Estimating the Effect of Leisure on Judicial Performance*

Tom S. Clark[†] Benjamin G. Engst[‡] Jeffrey K. Staton[§]

February 19, 2018

Abstract

Past research suggests that natural preferences for leisure influence the way in which federal judges carry out their work. We consider the extent to which leisure incentives reduce the speed with which judges work and the quality of their output. We take advantage of a natural experiment caused by an annual sporting event that creates differential distractions across judges. Using a difference-in-differences design, among federal court of appeals judges we show that a judge's alma mater's participation in the NCAA Mens' Basketball Tournament both slows the rate at which opinions are drafted and ultimately undermines opinion quality, even accounting for the additional time judges spend writing the opinion. The findings suggest that leisure incentives influence important normative concerns for swift and high quality justice.

*We thank Chuck Cameron, Adam Glynn, Thomas Gschwend, Kosuke Imai, John Kastellec and Royce de R. Barondes for helpful suggestions and guidance.

[†]Department of Political Science, Emory University

[‡]Center for Doctoral Studies in Social and Behavioral Sciences, University of Mannheim

[§]Department of Political Science, Emory University

1 Introduction

One of the most significant lessons of the social science of law and courts during the 20th century might be summarized as follows: "judges are people, too." A decades-long debate about the role of ideology in judicial decision-making was primarily about whether, how much and under what conditions judges reference their own world views when exercising their authority (e.g. Segal and Spaeth 2002, Epstein and Knight 1998, Maltzman, Spriggs and Wahlbeck 2000). Though ideological models provide useful summaries in many contexts, more broadly, judges have come to be understood as motivated by largely the same set of concerns that motivate individuals in the workplace generally, including reputations, salaries, career advancement, and of course, personal satisfaction (Epstein, Landes and Posner 2013, Bainbridge and Gulati 2002, Ash and MacLeod 2015). Critically, like all workers, judges must balance these typical career-related interests against a preference for leisure (e.g., Posner 1993). In this paper, we ask whether an increase in the marginal utility for time spent on leisure activities causes a decrease not only in the rate at which judges work, but also in the quality of work begun during periods of distraction.

That judges, or anyone for that matter, would put in fewer hours on projects begun in periods when they are naturally drawn to non-work activity should not be surprising or necessarily alarming. Still, there are two reasons why understanding the effects of leisure incentives is important. The first is that leisure incentives might produce delays in the administration of justice. Although it is certainly possible to devote more hours to a project tomorrow when fewer hours are devoted to it today in order to ensure that justice is carried out in a timely fashion, the constant flow of cases in most court systems makes it possible that leisure incentives can result in delays that would be considered inappropriate. A second, arguably more important reason, is that leisure incentives may undermine the quality of justice itself. As we discuss below, expectations about the effect of leisure on work rate would seem fairly straightforward. Judges should devote less time to managing cases when they are distracted by personal concerns. The key question with respect to the rate at which judges work is whether leisure incentives ultimately produce delays in the resolution of cases. Turning to the effect of leisure on quality, given extant theoretical claims, our expectations are less clear. For example, judges facing a near-constant inflow of cases for resolution might devote less time to projects when leisure is particularly attractive and fail to fully offset this reduction in effort at a later date, when leisure interests are less pressing. In this sense, leisure incentives would undermine quality. On the other hand, when faced with an increased desire for leisure, judges might merely opt to slow down their work, simply taking longer to resolve cases, working harder to catch up when they are not otherwise distracted, and maintaining quality at the same level. Here leisure would surely result in delays, but delays are the cost of quality control. And finally, by slowing down the rate of work, judges might also even produce more high quality work than they would have produced at a higher work rate given that they would be taking longer to consider alternative options.

Each of these possible strategies implies different empirical patterns in the relationship between an incentive for leisure and the judicial work product, but they also highlight the importance of some very basic empirical questions. Do leisure incentives produce delays? Do leisure incentives undermine or enhance the quality of judicial work? The importance of these questions follows from the normative questions that they invite. If we observe a delay that we believe is caused by leisure, should we be alarmed? It is hard to answer that question if we do not know whether leisure also influences quality, as we might be willing to suffer some delays in order to sustain quality. Our tolerance for delay may be different if we also believe that leisure incentives undermine the quality with which judges resolve cases. Our paper provides an empirical basis for answering these questions.

Our study offers several contributions to research on the role of leisure in judicial behavior. Prior studies have suggested a number of ways in which the preference for leisure influences judicial behavior (e.g. Cohen 1991, Klein and Hume 2003, Posner 1993, Bainbridge and Gulati 2002). A significant challenge to evaluating expectations related to leisure incentives is that these incentives are fundamentally latent. For this reason scholars have largely adopted caseload as a proxy (Epstein, Landes and Posner 2013). Specifically, heavy caseloads raise the possibility that a judge may not be able to devote enough time to family, friends or pastimes as would be desired. Judges who are busier at work confront particularly strong incentives to find ways to protect their leisure time. Conceptualized in this way, Cohen (1991) finds that U.S. federal district court judges with heavy caseloads were more likely to find unconstitutional the sentencing guidelines developed by the U.S. Sentencing Commission, ostensibly because judges believed that the guidelines would undermine plea bargaining and result in a greater number of trials. Cohen (1992) also finds that busy district court judges issued increasingly punitive sentences in antitrust criminal cases to defendants who plead not guilty, suggesting an effort to incentivize plea bargaining as a way to reduce work in time-consuming trials. Similarly, Epstein, Landes and Posner

(2013) find that judges with higher caseloads were more likely to take advantage of the Supreme Court's decision in Ashcroft v. Iqbal (2009), which made it easier to dismiss discrimination lawsuits aimed at high government officials.

Our first contribution is to focus on quality. Promoting more plea bargaining or dismissing more lawsuits will surely save judges' time at trial, but whether judges should promote more or less plea bargaining or admit more lawsuits in order to produce high quality justice is unclear. Our goal will be to develop a method for evaluating not only the way that leisure incentives influence the rate at which judges work but also the quality of their work. Second, we will focus on the behavior of judges on the U.S. Courts of Appeals. The majority of prior research has looked at district court judges, where many scholars argue that the incentives to protect leisure time are heightened. Still, relative to the Supreme Court of the United States, leisure incentives may still play a role at least at the margin for appellate court judges (Drahozal 1998). Further, because appellate court judges set precedent for relatively large jurisdictions, it is particularly important to know if and how leisure preferences influence quality.

Third, existing measurement approaches, which rely on differences in the caseloads across judicial districts, identify effects of the context in which judges operate. They identify the effect of increasing incentives to protect leisure time at the judge level only under the assumption that on average busier districts cause greater leisure pressures at the level of the judge. While we believe that this is a reasonable assumption, our research design attempts to get a judge-specific measure of leisure incentives.

Fourth, the credibility of the causal claims in existing studies depend on the assumption that differences in caseload across districts reflect exogenous sources of variation in an analysis of judge behavior. While we do not take a strong position on whether this is a valid assumption, it is worth considering that different types of people, people with fundamental differences in their preferences for how to trade-off work and leisure, might be attracted to jobs in different districts. Two possibilities seem plausible. On the one hand, busy districts might be particularly attractive to candidates with weak leisure preferences, and if so, existing results might underestimate the effect of caseload. On the other hand, busy districts might be more likely to attract candidates who are more willing to compromise legal principles in protection of leisure time, in which case existing results would overestimate the effect of caseload. Our approach, which relies on a natural experiment analyzed via differences-in-differences estimation will rule out by design these forms of sample selection bias, resulting in more credible empirical claims.[1]

[1] As we develop below this is immediately obvious if variation in caseload is largely cross-sectional, and invariant over

Our research design relies on a natural experiment introduced by the annual NCAA Men's Basketball Tournament, one of the most popular, attention-getting sporting tournaments in the United States, and one that creates opportunities for leisure even among people who do not pay much attention to basketball at all. The tournament it is credited with substantial drops in productivity in the for-profit private sector[2], and so there is reason to believe that, if judges behave in many of the same ways as workers generally, the tournament will serve as a useful source of variation in our context. We collect data on the teams participating in the NCAA Tournament each year and match judges to the institutions at which they received their undergraduate degrees. Using a differences-in-differences design, we examine the effect of a judge's alma mater participating in the Tournament and show that the effect of having a team in the Tournament is to delay the time it takes a judge to write opinions for cases heard during the Tournament and to lower quality opinion.

Although we find that the Tournament represents a useful opportunity to evaluate the effects of leisure, we are not interested in learning about the effects of a basketball tournament *per se*. The Tournament represents a window through which we can evaluate leisure incentives. We certainly believe that the effects are interesting but their real interest derives from their ability to suggest to us the effects of other sources of leisure, many of which will impose stronger and more frequent forces on judges.

In the next section, we outline the theoretical framework underlying our expectations regarding judicial leisure and opinion writing. We then describe our empirical strategy and evaluate the effect of a judge's alma mater's participation in the tournament on the amount of time it takes a judge to write an opinion. We next evaluate the effect of the team's participation on the quality of the opinions the judges write. Finally, we offer concluding remarks about the consequences of judicial insulation from accountability in light of the evidence and extant theory about judicial leisure.

2 Theoretical Background

We will assume that judges care about their professional reputations among a potentially diverse group of observers (Baum 2009, Miceli and Coşgel 1994, Posner 1993). When processing and resolving cases, judges are evaluated on a number of elements of their decision-making. Among those elements, two are of particular importance: the quality of their decisions and the speed or efficiency with which they

time across districts; however, our design is also robust to variation that is time variant.

[2] http://www.challengergray.com/press/press-releases/its-march-madness-years-madness-could-cost-19b

resolve cases (Garoupa and Ginsburg 2010, Miller 2004). The job of writing a high quality opinion takes time (Choi, Gulati and Posner 2012, Gulati and McCauliff 1998), and given that reputations are related to opinion quality, time for federal judges is a precious resource. Unfortunately, the work of appellate court judges takes place in the context of a more or less constant flow of cases, which must ultimately be resolved, and where there is a general concern for avoiding too high a backlog (Reinhardt 1993). Crucially, as Ash and MacLeod (2015) point out, a judge's incentive to pursue leisure interacts with her ability to manage which cases she hears and the flow of judicial work she encounters. Others have similarly shown that judges are susceptible to non-judicial distractions that can prime their perspectives even shape *how* they vote (e.g., **?**). One of the primary goals of our empirical analysis below is to exploit the NCAA Tournament as a shock to judges' leisure incentives.

