University of Cincinnati Law Review

Volume 88 | Issue 2 Article 1

January 2020

Fractured Justice: An Experimental Study of Pretrial Judicial Decision-Making

Prentiss Cox
University of Minnesota Law School, coxxx211@umn.edu

FRACTURED JUSTICE: AN EXPERIMENTAL STUDY OF PRETRIAL JUDICIAL DECISION-MAKING

Prentiss Cox

ABSTRACT

This Article reports on a controlled empirical examination of what happens when judges exercise their broad discretion to resolve non-dispositive pretrial disputes. Experimental studies of judicial decision-making are rare, and experimental studies of pretrial dispute resolution by predominantly state court judges are non-existent. Data was collected from the presentation of nine simulated disputes about discovery and litigation management in 166 chambers conferences before 61 volunteer judges.

The study results indicate that judges faced with the same facts usually fracture in deciding the dispute. This variation is not attributable to groups of judges deciding in a consistent way for some parties. Very few judges showed marked tendencies to defer decisions or rule for one type of party. Nor did judicial characteristics, such as gender, experience or measures of ideology, explain the simulation results. These results are surprisingly consistent with another experimental study on district court judicial decision-making in criminal sentencing, which found a similar pattern of fractured outcomes.

The study results also challenge the assumption that the addition of factors or presumptions in procedural rules produces more predictable and consistent decision-making, instead finding that the disputes guided by the vaguest of standards created more agreement among judges than when the relevant rule contained a multiple factor test or even a presumptive outcome. Given the varying results when judges exercise their discretion to determine disputes, changes to the civil procedure rules to limit judicial discretion advocated by some scholars would require an upheaval in current judicial decision-making practice.

I. INTRODUCTION

A fundamental choice in the design of law is when to make a firm rule and when to allow the exercise of discretion by a judge or jury. Civil procedure in the United States superficially is composed of rules—hence, the Federal Rules of Civil Procedure—but is predominantly a collection of more general standards with a few inflexible commandments. Judges typically resolve pretrial litigation disputes either by reference to these standards or by the use of standards to reverse presumptive outcomes in

the rules.[1] With these standards comes the exercise of judicial discretion in creating decisions in the specific circumstances of each dispute.[2]

Civil litigation in our courts has evolved from a process designed to culminate in trial to a series of pretrial events that effectively dispose of all but a tiny number of disputes. Since the adoption of the Federal Rules of Civil Procedure, the number of trials has dropped to a quarter of its previous level.[3] At the same time, non-dispositive pretrial rulings have increased substantially.[4] The importance of pretrial procedure combined with the broad discretion afforded trial court judges under the civil procedure rules means understanding dispute resolution outcomes in a pretrial setting is key to understanding how litigation functions in practice.

This Article reports on a natural experiment that sheds light on what happens when judges exercise their discretion to resolve pretrial disputes about discovery and litigation management. All first-year students at the University of Minnesota Law School take a course entitled "Law in Practice." This course uses simulations to introduce the application of law to resolve client problems in dispute resolution and transactional matters. One of those simulations is a chambers conference conducted by judges, mostly state district court and federal magistrate judges. The judges are presented with issues designed to reflect typical discovery and other non-dispositive pretrial disputes that arise in civil litigation. For this study, we retained the results of 166 of these simulations resolving 400 pretrial issues over a seven-year period.

The study here primarily examines whether judges reach consistent decisions when faced with the same fact patterns, and if not, whether varying outcomes reflect different approaches to resolving pretrial disputes or different characteristics or experience of the judges. The answers to these questions can inform the expectations of litigators about the predictability of resolving pretrial disputes, contribute to our general

The author is appreciative of guidance from Herbert Kritzer, June Carbone and Paul Vaaler, and for the research assistance of Scott Dewey and Grant Abrams.

1. *See infra* Part II.B.

2. Alexandra D. Lahav, *Procedural Design*, 71 VAND. L. REV. 821, 861 (2018) (observing that "[t]he soul of the Federal Rules, it might be said, is judicial discretion, and discretion may be the price for their (relative) simplicity.").

3. Todd D. Peterson, *Restoring Structural Checks on Judicial Power in the Era of Managerial Judging*, 29 U.C. DAVIS L. REV. 41, 62 (1995) (establishing that in 1938 the trial rate was "approximately 20% of the cases filed," but that "by 1990, 4.3% of the filed civil cases resulted in trials."). *But see* Herbert M. Kritzer, *The Trials and Tribulations of Counting "Trials"*, 62 DEPAUL L. REV. 415 (2013) (noting Marc Galanter's research and related scholarship on "the vanishing trial," and arguing that the reduction in trials is not as straightforward as prior data suggest, especially when considering state court litigation).

4. Stephen C. Yeazell, *The Misunderstood Consequences of Modern Civil Process*, 1994 WIS. L. REV. 631, 637 (1994) (observing that "non-dispositive decision making has grown substantially over the past fifty years, necessitated and created by changes in the rules of procedure.").

understanding of how judicial decision-making occurs, and provide insight into how we shape future amendments to the rules of civil procedure.

The study results indicate that judges faced with the same facts do not reach consistent outcomes across multiple issues. Even when decisions are grouped into four broad categories, only three of the nine issues decided resulted in a clear majority agreement among the judges on one outcome category. Some of the issues were resolved with a closely divided fracturing across all four categories. This variation in outcomes is not attributable to groups of judges deciding in a consistent way for some parties. Very few judges showed marked tendencies to defer decisions or rule for one type of party. Rather, varying outcomes occurred because judges generally decided multiple simulations based on their own unique patterns of decisions to defer, rule wholly for plaintiff or defendant, or reach compromise positions. Nor did judicial characteristics, such as gender, ideology or experience, explain the patterns in the simulation results.

This Article provides three contributions to scholarship in the areas of civil procedure and judicial decision-making. First, it is the only empirical scholarship examining how judges use chambers conferences to resolve non-dispositive pretrial disputes, even though such conferences are a common form of judicial engagement in litigation. The study results show that judges commonly use chambers conferences to resolve pretrial issues rather than defer the decision and are as likely to rule for a particular party as to reach a compromise solution to resolve the dispute.

Second, the study contributes to the empirical literature on judicial decision-making. The results here are in substantial accord with other experimental studies evaluating trial level judicial decision-making. In particular, the study results here are very similar to the decision-making variation shown in a well-known experimental study in criminal sentencing. With the exception of a very small number of judges at the margins, variation in study outcomes reflect judicial differences in exercise of discretion that do not consistently favor one category of party over another, but rather show a splintering of results by discrete decisions. The study results show neither a widely shared consensus about appropriate outcomes for individual issues nor a clear majority/minority grouping across issues. Nor do the results of these trial court level decisions yield to a simple explanation for difference by personal characteristics or experiences of the judges or proxies for judge ideology.

Third, civil procedure scholars disagree about whether the current emphasis on judicial discretion in managing civil litigation is preferable

to a system of more determinative rules.[5] This study provides unique experimental data about how trial court judges use their current broad discretion in pretrial matters. Because judges reach varying results, and often reach compromise solutions to the dispute, a transition to a system of rigid rules will wrought an upheaval in pretrial practice. Furthermore, the pretrial disputes in the study guided by the vaguest of standards in the federal rules created more agreement among judges than disputes where the relevant rule contained a multiple factor test or even a presumptive outcome. Accordingly, the study results suggest that adding factor tests and presumptions may not be enough to eliminate varying outcomes; rather, only explicitly determinative rules with no discretion to depart would achieve outcome consistency.

This Article describes the simulation experiment and methodology of data collection, summarizes the simulation results, and looks at the meaning of these results for understanding our system of pretrial litigation. Part II discusses the empirical and theoretical scholarship to which this Article contributes. Part III describes the simulations that form the study data, the methodology of producing data on the simulation results, and the study limitations. Part IV presents the results of the study. Finally, Part V discusses the meaning of the results for scholars of our litigation system and for judges and attorneys who engage in litigation.

II. SCHOLARSHIP ON JUDICIAL DECISION-MAKING IN CIVIL PROCEDURE

The empirical scholarship on judicial decision-making is vast. The first subpart summarizes existing scholarship and places the study reported here in a largely unexamined corner of this research. The second subpart looks at theoretical legal scholarship on trial court decision-making in discovery and litigation management, with a focus on the policy implications of allowing trial courts substantial discretion over such matters.

A. Empirical Scholarship of Judicial Decision-Mmaking

In the late 1940s, political scientist C. Herman Pritchett gathered data examining ideology as a determinant of Supreme Court case outcomes; then a heretical notion and now a view widely accepted among average citizens.[6] This scholarship is now produced by social scientists,

5. *See infra* Part II.B.

