Estimating the Effect of Leisure on Judicial Performance*

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

November 14, 2015

Abstract

Past research suggests federal judges confront incentives that undermine the speed and quality with which they resolve cases when leisure interests are particularly strong. Alternatively, the selection process for federal judges, which seeks to identify intrinsically motivated individuals, as well as judges’ own desire for prestige—commonly tied to the quality of their work—may considerably mitigate this concern. The challenge with empirically evaluating these claims lies in measuring latent preferences for leisure. We take advantage of an annual sporting event that creates differential distractions across judges. Using a difference-in-differences design, we show 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 have implications for the tradeoff between judicial independence and accountability and reinforce the importance of the recruitment of high quality judges.

*We thank Chuck Cameron, Adam Glynn, Thomas Gschwend, Kosuke Imai, and John Kastellec 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.” Indeed, the decades-long debate about the role of ideology in judicial decision making was primarily about whether judges reference their own world views when exercising discretion. Scholars from the law and economics tradition have been particularly interested in how judges approach their jobs similarly to workers in more typical firm settings, in light of the institutional differences between traditional firms and the judiciary. Prominent in this area of inquiry has been the consideration of whether judicial institutions incentivize judges to produce less and lower quality work than they are capable of producing, especially when the draw of leisure activities is particularly strong (e.g., Posner 1993).

However, as we discuss below, what we should expect to observe when judges confront an increased marginal benefit from leisure, given extant theoretical frameworks, is unclear. For example, when faced with an increased desire for leisure, judges may opt to slow down their work, simply taking longer to resolve cases, working harder to catch up when they are not otherwise distracted. Alternatively, judges, facing a near-constant inflow of cases for resolution might decide to skimp on the quality of their decision-making. Finally, perhaps the selection process by which judges are chosen produces judges who are intrinsically motivated to produce high-quality work. If so, then they may use the ability to delay their work in order to ensure high-quality decision-making. Each of these potential strategies for judges implies different empirical patterns in the relationship between an incentive for leisure and judicial work product.

A significant challenge to evaluating these expectations, though, is that it is difficult to directly observe judges’ latent incentives. When does a judge’s marginal utility to leisure increase? We propose a research design to empirically identify the effect of a heightened marginal benefit from leisure on judicial work product. The NCAA Men’s Basketball Tournament is one of the most popular, attention-getting sporting tournaments in the United States. Each year it attracts considerable attention, even from among non-sports fans. Critically, it is credited with substantial drops in productivity in the for-profit private sector. We collect data on the teams participating in the NCAA Tournament each year and match judges to their alma maters. 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 (1) to delay (considerably)
the time it takes a judge to write opinions for cases heard during the Tournament, (2) to produce a lower quality opinion, and yet is not (3) mitigated by judges ability to trade off timing and opinion quality.

The implications of these findings are considerable. As we describe in the conclusion, the empirical results provide new insight into an important debate about the consequences of judicial institutions and debates from the industrial organization literature about the efficiency and performance of politically insulated judges. Moreover, the findings have broader implications for understanding performance by politicians and bureaucrats who are not often subject to strong performance incentives.

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 Judicial Motivation and Leisure

The primary judicial task involves providing dispositions to cases—indicating whether Party A or Party B wins. Judges do this in the context of writing opinions, which provide rationales for these dispositions, which themselves follow from interpretations of past opinions, statutes, regulations, etc. On occasion, these opinions also develop new rules of law. Critically, this happens 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 back log (Reinhardt 1993).

When processing and resolving these cases, judges are evaluated on a number of elements of their decision-making. Among those elements, two are of particular import: the quality of their decisions and the speed or efficiency with which they resolve cases. The quality of judicial decision-making is of paramount importance in a number of public and scholarly evaluations of judges and the courts. When selecting judges for the federal bench, a nominee’s skills, preparedness and demonstrated legal capacity is a central consideration (Epstein et al. 2005). And, in studies of judicial performance and decision making, scholars are often interested in the incentives different institutional features create for judges
to invest effort into the quality of their decision making (e.g., Posner 1993, Lax and Cameron 2007, Clark and Carrubba 2012, Goelzhauser and Cann 2014, Ash and MacLeod 2015).