According to the US Administrative Office of the Court's Federal Judicial Caseload Statistics report, as of 2014, roughly 300-500 cases *per judge* were filed in the US courts of appeals, which is roughly the rate at which the judges terminated cases. However, there was also a backlog in each circuit of roughly 200 cases per judge. A great deal of public, political, and scholarly attention has been paid to the rate of case resolution and case backlogs on the courts. For example, the Judicial Conference routinely requests additional judgeships to deal with caseload problems.[3] The Administrative Office of the United States Courts is required to provide an annual report of statistical information, which always includes an accounting of caseload and backlog, identifying particular circuits which are lagging behind.[4] This information is commonly reported on by media outlets,[5] where particular judges are not too infrequently shamed.[6] And critically, increasingly caseloads have potential implications for judicial quality. Consider Judge Reinhardt's (1993) plea to the U.S. Congress, which develops the consequences of an increased caseload.

> Simply put, our federal court system is too small for the job. We seem to assume that judges can perform the same quality of work regardless of the number of cases they are assigned. That simply is not correct. Most of us are now working to maximum capacity. As a result, when our caseload increases, we inevitably pay less attention to individual cases. . . Those who believe we are doing the same quality work that we did in the past are simply fooling

[3] `http://www.uscourts.gov/file/document/2015-judicial-conference-judgeship-recommendations`
[4] 28 U.S.C. §604(a)(2)
[5] See `http://www.wsj.com/articles/in-federal-courts-civil-cases-pile-up-1428343746`.
[6] See `http://www.nytimes.com/1995/04/17/nyregion/awaiting-judgment-special-report-new-york-s-clogged-us-courts-delaying-civil.html`.

themselves. We adopt more and more procedures for "expediting" case, procedures that ensure that individual cases will get less attention.

On the surface it would appear that the goals of writing high quality opinions and ensuring that backlog is kept at a reasonable level, may be in tension with each other. If judges confront an increase in leisure incentives work rate should drop. Similarly, we might expect a reduction in quality as well. The question is whether this logic is clear upon a more careful consideration.

2.1 Judges as non-profit workers

Posner (1993) claims that because judges (like all others) desire leisure, work rate and quality will suffer when judges are distracted. Crucial for Posner's claim is that judges are best thought of as analogous to non-profit workers. Workers in the for-profit sector are commonly judged on the quality of their work, where quality is easily observable. For these workers, increased time devoted to leisure must be offset by increased time on work, so that quality can be kept constant. The failure to adjust in this way results in a less competitive product and potential loss of employment. At the very least, it will mean a loss in reputation. In contrast, Posner claims that the quality of work produced by federal judges is relatively unobservable. Judges are not paid by the number of cases they resolve, the quality of their decisions, or the number of hours they devote to resolving a dispute.[7] Of course, judicial work product is not truly unobservable. The public, politicians, the media, and scholars all frequently comment on both judicial quality and efficiency. The crucial point, however, is that between quality and amount of work completed, the quality of judicial work is relatively harder to discern than the amount of work judges do. Case backlog and the length of time cases take to reach conclusion is readily apparent to anyone with even a passing familiarity with the judiciary. Quality is far harder to measure.

If quality and work rate are not equally observable, increasing leisure incentives should decrease the work rate of a judge while distracted by personal interests, as well as the quality of judicial outputs produced while distracted. This follows because since quality is less observable than backlog, to avoid a backlog, decreases in effort when a worker is distracted by personal interests need not be offset by increases in effort at a later date(s). Instead, effort can be kept constant and quality can be sacrificed

[7] Of course, a market for law could, in principle, result in better quality. Early common law courts were based on a model of judges being paid by the volume of cases they resolved. However, it is notable that such systems also tended to have multiple judicial systems, which created competition among judges for business, thereby undermining incentives judges had to just "churn" cases without regard for the quality of their work (see, generally Blatcher 1978).

in order to avoid a backlog.

2.2 Alternative arguments

There are several plausible alternative arguments. Here we consider three, which emphasize (1) the selection mechanism for federal judges, (2) the management of workflow, and (3) the judge's staff.

2.2.1 Judicial Section

Judges are selected via a relatively searching process. They are recruited typically from among lawyers with considerable and laudable histories of work (e.g., Epstein et al. 2005, Savchak et al. 2006). Through this process, it may be possible to select individuals with a strong taste for high quality judicial work. If this is true, it may be possible to completely offset the concerns for quality created the leisure incentives. It may be that federal judges either are not commonly distracted by leisure concerns, or that if they are, they behave like for-profit workers whose product quality is observable, i.e., they offset temporary reductions of effort in the present with temporary increases in the future. Indeed, recent research suggests judges have an intrinsic desire to write high-quality opinions, even while constraints such as time pressures inhibit their ability to pursue that goal (Ash and MacLeod 2015). If this kind of selection process holds, the empirical implication is that while we may observe a decrease in work rate when leisure incentives increase, we should not observe a decrease in quality.

2.2.2 Managing Workflow

Judges might manage the tradeoff between backlog and quality in one of two ways, which would suggest different empirical implications. First, consider the ability to smooth out effort reductions over time. It may be possible that leisure undermines quality, but by smoothing, it will be difficult or impossible to observe it (e.g., Clark and Strauss 2010). For example, suppose a judge with a constant flow of one incoming case per day. Now, suppose that, when using the appropriate amount of effort, that judge can only complete one case per day. If on any given day she chooses to devote additional time to leisure, she will have a problem. She can either (a) delay resolving the case, allocating some of the next day's time to the current case or (b) apply sub-optimal effort to resolving cases. If she delays resolving the case, she creates a delay in her case resolutions. Either this delay will perpetuate through time, because she will then have to push off future cases accordingly, or it will cause her to decrease the quality of future cases she decides. If she decides to apply sub-optimal quality to her work, she could either simply decide the current case with lower quality, or she can balance delays and quality, spreading out the quality reduction over time, marginally decreasing the quality with which she decides the current and future cases. Assuming that judges always have some form of personal distraction at some point in a year, if this alternative approach is taken, then it will be impossible to observe a quality effect, as all judges will have similarly smoothed out the effects of their leisure incentives over the course of any given time period. Thus, again whereas we might expect to observe a reduction in work rate associated with a particular source of leisure, this kind of quality management would imply that it will be highly difficult to detect a decrease in quality.

A second possibility is that by slowing down work rate, leisure incentives give judges more time to contemplate solutions to legal problems. With an increased window working on an opinion, the effort needed to write the same high quality opinion may be less than it would have been with a smaller window. In this sense, there may be no tradeoff between backlog and quality. For sure, judges would work less when distracted by leisure, but quality would not be harmed. Indeed, it might be improved.

2.2.3 The Role of Clerks

Federal court of appeals judges do not work alone. They manage a staff typically consisting of a secretary, a team of clerks and sometimes externs or volunteer law students. It is well-documented that clerks play a number of important roles in the work that judges do, from the development of bench memos, the preparation of non-legal materials like speeches and the drafting and editing of judicial opinions (Peppers, Giles and Tainer-Parkins 2014). Indeed, 98% of court of appeals judges in Peppers, Giles and Tainer-Parkins's (2014) study report using clerks to conduct legal research; 88% use clerks to develop bench memos; and, 95% use clerks to draft opinions.

To understand whether and how clerks might influence the effect of a judge's leisure incentives, it is useful to summarize how clerks are typically used. Gulati and Posner (2016) suggest that federal court of appeals judges operate under one of three basic management models. One, which is increasingly rare, is that of the "Authoring Judge." An authoring judge does what the model says she does – she authors her opinions. The clerks of an authoring judge do not typically prepare bench memos; however, they provide important assistance to the judge in the form of legal research and editing.

The second and most common model is that of the "Editing Judge." The clerks of editing judges, when assigned to particular cases, commonly prepare bench memos and attend oral arguments when possible. Judges meet with these clerks following the post-oral argument conference. These clerks are typically expected to draft the initial opinion. The judge reviews the opinion and may ask for additional edits. Although the editing judge's clerk is involved in many aspects of the development of the judicial opinion, it important to stress that it is the judge who guides the process. Gulati and Posner (2016) write:

> A distinction worth noting at this point between the editing judges and the authoring judge ... is that the editing judges made clear to us that they specify outcomes to their clerks and then tell them to explain and justify that outcome in the opinion draft. Few clerks, and particularly not the ambitious ones, are going to come to the judge and tell him that the arguments in favor of that outcome are simply not good enough and that therefore the judge should change his vote.

A final model is that of the "Hierarchical" or "Delegating Judge." The delegating judge manages a staff led by a head clerk, sometimes one of the four clerks that the judge hires in a year but more often a "permanent clerk." This head clerk is responsible for managing the other clerks, who largely relate to the judge through the head. As is the case with the editing judge model, delegating judges allow clerks to develop draft opinions – the judge primarily serves as an editor, again, with the additional layer of hierarchy established by the head clerk.

If clerks are *really in charge* of the production of opinions, the leisure distractions of the judge need not have any effect on work rate or quality, unless leisure interests of clerks happen to coincide commonly with those of the judge. Of course, although clerks are important pieces of the process by which judges produce opinions, it is hard to imagine a theory of the judicial hierarchy in which clerks are fully in charge. Judges are ultimately responsible for their opinions. That said, because clerks take on a good deal of the work involved in producing an opinion, typical management structures for judicial staff may attenuate the effect of a judge's personal increase in leisure incentives. If this is true, it will be more difficult to observe an effect of a judge's leisure preferences than it would be absent the role of clerks.

A second implication of the role of clerks deals lies in the interpretation of the findings we report. Specifically, how might we interpret findings showing that judges work at lower rates and produce lower quality work when they are distracted by leisure concerns? For one, it may be most appropriate to interpret effects as dealing with individual chambers rather than judges. If judicial opinions are largely produced by teams, then we might attribute a negative effect of a judge's leisure on work rate to her entire team. Further, the mechanisms connecting leisure to outputs are likely to differ subtly across management models. For example, the effect on an authoring judge will be more directly related to the judge's own writing, whereas the effect under the other two models will relate more to the judge's ability to edit well. When a judge is distracted, final products may more validly reflect the work of clerks and that could account for a decrease in quality.