6. Herman C. Pritchett, *Divisions of Opinion Among Justices of the U.S. Supreme Court*, *1939-1941*, 35 AM. POL. SCI. REV. 890 (1941); HERMAN C. PRITCHETT, THE ROOSEVELT COURT: A STUDY IN JUDICIAL POLITICS AND VALUES, 1937-1947 (Macmillan 1948). *See* Lee Epstein, *Some Thoughts on the Study of Judicial Behavior*, 57 WM. & MARY L. REV. 2017, 2019-22 (2016) (describing Pritchett's work

psychologists, legal academics, and more recently by scholars with more than one of these credentials or affiliations, as faculty holding both a J.D. and Ph.D. increasingly fill the ranks of the legal academy.[7] The animating concern of ideology explaining federal appellate court decision-making continues today as a primary focus of empirical legal scholarship. Underlying this work is a measuring of the realist challenge to the formalist notion that law dictates a correct result. Although few, if any, legal scholars would claim today to subscribe to a strict formalist view of judicial outcomes, the degree to which law creates a clearly correct result in a given dispute and the particular factors that otherwise account for judicial resolution of disputes are tested in this vein of empirical scholarship.[8]

Scholarship on the influence of ideology has been joined by countless other empirical studies attempting to explain why and how judges make the decisions they make, including identifying the explanatory value of race, gender, and other judicial characteristics on judicial outcomes.[9] While study of judicial decisions has grown far beyond ideological motives, the type of judges studied remains fairly narrow. United States Supreme Court decision-making studies are voluminous, of course, as are studies of the decisions of federal appellate judges.[10] State courts are less frequently studied, and the studies that exist mostly concern state appellate courts and focus substantially on the impact of state judicial elections on outcomes.[11] Similarly, trial court decision-making is less

and the reaction to it). Quinnipiac University Poll (May 2019) (finding 59% of U. S. people polled agreed that the Supreme Court is "too influenced by politics."), https://www.pollingreport.com/court.htm

7. Michael Heise, *An Empirical Analysis of Empirical Legal Scholarship Production, 1990-2009,* 2011 U. ILL. L. REV. 1739, 1747-48 (2011).

8. *See* Joshua B. Fischman, *Reuniting "Is' and "Ought' in Empirical Legal Scholarship*, 162 U. PA. L. REV. 117, 146-55 (2013).

9. *See* Epstein, *supra* note 6, at 2047-50 (categorizing scholarship on the impact of judicial characteristics on case outcomes); Gregory C. Sisk, *The Quantitative Moment and the Qualitative Opportunity: Legal Studies of Judicial Decision Making*, 93 CORNELL L. REV. 873, 876 (2008) (describing empirical work on judicial decision-making); LAWRENCE BAUM, THE PUZZLE OF JUDICIAL BEHAVIOR (U of Michigan Press 1997) (describing empirical work on judicial decision-making and including a reference listing empirical studies on the topic); Neal C. Tate & Roger Handberg, *Time Binding and Theory Building in Personal Attribute Models of Supreme Court Voting Behavior, 1916-88*, 35 AM. J. POL. SCI. 460 (1991); Joel Grossman, *Social Backgrounds and Judicial Decision-making*, 79 HARV. L. REV. 1551 (1966).

10. Heise, *supra* note 7, at 838 ("as is generally true with much judicial decision-making literature, many of the studies, finding ideology as an influential variable focus on the Supreme Court or, to a lesser extent, federal courts of appeal. Whether any findings of significance in these settings might hold for judges at the trial court level is unclear.").

11. LEE EPSTEIN, WILLIAM LANDUS & RICHARD POSNER, THE BEHAVIOR OF FEDERAL JUDGES: A THEORETICAL AND EMPIRICAL STUDY OF RATIONAL CHOICE 97-98 (Harvard University Press 2013) (colleting empirical studies of state court judicial decisions). *See, e.g.,* HERBERT KRITZER, JUSTICES ON THE BALLOT: CONTINUITY AND CHANGE IN STATE SUPREME COURT ELECTIONS (Cambridge University Press 2015); Scott R. Bauries, *Is There an Elephant in the Room?: Judicial Review of Educational Adequacy and the Separation of Powers in State Constitutions*, 61 ALA. L. REV. 701, 719 (2010)

studied. The existing studies focus on federal district court compliance with appellate directives or criminal sentencing,[12] and rely on trial court opinions, which are primarily issued on merit decisions.[13] Studies of decision-making on disputed issues by state trial court judges in civil matters are rare.[14]

Empirical work of matters relevant to the study presented here— how judges resolve discovery and other pretrial non-dispositive issues—is limited. Extensive data collection and analysis has been conducted by the Federal Judicial Center and others using docket data and surveys of judicial personnel and litigants looking at the burden of discovery on the courts and litigants, or that relate to case and docket management.[15] The author is unaware of any empirical study of how judges use chambers conferences to make decisions on pretrial non-dispositive matters in a chambers conference setting.

Empirical judicial decision-making studies rely mostly on large sets of data coding the outcomes of written judicial opinions on the merits of a case. The obvious limitation in this work is that specific case variation makes it difficult to assert definitively that a particular judge, or even a group of judges with a particular set of ideological preferences and personal characteristic, would have reached outcomes different from judges with opposing preferences and characteristics if given the same exact cases.[16]

An alternative approach that addresses these limitations is the use of

(observing that few studies of judicial decision-making employing attitudinal modeling have addressed state court decision making).

12. *See* EPSTEIN ET. AL, *supra* note 11, at 96-97 (colleting empirical studies of U. S. district courts); Diego A. Zambrano, *Judicial Mistakes in Discovery*, 113 NW. U. L. REV. 218 (2018) (describing empirical studies on trial court non-compliance with appellate directives and reporting on a study of compliance by trial courts with the changes to Rule 26(b) of the Federal Rules of Civil Procedure).

13. Pauline T. Kim et. al., *How Should We Study District Judge Decision-Making?*, 29 WASH. U. J.L. & POL'Y 83, 94-102 (2009) (surveying the literature on district court decision-making, noting that they predominantly use written opinions, and arguing for a focus on decisions rather than opinions and for attention to procedural rulings).

14. *But see generally* KENNETH DOLBEARE, TRIAL COURTS IN URBAN POLITICS: STATE COURT POLICY IMPACT AND FUNCTIONS IN A LOCAL POLITICAL SYSTEM (John Wiley & Sons 1967).

15. *See* Thomas E. Willging, *Past and Potential Uses of Empirical Research in Civil Rulemaking*, 77 NOTRE DAME L. REV. 1121, 1127-33 (2002) (surveying and describing empirical studies on judicial pretrial management of cases); Judith A. McKenna & Elizabeth C. Wiggins, *Empirical Research on Civil Discovery*, 39 B.C. L. REV. 785, 788 (1998) (summarizing empirical work on the use of discovery in civil litigation).

16. Joshua Fischman, *Measuring Inconsistency, Indeterminacy and Error in Adjudication*, 16 AM. L. & ECON. REV. 40, 42 (2014) (observing that "inconsistency . . . cannot be precisely estimated with the kinds of observational data that are typically available. Studies that compare rates at which judges reach various outcomes can only identify a range of feasible levels of inconsistency, even when cases are randomly assigned. More precise measurement of inconsistency requires data in which judges' decisions in the same cases are simultaneously observable.").

experimental studies. Experimental studies of courts have included random case assignment between a control group and cases assigned using different litigation management practices.[17] Psychologists have used surveys of judges to examine the extent to which well-known heuristics apply to federal judges as decision-makers.[18]

The study here lies in a subset of experimental scholarship in which an identical fact scenario is given to a large number of decision-makers in a controlled setting. This type of empirical work eliminates the problem inherent in studies using large sets of data from actual cases in which the "judges are never observed deciding the same case."[19] Perhaps because such experimental studies aim to remove case variance as a factor in differing outcomes, such studies have sought to measure variance in outcome as a primary purpose of the study, rather than focus on correlations between judicial characteristics and a series of case outcomes over time.

The most influential of such work is a 1974 study in which federal district court judges in the Second Circuit were asked to impose mock sentencing based on identical pre-sentence reports in twenty cases.[20] The study found substantial variation in sentencing in all but three of the twenty cases.[21] It concluded that the disparity was not substantially explainable by the judge's district or the judge's length of service, and that while the judges collectively created a broad range of results in each mock case, only a few judges sentenced consistently in the same direction.[22] This study had a substantial influence on the introduction of sentencing guidelines in an effort to create uniformity in sentencing outcomes.[23]

A similar level of disparity in sentencing was found in an experimental study in 1977 which presented sentencing files in five cases to forty-seven Virginia state court judges[24] and a 2001 study that asked fifty-two

17. *See* Willging, *supra* note 15, at 1127-34 (describing experimental studies by the FJC and others using differential case assignments).