At the same time, 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.\footnote{http://www.uscourts.gov/file/document/2015-judicial-conference-judgeship-recommendations} The Administrative Office of the United States Courts is required to provide an annual report of statical information, which always includes an accounting of caseload and backlog, identifying particular circuits which are lagging behind.\footnote{28 U.S.C. §604(a)(2)} This information is commonly reported on by media outlets,\footnote{See http://www.wsj.com/articles/in-federal-courts-civil-cases-pile-up-1428343746.} where particular judges are not too infrequently shamed.\footnote{See http://www.nytimes.com/1995/04/17/nyregion/awaiting-judgment-special-report-new-york-s-clogged-us-courts-delaying-civil.html.}

Moreover, judges, especially appellate judges, handle cases of immense importance, and their decisions can dramatically affect individual lives and sometimes influence policy that touches nearly everyone in their jurisdiction. It is therefore important to understand how judicial institutions translate normal human motivations into behavioral incentives for judges. Scholars working in the law and economics tradition have relied extensively on theories of the firm, as well as other organizational models, to generate expectations about how judges respond to different conditions. Chief among those that have been investigated is the incentive for leisure. Posner (1993) most notably outlines a list of activities from which judges derive utility, among which are power and influence, prestige, and leisure. Posner claims that because judges (presumably like others) desire leisure, they will not perform as well in their judicial functions as they would absent that incentive.

Crucial for Posner’s claim is that judges are best though of as analogous to non-profit workers. Posner claims their work product is highly unobservable, and their institutional arrangements do not create incentives for them to produce at maximum capacity. For example, judges are not paid by the number of cases they file, the quality of their decisions, or the number of hours they devote to resolving a dispute.\footnote{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).} As our discussion above suggests, it is not necessarily the case that judicial work product is truly unobservable. The public, politicians, the media, and scholars all frequently comment on both
judicial quality and efficiency. However, what is true is that between the two dimensions, the quality of judicial work is relatively harder to discern than the efficiency, or speed, with which judges resolve cases. The 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, and judges themselves often discuss publicly the number of cases they confront.

With these observations in hand, we return to the motivating question—what happens when judges face a heightened marginal benefit from leisure? How will they respond? Posner claims that judges will shirk on their professional responsibilities, taking advantage of judicial insulation from political oversight, carrots, and sticks. But, how should that manifest in terms of the time to case resolution and the quality of decision making? To generate more concrete expectations, it is useful to provide deeper background on the judicial context and judges’ motivations.

To begin, judges face a constant flow of cases and have a distinct distaste for backlog. 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. This work flow interacts importantly with the observability of quality. When quality is readily observable, in order to avoid a backlog, a worker must give greater effort after personal matters are resolved in order to ensure that product quality remains high. Productivity drops corresponding to increases in the marginal utility of leisure are offset by increases in effort on days when there are less pressing personal matters—i.e., workers can shift their professional efforts to other days or times. When quality is unobservable, however, workers are presented with an alternative strategy. To avoid a backlog when quality cannot be observed, decreases in effort when a worker is distracted by personal interests need not be offset by increases effort at a later date(s). Instead, effort can be kept constant and quality can be sacrificed in order to avoid a backlog. 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...
who believe we are doing the same quality work that we did in the past are simply fooling themselves. We adopt more and more procedures for “expediting” case, procedures that ensure that individual cases will get less attention.

Importantly for our discussion, the most immediate consequence of a decrease in effort today is an increase in tomorrow’s caseload.

At the same time, 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 nature of judicial work. 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).

Finally, although leisure incentives might decrease quality, judges are able to smooth out their effort reductions in ways that make it difficult or impossible to observe (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.

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”).

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.\(^6\) 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.\(^7\) For 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. Thus, the Tournament represents a particularly useful way of evaluating productivity under personal distraction.

The effect of the NCAA Tournament on workplace productivity in the US is something that has been often discussed in popular culture. 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.\(^8\) (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 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

\(^7\)In surveys conducted by CareerBuilder.com, during recent years between 15% and 20% of respondents report participating in such pools.
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.\(^9\)

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.\(^10\) To estimate the causal effect of Tournament-induced distraction on judges’ performance, we use a difference-in-differences design, which suggests strong evidence of a meaningful effect on the amount of time it takes a judge to prepare a written opinion in a case. 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.

\(^9\)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.

\(^{10}\)A reasonable concern is that clerks (and not judges) are doing the lion share of the work. The difference-in-differences design with year- and judge-fixed effects accounts for judge specific clerk characteristics, assuming that Tournament Authors do not disproportionately assign excess work to their clerks during the Tournament season. In the event that this is what happens, that judges distracted by the Tournament are able to offload excess work on their clerks, then the process creates a bias against finding a difference in the differences between Tournament Authors and Non-Tournament Authors.
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. 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 32,586 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.