Finally, without carefully measuring the management style of every judge in our sample (we do not), the effects we estimate are best thought of as average effects across the three models. It is not clear ex ante which management model is likely to be more affected by a leisure distraction. It is possible that the effects will be stronger on authoring judges because they are in more control of more aspects of the final product. Yet, it is also possible that the writing process sharpens the attention of a judge in ways that editing does not; and, if that is true, the effects would be stronger among the editing and delegating judges.

2.3 Summary

As these features of the judicial process illustrate, it is not clear what to expect from the observation that judges occasionally face a heightened marginal benefit from leisure. We might expect judges to simply slow down their decision making, accepting a backlog in order to retain high quality work product. Alternatively, we might expect judges to sacrifice the quality of their opinions in order to keep apace with their constant flow of cases. Or, we might expect a smoother tradeoff between the two dimensions of their decision making. Unfortunately, the theoretical ambiguity has not been allayed by powerful empirical strategies. Because the theoretical model itself turns on a difficult-or-impossible-to-observe factor—the marginal utility associated with leisure—scholars have had to rely at best on crude proxies for judicial motivations to work, such as career interests. Our primary goal is to offer a powerful design in order to estimate the causal effect of a judge's leisure incentives.

3 Analyzing the Effects of Judicial Leisure

To study the effect of a marginal increase in the utility judges place on leisure we identify a salient feature of American culture that leads to distraction from the daily work-flow by drawing large public attention but affects workers differentially each year. These features make the event measurable compared to private events which are less frequently reported. Specifically, we examine how judges at the US Courts of Appeals are affected by the annual NCAA Division I Men's Basketball Championship (hereinafter, "NCAA Tournament" or "Tournament"). Importantly, given the nature of the event and the common activities that take place around it, the NCAA Tournament is likely to influence not only individuals who typically enjoy watching college basketball (or even sports) but a large proportion of the U.S. workforce.

The NCAA Tournament takes place in late-February and March each year and ends no later than very early April. The number of teams competing in the Tournament has varied over time, but between 1985 and 2000 the number was fixed at 64. (We limit our empirical analysis below to this time period.) The tournament is played in a series of single-elimination rounds. There are several methods by which a team can qualify for participation in the Tournament, and once qualified, they are "seeded" into a bracket and divided into four regions throughout the United States. The Tournament begins mid-day on a Thursday, and the first two rounds are played continuously through Sunday of that week. The third round (the "Sweet Sixteen") is played in the evenings of the Tournament's second Thursday and Friday, with fourth round games being played over the weekend. The Tournament's semi-finals and finals (the "Final Four") are played Saturday of the third weekend and the subsequent Monday night.

The NCAA Tournament is a very popular sporting event. The average rating for an NCAA Tournament game is historically between a 6 and 7, which is comparable to the CBS Evening News. An estimated 28% of televisions in the United States were tuned in to watch the 2015 championship game between Duke and Wisconsin. Moreover, according to the NCAA 350 million social impressions of the Tournament where shared on Facebook and Twitter and 17.8 million hours of live video where watched online in 2015.[8] This highlights the large societal impact of this sporting event. In addition to the Tournament's games, it is common for workers to take part in betting pools, where individuals fill out "brackets" and success is a function of how well participants predict outcomes along the way.[9] For

[8]http://www.ncaa.com/news/ncaa/article/2015-04-07/2015-ncaa-tournament-has-highest-average-viewership-22-years

[9]In surveys conducted by CareerBuilder.com, during recent years between 15% and 20% of respondents report partici-

example, in a widely publicized survey, Challenger, Grey & Christmas estimated that 60,000,000 American workers were expected to participate in a "March Madness Pool" in 2015. Importantly, individuals who do not commonly follow basketball or even sports at all often participate in office pools. Some even suggest that a lack of familiarity with college basketball is a necessary condition for success in an office pool!

The effect of the NCAA Tournament on workplace productivity in the US is something that has been documented. Assuming each of the individuals Challenger, Grey & Christmas estimate are participating in a pool spends just one hour of work time following the NCAA Tournament, the cost to employers would be over $1.9 billion.[10] (Indeed, the Challenger, Grey & Christmas survey found that 56% of respondents indicated they would spend at least one hour of work time on the March Madness pool.) According to Fortune Magazine, people spent a collective 664 million hours just watching television broadcasts of NCAA Tournament games. Similarly, Clotfelter (2012) finds that downloads of academic articles through JStor at university libraries drops sharply (about 6%) during the first week of the NCAA Tournament. These estimates dovetail with those from other sporting mega-events. For example, Lozano (2011) finds there is a considerable reduction in the hours individuals work during the World Cup. However, and critically, he finds that the effect is concentrated among salaried workers, as opposed to hourly workers. The judges we study are, in many ways, even more immune to the professional pressures that differentiate salaried and hourly workers.

Crucially, attention to the NCAA Tournament is almost surely a function of whether one's team is participating in the Tournament. When one's *alma mater* is selected to play in the Tournament, the team receives considerable national media attention. Moreover, a team's success is linked to other teams' performance, creating an incentive for fans to watch games involving other, competing teams. Consistent with this claim, Clotfelter also shows that the effects during later weeks in the Tournament are most pronounced when the victor is a surprise.[11]

The Tournament can influence individuals who are neither watching games, commenting on games on social media, nor participating in office pools. Because media coverage of colleges and universities is highly correlated with whether the institution is in the Tournament, a person whose school makes it

pating in such pools.

[10]http://www.challengergray.com/press/press-releases/its-march-madness-years-madness-could-cost-19b

[11]Clotfelter takes this as evidence that individuals rationally anticipate their teams participation and so adjust their work schedules accordingly. As a consequence, only those "surprised" by their team's success should have a noticeable decrease in productivity.

is likely to be exposed to much more information about their school than usual. In fact, depending on the school he or she many *only* see national coverage of their alma mater during this period when their school is playing. Increases in the salience of one's school can serve as reminders of past friendships, past mentors, and a variety of forms of nostalgia. It can encourage somewhat time consuming searching on social networking sites, reconnections, in addition to a variety of potential social activities ultimately unrelated to the Tournament itself. Importantly, because when schools advertise during games, they present information on the academic and social aspects of their institutions, largely setting aside athletic factors which sell themselves through the team's participation. This type of advertising is especially appealing to non-sports fans. Thus even the casual or non-sports fan who happens to tune in for a few minutes is pulsed with reminders of broad features of their alma mater. In all of these ways, the Tournament represents a particularly useful way of evaluating productivity under a form of personal distraction that can influence people with diverse leisure preferences.

Our design relies on the exogeneity of a judge's *alma mater* with respect to its appearance NCAA Tournaments years after a judge would have graduated. To estimate the causal effect of Tournament-induced distraction on judges' performance, we use a difference-in-differences design. We also assess the causal effect of the Tournament on the quality of the opinion a judge writes. We show that opinions written by judges with teams in the Tournament during the Tournament are more negatively cited than other opinions. We again employ a difference-in-differences estimator to account for other factors driving that relationship. What is more, mediation analysis suggests that the extra time judges take to write their opinions does not mitigate against that effect. We now turn to our two empirical analyses.

4 The Effect of the Tournament on Delays in Decisions

Our first analysis considers the effect of judges' leisure incentives on the speed with which they decide cases. Specifically, we examine how interest in the NCAA Tournament affects the amount of time an opinion author takes to issue an opinion after the case has been heard. Our expectation is that the NCAA Tournament will differentially affect judges whose alma maters are participating in the Tournament that year by increasing the time it takes for them to prepare and publish their opinions.

4.1 The data

To test our expectation we require data on (1) the timing of judicial decisions, (2) teams' participation in the NCAA Tournament, (3) the judges participating in and authoring each case, and (4) judges' school affiliations.

Case data. We collect the text of all decisions included in the *Federal Reporter*, from volume 797 of the second volume (F.2d) through volume 529 of the third volume (F.3d). This includes all decisions from 1985 through 2005. We obtain the texts from `bulk.resource.org`, which provides all judicial opinions formatted in `html`. We limit our attention, though, to cases decided between 1985 and 2000, when the format of the NCAA Tournament was constant. We then extract from each html file (1) the date of oral argument (if provided in the header for the opinion), (2) the date of decision, (3) the judges hearing the case, and (4) the case citation.[12] In the `html` code, the line with the dates of argument and decision is specifically tagged, making it easy to extract that text. We then write a `regular expression` to parse that text into the date of oral argument and the date of decision. In many instances, oral argument is not given in a case or is not reported. In some instances, courts report the date a case was submitted to the court. In even rarer instances, the date a case was decided is not formatted in a standard way and so is not easily extracted from the html code. This process yielded 9,309 cases for which we have the date of argument and the date of the decision between the years 1985 and 2000. Our variable of interest, $TimeToDecision_i$ is the number of days between oral argument and the opinion in case i. When either of these pieces of information is unavailable from the *Federal Reporter* file, we code this variable as missing. Figure 1 shows the distribution of the logged number of days from oral argument to the decision.[13]

To identify the judges hearing a case, we also write a `regular expression`. Fortunately, the Federal Reporter has standardized how it identifies the judges hearing a case, by reporting "Before" in the header, followed by the judges' names in capital letters. We search through the header to the decision for a line matching those parameters, focusing only on cases where three judges are hearing the case, and extract the three names. When we cannot identify three judges hearing the decision, or when

[12] Commonly-used databases, such as the US Appeals Courts Database (Songer 1999) do not contain oral argument dates.

[13] 99% of the cases are decided in less than 300 days. Among the cases that took longer than 300 days to decide, the average time to decision is 413 days. We exclude these cases because they are such extreme outliers they could unduly influence our analysis.