18. *See* Chris Guthrie et. al., *Blinking on the Bench: How Judges Decide Cases,* 93 CORNELL L. REV. 1 (2007); J. Rachlinski et. al., *Inside the Bankruptcy Judge's Mind*, 86 B.U. L. REV. 1227, 1230 (2006); Andrew J. Wistrich et. al., *Can Judges Ignore Inadmissible Information? The Difficulty of Deliberately Disregarding,* 153 U. PA. L. REV. 1251, 1258 (2005); Chris Guthrie et al., *Inside the Judicial Mind,* 86 CORNELL L. REV. 777 (2001).

19. Fischman, *supra* note 8, at 150.

20. Anthony Partridge & William Eldridge, *The Second Circuit Sentencing Study: A Report to the Judges of the Second Circuit* 1-4 (1974) (hereinafter "Second Circuit Sentencing Study").

21. *Id.* at 38-39.

22. *Id.* at 23-40.

23. Stephen Breyer, *The Federal Sentencing Guidelines and the Key Compromises Upon Which They Rest*, 17 HOFSTRA L. REV. 1, 4 (1988) (citing congressional records indicating that Congress relied on the Second Circuit Sentencing Study in adopting federal sentencing guidelines).

24. William Austin & Thomas Williams III, *A Survey of Judges' Responses to Simulated Legal Cases: Research Note on Sentencing Disparity*, 68 J. CRIM. L. & CRIMINOLOGY 306 (1977).

Spanish judges to review the transcript of a single actual rape trial and impose a sentence.[25] A 1984 study in the Netherlands presented nine mock cases to 114 judges in the civil litigation context and also found nothing to substantiate the hypothesized explanatory effect of personal characteristics on outcomes, but the results nonetheless showed variation in outcomes.[26]

Experimental studies presenting the same fact pattern to multiple judges are rare.[27] Such research is costly and time-consuming, and requires obtaining the cooperation of a large number of judges, which perhaps accounts for its comparatively infrequent use.[28] And these studies come with their own set of limitations, of course. Creating data in controlled conditions lacks the consequence of reaching outcomes with the rights of actual people at stake.[29] Therefore, the artificial conditions of the experiment that control for case variations necessarily fail to replicate the real-world conditions of the decision-maker.

In sum, the study reported here rests in a particularly dark corner of empirical legal research. It uses experimental data. The decisions studied are at the trial court level on non-dispositive, pretrial litigation disputes. The participating judges are predominantly state trial court judges, a largely unstudied group. And the decisions, or deferrals of decisions, occur during informal chambers conferences, about which little has been published in any respect.

B. Scholarship on Trial Court Discretion in Pretrial Matters.

This subpart briefly summarizes the theoretical scholarship on the role

25. Ramon Arce, Francisca Fariña, Mercedes Novo & Dolores Seijo, *Judges' decision-making from within*, published in En R. Roesch, R. R. Corrado & R. J. Dempster (Eds.), *Psychology in the courts: international advances in knowledge* 195-206 (Routledge 2001).

26. Peter J. Van Koppen; Jan Ten Kate, *Individual Differences in Judicial Behavior: Personal Characteristics and Private Law Decision-Making*, 18 LAW & SOC'Y REV. 225 (1984).

27. Fischman, *supra* note 8, at 150; Kevin Clermont & Theodore Eisenberg *Litigation Realities*, 88 CORNELL L. REV. 119, 126 (2002) (noting that experimental "methods have long been possible, and for just as long they have gone rarely employed.").

28. Clermont and Eisenberg name those conducting experimental studies as "the real heroes of empirical research," but describe such research as "a drag," and note the mismatch between the training of various types of scholars conducting experimental research on judicial decision-making and the professional rewards for such research. Clermont & Eisenberg*, supra* note 27, at 126. Fischman, *supra* note 16, at 59 (noting that as to simulated decision-making studies, "judges may be loath to cooperate, especially if the research could support reforms that the judges oppose.").

29. *See* Epstein, *supra* note 6, at 2071 ("I take the experimental evidence quite seriously, but some members of the legal community (especially judges) do not; they complain that the experiments are artificial and do not capture the real courtroom environment."); Heiss, *supra* note 7, at 846 ("Whatever gains might be realized for comparability purposes, however, are off-set by losses in authenticity."); Fischman, *supra* note 16, at 58-59 (collecting scholarship criticizing the validity of simulated decision-making studies).

of trial court exercise of discretion in decision-making on discovery disputes and other non-dispositive pretrial matters. Legal scholars agree that pretrial procedure has become more important over time, and they agree that trial court judges exercise all but unchecked authority to decide pretrial issues. As Professor Stephen Yeazell has observed, "[c]ourts now devote the bulk of their civil work to such pretrial tasks: ruling on discovery disputes, deciding joinder issues, conducting pretrial and settlement conferences . . . This work is important, required, and often practically dispositive."[30] And as Professor Yeazell and others have further elaborated, such decisions are rarely rigorously reviewed at the appellate level.[31]

Scholars divide on whether the broad discretion given to trial court judges to manage litigation in the pretrial phase improves or hinders the efficient and fair resolution of civil litigation matters.[32] This debate substantially mirrors the well-known scholarship on law as a form of rules versus standards.[33] Advocates for rule-based decision-making express concern over the competency of judges to make effective pretrial decisions given their information limits and the strategic maneuvering of the parties, resulting in unfair or arbitrary decisions that are more costly to obtain.[34] Those inclined toward broad judicial pretrial discretion argue that this system allows for efficient and fair procedures appropriate to differing case conditions.[35]

---

30. Yeazell, *supra* note 4, at 639.

31. *Id.* at 660-64; Elizabeth G. Thornburg, *The Managerial Judge Goes to Trial*, 44 U. RICH. L. REV. 1261, 1295 (2010). As Professor Robert Bone put it, "it is only a slight exaggeration to say that federal procedure, especially at the pretrial stage, is largely the trial judge's creation, subject to minimal appellate review." Robert Bone, *Who Decides? A Critical Look at Procedural Discretion*, 28 CARDOZO L. REV. 1961, 1962 (2007).

32. The scholarship occurs in two over-lapping descriptions: evaluations of whether the Rules of Civil Procedure should be reformed to limit judicial discretion over pretrial disputes; and whether "managerial judging," or proactive judicial involvement in cases, is a worthwhile practice. Both of these strands of scholarship are cited below.

33. *See* Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557 (1992). As Professor Robert Effron observed, "[t]his tension between the benefits of ex ante rules specification and ex post decision-making is not susceptible to a unitary solution across all rules and procedural devices." *Reason Giving and Rule Making in Procedural Law*, 65 ALA. L. REV. 683, 689 (2014).

34. *See, e.g.,* Thornberg, *supra* note 31, at 1300 ("[t]he result is a hodgepodge of potentially inconsistent rulings that vary from judge to judge and case to case."); Bone, *supra* note 31, at 1964 ("[R]ulemakers should be much more skeptical of delegating (procedural) discretion to trial judges and should seriously consider adopting rules that limit or channel discretion more aggressively."); Todd D. Peterson, *Restoring Structural Checks on Judicial Power in the Era of Managerial Judging,* 28 U.C. DAVIS L. REV. 41, 46 (1995) (arguing as to pretrial litigation management, "that we must limit judges' unchecked powers or restrain their use of this power.").

35. *See, e.g.,* Pauline T. Kim, *Lower Court Discretion*, 82 N.Y.U. L. REV. 383, 438 (2007) ("district courts are in a better position to make decisions regarding the management of litigation."); Richard L. Marcus, *Slouching Toward Discretion,* 78 NOTRE DAME L. REV. 1561, 1564 (2003) ("[a]lthough this examination confirms the views of the critics that discretionary activity has indeed increased (in pretrial procedure), I view these developments with less alarm than some others.").

The question of judicial discretion in implementing procedural rules is at the heart of the fairness of those rules in practice. Fairness logically dictates "that cases with relatively similar facts ought to reach relatively similar outcomes."[36] As Professor Robin Effron puts it:

> The push and pull between the appeal of strict procedural rules with consistent outcomes and the realities of the costs and practical difficulties of such rules are reflected in the different attitudes that rulemakers take toward crafting the rules that implement procedural devices. While there are some rules in which consistency is deliberately subordinated to other procedural values, in other instances, the extent to which uniformity should be enforced or imposed is either unstated or clearly aspirational. It is in these situations that tensions in interpreting and applying procedural rules can arise. Although judicial discretion is not the only procedural value to compete with uniformity, it looms large in the debate ….[37]

The study presented here provides an empirical reference point for evaluating this debate by offering experimental data on how judges differ when actually exercising discretion concerning the same facts and procedural issue.

III. STUDY METHODOLOGY AND LIMITATIONS

This Part explains the Law in Practice course and its simulations from which the study data was derived. This Part also describes how the data was coded from the reported simulation results and discusses the study's limitations.