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 a decision is decided by more than three judges (e.g., en banc decisions), we exclude the case from our

\[11\] Commonly-used databases, such as the US Appeals Courts Database (Songer 1999) do not contain oral argument dates.
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.

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 on last name.

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 8,934 cases. The vast majority
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).

<table>
<thead>
<tr>
<th></th>
<th>Case Not Heard During Tournament</th>
<th>Case Heard During Tournament</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Tournament Author</td>
<td>5,675 (63.69%)</td>
<td>1,472 (16.52%)</td>
</tr>
<tr>
<td>Tournament Author</td>
<td>1,382 (15.51%)</td>
<td>382 (4.29%)</td>
</tr>
</tbody>
</table>

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 23 cases our data indicated that a decision took less than a day. While this is unlikely we excluded these cases leaving as with 8,911 cases to analyze.

We code the variable $Tournament_{[i]}$ 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 $MarchMadness_{[i]}$ 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.\textsuperscript{12}

In our data, 1,764 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

\textsuperscript{12}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.
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 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. Non-Tournament Authors need 139 days on average to take a decision while Tournament Authors need only 127 days in month other than February or March (compare black triangle in Figure 3). This picture changes completely during the NCAA Tournament. Non-Tournament Authors are faster in taking decisions with 123 days on average compared to Tournament Authors who take 151 days on average (compare gray triangle in Figure 3). Hence, while Non-Tournament Authors are even faster than Tournament Authors during month other than February or March Tournament Authors take the longest time to a decision over all during the NCAA Tournament.

We model $\text{TimeToDecision}_i$ as a function of (1) whether the alma mater of any judge writing a decision is in the Tournament the year the case was argued, (2) whether the case was heard during
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.

February or March (i.e., during the time of the Tournament), and (3) the interaction of those two variables. Our expectation is that the interaction will have a positive relationship to TimeToDecision, as judges from Tournament teams will work more slowly during the Tournament months. Recall from Figure 1 that the dependent variable, the number of days between oral argument and the decision is log-linear, which is typical of count data such as these. To model this variable, we employ a Poisson specification. Formally, we assume

\[
\log (\mathbb{E}[\text{TimeToDecision}_i|\mathbf{M}_i]) = \alpha + \beta_1 \text{Tournament}_{a[i]}y[i] + \beta_2 \text{MarchMadness}_{t[i]} + \beta_3 \text{Tournament}_{a[i]}y[i] \times \text{MarchMadness}_{t[i]} \tag{1}
\]

where \(\mathbf{M}_i\) is a matrix of covariates. We also estimate the model including (i) year fixed effects, (ii) author fixed effects, and (iii) both author and year fixed effects. The primary results are reported in Table 2.

These estimates show that we estimate a positive and statistically significant effect of the interaction between \(\text{Tournament}_{a[i]}y[i]\) and \(\text{MarchMadness}_{t[i]}\) 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
<table>
<thead>
<tr>
<th></th>
<th>Main Model</th>
<th>Year Fixed Effects</th>
<th>Author Fixed Effects</th>
<th>Author &amp; Year Fixed Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tournament_a[i]t[i]</td>
<td>−0.09</td>
<td>−0.06</td>
<td>−0.05</td>
<td>−0.05</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>March Madness_i</td>
<td>−0.13</td>
<td>−0.08</td>
<td>−0.08</td>
<td>−0.07</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Tournament Judge × Tournament Case</td>
<td>0.30</td>
<td>0.27</td>
<td>0.21</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.006)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Intercept</td>
<td>4.93</td>
<td>6.05</td>
<td>3.98</td>
<td>3.49</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.008)</td>
<td>(0.08)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>AIC</td>
<td>729,880</td>
<td>692,346</td>
<td>548,830</td>
<td>543,568</td>
</tr>
<tr>
<td>N</td>
<td>8,911</td>
<td>8,911</td>
<td>8,911</td>
<td>8,911</td>
</tr>
</tbody>
</table>

Table 2: *Empirical models of number of days between oral argument and decision on federal court of appeals, 1985-2000.* Cell entries are Poisson regression coefficients (standard errors in parentheses).
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.”