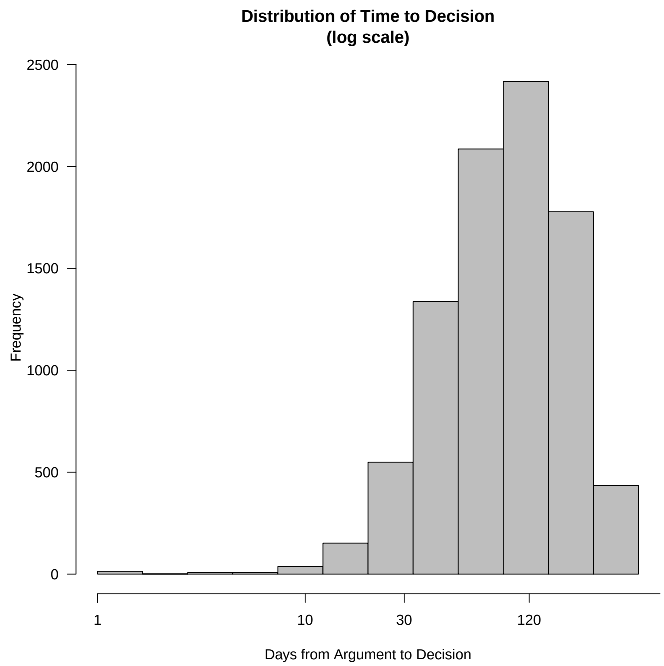

Figure 1: *Distribution of days from oral argument to decision, 1985-2000.* Figure shows the distribution of (logged) days from oral argument to the decision in a case. Note, the figure shows the natural log of the number of days, whereas the values reported on the x-axis are on the linear scale.

a decision is decided by more than three judges (e.g., *en banc* decisions), we exclude the case from our data. The case citation for each decision is also specially tagged in the `html` code, and we extract that citation from the raw code.

Judge data. To identify the alma mater for each judge, we rely on the Federal Judicial Center's Biographical Directory of Federal Judges, which is available for download from their website `http://www.fjc.gov` and is updated daily. The database contains extensive biographical information for every Article III judge in US history, including all of the post-secondary schools each judge attended. From these data, we extract all judges serving after 1985 (the start of our case data) and match judges by their last names, circuits, and dates of service (i.e., a judge's dates of service must include the date a case was decided). There is a small number of instances of judges deciding cases having multiple last-name matches in the biographical data, which we handle manually, matching on first names when the last name does not uniquely identify the judge.

	Case Not Heard During Tournament	Case Heard During Tournament
Non-Tournament Author	6,723 (75.45%)	1,738 (19.50%)
Tournament Author	273 (3.06%)	84 (0.94%)

Table 1: *Distribution of cases by NCAA Tournament alma mater separated by authorship status.* Rows distinguish panels with Tournament Authors from those panels with Non-Tournament Authors. Columns distinguish cases heard during February and March from those heard during other months. Each cell shows the raw number of cases in our sample heard by each pair of conditions (along with percentages).

NCAA Tournament data. To identify all teams playing in the NCAA Men's Tournament, we reference the Wikipedia webpages for each year's tournament. These pages contain tables of teams playing in each regional division of the Tournament. Using the `XML` package (Lang 2013) for `R`, we extract those tables and create a database of the 64 teams playing in the Tournament.

In total we were able to identify all of the variables of interest for 9,309 cases. The vast majority of missing data occurs in when cases decided *per curiam*, which means no author is identified. Other instances occur when judges' names are misspelled and ambiguous, making it impossible to match them with their records in the Federal Judicial Center's data. Furthermore, in 26 cases our data indicated that a decision took less than a day. While this is unlikely we excluded these cases and in addition we excluded 465 cases that took more than 300 days affecting march madness in two years. This leaves as with 8,818 cases to analyze.

We code the variable $Tournament_{a[i]y[i]}$ as 1 if the author in case i has an alma mater participating in the NCAA Tournament in year case i was heard, and 0 otherwise. We code the variable $MarchMadness_{t[i]}$ as 1 if case i is heard (has oral argument) during the month of February or March. We choose this coding, because it is those cases whose work is likely to be disrupted by distraction due to the NCAA Tournament. Cases heard during February are in the process of being written during the Tournament, which usually takes places during the middle of March, as are cases heard during March. Cases heard during January are likely to be (at least nearly) completed by the time of the NCAA Tournament.[14]

[14]The findings we report below are robust to relaxing these coding rules in sensible ways. We discuss those robustness checks as we present our results.

In our data, 357 cases were authored by a judge whose alma mater participated in the NCAA Tournament during the year the case was heard. We call these judges—judges who author opinions in cases heard during a year their teams were in the NCAA Tournament—Tournament Authors while we call the judges authoring the remaining 7,147 cases Non-Tournament Authors. (Note, a given judge can be both a Tournament Author and a Non-Tournament Author, as s/he is a Tournament Author only in the years his/her alma mater was in the Tournament.) Table 1 compares cases during the Tournament to those decided not during the Tournament separated by authorship status.

4.2 Identification strategy

To estimate the causal effect of a judge's alma mater participating in the Tournament, we rely on a differences-in-differences design. In particular, we consider the difference between the time from oral argument to decision for cases that were heard in months other than February and March (i) with a Tournament Author and (ii) without a Non-Tournament Author. We then consider that difference—between Tournament Authors and Non-Tournament Authors—for cases heard during February or March. The difference between those two quantities (the two differences) is our quantity of interest. This differences-in-differences design captures the expectation that Tournament Authors take longer to write a decision than Non-Tournament Authors during February or March. By focusing on how the difference between these two types of Judges changes in February or March, we can account for other underlying differences between the Judges. These should be constant over time.

This design is particularly important in this setting. As described above, there have been many studies of whether labor markets and firms decrease in productivity during major sporting events such as the NCAA Tournament. However, if we simply asked whether the judiciary slows down, then we cannot evaluate the normative and theoretical issues at hand here—whether individual judges behave differently as they place increased weight on dimensions of their utility other than work, such as leisure.

The key identifying assumption involves common trends. The common trends assumption holds that any differences between the two groups (here, Tournament Authors and Non-Tournament Authors) would be the same across the two periods of observation (here, cases during the NCAA Tournament and cases heard at other times of the year) if there were no treatment effect. Figure 2 summarizes the common trends assumption and how our hypothesis relates. The lower black line shows the hypothetical difference between the time to decision for cases heard in months other than March and cases heard

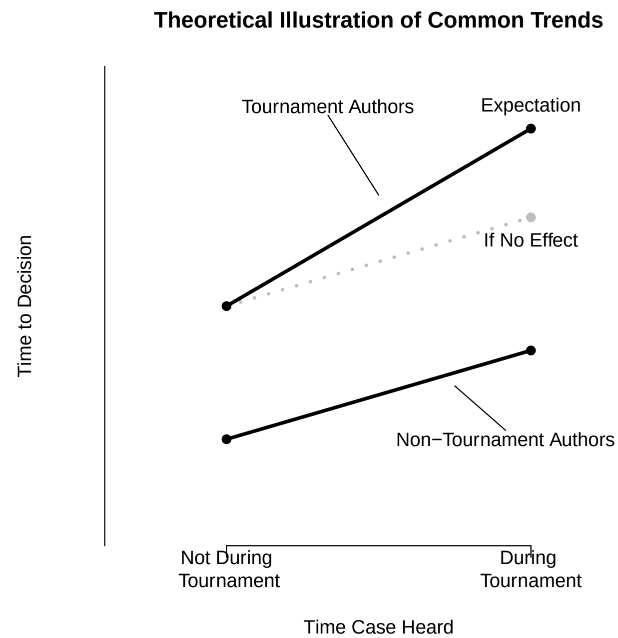

Figure 2: *Illustration of hypothetical common trends and our theoretical expectation.*

during March, by authoring judges without a team in the NCAA Tournament. The grey line shows our assumption about authoring judges with a team in the Tournament. If there is no causal effect of having your team in the Tournament, then those judges' behavior should *change* between March and other months *the same as* the behavior of other judges. Note, the model does not assume no difference for Non-Tournament Judges in the two time periods. Rather, it assumed only that the differences between the two groups across the time periods will be similar but for the causal effect of having a team in the Tournament. The upper black line shows, by contrast, our expectation that the Tournament Authors will exhibit a larger delay in deciding cases during March, relative to other months, than Non-Tournament Authors. Note, as well, we do not assume a strict increase in delay—just that any change in delay will be towards longer delay among Tournament-Team Judges.

4.3 The empirical model

Figure 3 compares the length a decision takes during month other than February and March to the period of the NCAA Tournament. The black lines show the distribution of the time to decision for cases not heard during the Tournament, and the black triangles show the means for those distributions. In both the left- and right-hand panels, the mean for these distributions is 99 days. In other words,

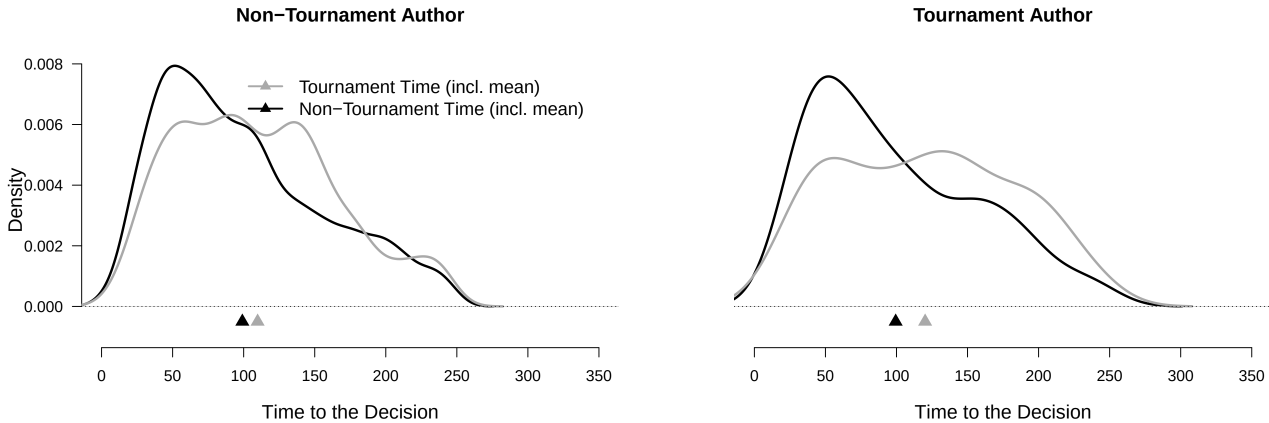

Figure 3: *Distribution of days to a decision in month other than February and March (black lines) and during the NCAA Tournament (gray lines) including mean values (triangles) separated by Tournament Authors and Non-Tournament Authors.* The figure highlights only the major parts of the heavily right skewed distribution of days to a decision.

both Tournament Authors and Non-Tournament Authors write their opinions, on average, in 99 days, when a case is not heard during the Tournament. This picture changes during the NCAA Tournament. The grey lines show the distribution of the time to decision for cases heard during the Tournament, and the grey triangles show the means for those distributions. Non-Tournament Authors are faster in taking decisions with 110 days on average compared to Tournament Authors who take 120 days on average. Hence, in non-tournament times it takes Non-Tournament Authors and Tournament Authors equally long to make decisions. While both groups of authors slow down during tournament times the tournament authors are on average slower compared to their colleges.