A. Law in Practice Chambers Conference Simulations

All first-year students at the University of Minnesota Law School are enrolled in a course entitled "Law in Practice." Students engage in a series of simulations based on two case files—one litigation file, which draws from the facts of a litigated case, and one transactional file. The facts for each case file are provided to the students over the course of the semester on a weekly basis. For the litigation case file, which is of sole relevance here, the facts are presented in sequence as they would be revealed to the attorney in a typical litigation matter, starting with the client interview and then changing as documents and other interviews are conducted and as information is obtained in discovery.

The course has two components—a "law firm" classroom section and a smaller "practice group." A full-time law school faculty member teaches

36. William B. Rubenstein, *The Concept of Equality in Civil Procedure*, 23 CARDOZO L. REV. 1865, 1893 (2002).

37. Effron, *supra* note 33, at 695.

the weekly law firm class for a section of students, from forty to fifty students, in which students learn the law underlying the legal disputes in the case files and the concepts used in the performance of skills needed in legal practice. For example, in one of the litigation case files, the class studies the federal Age Discrimination in Employment Act ("ADEA"), discusses how to analyze the merits of an ADEA claim under the known facts and circumstances of the litigation file, reads about distributive negotiating theory, then uses that merits analysis and the negotiating theory to derive a proposed settlement offer for client consideration.

Students also participate in a "practice group" of approximately eight students, which is supervised by a practicing attorney acting as an adjunct instructor of the law school. The primary focus of the practice group is conducting a series of simulations based on the case files. For the litigation case file, students conduct a client interview and witness interview, take and defend a deposition of a third-party witness, argue issues arising from pretrial disputes before a judge in a chambers conference, and participate in a mediation conducted by a qualified neutral. The data for this study are derived from the chambers conference simulation.

The Law in Practice course has used the following three litigation case files: (1) the "Flores" file is a joinder case involving a claim of wage underpayment filed by eleven former janitors against a local grocery chain for acting as a joint employer in misclassification of the janitors as independent contractors; (2) the "Aldrich" file is a suit by a smaller regional bank against a larger bank for fraud by non-disclosure of information as part of loan refinance by a business borrower; and (3) the "Muessig" file is an age discrimination suit by a high school art teacher against a private school. Over the six years in which these files were used in the course, the Flores file was used once (2018), the Aldrich file was used twice (2013 and 2016), and the Muessig filed was used four times (2014, 2015, 2017 and 2019). All of these files area loosely based on the facts in a reported case.

B. Chambers Conference Simulations

The chambers conference simulations occur after students conduct interviews and a deposition, and after students received information about initial discovery responses, including document production, interrogatory responses, and excerpts from deposition transcripts. This subpart describes the issues that students were instructed to present for resolution in the chambers conference and describes the conduct of the chambers conference simulations by the volunteer judges.

1. Issues Presented for Resolution in the Chambers Conferences

Each of the three case files present a discrete set of issues to be resolved at the chambers conference. The issues were designed to reflect typical pretrial disputes. The number of issues varied for each case file, with four issues in the Flores case, three issues in the Aldrich case, and two issues in the Muessig case. Thus, judges determined a total of nine different issues in the various chambers conference simulations.[38]

Five of the issues involve a dispute about production of requested documents resolvable by reference to the "proportionality" standards provided in Federal Rule 26(b)(1).[39] The information requested clearly was relevant for purposes of discovery, so the question presented to the judge was whether the requested document production met the requirement in Rule 26(b)(1) that the request be:

> proportional to the needs of the case, considering the importance of the issues at stake in the action, the amount in controversy, the parties' relative access to relevant information, the parties' resources, the importance of the discovery in resolving the issues, and whether the burden or expense of the proposed discovery outweighs its likely benefit.[40]

The other four issues presented a variety of pretrial matters: whether the eleven plaintiffs in the Flores case should be allowed to amend their complaint to add three additional plaintiffs after the date indicated in the Scheduling Order for amending pleadings;[41] whether the plaintiff in the Aldrich case should be allowed to take more than the presumptive limit of ten depositions;[42] whether the court should adopt a protective order proposed by the defendant in the Aldrich case;[43] and whether the judge in the Muessig case should grant the defendant's request that the plaintiff produce her medical records given her request for "garden variety" emotional damages.

Appendix A describes in more detail the nine issues presented during the chambers conferences. Table 1 below provides a summary of each of the nine issues.

---

38. During the period of the simulations, Minnesota courts employed civil procedure rules on eight of the nine issues that were materially the same as the Federal Rules of Civil Procedure. *Compare* Fed.R. CIV. P. 15(a)(2) and MINN. R. CIV. P. 15.01; FED. R. CIV. P. 26(b)(1) and MINN. R. CIV. P. 26.02(b) (conformed to federal rule in 2006). On one issue—the number of permitted deposition—the Federal Rules use a presumptive limit of ten, FED. R. CIV. P. 30(a)(2)(A)(i), while the analogous Minnesota rules have no such limit, MINN. R. CIV. P. 30.

39. FED. R. CIV. P. 26(b)(1).

40. *Id.*

41. Amendment of complaints following the initial pleading period is governed under FED. R. CIV. P. 15(a)(2).

42. FED. R. CIV. P. 30(a)(2)(i) establishes a presumptive limit of ten depositions per side.

43. The grant of a protective order is governed under FED. R. CIV. P. 26(c).

Table 1: Description of Issues in Chambers Conference Simulations

#	Case	Issue Type	Issue Description	Notes re: Outcomes
1	Flores	Amend Pleading	Plaintiff demand to add three plaintiffs after deadline in Sched Order for Complaint amendment.	No reasonable option exists for compromise on this issue.
2	Flores	Document Production Proportionality	Plaintiff demand to produce requested outsource contracts from other types of vendors.	
3	Flores	Document Production Proportionality	Plaintiff demand to produce communications related to above contracts.	Decision for Defendant on Issue on #2 Decision for Defendant on Issue #3.
4	Flores	Document Production Proportionality	Plaintiff demand to produce records of personnel policy violations by employees other than plaintiffs.	
5	Aldrich	# of Depositions Permitted	Plaintiff demand to take more than ten depositions.	Plaintiff requests 18. Defendant wants 10 limit.
6	Aldrich	Terms of Protective Order	Defendant demand for terms of proposed protective order.	Plaintiff does not have clear position on desired outcome

				priorities.
7	Aldrich	Document Production Proportionality	Plaintiff demand to produce records of violations of defendant lending policy with banks other than plaintiff.	
8	Muessig	Document Production Proportionality	Plaintiff demand to produce entire personnel files of employees other than plaintiff.	
9	Muessig	Document Production + Resolution of uncertain legal issue	Defendant demand for plaintiff's medical records related to claim for emotional distress damages, with law unclear on whether type of damages claim matters.	Plaintiff wants no production but be allowed to seek "garden variety" emotional damages. Defendant wants records or drop any claim for emotional damages.

With only one exception, the parties had preferred outcomes on each of the issues.[44] For judges, each issue was designed so that it was reasonable for a judge to resolve the issue with more than one outcome.

Over the six-year period included in this study, 166 chambers conference simulations were completed with the participation of 61 judges. The four issues in the Flores case were presented in 24

44. The exception is the plaintiff had no clear position on issue #6 as to what constitutes success in opposition to the proposed protective order submitted by the defendant. *See infra* Appendix A.

simulations; the three issues in the Aldrich case were presented in 40 simulations; and the two issues in the Muessig case were presented in 102 simulations. Thus, judges were presented with a total of 420 issues for resolution in these conferences. For twenty of these issues, the resolution was missing on the reporting forms or too incomplete or unclear to code. Thus, we have data on the resolution of exactly 400 issues presented in the 166 chamber conferences.[45]

2. Judge Participation in Chambers Conferences

No procedure or outcome was prescribed for the simulation. Judges were instructed that "[w]e do not have a prescribed outcome for your chambers conference. Please handle the matter as is warranted by the facts presented and the arguments of the students." For the two case files with a state court venue, the judges were told that the state civil procedure rules were identical to the Federal Rules of Civil Procedure. Judges were given a summary of the issues to be considered in the simulation, the key facts related to those issues, and relevant case law, if any. Judges were instructed that "[t]his summary is all that you need to read to adequately conduct the simulation." Judges also were given documents provided to the students that form the basis of the dispute, including pleadings, communications between counsel, case law, and prior court orders.

Two students for each side appeared before the judge at the chambers conference. The practice group instructor observed the simulation. Judges typically engaged in a discussion with the students about the conduct of the simulation following the formal simulation. The course provided the participating students with a form identifying each issue that should have been presented for that chambers conference simulation and asking the students to state the outcome for that issue. Students from the opposing sides cooperatively completed these forms and returned them to the program administrator.