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 99% confidence intervals. The effects are separated by Tournament and Non-Tournament Authors overwhelmingly supporting our hypothesis. Tournament Authors need about 127 days on average for a case during non-Tournament month. This time sharply increases during the Tournament to 151 days on average. On average, Non-Tournament Authors need longer time compared to Tournament Authors when writing opinions during Non-Tournament month, taking 139 days on average. However, during the Tournament they are not just faster in writing opinions but they are also faster than Tournament Authors during month other than March and February. Hence, on average Non-Tournament Authors need about 123 days during the time of the NCAA Tournament. In fact, the difference-in-differences estimate is an average of a 40-day change in the time to decision.

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—$\hat{\beta}_3 = 0.31$, $se = 0.01$. By contrast, if we conduct a “placebo” test and code only cases heard during April as being cases during the NCAA Tournament, we find no effect at all—$\hat{\beta}_3 = 0.0003$, $se = 0.01$. 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—$\hat{\beta}_3 = 0.19$, $se = 0.01$. 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.

These findings are consistent with the anticipated effects of a marginal increase in a judge’s leisure utility. Judges who are distracted by the NCAA Tournament slow down in their work production. Our design allows us to make inferences about the causal effect of the NCAA Tournament by comparing
judges subject to the leisure incentive—the Tournament Authors—to those not subject to it—the Non-Tournament Authors—to account for any systematic patterns associated with any differences between cases heard during the NCAA Tournament and those not heard during the NCAA Tournament. The pattern we observe from Non-Tournament Authors predicts that Tournament Authors should prepare the opinions from February and March cases in roughly 115 days; however, these judges actually take, on average, more than 150 days to write their opinions.

### 4.4 Threats to inference

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 do 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).

Let

$$A_{ai} = \exp(\beta_1 a_{Tournament_{i[i]} + \beta_2 a_{MarchMadness_{i[i]} + \beta_3 a_{Tournament_{i[i]} \times MarchMadness_{i[i]}}}}$$

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

$$\Pr(\text{Author}_i = a | \text{M}_i) = \frac{A_{ai}}{\sum_{k=1}^{3} A_{j[i[k]]i}}$$

(2)

where $j[i[k]]$ is the $k$th judge hearing case $i$, $\text{Author}_i$ is the identity of the author of the majority opinion in case $i$, and, as above, $\text{M}_i$ 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 $\beta_3$ 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
<table>
<thead>
<tr>
<th>Variable</th>
<th>β</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tournament(_{a[i][y][i]})</td>
<td>0.16</td>
<td>(0.03)</td>
</tr>
<tr>
<td>MarchMadness(_{t[i]})</td>
<td>0.00</td>
<td>(1.41)</td>
</tr>
<tr>
<td>Tournament(<em>{a[i][y][i]}) × MarchMadness(</em>{t[i]})</td>
<td>0.03</td>
<td>(0.07)</td>
</tr>
</tbody>
</table>

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.

A final 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.

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 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 searched Westlaw for each case in our data 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. For each case, we performed a KeyCite search, which uses Westlaw’s databases to identify every subsequent case citing the case we search (including unpublished decisions). Westlaw returns a report that divides all subsequent citations into discrete, and mutually exclusive categories. 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). 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

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. Our identification strategy allows us to assess the causal pathway of distraction—whether some or any of the effect of leisure on quality runs through, or is mitigated by, delay or instead whether there is simply a direct effect of leisure on opinion quality.

Given we have (competing) expectations about the effect of a delay on the quality of a judicial
opinion, as well as an expectation about the direct effect of judicial distraction on opinion quality, we need an empirical strategy to identify the path illustrated in Figure 6. This is, we want to identify both the average direct effect (ADE) of the Tournament (T) on opinion quality (Q) and the average causal mediation effect (ACME) of the Tournament that is mediated (M) through its effect on time judges take to render their opinions. Taken together the sum of ADE and ACME is the average treatment effect (ATE).

A traditional way of estimating the total effect of our treatment variable, when we suspect it is mediated by another variable (here, the time to decision), is known as linear structural equation modeling (LSEM). The method is to simultaneously estimate the direct effect of the Tournament status on the quality of the opinion, the effect of the Tournament status on the time to decision, and the effect of the Tournament status on opinion quality, conditioning on (controlling for) the time to decision. One can then use the estimated coefficients from the three models to calculate the direct effect of Tournament status on opinion quality, and the effect of Tournament status on opinion quality that is mediated by (contained in) any effect of the time to decision on opinion quality. However, Imai, Keele and Tingley (2010) show that this estimate is valid under strong assumptions only, including that the relationship between the treatment and the mediator is linear. This assumption is not only strong but in our case inappropriate.