We model $TimeToDecision_i$ as a function of (1) whether a panel judge's alma mater is in the Tournament the year the case was argued, (2) whether the case was heard during February or March (i.e., during the time of the Tournament), and (3) the interaction of those two variables. We estimate the model coding $Tournament_{a[i]y[i]}$ as 1 two different ways—if the opinion author's alma mater is in the Tournament and if any judge on the panel has an alma mater in the Tournament.[15] This helps

[15]One might also propose a third option, which is the calculate the proportion of the judges who have alma maters in the Tournament in a given year. The only possible values for the mean of the TournamentAuthor variable are 0, .33, .67, and 1. (These correspond to the situations where 0, 1, 2,and 3 of the judges have alma maters in the Tournament, respectively.) There are no instances in which the variable takes on the value 1—i.e., there are no observations in which all three judges have alma maters in the Tournament. Further, in the data, there are only 70 instances in which two of three judges have alma maters in the Tournament. Of those, only 8 cases occur during the Tournament. Therefore, for only 8 of the 8,818 observations (less than 0.1% of the data) could the interactive term, capturing the difference-in-differences be coded differently from the "Any Judge." (The mean time to decision among those 8 observations is 138.38 days, which is in line with the mean for observations where any judge is the author and the case is heard during the Tournament—132.19

capture the notion that peer effects may reach to the collective workflow among judges on a panel and slow down the opinion-writing process. Our expectation is that the interaction will have a positive relationship to $TimeToDecision_i$, as judges from Tournament teams will work more slowly during the Tournament months. We employ a linear difference-in-differences model. Formally, we assume

$$\begin{aligned} TimeToDecision_i &= \alpha + \beta_1 Tournament_{a[i]y[i]} \\ &\quad + \beta_2 MarchMadness_{t[i]} \\ &\quad + \beta_3 Tournament_{a[i]y[i]} \times MarchMadness_{t[i]} + \varepsilon_i \end{aligned} \tag{1}$$

where ε_i is a Normally-distributed random variable with mean 0 and standard deviation σ_ε. Further, we estimate the model including both (i) year fixed effects and (ii) author fixed effects. Finally, we cluster the standard errors at the level of the unique combination of three judges deciding each case (i.e., the panel) to avoid any downward bias in uncertainty that might result from different numbers of observations from the circuits. We also consider alternative specifications that include covariates for demographic features of the panel; details of other specifications are reported in the appendix. We estimate each model in a frequentist framework. Further, we also consider estimating the models using a negative binomial specification, rather than a Poisson model. The substantive findings we report hold in these alternative models. The primary results are reported in Table 2.

These estimates show that we estimate a positive and statistically significant effect of the interaction between $Tournament_{a[i]t[i]}$ and $MarchMadness_{t[i]}$ and the time it takes to render a decision, across all of our specifications. This means that the difference in the time it takes to render a decision between judges with teams in the Tournament and judges without teams in the Tournament *increases* for cases heard during the Tournament. Importantly, the difference-in-differences design accounts for any effect of either (i) the Tournament itself or (ii) being from a Tournament school on the speed with which judges render decisions. This estimate captures only the effect of the case being heard during the Tournament on judges from Tournament teams. And, the models with fixed effects for the author's identity and the year of the case mean the estimate is identified from variation in individual judges' "Tournament status." Further, in the final model, we see that the effects we detect are consistent when we re-code

days. These two means are not statistically distinguishable from each other.)

	Author Only		Any Judge	
	Year FEs	Author & Year FEs	Year FEs	Author & Year FEs
Tournament$_{a[i]t[i]}$	-0.59	-1.23	12.17***	8.89***
	(4.57)	(4.61)	(3.17)	(1.97)
March Madness$_i$	18.06***	16.37***	10.96**	12.45***
	(2.51)	(2.61)	(5.12)	(4.00)
Tournament$_{a[i]t[i]}$ × March Madness$_i$	13.23**	7.75	20.57***	11.64**
	(4.42)	(5.93)	(6.90)	(4.67)
Intercept	102.04***	80.24***	83.90***	43.61***
	(5.24)	(1.20)	(6.61)	(1.97)
R^2	0.09	0.33	0.84	0.90
N	8,818	8,818	8,818	8,818

Table 2: *Empirical models of number of days between oral argument and decision on federal court of appeals, 1985-2000.* Cell entries are OLS regression coefficients (clustered standard errors in parentheses). $^{***}p \leq 0.01$; $^{**}p \leq 0.05$; $^{*}p \leq 0.10$

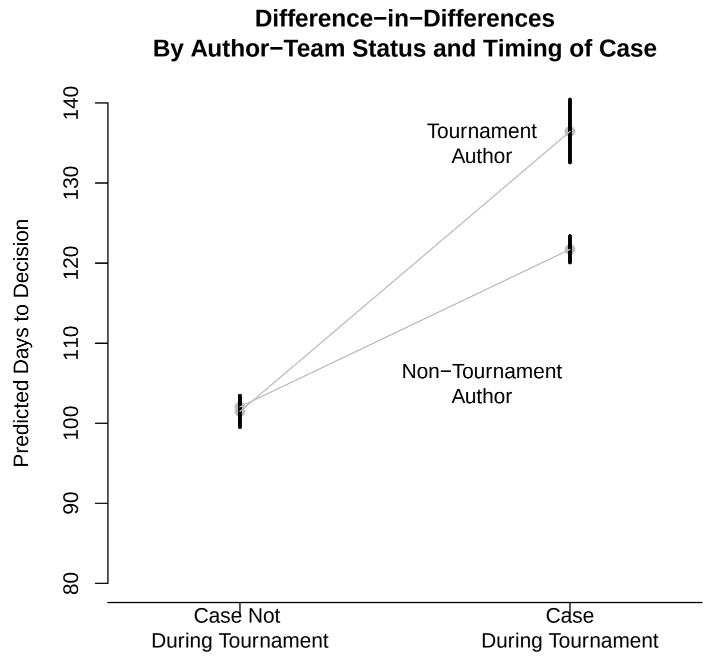

Figure 4: *Mean expected days (including 95% confidence intervals) Tournament and Non-Tournament Authors need to write an opinion during the Tournament compared to Non-Tournament month.*

the data to measure whether any judge on a panel had an alma mater participating in the Tournament.

To illustrate the effect of the NCAA Tournament on the Tournament Authors, Figure 4 summarizes the mean expected days a decision will take during the NCAA Tournament and in other month. The points show point estimates, and the black bars show 95% confidence intervals. The effects are separated by Tournament and Non-Tournament Authors overwhelmingly supporting our hypothesis. It takes both kinds of judges roughly 99 days, on average, to produce opinions for cases heard during non-Tournament months. During the Tournament, that number increases to 109 days, on average, for the Non-Tournament authors; however, it increases to just over 120 days for Tournament authors. That difference—between 109 and 120 days is our estimate of the causal effect of having an alma mater's team participating in the Tournament. In other words, the effect of having one's team in the Tournament is to delay the publication of an opinion by 21 days.

What is more, these findings are fairly robust to alternative coding schemes for our key variables. For example, if we code only cases heard during March as being cases decided during the NCAA Tournament, we still estimate a positive, statistically significant interactive effect—$\hat{\beta}_3 = 0.14$, $se = 0.03$. By contrast, if we conduct a "placebo" test and code only cases heard during any given month, we find no effect at all for all months, other than the Tournament months. Moreover, we still find a positive, though substantively smaller effect if we include cases heard during January along with cases heard in February or March—$\hat{\beta}_3 = 0.15$, $se = 0.02$. This is consistent with our expectation above that the causal effect of the NCAA Tournament is strongest for cases heard during February and March, as those are the cases where the mechanism—distraction by the Tournament affecting effort on opinion-writing—is at work. We report the full results of these tests in the appendix.

4.4 Threats to inference

Can judges avoid their work when tempted with leisure? One might worry that when a panel includes a judge whose alma mater is in the Tournament, his or her colleagues would (in the collegial spirit) not ask the judge to take on the responsibility for drafting the majority opinion. If this were true, then the Tournament treatment would not be applied as suspected, threatening the causal inference we draw. There are, however, theoretical reasons to doubt this possibility. Generally, panels of judges sit together for a fixed period of time, hearing a (sometimes alleged) randomly-assigned set of cases. Because of the temporary nature of the panel, there is a strong norm of equity in the workload among the judges, and oftentimes the opinions are assigned before the cases are even heard. If this norm is binding, for whatever reason, then the possibility that the Tournament does not affect the panels as suspected is mitigated. Further, relative to some other kinds of personal distractions, including those that are difficult and serious (e.g., the illness of a family member), the NCAA Tournament is unlikely to be the kind of personal distraction that would warrant a relaxation of typical work norms.

What is more, there is empirical evidence that judges not get to shirk from their authorship responsibilities when their teams are playing in the Tournament. Using our data, we estimate a conditional logit model, where for each case, there is a choice of to whom to assign the opinion, from among the three judges participating, subject to the constraint that only one judge can be assigned the opinion. The conditional logit model allows us to include covariates that vary at the level of the alternatives, rather than the choice itself. Here, we have whether each judge's alma mater is in that year's Tournament, which varies at the level of the judge (alternative) rather than at the level of the panel (the choice to whom to assign the opinion).

Variable	$\hat{\beta}$
$Tournament_{a[i]y[i]}$	0.16 (0.03)
$MarchMadness_{t[i]}$	-0.01 (1.41)
$Tournament_{a[i]y[i]}i \times MarchMadness_{t[i]}$	0.03 (0.07)
N obs	84,879
N of events	24,923

Table 3: *Predictors of whether a judge is selected to author an opinion.* Conditional logit model in which the choice is among the three judges hearing a case, and the predictors are each judge's alma mater's participation in the Tournament, whether the case is heard during the Tournament, and the interaction of those two.

Let

$$A_{ai} = \exp\left(\beta_{1a}Tournament_{a[i]y[i]} + \beta_{2a}MarchMadness_{t[i]} + \beta_{3a}Tournament_{a[i]y[i]} \times MarchMadness_{t[i]}\right)$$

Formally, we assume the probability that judge a is selected to author the opinion in case i as follows:

$$\Pr\left(Author_i = a|\mathbf{M}_i\right) = \frac{A_{ai}}{\sum_{k=1}^{3} A_{j[i[k]]i}} \tag{2}$$

where $j[i[k]]$ is the kth judge hearing case i, $Author_i$ is the identity of the author of the majority opinion in case i, and, as above, $\mathbf{M}_i$ is a matrix of covariates.