The judges overwhelmingly were trial courts judges, including forty-two state district court judges, five federal magistrate judges; five state court referees; one federal district court judge; and two administrative law judges. The remaining six judges mostly were from state appellate courts. The mean tenure of the judges was ten years and the median tenure was eight years, with a range from one to thirty-two years on the bench.

45. One of the 61 judges conducted only one simulation and both of the issues presented to this judge were among the 20 issues we were unable to code. Therefore, the data presented here is recorded for 60 judges rather than 61 judges.

C. Study Issues and Data Analysis

This study looks primarily at the extent to which judges diverge in deciding the same discovery and other pretrial, non-dispositive issue when presented with the same fact pattern and law in the context of a chambers conference. The course provided the participating students with a form identifying each issue that should have been presented for that chambers conference simulation and asking the students to state the outcome for that issue. Students from the opposing sides cooperatively completed these forms and returned them to the program administrator. The data on simulation results used in this study was derived exclusively from these forms. The results for each issue in each simulation were coded into numerous specific outcomes that commonly occurred for each issue. These results were then grouped into the following four broad categories:[46]

(1) plaintiff prevailed (i.e., plaintiff achieved its desired outcome);
(2) compromise (i.e., the result was not the outcome desired by either party);
(3) defendant prevailed (i.e., defendant achieved its desired outcome); or
(4) the judge deferred a decision on the issue.[47]

In addition, this study examined whether differences in outcomes could be explained by identifiable characteristics of the judges making the decision, which is the type of causative question of concern in most empirical studies of judicial decision-making. To answer this question, data was coded on the following thirteen characteristics of each of the judges who participated in the simulations: number of years in judicial office; whether the judge is currently on the bench or retired; the particular district court for state trial court attorneys; gender; race; prior attorney employment as a prosecutor, public defender, non-profit attorney, civil government attorney, legislator, or attorney in a major law firm;[48] the governor or president who appointed the judge, and the party of the appointer, if any; and a measure of ideological affiliation based on

46. For 20 of the 364 issues that were to be presented at the chambers conference, no outcome within these four categories occurred and thus these 20 issues were excluded from the results. The reasons for no recordable outcome included the following: the issue not discussed by the students at the simulation; the students resolved the issue without direction by the judge; the judge took no recordable action; or the response on the form was unclear. We also coded whether the judge *sua sponte* discussed settlement of the case with the parties.

47. Judicial deferral of the issue took the following forms: the judge directing the parties to continue discussions on the matter in an attempt to resolve, the judge took the issue under advisement, or the judge directed the parties to file a motion or otherwise submit briefs on the matter.

48. A major law firm was defined as a law firm with more than 100 attorneys in 2018, regardless of when the judge was part of the firm. For judges who worked for law firms that have merged, the number of attorneys in the surviving firm was used.

publicly available data on contributions to candidates for elected office. All of this information was obtained solely from publicly available data.[49]

D. Study Limitations

Two obvious limits for the use of the study data are: (1) that it is a simulation and thus an artificial environment, and (2) that students appeared at the chambers conference and made arguments of varying substance and quality that influenced the outcome.

The former limitation mostly is what it is; experimental studies have the advantage of presenting the same case to multiple judges and the disadvantage of an artificial environment. Two factors, however, weigh in favor of the accuracy of these results for real events compared to other experimental studies. First, unlike the other experiential studies, the chambers conference simulations attempted to mimic an actual litigation event rather than present a paper case file. Also, unlike other experimental studies, there is some consequence here, albeit evaluation of students rather than real consequence for litigants or the public.

The presence of students, however, may exacerbate the second limitation regarding the relative impact of student argument on the outcome. Unlike other experimental simulation studies, judges could be influenced by student arguments even though the underlying facts for each like simulation are the same. Yet this could be seen as creating a more realistic setting because it parallels actual practice. Attorneys appear at chambers conferences and argue for positions, thus influencing outcomes.[50]

IV. STUDY RESULTS

This Part presents the study results in three subparts. The first subpart presents an overview of the results. The second subpart provides data illuminating the degree of variation among judges as a determinant of these results. The third subpart looks at whether this variation can be

49. Biographical data was available on court websites for all state and federal court judges in Minnesota allowing to code all of the fields for all of the judges other than the data on political contributions. The data used for the political contribution field was derived from Adam Bonica, *Database on Ideology, Money in Politics, and Elections: Public version 2.0*, STAN. UNIV. LIBRARY (Mar. 11, 2016, 2:01 PM) ("DIME"), https://data.stanford.edu/dime. Data was extracted from the DIME database for political contributions from 1980 through 2014. Data for this field was available only for 29 of the 61 judges.

50. The author's discussions with the many practicing attorneys who supervise the chambers conference simulations provides anecdotal support for the conclusion that the volunteer judges attempt to create an experience similar to a chambers conference in an actual case, as stated in the instructions to the judges.

explained by characteristics of the judges.

A. Overview of Results

Table 2 provides the simulation results by issue for the four categories of outcomes.

Table 2: Outcomes by Four Categories by Issue

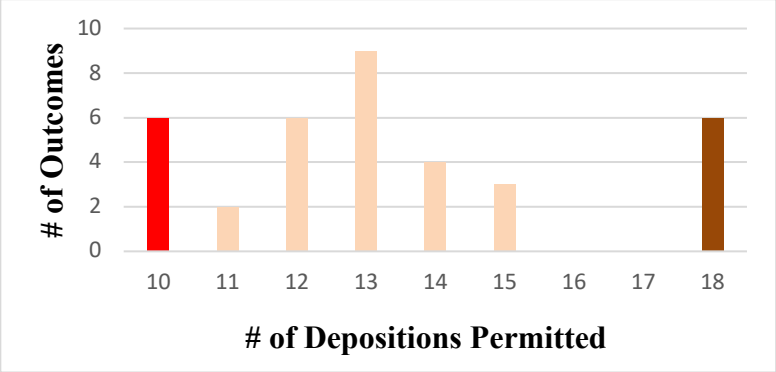
	Plaintiff Prevails	Compromise	Defendant Prevails	Deferred Decision
Issue #1 (n=23)	74%	0%	17%	9%
Issue #2 (n=24)	25%	25%	25%	25%
Issue #3 (n=22)	9%	18%	50%	23%
Issue #4 (n=23)	17%	35%	22%	26%
Issue #5 (n=37)	16%	65%	16%	3%
Issue #6 (n=40)	0%	50%	23%	28%
Issue #7 (n=39)	3%	44%	21%	33%
Issue #8 (n=97)	22%	67%	0%	11%
Issue #9 (n=95)	50%	20%	20%	10%
TOTALS (n=400)	**105 26%**	**163 41%**	**68 17%**	**64 16%**

The overall variation in results is greater than apparent in the summary data because the compromise category is broadly constructed to capture any outcome that is not a ruling wholly for one of the parties or a deferred decision. Issue #5 provides an easy example because the results are numeric. The parties were disputing the number of depositions that would be allowed for the plaintiff. The defendant sought the default limit set in Rule 30, which is ten depositions.[51] The plaintiff sought a total of eighteen depositions. For the thirty-nine simulation decisions reported here, three

51. *See* FED. R. CIV. P. 30(a)(2)(A)(i).

were deferred outcomes and the remaining thirty-six simulation results were as presented in Figure 1.

Figure 1: Number of Depositions Allowed as Outcome of Issue #5



B. Examining Variation of Outcomes by Judge

This subpart examines the question of whether the variation in outcomes is a result of groups of judges consistently reaching the same decisions compared to other groups of judges. Variation in judicial decision-making was measured in three ways: variations in propensity to defer decisions, variation in decisions that favored plaintiff or defendant relative to the mean decision of all judges on the issue; and the likelihood of the judge determining the case consistent with the plurality. Each of these types of variation are explained below. In short, outcome variation is not substantially attributable to some judges consistently deferring or ruling in one direction, although there are a few judges that showed such tendencies.

1. Deferral of Decisions

Deferrals accounted for 64 of the 400 decisions, or 16%. Thirty-one of the sixty judges, or about half, did not defer on any decision. Thirty-three judges decided five decisions or more, and only six (18%) of these judges deferred more than one-third of decisions. These six judges accounted for twenty-two (34%) of the total of sixty-four deferrals.

Table 3 shows the number of judges who deferred by the total number of decisions made by the judge.