We employ, again, a difference-in-differences strategy but in conjunction with the method proposed by Imai et al. (2011). This will allow for estimating causal paths that may run through a mediating variable whereby we are able to use our non-parametric identification results to estimate the ADE and ACME independent of the particular statistical model (Imai et al. 2011, 773). Here, the causal path is from the Tournament to opinion quality, where we suspect part of the causal path may run
through the time to decision variable. The intuition behind Imai et al.’s method is direct. We first estimate a model of the time to decision, which is our *mediating* variable. That is, we suspect part of the effect of the Tournament on opinion quality runs through the time to decision. We then estimate a model of opinion quality as a function of our Tournament variables, the mediating variable (time to decision), and all other covariates in our time to decision model. With these two models estimated, we can predict the quality of the opinion using the time to decision predicted when an opinion author’s team is in the Tournament and the Tournament is taking place. We can also predict the quality of the opinion when that condition is not met—i.e., when either the opinion author does not have a team in the Tournament, the Tournament is not taking place, or neither (our interactive term is 0). The average difference between those two predictions is the average causal mediation affect (ACME), where the average treatment effect (ATE) is simply the sum of the ACME and the average direct effect (ADE) of the Tournament status on the opinion quality.

5.3 The empirical model

We estimate two separate equations. Where $M_i$ is a matrix of covariates including $Tournament_{a[i]y[i]}$ and $MarchMadness_{t[i]}$, $X_i$ is a matrix of indicator variables for authors and years, and $\xi_1$ and $\xi_2$ are vectors of parameters to be estimated, we estimate

$$\log (E[TimeToDecision_i|M_i, X_i]) = \beta_0 + \beta_1 Tournament_{a[i]t[i]}$$
$$+ \beta_2 MarchMadness_{t[i]}$$
$$+ \beta_3 Tournament_{a[i]y[i]} \times MarchMadness_{t[i]}$$
$$+ \xi_1^T X_i$$

$$\log (E[NegativeCitations_i|M_i, X_i]) = \gamma_0 + \gamma_1 Tournament_{a[i]y[i]}$$
$$+ \gamma_2 MarchMadness_{t[i]}$$
$$+ \gamma_3 TournamentAuthor_{a[i]y[i]} \times MarchMadness_{t[i]}$$
$$+ \gamma_4 \log (TimeToDecision_i)$$
$$+ \xi_2^T X_i$$

We estimate the models using only observations for which we have all of the covariates for each model—i.e., we exclude observations from our estimation of Equation (3) observations for which we were unable to collect citation data. Consistent with our analysis above, the estimated effects here
are very robust to any particular specification of the model using fixed effects. We report here two specifications—without fixed effects and with Author fixed effects, as these demonstrate the range of estimates we recover from various specifications. We focus our discussion here on the model without fixed effects to facilitate analysis of the effect that is mediated by the time to decision.

Table 4 summarizes the estimates from these models. Before considering the mediating effect of time on opinion quality, we first note the positive coefficient associated with the interaction of \( \text{Tournament}_{a[i]} \times \text{MarchMadness}_{t[i]} \). Given the two models are estimated separately, this quantity corresponds to a difference-in-differences estimate in a model where we also control for Time to Decision\(_i\). The coefficient indicates, then, that the difference-in-differences is positive—the difference between negative citations of opinions written by Tournament Authors and opinions written by other authors \emph{increases} during the NCAA Tournament. On its face, this is strongly suggestive of a decrease in opinion quality when judges have an extra leisure incentive.