We estimate this model using all panels for which the opinion author is specified (e.g., the decision is not "per curiam" or co-authored) and for which we have clearly matched the author to the panel. This results in $24,923$ cases. The results from estimating Model 2 are reported in Table 3. The evidence here is clear. The interactive effect—i.e., our estimate of β_{3i} is substantively very close to zero and statistically indistinguishable from zero. Taken in conjunction, the organization of the courts, the norms of equity in opinion-writing, and these empirical results indicate that judges are not able to strategically avoid opinion-writing responsibilities during the Tournament and therefore are unable to avoid the treatment effect. Taken in conjunction, these two empirical test show that the effect of a judge's team participating in the Tournament is associated with a sharp increase in the amount of time a judge takes to issue an opinion, and that effect cannot be attributed to strategic avoidance of writing opinions.

Are judges from Tournament-participating alma maters different? One might also be concerned that judges who attended schools that often participate in the NCAA Tournament—big state schools, and other athletic powerhouses—are different in other characteristics that might exacerbate the slow-down we observe in March for reasons unrelated to leisure. We consider a handful of demographic characteristics of our judges across those who are Tournament Judges, and Non-Tournament Judges. Table 4 summarizes these results. Across virtually all of these characteristics, the two groups are identical. The average judge in each group was born in 1935. In each group, men constitute 68-69% of the observations. In the Tournament and Non-Tournament judge groups, 94% and 83% are white, respectively, suggesting a slight, though statistically insignificant difference. We also consider the ABA rating, which is a 4-point scale rating how qualified the judge is for the position when nominated. (Some judges serve in multiple federal courts and so have ABA ratings for each nomination. We use here the ABA rating from their first post, whether it was the Court of Appeals or another court; using the rating from their Court of Appeals confirmation process yields identical results.) The statistics here indicate, again the two groups are identical. In addition, we consider whether the Senate confirmed the judge on a voice vote, as a proxy, perhaps, for less well quantified metrics of the judge's temperament, political divisiveness, or quality. Here, we find among the Tournament and Non-Tournament judges, similar rates of voice-votes, 85% and 93%, respectively. Again that difference is not statistically significant. Finally, we consider the proportion of judges who are Republicans, as opposed to Democrats, and find that there is no statistical difference between the two groups.

Characteristic	Tournament Judges	Non-Tournament Judges
Birth Year	1936 (17.1)	1935 (15.9)
Proportion Male	0.68 (0.5)	0.69 (0.7)
Proportion White	0.94 (0.2)	0.82 (0.4)
ABA Rating	1.27 (0.8)	1.43 (0.7)
Senate voice vote	0.85 (0.4)	0.93 (0.3)
Proportion GOP	0.30 (0.5)	0.34 (0.5)

Table 4: *Comparing judge characteristics between judges with Tournament teams authoring opinions during the Tournament ("Treatment" Group) and other judges ("Control" Group). Cells show mean values with standard deviations in parentheses.*

These data suggest that there are no significant differences between the two groups of judges that might give rise to a spurious correlation in the difference-in-differences design and therefore that we have adequate balance in the kinds of judges in each of our categories.That said, it is absolutely critical to notice that these features, i.e., age, race, gender, perceived qualifications, political support, and ideological orientation do not vary at all over time, much less across the periods we compare (the time of the tournament and the time when there is no tournament). A notable feature of the difference-in-differences design is to be able to account for time invariant heterogeneity in potentially confounding features. To the extent that one worries about baseline differences between the groups, the difference-in-differences design addresses the concern.

Are published opinions special? Another concern is that we focus only on published opinions. Our design assumes judges cannot offset their leisure incentives by changing their workload in the set of unpublished (ie, less important) decisions. Suppose their workloads in unpublished cases shifts, decreasing their workload on these less important cases. This would allow judges to offset the consequences of leisure distraction, biasing against the result we find. In so far as we constrain the implications of our findings to published opinions, we are restricting our findings to the cases that will be read by the legal community and most likely to influence future law.

As describe above, there are multiple dimensions along which judicial work can be evaluated, and a marginal increase in the desire for leisure may involve simply trading off, for example, the amount of time it takes to reach a decision in order to maintain the quality of the judicial product. We turn now to a consideration of how the Tournament affects judicial opinion quality.

Is it really the clerks? As we describe in the theoretical section it may be appropriate to interpret the effects we find as reflecting the work of a judge's entire staff. But one might wonder if there is something about the clerks that the judge hires that might explain the outcome. This is a particular concern here in that clerks may change over the course of the year, potentially in ways that confound the inference we have drawn. A natural concern is that it is the distraction of the clerks that is causing the delay. For this to be the source of confounding, at the very least, it must be that federal court of appeals judges are particularly likely to hire clerks who attended their own undergraduate institutions. Judges do take considerable interest in the quality of a potential clerk's education; however, the typical concern is with a candidate's law school. Peppers, Giles and Tainer-Parkins (2014) report that 91% of the judges in their sample mentioned law school ranking as a factor that they consider when hiring clerks.

Fully 66% suggested that law school ranking was the most important factor. In contrast, only 44% of judges in the sample thought to even mention the quality of a candidate's undergraduate institution as possible relevant factor and only 5% reported that it was the most important factor. It would seem unlikely that this possible source of confounding is problematic.

Table 5 considers the possibility more carefully. To track down the undergraduate institutions of the clerks working for federal court of appeals judges, we consulted the Judicial Yellow Book (Book 2003). We began with the earliest available hard copies at the University of Georgia's Law Library. Specifically we began with the Fall volume for 1996. We also considered the Spring volume for 1997.[16] Across all circuits, including the federal circuit, we identified data on 664 clerks and their judges. As Table 5 details, educational information is commonly reported for judges. Consistent with Peppers, Giles and Tainer-Parkins's (2014) survey, although the law school information of clerks is reported more often than not, undergraduate institution information is routinely ignored. Of the 664 clerks in the Yellow Book during this period, there is undergraduate information on only 53 clerks. Even among these clerks, there just is very little evidence that judges and clerks commonly have similar alma maters. Indeed, there are only four cases of a match between the undergraduate institutions of clerks and their judges. And in every case, the alma mater is Harvard University, a school that was never in the NCAA tournament during the period of our study.

Data Availabilty in Yellow Book (%)		Institutional Matches btw. Judges & Clerks (%)	
Judge Undergraduate Institution	95	Judge Undergraduate - Clerk Undergraduate	8
Judge Law School	99	Judge Undergraduate - Clerk Law	8
Clerk Undergraduate Institution	8		
Clerk Law School	69		

Table 5: *Data availability for the educational institutions of 664 clerks and their judges reported in the Judicial Yellow Book (Fall 1996 and Spring 1997). Whereas information on judges' undergraduate institution, law school and clerk law school are commonly reported, the undergraduate institution of the clerks were reported in only 8% of cases. Of the clerks with undergraduate information available (53 clerks out of 664), only 4 or 8% matched their judge.*

Finally, the difference-in-differences design with year- and judge-fixed effects accounts for judge specific clerk characteristics, assuming that Tournament Authors to do not disproportionately assign excess work to their clerks during the Tournament season. If this is ultimately what happens, this

[16] After speaking with other scholars and reviewing the results of the first two volumes, we concluded that further research, i.e., beyond Spring 1997, was unnecessary.

process creates a bias against finding a difference in the differences between Tournament Authors and Non-Tournament Authors.

5 The Effect of the Tournament on Judicial Quality

As we outlined above, there are multiple ways in which judges might respond to heightened leisure incentives. One possibility might be to simply slow down their rate of case resolution, taking longer to write their opinions, in which case the preceding result is all we might expect to find. On the other hand, we might also expect that judges sacrifice the quality of their decisions in order to avoid too large effects of their delays on case backlogs. Or, alternatively, we might expect a smoother balancing of timing and quality of decisions. In this section, we investigate the consequences of the Tournament for judicial opinion quality and consider these multiple causal pathways—a direct effect of the Tournament on opinion quality as well as the possibility that taking longer to resolve cases mediates the deleterious effects of judicial distraction on opinion quality.

5.1 The data

How can we best study judicial opinion quality? The concept is notoriously elusive, and a veritable cottage industry has developed around the goal of measuring it. An increasingly common metric is to use citation patterns to proxy for opinion quality (e.g., Posner 2000, Choi and Gulati 2007, Ash and MacLeod 2015). Citations to opinions come in many forms, some of which simply relate a case to past cases on factual grounds, some describe past cases to provide doctrinal context, some praise past cases, and some criticize past cases. Here, we rely on the number of *negative* citations to an opinion as a proxy for judicial quality (or, really, lack thereof). We prefer to use negative citations because as contrasted with positive citations, there is little ambiguity in what is a negative, adverse citation. By contrast, positive citations often include a hodgepodge of forms of citation, many elements of which are likely not driven by judicial quality. In conjunction with the data collected for the above analyses, citation data will allow us to evaluate the effect of the Tournament on judicial opinion quality.

Citations to decisions. To collect data on citations, we performed a KeyCite search on a sample of all court of appeals decisions, which uses Westlaw's databases to identify every subsequent case

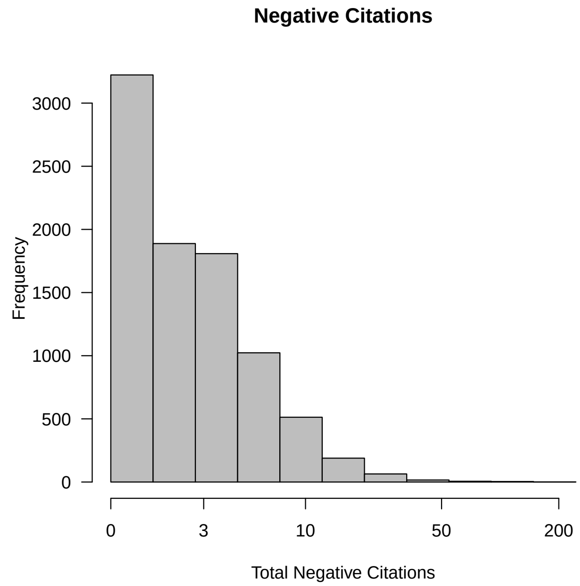

Figure 5: *Distribution of negative citations to precedent cases.*

citing the case we search (including unpublished decisions). Westlaw returns a report that divides all subsequent citations into discrete, and mutually exclusive categories. We then identified cases from our sample of KeyCite reports that are also in our data and for which we have the full set of other relevant covariates—the judges on the panel, the author of the opinion, and the author's alma mater. Most important for our purposes is the category "Negative Cases", which are cases that cite the searched case adversely. This category is a relatively narrowly-defined classification, which means a subsequent case criticizes, distinguishes, overrules, etc., the case. We are able to identify the total number of Negative Citations for 5,449 of the cases for which we have a complete set of the other covariates (a non-per curiam author, the date of argument, and the date of decision).[17] For each case, we code our primary variable of interest, $NegativeCitations_i$ as the total number of Negative Citations for each case, which is summarized in Figure 5.