Table 3: Number of Judges by Percentage of Deferral Decisions by Total Number of Decisions

# of Decisions by Judge	% of Decisions Deferred			
	0%	**7-24%**	**25-33%**	**34+%**
<5 (n=27)	20	0	4	3
5-10 (n=16)	4	6	2	4
>10 (n=17)	7	5	3	2
TOTAL	31 (52%)	11 (18%)	9 (15%)	9 (15%)

2. Variation by Decision in Favor of Plaintiff or Defendant

Second, the data show only a few judges who made decisions that heavily lean toward the plaintiff or defendant. To evaluate this type of variation, each decision was coded as 1 if ruling wholly for defendant, 0 for a compromise or deferral, and -1 if the decision was wholly for the plaintiff. For each issue, the average score was calculated. Then, for each issued decided by a judge, the deviation from the average score for all judges deciding the issue was calculated. The "mean variation score" for each judge was the mean of the deviation scores across all of that judge's decisions. This method of evaluating variation roughly corresponds to the analysis used in the Second Circuit Sentencing study.[52]

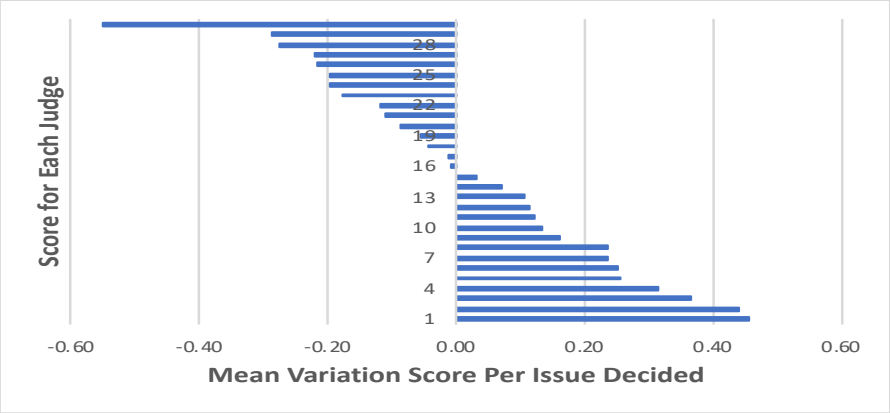
For example, on issue #9, 48 judges decided wholly in favor of plaintiff, 28 judges reached a compromise decision or deferred decision, and 19 judges decided in favor of defendant. The mean score for this issue was -.3053, calculated as follows: [(48*-1)+(28*0)+(19*1)]/[48+28+19]. A judge received a variation score for that issue as follows: a decision in favor of a plaintiff received a score of -.6947, a compromise or deferred decision received a score of .3053, and a decision in favor of defendant received a score of 1.3053. Therefore, the sum of all judges scores was 0. A judge who decided ten issues with a summed variation score of 1.74 would receive a mean variation score of .174 per issue.

A judge consistently deciding for plaintiffs or defendants would have a score of about one or minus one, depending on which issues they decided. For example, the mean scores on issue #5 through issue #9 were 0.0, .2250, .1795, -.2165, and -.3053. A judge deciding only these five issues and resolving each in favor of the defendant would have a mean variation score of 1.0235, calculated as follows: [(1+.7750+.8205+1.2165+1.3053)/5].

52. Second Circuit Sentencing Study, *supra* note 20, at 36-37.

Figure 2 shows the mean variation score per issue for each judge who decided at least five issues. Only four of the thirty-three judges (12%) had a mean variation score by issue of greater than .3 or lesser than -.3. Of the seventeen judges with more than ten decision, only one (6%) had a mean variation score by issue of greater than .3 or lesser than -.3.

Figure 2: Mean Variation Score Per Decision for Each Judge with 5 or more decisions



3. Agreement with Plurality

A third question one can ask regarding variation is whether the differing outcomes reflect one group of judges consistently agreeing and the remaining group consistently reaching a minority decision. Again, the data does not show this pattern. To evaluate this type of variation, a plurality result was identified for each issue and then a judge's result for each issue was compared to determine if the judge reached the plurality result on that issue.

Judges reached a plurality result in 210 of the 400 issues determined, or 53% of the decisions. Considering only the judges with five decisions or more, the result is the same: 53% (176 out of 331) of those judges reached the plurality decision. No judge making five decisions or more reached the plurality outcome on all issues, and only one judge making more than five decisions failed to reach a plurality outcome on any issue. Twenty-three of the thirty-three judges (70%) agreed with the plurality in 40% to 70% of the cases.[53]

53. Interestingly, there was no overlap between the ten judges who did not fall in this 40% to 70% range and the four outlier judges with mean variation scores of more than .3 or less than -.3 noted in the prior subpart.

C. Relationship between Outcome Variation and Judicial Characteristics

No statistically significant relationship was found between the simulation outcomes and any of the identified characteristics of judges. Neither proxies for ideology,[54] the gender or race of the judges, nor prior employment of the judges showed a significant relationship to the simulation outcomes. Perhaps more surprisingly given the focus on pretrial disputes, neither the number of years on the bench nor the type of judgeship held offered a viable explanation for the varying simulation outcomes.

A chi-square test was run to determine if there was a statistically significant relationship at the .05 significance level between the outcomes of the simulations and the thirteen categories of data gathered reflecting judicial characteristics and experience.[55] The test was run for the decision on whether to defer for overall deferral percentage and also for each of the nine individual issues and found no statistically significant relationship with the thirteen categories of data reflecting judicial characteristics and experience. The test was then run for the three categories of decision when the judge did not defer (pro-plaintiff, compromise, pro-defendant) for each of the nine individual issues, and found only one of these 117 tests showed any statistically significant relationship.[56] The judicial characteristics also were tested against the derived measures for outcome variation totaled across all nine issues and no statistically significant relationship was found.

V. ANALYSIS AND MEANING OF STUDY RESULTS

This Part elaborates three contributions from this study for legal scholarship. Subpart A looks at the meaning of the study results for understanding how chambers conferences work to resolve discovery issues. Subpart B notes the similarity of the finding on outcome variance as compared to other experimental studies of judicial decision-making in different contexts. Subpart C discusses the relevance of this experiment for the civil procedure scholarship on whether to limit judicial discretion in pretrial matters.

54. *See supra* note 48 and accompanying text for data used to measure ideology.

55. See *supra* Part III.C. for a list of the thirteen categories of data.

56. Judges with private practice experience were more likely to make a pro-plaintiff decision on Issue #9. This result seems anomalous and likely to have little explanatory value in and of itself.

A. Chambers Conferences Result in Decisions on Pretrial Issues

Litigators regularly engage in chambers conferences to resolve discovery and litigation management disputes.[57] Yet no research appears to exist concerning how judges conduct such conferences. The study findings here suggest two conclusions about how judges resolve pretrial non-dispositive disputes in chambers conferences.

First, judges routinely direct outcomes when the parties appear before them in informal chambers conferences rather than just set the table for later resolution. Only sixteen percent of the conferences ended in a deferred decision, even when deferral was broadly defined to mean anything from telling the parties to work out a result to scheduling a motion hearing to decide the matter. And, deferral of decisions was not substantially attributable to the decisions of a few judges averse to making a decision.

Second, judges are not shy about directing a result for one party in chambers conferences. Litigation events that present the court with a decision on the merits generally require a binary choice, such as a grant or denial of summary judgment motion, or affirmance or reversal on appeal. Chambers conferences, on other hand, typically present issues that allow for a compromise ruling. Nonetheless, judges in the study were about equally likely to direct a compromise decision as to decide wholly for one party or the other. Judges wholly accepted the plaintiff's position or the defendant's position in 43% of the simulations (plaintiffs prevailing in 26% of the conferences and defendants prevailing in 17% of the conferences); compromise decisions constituted 41% of the outcomes. Decisions for one party or other outnumbered compromise for five issues; compromise decisions outnumbered decisions wholly for one party in the other four issues.

B. Variation in Chambers Conference Outcomes Show Similar Patterns to Other Experimental Studies of Judicial Decision-making

Even when considering the results at the level of four broad categories of decisions, the judges directed widely varying resolutions on each of the nine issues presented. Judges reached a clear majority result (51% or more) in three of the nine issues presented. On three other issues, half of the judges reached the same outcome.[58] In the remaining three issues, the judges did not reach even a 50% level of agreement among the four

57. Peterson, *supra* note 3, at 67. ("Judicial management of the pretrial process entails early and multiple pretrial conferences . . .").

58. In two simulations, exactly half of the judges reached the same result, and in a third simulation 50.5% reached a common result.

categories. Notably, on issue #2, 24 judges split evenly among the four options.

In addition to failing to consistently reach a common outcome, judges typically fractured as to the non-plurality result. On issue #9, for example, forty-eight of the ninety-five judges (50.5%) decided for the plaintiff, but the remaining forty-seven decisions were spread out among the other three categories, with nineteen decisions each for compromise and for the defendant, and nine deferrals. Of the thirty-four possible outcomes over all the issues,[59] only one—ruling for the defendant on issue eight—was not a result that occurred in the simulations.