To understand the meaning of this estimate, it is instructive to consider competing causal paths. As noted, one such path may be direct—judges may simply do poorer quality work when they are distracted by the Tournament. Another such path is indirect, or mediated; judges may take longer to do their work and as a consequence might offset the deleterious effects of distraction on quality. To assess this possibility, we employ the algorithm proposed in Imai et al. (2011) to calculate the direct effect of Tournament status on opinion quality and the effect on opinion quality that is accounted for by its effect on the time to decision. In particular, we calculate the predicted value of \( \text{TimeToDecision}_i \) when \( \text{Tournament}_{a[i]} \times \text{MarchMadness}_{t[i]} = 1 \) and the predicted value of \( \text{TimeToDecision}_i \) when \( \text{Tournament}_{a[i]} \times \text{MarchMadness}_{t[i]} = 0 \). We then calculate the predicted value of \( \text{NegativeCitations}_i \) using the two different predicted values of \( \text{TimeToDecision}_i \). This quantity is the ACME. We estimate as well as the ACME and ADE of Tournament status on the number of Negative Citations. Our estimates of the ACME and ADE are summarized in Table 5. The cell entries are estimated effects of the Tournament status on the number of negative citations, with 95% quasi-Bayesian confidence intervals reported below, using robust standard errors. We estimate these effects using the \text{mediation} package developed for R by Tingley et al. (2014).

The first thing to note is that the estimate of the direct effect of the Tournament (for Tournament Authors) is positive—even accounting for any effect that runs through the time to decision, there is a direct negative effect of the interaction of \( \text{Tournament}_{a[i]} \times \text{MarchMadness}_{t[i]} \) on the number of
<table>
<thead>
<tr>
<th></th>
<th>No Fixed Effects</th>
<th>Author Fixed Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time to Decision</td>
<td>Negative Citations</td>
</tr>
<tr>
<td></td>
<td>(OLS)</td>
<td>(Poisson)</td>
</tr>
<tr>
<td></td>
<td>Time to Decision</td>
<td>Negative Citations</td>
</tr>
<tr>
<td></td>
<td>(OLS)</td>
<td>(Poisson)</td>
</tr>
<tr>
<td>TournamentAuthor&lt;sub&gt;i&lt;/sub&gt;</td>
<td>-0.15 (0.03)</td>
<td>-0.26 (0.02)</td>
</tr>
<tr>
<td>MarchMadness&lt;sub&gt;t[i]&lt;/sub&gt;</td>
<td>-0.19 (0.03)</td>
<td>-0.14 (0.02)</td>
</tr>
<tr>
<td>TournamentAuthor&lt;sub&gt;i&lt;/sub&gt; × MarchMadness&lt;sub&gt;t[i]&lt;/sub&gt;</td>
<td>0.40 (0.07)</td>
<td>0.27 (0.05)</td>
</tr>
<tr>
<td>log(Time to Decision)</td>
<td>4.65 (0.01)</td>
<td>0.17 (0.01)</td>
</tr>
<tr>
<td>Intercept</td>
<td>12,609</td>
<td>40.396</td>
</tr>
<tr>
<td>N</td>
<td>5449</td>
<td>5449</td>
</tr>
</tbody>
</table>

Table 4: Empirical models of effect of Tournament status on opinion quality. The two models are estimated separately. The Negative Citations model has the total number of negative citations to each case as the dependent variable, and cell entries are Poisson coefficients. The Time to Decision model has the logged number of days from oral argument to the decision as the dependent variable, and cell entries are OLS coefficients. Standard errors in parentheses for both models.
Effect Estimate

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average causal mediation effect (ACME)</td>
<td>0.23 [0.13, 0.38]</td>
</tr>
<tr>
<td>Average direct effect (ADE)</td>
<td>0.96 [0.08, 2.10]</td>
</tr>
<tr>
<td>Proportion of effect mediated</td>
<td>0.19 [0.09, 0.60]</td>
</tr>
</tbody>
</table>

Table 5: Estimated direct and mediated effect if Tournament status on opinion quality.

negative citations to an case. This evidence suggests that judges do in fact compromise the quality of their opinions when they are faced with an additional leisure incentive.

What is more, being able to take longer to write an opinion does not appear to offset that effect in any way. In all of the specifications we have considered, the point estimate for the effect of the Tournament that runs through the time to decision is actually positive—i.e., worse quality. In some specifications that point estimate is statistically distinguishable from 0, and in others it is not. However, in none of the specifications do we find any evidence for an effect whereby taking longer to reach a decision, as caused by the Tournament, offsets the deleterious effects of distraction on opinion quality.

In fact, the effect of the Tournament that runs through the time to decision is positive, indicating the delay in decision caused by the Tournament causes itself additional negative citations. Thus, quite contrary to the expectation that judges can offset the effect of distraction on quality by simply delaying their work, these data suggest not. Importantly, though, the magnitude and significance of that particular effect are sensitive to model specification. We therefore can only conclude there is no evidence that taking longer to issue decisions offsets the effect of distraction on opinion quality, while there is suggestive evidence that the effect that runs through the time to decision may exacerbate the adverse effects of distraction on opinion quality.