5.2 Identification strategy and empirical model

As in the preceding analyses, the appropriate research design here is the difference-in-differences design. We model the same empirical model, where now our dependent variable is the number of negative

[17]As we discuss in the conclusion, this is but one possible measure of opinion quality. However, because a negative treatment is such a significant and rare occurrence, we believe it is a particularly powerful measure of opinion quality.

citations an opinion receives, rather than the number of days it takes to produce. We estimate

$$\begin{aligned} NegativeCitations_i &= \gamma_0 + \gamma_1 Tournament_{a[i]y[i]} \\ &\quad + \gamma_2 MarchMadness_{t[i]} \\ &\quad + \gamma_3 TournamentAuthor_{a[i]y[i]} \times MarchMadness_{t[i]} \\ &\quad + \xi^T \mathbf{X}_i + \varepsilon_i \end{aligned} \tag{3}$$

where, as before, ε is a Normally-distributed random variable with mean 0 and standard deviation σ_ε. For each model we exclude a small number of observations where the number of citations is an extreme outlier.

We report the results of estimating this model in the first column in Table 6. Here, we find a positive correlation between the interaction of $Tournament_{a[i]y[i]}$ and $MarchMadness_{t[i]}$ and the number of negative citations an opinion receives. These results indicate, then, that just as the time it takes to write a decision, the quality of the work product itself is affected by a judge's distraction by the March Madness Tournament. Substantively, these results indicate that in a month not during the Tournament, an opinion written by a judge whose alma mater is not playing in the Tournament will receive, in expectation, 1.9 negative citations. Opinions written during the Tournament by the same judge receive, in expectation, 2.8 opinion, a 50% increase. Again, most of our results hold if we employ a negative binomial specification. The result in the "Positive Cites" model does not.

It is important to note, of course, that these effects are not limited to the metric of opinion quality we have selected. In the next three columns of Table 6, we report estimates of the model using alternative metrics of quality. First, we separate out citation types. In particular, we examine the number of Negative citations, excluding "Distinguishing" citations, which are often not critical of an opinion but simply setting out to set the opinion aside as not on-point. We also consider the total number of all citations (negative, positive, and neutral citations) and the total number of positive citations. If the Tournament is having an effect on opinion quality, we should expect to see a positive effect on negative citations, and a negative effect on positive citations. The expectation concerning all citations is more ambiguous, though one might expect a lower quality opinion is cited less often, overall.

The results from those models corroborate the finding from the first model. The Tournament has

	Negative Cites	Negative Cites (w/o Distinguishing)	Positive Cites	Total Cites
Tournament$_{a[i]t[i]}$	-0.41	-0.04	-3.64	-2.46
	(0.33)	(0.19)	(11.79)	(10.39)
March Madness$_i$	-0.46^{*}	-0.20	6.42	32.96^{***}
	(0.27)	(0.15)	(9.10)	(7.94)
Tournament$_{a[i]t[i]}$ $\times$ March Madness$_i$	1.35^{*}	1.00^{**}	-12.48	-39.64^{*}
	(0.74)	(0.42)	(24.68)	(21.74)
AIC	0.01	0.01	0.28	0.25
N	1424	1424	1437	1429

Table 6: *Empirical models of number of citations to federal court of appeals decisions, 1985-2000.* Cell entries are OLS regression coefficients (standard errors in parentheses). $^{***}p \leq 0.01$; $^{**}p \leq 0.05$; $^{*}p \leq 0.10$

a positive effect on the number of negative citations, even when we exclude "distinguishing" citations. The Tournament has a negative effect on both the number of positive citations and the total number of citations. Taken together, these results suggest the Tournament cases a drop in the impact or quality of a published opinion.

One might be worried that other things, in particular the length of time to write a decision, affect the rate of negative citation. In general, it is inappropriate to include the time to decision as a covariate, as that is post-treatment to the March Madness variables and therefore introduces bias. However, when we estimate such a model, we still find a positive effect of the Tournament interaction and the number of negative citations.

Further, one might be concerned that the time it takes to write a decision mediates all of the effect on negative citations—that there is no direct effect of distraction on the rate of negative citation. Judges simply take longer to write their opinions (as we saw above) and that affects the rate of negative citation. The challenge for an empirical analysis is to assess whether any effect of the Tournament on opinion quality can be offset by taking longer to complete an opinion. As we described above, judges do not necessarily face strict deadlines for their decisions, and so if they want to avoid the deleterious effects of a marginal increase in leisure utility on the quality of the decisions, they could just always delay their work and take the time necessary to write a decision of sufficient quality. As we saw above, there is certainly an effect of a desire for leisure on the time it takes a judge to do her work. At the same time, the caseload pressure judges face suggests there is a limit to how they can delay, and so it is possible that they cannot fully offset the effect of distraction through delay.

6 Discussion and Conclusion

The findings we report here have implications for at least three areas in which the literature on judging has grown in recent years. First, the findings speak to the growing literature, with roots in Posner's (1993) account of the judicial process, on the various incentives judges face. Second, the analysis and findings provide groundwork for the analysis of multiple components of the judicial work product—speed and quality—and highlight the value of articulating the range of options judges have when deciding how to respond to changes in their incentives. Third, the findings show how empirical approaches to studying judicial leisure, which are often deployed at the trial court level, can be used to study appellate judging, where the theoretical framework that dominates the literature often applies.

A broader model of judging. The literature on judicial behavior has evolved considerably over the last century. Whereas once debates centered on whether and precisely how ideological beliefs influence judicial work, ideology is now understood to be just one of many forces that operate on the judicial behavior. As we have come to think about judges as motivated by general career concerns, various claims have been made about how judges, as laborers in a market, will change their work behavior when they increasingly value leisure time (e.g., Posner 1993). A central challenge to the ability to move forward research on the role of leisure is empirical. How do we measure a judge's interest in leisure, let alone variation in leisure incentives within a given judge over time or across judges at at given point in time? Our research design offers an opportunity for assessing the consequences of heightened incentives for leisure. The NCAA Men's Basketball Tournament is one of the most popular sporting events in the country every year, and reports often demonstrate a massive effect of the Tournament on workplace productivity in the private sector. By leveraging the selection of teams into the Tournament exogenous of individual judges' workloads, our difference-in-differences design shows that when judges' alma maters appear in the Tournament, they divert effort away from judging, as predicted by the literature on judges in the labor market. Crucially, the causal identification of our research design is stronger than past studies, because we can leverage exogenous shocks to judges' leisure interests that vary over time and across individual judges.

Of course, our analysis is in many ways only a single step forward. Our empirical leverage on judicial leisure is limited—we exploit exogenous variation in a judge's leisure incentive from the NCAA Tournament. Surely, judges face much more significant and common increases to their leisure incentives, such as when a child is to marry, at the death of a loved one, or because of other personal interests. A richer investigation of the effects of leisure would employ designs that could exploit exogenous variation in other, potentially more powerful, shocks in leisure incentives.

Further, what we cannot address, and what the two strands of literature need, is a comprehensive model that directly evaluates how judges balance their myriad interests. Epstein, Landes and Posner (2013) lay out many of interests a judge might have and evaluate many of them empirically. However, what remains elusive is a comprehensive theoretical model that identifies how those incentives interact and is subject to empirical evaluation. While we do not propose a theoretical innovation, we expect the empirical strategy we identify will contribute to on-going efforts to elaborate a more holistic model of judging.

Are the effects we find important? By synthesizing existing claims about the many forces that influence judicial behavior, we have shown that the existing literature unequivocally predicts that judges will work more slowly when confronted with a heightened interest in leisure but that the literature makes inconclusive predictions with respect to the quality of the work judges produce. Moreover, our empirical strategy provides analytic leverage with respect to both components. We are naturally led to ask what all of this means? Are the effects we observe important?

Unfortunately, there is no definitive answer to this question. Different people will have different opinions about how much of a delay in case resolution is tolerable. And people will differ over whether a particular delay is tolerable in light of the effect of leisure on quality. That said, we are confident that we can place some structure on our normative evaluations. If delay is the cost of quality, we are in a position to ask how much of a loss in quality we would be willing to sustain in order to ensure that cases are not delayed when judges are distracted. If we suppose that there is this kind of tradeoff, then it becomes easier to evaluate the substantive significance of our findings. What we find is that changes in leisure incentives result in a both in delays in the completion of cases and the reduction of quality. As far as we can tell, judges are not fully trading off one harm for another. They would appear to be balancing the harms simultaneously. More to the point, if we continue to assume that judges both want to resolve cases without delay and to write high quality opinions, we are led to conclude that the delays we are observing are likely shorter than they would be if judges were willing to give even more on the quality dimension. Similarly, the quality reduction we observe is lower than it would be if judges were willing to eliminate delays. In evaluating the substantive significance of particular findings, it is critical to keep this tradeoff in mind.

With respect to workrate, we find that when a judge's alma mater participates in the Tournament, he takes longer to write the opinions for which he is responsible. In particular, during months other than those during which the Tournament is taking place, there is no difference in the speed with which judges from that year's Tournament teams work and the speed with which judges whose team are not in that year's Tournament work. However, during the months of the Tournament, those judges take between 11 and 20 days longer to complete their opinions. The 95% confidence intervals include effects as long as 37 days.[18] The absolute magnitude of the estimated effect may not seem particularly long, but it is commensurate to what one might reasonably expect given the nature of the variation we exploit. Were we to have analytic leverage on other, more powerful shocks to a judge's leisure incentives, one might expect more substantively meaningful effects. The insight from our analysis is instead the credible claim of a causal effect of leisure on the judicial work product.