Using variation from the mean as a measure of judicial decisions in favor of plaintiff or defendant, as described above, it appears that while judges reach varying results on individual issues, they do so primarily by deciding relatively favorably for plaintiffs in some cases and relatively favorably for defendants in other cases. The aggregate result is that the total variation from the mean score for an overwhelming number of judges hovers near zero. There are a few outliers, as shown in Figure 2. Two of the thirty-three judges with five or more decisions had notably higher scores, indicating a pro-defendant slant, and at least one of the judges had a particularly high negative score, indicating a pro-plaintiff slant. But the variation mostly reflects decisions in different directions on various issues that average out to scores, with little indication of a strong overall slant in favor of either the plaintiff or defendant across multiple determinations.

The result of substantial variation in outcomes when judges decide the same simulated dispute is consistent with the finding of other experimental studies. In the Second Circuit Sentencing Study, the authors found "a wide range of disagreement among Second Circuit judges about the appropriate sentences in the twenty cases. Substantial disagreement persists, moreover, even if the extremes of the distribution are ignored."[60] In the study of Spanish judges asked to determine the merits of a criminal trial, the researchers found that, "[s]trikingly, of the 52 judges under study, half reached a not-guilty verdict, and the other half a guilty verdict."[61] And the Virginia sentencing study concluded that "it is clear that when legal cases are equalized within offense categories, judges still show substantial disparity on all three criteria," with the criteria being verdict, sentencing mode, and magnitude of penalty.[62]

59. There were four outcomes for nine issues, which would result in thirty-six outcomes, but two of those outcomes (finding a compromise on issue #1 and ruling for plaintiff on issue #6) were not possible due to the design of the simulation problem. *See infra* Appendix A.

60. Second Circuit Sentencing Study, *supra* note 20, at 9.

61. Arce, et. al., *supra* note 25, at 200.

62. Austin & Williams, *supra* note 24, at 309.

Perhaps more importantly, the simulation results here are remarkably similar to a key finding of the Second Circuit Sentencing Study. In that study, judges were ranked against each other from most lenient to most severe for each case file on which they imposed a hypothetical sentence.[63] The researchers then averaged the judges ranks in an attempt to determine "whether the disparity observed is a function of some judges habitually rendering relatively severe sentences while others habitually render light ones."[64] They found that "most of the judges had average ranks quite close to the center," although there were a few outliers. The researchers concluded that:

> The overwhelming majority of the Second Circuit judges are sometimes severe relative to their colleagues and sometimes lenient. If there are indeed "hanging judges" and lenient ones -- and it would appear that there are a few -- their contribution to the disparity problem is minor compared to the contribution made by judges who cannot be so characterized.[65]

The finding of a total lack of statistically significant correlation between the varying judicial outcomes and the thirteen characteristics coded for each judge is also consistent with other empirical studies of trial court decision-making. While studies have found one or more of these factors to have explanatory value for judicial outcomes, studies also show less discernible influence of ideological or personal characteristics on judicial decision-making at the trial court level.[66] As Michael Heise summarized this research, "precisely when lack of comparability is reduced by holding to a case-specific scenario, political affiliation as a predictor declines."[67]

C. Reform of Civil Procedure to Reduce Judicial Discretion Would Substantially Disrupt Current Practice

The study results have at least two implications for the scholarly debate between the advocates of civil procedure rules with high levels of judicial discretion and those seeking to restrict that discretion. First, the high level of decisional variance suggests that limiting judicial pretrial discretion would be a wrenching process. Second, adding nudges and presumptions to the Rules may not do much to obtain the results desired by those who prize consistency in outcomes.

---

63. Second Circuit Sentencing Study, *supra* note 20, at 36-40.

64. *Id.* at 23.

65. *Id.*

66. Herbert Kritzer, *Polarized Justice? Changing Patterns of Decision-Making in the Federal Courts*, 28 KAN. J.L. & PUB. POL'Y 309, 356-57 (2019).

67. Heise, *supra* note 7, at 838.

1. Limiting Discretion Will Not Come Easy

The clear finding of the study is that it is not the case that judges faced with the same issue and facts will reach the same result. This may be seen as support for a concern about arbitrarily uneven outcomes in application of the civil procedure rules. But this fracturing also means that imposition of a rigid set of decision-making rules would be a notable change from current pretrial practice. Advocates of consistent outcomes would have to accept that developing more determinative rules likely would result in an upheaval—for better or worse—in pretrial litigation outcomes.

A rule with a clearly determined outcome may be even more disruptive than just reversing a fractured result because the bright-line rule may be the exact opposite of the majority judicial position when exercising discretion. For example, issue #1 presented the problem of whether a plaintiff should be allowed to amend a complaint after expiration of the date for pleading amendment in the court's scheduling order. Unlike seven of the nine issues, there were only three possible outcomes on this issue because a compromise result was not a reasonable position.[68] Issue #1 resulted in the highest percentage agreement, with 74% of judges accepting the plaintiff's request to amend the complaint. The obvious bright-line rule here, however, is to prohibit further complaint amendment after the scheduling order deadline—a position taken by just 17% of the judges.

Another reason to pause in considering adoption of rigid pretrial rules is the large number of compromise decisions. In addition to the 41% of decisions resulting in compromise, another 16% of decisions were deferrals. For example, two different files contained a dispute over the production of personnel records of employees not directly involved in the claims at issue. The defendants in both files argued that production of the records intruded on employee privacy concerns and was too costly. The plurality result was a compromise in both files, but one file (issue #4) showed a close fracturing across all four categories while the other file (issue #8) showed a strong majority for compromise with none of the ninety-seven decisions for the defendant—the only possible result among the nine issues for which no judge decided. It is difficult to imagine a bright-line rule, even one specific to production of personnel records in litigation, that would not constitute a disruption of these outcomes. Such a rule might direct an outcome for one party or the other or it might adopt one version of a compromise reached, but it almost certainly would involve a wholesale overturning of these results.

68. *See infra* Appendix A.

2. Presumptive and Detailed Rules May Not Do Much to Restrict the Exercise of Discretion in Pretrial Rulings

The results of this study also suggest that trying to reach more predictable outcomes through nudges and presumptions might be futile. The Federal Rules of Civil Procedure are uneven in permitting judges to exercise discretion employing broad standards versus precise rules. In some decisions, judges are given only the vaguest standards for their decisions, thereby explicitly granting the judge broad discretionary authority. For example, pleading amendment after initial filings is allowed over the objection of an opposing party "when justice so requires."[69] Judges make other pretrial decisions under standards that prescribe certain factors to consider but provide discretion to weigh these considerations as they see best in directing an outcome. Most discovery disputes present a question of proportionality for which the trial court is directed to apply a six factor test.[70] Finally, judges sometimes exercise discretion only when asked to reverse a presumptive outcome prescribed in the rules, such as a request to allow a party to take more than ten depositions and serve more than twenty five interrogatories absent opposing party agreement.[71] A mix of these types of discretionary decisions exists in the issues determined in the study simulations.

Judges are notoriously averse to limits on their exercise of discretion, "particularly in this area of managing the cases that come before them."[72] Judges in the study results reported here did not appear to be constrained in exercising that discretion when confronted with an issue that required application of a civil procedure rule containing a specific presumption or detailed criteria.

Issue #5 required the judges to determine whether the plaintiff could exceed the presumptive limit of ten depositions found in Rule 30(a)(2), which is permitted if the judge determines that doing so is consistent with

69. FED. R. CIV. P. 15(a)(2).

70. FED. R. CIV. P. 26(b)(1).

71. FED. R. CIV. P. 30(a)(2); FED. R. CIV. P. 33(a)(1); FED. R. CIV. P. 26(b)(2).

72. Judith A. McMorrow, *The (F)Utility of Rules: Regulating Attorney Conduct in Federal Court Practice*, 58 SMU L. REV. 3, 44–45 (2005). Judicial resistance to limits on discretion has been observed a variety of contexts. *See, e.g.,* Mary Berkheiser, *The Fiction of Juvenile Right to Counsel: Waiver in the Juvenile Courts*, 54 FLA. L. REV. 577, 617 (2002) ("[t]he ineffectiveness of duly enacted statutes and court rules in ensuring the validity of juvenile waivers demonstrates the resistance of juvenile court judges to externally imposed limits on their discretion."); Daniel J. Freed, *Federal Sentencing in the Wake of Guidelines: Unacceptable Limits on the Discretion of Sentencers*, 101 YALE L.J. 1681, 1719-21 (1992) (describing the negative reaction of federal district court judges to the imposition of criminal sentencing guidelines); Fred C. Zacharias, *A Critical Look at Rules Governing Grand Jury Subpoenas of Attorneys*, 76 MINN. L. REV. 917, 935–36 (1992) (". . . judges steadfastly have resisted imposing limits upon grand jury secrecy, discretion to cast a wide investigative net, and prerogative to act free of outside interference.").

Rule 26(b) principles. Only six of the thirty-seven (16%) of judges upheld the presumptive limit of ten depositions, with another six allowing the plaintiff's request of eighteen depositions and twenty-four judges finding a compromise number of depositions between the two positions.

