expected correlations between case numbers and the judge-level covariates of senior status; chief status; and first, second, and last year on the bench.[201] In both models, we drop visiting judges. In the first model, we drop observations for senior judges; chief judges; and each judge's first, second, and last years on the bench. In the second model, we retain observations for the latter judge types but include those attributes as covariates, interacting them with gender. The idea here is that these statuses are correlated with gender, likely affect the number of cases a judge participates in, and might also affect female and male judges differently. The second model is our preferred one (because it captures more of the data). In both models, we also control for the party of the President who appointed the judge, the judge's tenure (years on the court at the time of decision), and year and court (interacted). The results are reported in Table 7.

---

[200] *See supra* note 187.

[201] *See supra* note 196.

TABLE 7: AGGREGATE EFFECTS OF JUDGE GENDER AND RACE ON CASE COUNTS

	Model 1	Model 2
Male	10.75**	17.94**
	(2.8)	(4.23)
Asian American	3.76	18.44
	(13.3)	(21.64)
Hispanic	7.0	-6.93
	(6.19)	(9.01)
White	5.76	-5.09
	(3.48)	(6.17)
Democrat	-1.15	-0.88
	(2.94)	(3.99)
Tenure	-0.58**	-1.39**
	(0.20)	(0.22)
Circuit	Yes	Yes
Year (fixed effect)	Yes	Yes
Circuit x Year (fixed effect)	Yes	Yes
Chief/Senior/(1st/2d/Last)	Dropped	Interacted w/ gender
R^2	0.91	0.67
Observations	2,102	4,624

Note: *$p < 0.05$; **$p < 0.01$. The dependent variable is the number of cases in which a judge appeared in a given year. Reference category for gender is female, for race African American. Robust standard errors in parentheses. Substituting age for tenure led to very similar results. With standard errors clustered at the judge level, for Model 1, significance for male is at the $p < 0.05$ level; for Model 2, significance for male is at the $p < 0.1$ level.

We find a significant relationship between judge gender and number of cases heard in both models, while judge race does not have a significant effect in either. In Model 1, we find that a male judge can be expected to appear in about eleven more cases per year than a female one ($p < 0.01$). In Model 2 (the preferred model), we find similar results but of a stronger magnitude, with male judges predicted to appear in eighteen more cases per year.

3. Representation test: disaggregate level.

Third, we examine whether race and gender are associated with over- or underrepresentation in our data at the disaggregated, circuit-year level. Here, we use the same expected distribution

from the prior analysis assuming equal contributions of active (i.e., nonsenior, nonvisiting) judges in each circuit-year.[202]

In a substantial number of circuit-years, the number of times that male judges and white judges appear in cases depart from the numbers that would be expected if representation was balanced—that is, not systematically associated with gender and race—across judges. With respect to race, in over one-fifth of circuit-years, white judges appear at unexpected levels (at the $p < 0.05$ level).[203] With respect to gender, the effect is more widespread, with nearly a quarter of circuit-years departing from a balanced representation.[204] With respect to both gender and race, there is considerable heterogeneity in terms of which group is overrepresented across circuit-years.[205] In the aggregate, this heterogeneity roughly balances out in the case of race but is biased with respect to gender, indicating systematic underrepresentation of female judges within circuit-years in reported cases.[206]

---

[202] In addition to dropping visiting judges and senior judges from this analysis, we drop chief judges and judges in their first, second, or final years on the bench. *See supra* note 196.

[203] Of 198 circuit-years tested, 42 showed significant effects at the $p < 0.05$ level; 24 showed significant effects at the $p < 0.01$ level; and 16 showed significant effects at the $p < 0.001$ level.

[204] Of 195 circuit-years tested, 47 showed significant effects at the $p < 0.05$ level; 27 showed significant effects at the $p < 0.01$ level; and 22 showed significant effects at the $p < 0.001$ level.

[205] To estimate this heterogeneity, for each circuit-year we calculated an expected number of panel slots for the relevant demographic variables (gender or race) as well as a standard deviation. The observed distribution of panel slots is then represented in terms of deviations from the expectation. A normal distribution would be expected (approximately) absent some bias. By contrast, the observed distribution has unexpectedly fat tails, with both under- and overrepresentation according to the studied traits. For gender, of the 195 circuit-years tested, 44 circuit-years were more than 2 standard deviations away from the expected number, 26 circuit-years were more than 3 standard deviations away; and 17 circuit-years were more than 4 standard deviations away. For race, 39 circuit-years were more than 2 standard deviations from the expected number; 17 circuit-years were more than 3 standard deviations away; and 11 circuit-years more than 4 standard deviations away.

[206] More specifically, we compare the mean of the observed distribution described in the prior footnote with the expected mean, for both race and gender. For race, the expected and actual means are not statistically significantly different. For gender (female), the observed mean is 0.66 standard deviations below the expected mean ($p < 0.01$).