With respect to work quality, we find that judges' opinions do suffer when their authors face heightened incentives for leisure. In particular, the effect of an opinion author having his alma mater in the Tournament is to increase the number of negative citations to the opinion. Of course, this is but one possible measure of opinion quality, but the finding is important. Negative citations are rare, and so the finding that judicial leisure incentives cause an increase in negative treatment of an opinion is particularly striking. Moreover, as noted, the extant theoretical accounts of judicial responses to leisure incentives is equivocal on the implications for the quality of judges' work. Judges might slow down in order to maintain their work quality, or they might trade off on both dimensions, balancing a deterioration of quality and speed. We find evidence suggesting the latter. The substantive magnitude of these effects is less clear, as it can be hard to know what to interpret from a given change in citations to an opinion.

Expanding the model to appellate judges. One way in which we might make such headway relates to a third aspect of our analysis—considering the different positions in which various judges find themselves. As described above, much of the extant empirical literature focuses on trial judges. However, one might suspect the theoretical incentives identified in the literature more naturally apply to appellate judges. In particular, judges at the US Court of Appeals face little risk of review by their only superior court—the Supreme Court. Moreover, their more prestigious positions insulate them somewhat from the kinds of career and reputation concerns that might influence judges in lower courts. A comprehensive theoretical model of the judicial process would embed the various kinds of judicial positions to assess how institutional circumstances condition a judge's responsiveness to the various incentives and willingness to tradeoff various components of the judicial work product.

More important, the higher stature and consequences of appellate judging—especially within written opinions—suggests greater implications for the findings here than those in the context of lower courts.

[18] The confidence intervals for the interactive term in each of the four models reported in Table 2 are: (3.39,21.07), (-4.11,19.61), (4.48,36.67), and (2.30,20.98).

Delays in resolution of important questions of law have significant implications for the rule of law and justice. Deteriorating quality of appellate opinions can affect the law in many cases and even the extent to which the Supreme Court must intervene with law making in the circuits. Moreover, we might expect that the way judges respond to leisure incentives can influence the behavior of judges in lower courts, given most models of hierarchy and oversight predict that subordinates' behavior is conditioned by the preferences of their superiors. It is again critical to remember that the effects we identify are modest. Still, also bears repeating that the treatment is relatively weak. Future work should look for stronger sources of variation in leisure incentives.

Taken together these various implications from our study—the building upon literature that develops a broader model of judging, the incorporation of multiple components of the judicial work product, and the expansion of empirical studies to appellate courts—provide a step forward in the study of judicial incentives and performance. At the same time, we recognize our findings are just that—only a step. We anticipate future research will build upon our findings and research design as the literature moves towards a more comprehensive model of judging.

References

Ash, Elliott and W Bentley MacLeod. 2015. "Intrinsic motivation in public service: Theory and evidence from state supreme courts." *The Journal of Law and Economics* 58(4):863–913.

Bainbridge, Stephen M and G Mitu Gulati. 2002. "How Do Judges Maximize?(The Same Way Everybody Else Does-Boundedly): Rules of Thumb in Securities Fraud Opinions." *Emory Law Journal* 51.

Baum, Lawrence. 2009. *Judges and their audiences: A perspective on judicial behavior.* Princeton University Press.

Blatcher, Marjorie. 1978. *The Court of King's Bench 1450-1550: A Study in Self-help.* The Athlone Press.

Book, Judicial Yellow. 2003. "Who's Who in Federal and State Courts.".

Choi, Stephen J. and Mitu Gulati. 2007. "Ranking Judges According to Citation Bias (As a Means to Reduce Bias)." *Notre Dame Law Review* 82(3):1279–1309.

Choi, Stephen J, Mitu Gulati and Eric A Posner. 2012. "What do federal district judges want? An analysis of publications, citations, and reversals." *Journal of Law, Economics, and Organization* 28(3):518–549.

Clark, Tom S. and Aaron B. Strauss. 2010. "The Implications of High Court Docket Control for Resource Allocation and Legal Efficiency." *Journal of Theoretical Politics* 22(2):247–268.

Clotfelter, Charles. 2012. "March Madness and Productivity: The Real Story.".
URL: *https://hbr.org/2012/03/the-march-madness-really-destr/*

Cohen, Mark A. 1991. "Explaining Judicial Behavior or What's" Unconstitutional" About the Sentencing Commission?" *Journal of Law, Economics, & Organization* 7(1):183–199.

Cohen, Mark A. 1992. "The motives of judges: Empirical evidence from antitrust sentencing." *International Review of law and Economics* 12(1):13–30.

Drahozal, Christopher R. 1998. "Judicial Incentives and the Appeals Process." *SMU Law Review* 51:469.

Epstein, Lee and Jack Knight. 1998. *The Choices Justices Make.* Washington, D.C.: C.Q. Press.

Epstein, Lee, Jeffrey A. Segal, Nancy Staudt and René Lindstädt. 2005. "The Role of Qualifications in the Confirmation of Nominees to the US Supreme Court." *Florida State University Law Review* 32:1145–1173.

Epstein, Lee, William M Landes and Richard A Posner. 2013. *The Behavior of Federal Judges: A Theoretical and Empirical Study of Rational Choice.* Harvard University Press.

Garoupa, Nuno and Tom Ginsburg. 2010. "Reputation, Information and the Organization of the Judiciary." *J. OF COMP. L.* 4:226.

Gulati, Mitu and Catherine MA McCauliff. 1998. "On" Not" Making Law." *Law and Contemporary Problems* 61(3):157–227.

Gulati, Mitu and Richard A Posner. 2016. "Management of Staff by Federal Court of Appeals Judges, The." *Vand. L. Rev.* 69:479.

Klein, David E. and Robert J. Hume. 2003. "Fear of Reversal as an Explanation of Lower Court Compliance." *Law and Society Review* 37(3):579–606.

Lang, Duncan Temple. 2013. *XML: Tools for parsing and generating XML within R and S-Plus.* R package version 3.98-1.1.
URL: *http://CRAN.R-project.org/package=XML*

Lozano, Fernando A. 2011. "The Flexibility of the Workweek in the United States: Evidence from the FIFA World Cup." *Economic Inquiry* 49(2):512–529.

Maltzman, Forrest, James F. Spriggs, II and Paul J. Wahlbeck. 2000. *Crafting Law on the Supreme Court: The Collegial Game.* New York: Cambridge University Press.

Miceli, Thomas J and Metin M Coşgel. 1994. "Reputation and judicial decision-making." *Journal of Economic Behavior & Organization* 23(1):31–51.

Miller, Geoffrey P. 2004. "Bad Judges." *Tex L. Rev.* 83:431.

Peppers, Todd C, Michael W Giles and Bridget Tainer-Parkins. 2014. "Surgeons or Scribes: The Role of United States Court of Appeals Law clerks in Appellate Triage." *Marq. L. Rev.* 98:313.

Posner, Richard A. 1993. "What Do Judges and Justices Maximize? (The Same Thing Everybody Else Does)." *Supreme Court Economic Review* 3(1):1–41.

Posner, Richard A. 2000. "Is the Ninth Circuit Too Large: A Statistical Study of Judicial Quality." *Journal of Legal Studies* 29(2):711–719.

Reinhardt, Stephen. 1993. "A Plea to Save the Federal Courts: Too Few Judges, Too Many Cases." *ABA Journal* 79(1):52–54.

Savchak, Elisha Carol, Thomas G. Hansford, Donald R. Songer, Kenneth L. Manning and Robert A. Carp. 2006. "Taking It to the Next Level: The Elevation of District Court Judges to the U.S. Courts of Appeals." *American Journal of Political Science* 50(2):478–493.

Segal, Jeffrey A. and Harold J Spaeth. 2002. *The Supreme Court and the Attitudinal Model Revisited*. New York: Cambridge University Press.

Songer, Donald R. 1999. "The United States Courts of Appeals Database, 1925-1996." available at `http://www.as.uky.edu/polisci/ulmerproject`, accessed 28 November 2006.

Appendix

Here we report the results of the placebo tests described in the paper.

First, we replicate out analysis as a "placebo" test, coding each other month of the year as the period of the Tournament. We then re-estimate Equation (2) with the key variable, March Madness$_i$, coded, in turn, according to the placebo. We include fixed effects for the Author of the case opinion as well as for the year the case was decided. Standard errors are clustered on the three-judge combination that decided the case. We reported the OLS estimates of the difference-in-difference estimate, given by the key term—Tournament$_{a[i]t[i]}$ $\times$ March Madness$_i$—in Table 7 along with its standard error. We report results from the model that codes the panel as having a Tournament judge when any member of the panel has an ala mater participating in the Tournament. The alternative coding yields similar results, again with none of the difference-in-differences estimates statistically distinguishable from zero. We do not report the other estimated coefficients here for reasons of space. The key finding here is that none of these coefficients is statistically distinguishable from 0.

Month	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
January	9.86^{***}	2.96	23.84
	(2.92)	(10.00)	(12.60)
February	11.21^{***}	15.36^{*}	7.33
	(2.31)	(9.21)	(8.38)
May	13.84^{***}	6.05	−13.93
	(1.78)	(12.26)	(14.68)
June	8.51^{**}	-31.28^{***}	11.53
	(4.26)	(8.65)	(9.98)
July	11.03^{***}	−31.94	13.13
	(4.10)	(30.72)	(21.16)
August	11.79^{***}	3.82	−9.30
	(4.23)	(10.95)	(13.24)
September	11.13	−46.95	−12.00
	(4.09^{***})	(55.10)	(35.59)
October	13.79^{***}	8.50	−16.85
	(4.22)	(9.34)	(8.52)
November	12.70	11.16	−21.51
	(4.16)	(10.22)	(16.68)
December	11.92^{***}	3.96	−9.74
	(4.22)	(11.84)	(19.50)

Table 7: *Results of placebo test.* Table shows the results of 10 separate OLS models. We report the three main coefficients. The column with $\hat{\beta}_3$ gives the difference-in-difference estimate along with its standard error. In each model, we code the indicated month as the NCAA Tournament and estimate the model from Equation (2). For reference, in the true data, the estimated coefficient associated with the difference-in-difference estimate is $\hat{\beta}_3 = 11.64$, $se = 4.67$. $^{***}p \leq 0.01$; $^{**}p \leq 0.05$; $^{*}p \leq 0.10$