Professor Robert Bone, an advocate of constricting judicial pretrial discretion, employs the following example involving the ten deposition limit to raise this question about the impact of judicial discretion in determining pretrial disputes:

> The idea of including express numerical limits is sound, but making them presumptive and thus allowing exceptions undermines the strictness of the rule and interferes with the ability of parties to credibly pre-commit. Of course, the extent of interference depends on how readily trial judges grant exceptions, and it is difficult to know for sure how often this happens … In fact, if it turned out that trial judges actually granted exceptions only rarely, that fact would tend to support a strict rule, since a strict rule would get it right virtually all the time and also save the litigation costs of determining individual exceptions.[73]

While the study results do not answer the question of how often judges would be willing to depart from the deposition limit across a range of fact patterns, the results reported here show that at least in one simulated circumstance, few judges adhered to the rule presumption when presented a case with reasonable arguments for each side.

Other issues determined in the simulations similarly provide a suggestion that judges were no less likely (and perhaps more likely) to exercise discretion with varying results when applying rules that set forth detailed decision-making criteria. Of the remaining eight issues, the two decisions requiring application of a general principle created more agreement than the six issues requiring application of a rule with detailed criteria. Issue #1 invoked the Rule 15(a) standard that "[t]he court should freely give leave to amend when justice so requires." Issue #6 concerned a request for a protective order, which under Rule 26(c) may issue if the court "for good cause" determines that it is needed to "protect a party or person from annoyance, embarrassment, oppression, or undue burden or expense." These simulation issues required construing general principles that provide maximum discretion and minimum express direction to the trial court judge. The results on these two issues showed more than usual agreement among the decision-makers, with issue #1 resulting in the highest percentage common result (73.9%) and issue #6 resulting in the median percentage common result (50.0%).[74]

73. Bone, *supra* note 31, at 2008.

74. Both issues were designed to give reasonable arguments for each side, but issue #1 did not allow for a reasonable compromise result and no judge reached a compromise result, and issue #6 did not allow for a clear result in favor of plaintiff and so no such result was recorded. That these two issues

The remaining six issues involved some variant of a discovery "proportionality" dispute, which required application of six factors identified in Rule 26(b)(1) to determine whether to allow the party to obtain discovery.[75] Overall, these issues produced the most disagreement among the judges, accounting for all three of the issues resulting in less than a 50% common result. One of these six issues, issue #9, combined presentation of an uncertain legal issue as a predicate to determining the discovery proportionality issue, and that issue resulted in 50.5% of the decisions determined for plaintiff.

There are two reasons to be cautious in reaching conclusions about the generalizability of these particular results. First, comparing one set of issues here to another set of issues is apples and oranges, as the underlying fact patterns and equities likely have more to do with variation in outcomes than any purported limits on discretion in the civil procedure rule applicable to the dispute. Second, the best example above—the ten-deposition limit—is the one situation relevant here in which Minnesota state civil procedure rules differ from the federal rules, as Minnesota courts have no analogous limit on the number of depositions. Although judges were instructed to apply the federal rules and students were instructed to argue the existence of the federal rule limit, it may be that the state court judges who filled most of the ranks of volunteers for the simulations applied the state rules with which they are familiar. However, of the thirty-seven judges conducting this simulation, four were federal judges and all four of these judges reached a compromise result rather than adhering to the ten-deposition limit.

Limitations noted, the study results provide some evidence that nudges and presumptions may not do much, or do anything, to obtain the results desired by those who prize outcome consistency. Judges given discretion may make little distinction between deciding an issue under a general rule providing discretion and deciding an issue with a set of factors or a presumption that ultimately still allow for judicial exercise of discretion. Experimental studies designed around this particular question would be useful.

V. CONCLUSION

This experimental study presents data from a robust set of 400 pretrial

effectively had three options instead of four likely skewed the results in favor of higher levels of agreement.

75. Fed. R. Civ. P. 26(b)(1) provides: "Parties may obtain discovery regarding any nonprivileged matter that is relevant to any party's claim or defense and proportional to the needs of the case, considering the importance of the issues at stake in the action, the amount in controversy, the parties' relative access to relevant information, the parties' resources, the importance of the discovery in resolving the issues, and whether the burden or expense of the proposed discovery outweighs its likely benefit."

issues decided in 166 simulated chambers conferences presenting nine disputed issues. The results provide an empirical look at how judges conduct chambers conferences, and how judges vary in resolving discovery and other non-dispositive pretrial disputes. While practices vary, judges overwhelmingly use chambers conferences to resolve the disputed issues presented rather than postponing determination of the issue. The resolutions observed include both compromise results and rulings wholly in favor of one party or the other. Consistent with other experimental studies on judicial decision-making, judges vary widely in determining outcomes. Those disparities, however, are not explainable by groups of judges consistently ruling in an opposing direction or manner. Nor is the outcome variation explainable by reference to judicial characteristic or experience. These results should be useful to scholars of civil procedure and to attorneys who engage in litigation in our public courts.

APPENDIX A: DESCRIPTION OF NINE SIMIULATION ISSUES

A. Flores Simulation

Flores Simulation: Joinder of 11 former janitors who cleaned stores for a local grocery store chain claiming that grocer was joint employer with janitorial services company that hired the janitors and classified them as independent contractor.

Issue #1: Plaintiffs seek to add three additional former janitors as plaintiff in the lawsuit by filing a Third Amended Complaint. The deadline in the Scheduling Order for complaint amendment has passed by more than two months. Defendant consented to two prior amended complaints to add plaintiffs prior to the expiration of the Scheduling Order deadline, but objects to this amendment as untimely. A compromise solution on this issue was not substantially viable option because the three putative plaintiffs were all similarly positioned relative to the litigation and delay would exacerbate the harms claimed by Defendant from allowing amendment.

Issue #2: Plaintiffs demand production of written contracts with third-party service providers for the Defendant. Plaintiffs seek the contracts to show the difference in supervision and control of the workers for the janitorial services company and the workers for the other third-party service providers. Defendant objects that these documents are confidential commercial information and it would be burdensome to produce because it has over 30 such contracts with multiple versions operative at different times for the same service.

Issue #3: Plaintiffs demand production of documents that reflect communications from or to Defendant concerning the terms of or negotiations concerning any contract produced in response to the request identified in Issue #2. Accordingly, a ruling wholly for Defendant on Issue #2 (i.e., no documents will be ordered to be produced) necessitates a ruling wholly for Defendant on Issue #3.

Issue #4: Plaintiffs demand production of documents that identify a disciplinary action or other negative evaluation of job performance taken against employees of Defendant performing janitorial duties. Defendant objects to the request as burdensome and raises privacy concerns on behalf of its employees whose records would have to be revealed.

Aldrich Simulation: Smaller regional bank sues larger bank for fraud by failing to disclose as part of loan refinance that business borrower and its principal were engaged in fraud.

Issue #5: Plaintiff seeks to take eighteen depositions, which is eight over the limit of ten in Fed. R. Civ. P. 30(a). Plaintiff has deposed four individuals to date and has noticed but not completed three depositions of three other people. Plaintiff has identified to Defendant and the court eleven other people it wants to depose. Defendant views these depositions as a burdensome "fishing expedition" and will not consent.

Issue #6: Plaintiff demands Defendant's commercial lending policies. Defendant is willing to provide these policies only if Plaintiff agrees to tight restrictions on their use in a draft protective order provided by Defendant. Plaintiff does not consent to Defendant's proposed protective order, but also does not present a clear alternative to that order, and thus there is no wholly pro-plaintiff decision possible on this issue.

Issue #7: Plaintiff demands all Defendant documents that identify "any action taken by Defendant that violates the Defendant commercial lending policy." Defendant refuses to produce these documents for two reasons: burden and confidentiality. Defendant estimates that it will cost in excess of $500,000 to conduct a thorough review of its records to identify all documents relevant to this request.

B. Muessig Simulation

Muessig Simulation: Fired teacher sues private high school for violation of the federal Age Discrimination in Employment Act (ADEA).

Issue #8: Plaintiff demands Defendant produce the entire personnel file for all teachers who did not possess a master's degree. This information clearly is relevant because Plaintiff was fired for not possession a master's degree. Defendant has previously produced a spreadsheet with information about each of its teachers' educational credentials, date of hire and subject matter taught. Defendant objects to the request as burdensome and cumulative, and it also raises privacy concerns on behalf of its employees whose records would have to be revealed.

Issue #9: Defendant demands access to Plaintiff's medical treatment records, especially mental health records. In her Complaint, Plaintiff sought damages for embarrassment, humiliation or the like, but in her deposition, Plaintiff said her damage claim included loss for trouble sleeping as the result of her termination. Judges and students are instructed that the appellate courts in the jurisdiction have not ruled on the exact contours of a waiver of mental health treatment records when the plaintiff seeks emotional damages, and that the proper law is a matter of first impression.