These results suggest a remarkable effect of judicial distraction. Whereas, as noted above, one possible consequence of judicial distraction is that judges just work more slowly, these results suggest otherwise. They seem to work not just more slowly but also produce poorer quality opinions. As we suggested in our theoretical motivation, judges at the US courts of appeals cannot just slow down their work; cases come in steadily, and so judges are practically constrained in how much they can incorporate leisure into their choices by simply delaying their work. At some point, they must begin to trade off on other dimensions, and the evidence here suggests they trade off on the next most likely dimension—the
quality of their work.

6 Discussion and Conclusion

The findings we report here have implications for at least three substantive areas of inquiry in politics. First, they speak to organizational studies of how judicial institutions translate judges’ motivations into incentives for behavior. Second, the results have implications for normative debates about the desirability of judicial insulation from political insight. Third, our study has implications for political institutional design more broadly.

Judicial organization and professional incentives. Organizational approaches to studying the judiciary, especially those in the law and economics tradition, have long been concerned with the consequences of distraction—or, leisure—on judges’ work product. Judges, like all other people, derive utility from leisure and, relying on theories of the firm, these arguments have posited that judges will be more likely to shirk from their duties when confronted with an increased marginal benefit from leisure. Given their institutional protections (many judges, including federal judges, have long tenure, and most judges, even elected ones, are difficult to remove from office) and lack of financial incentives to work hard (judges in the United States are not rewarded based on the quantity or quality of their work), judges are in many ways analogous to non-profit firms, where we accept a certain amount of slack in exchange for the production of a difficult-to-evaluate work product.

However, just as in a firm producing widgets, judges have a choice of how to respond when confronted with a desire for leisure. They can slow down the production of widgets, make shoddier widgets, or smooth production over time to ensure even quality and sufficient average production rates. Our identification strategy allows us to estimate which of these avenues judges pursue when faced with a temporary increased marginal benefit to leisure. We show that judges slow down the rate of case resolution and also sacrifice quality. Moreover, the mediation analysis suggests that judges cannot completely mitigate the deleterious consequences of distraction on work quality through delaying case decisions, presumably because of the constant flow of cases and distaste for major backlogs in their work.
Normative concerns about judicial independence. Political institutions necessarily entail some degree of tradeoff between competing social goals. The scholarly and practical debate about judicial independence has, since its inception, contemplated multiple goals that are necessarily in tension with each other, including promoting judicial impartiality, ensuring majoritarian rule and democratic accountability, and facilitating judicial efficiency and quality. In favor of judicial insulation, scholars have argued that impartial judges are essential for overcoming commitment dilemmas that arise when the sovereign has enough power to rule effectively, and therefore are necessary for promoting economic growth and political freedom. In favor of democratic accountability, normative arguments about majority rule counsel against unfettered independence. Indeed, economic-minded scholars have cautioned that performance incentives are undermined when judges are freed from oversight and accountability.

The standard conclusion in the law and economics literature—that judges’ institutional protections facilitate shirking by judges who desire leisure—suggests a normative tension. Namely, the conclusion suggests that the normatively desirable benefits of judicial independence and protection from political meddling undermine the efficiency of the judicial branch. However, our findings highlight an alternative consideration about the relationship between judicial independence and judicial efficiency. Political insulation may promote lower quality judicial work. Of course, drawing attention to this possibility does not imply that the work of the federal judiciary is of low quality or that the problem cannot be mitigated. Our study does not suggest that judges are somehow doing a “bad job,” but only that they behave in the sort of ways that workers generally behave. What is more, by ensuring the prestige and independence of judicial posts, by recruiting individuals with reputations for strong work ethics (which we do), we can help alleviate the consequences of these very natural incentives to invest in leisure when leisure is particularly attractive.

Another implication of our study relates to work by Lax and Cameron (2007) and related models (e.g., Clark and Carrubba 2012, Clark and Staton 2015). In such models, judges have a preference for preparing high-quality decisions because the higher a decision’s quality, the stronger its control of future outcomes. Thus, control over policy creates an incentive to invest in high-quality opinions. If a judge does not benefit from judicial independence—i.e., if the judge thinks someone else controls decisions or that her opinions have no independent policy consequences—then she will not have an incentive to craft high-quality opinions. However, if a judge does anticipate being able to shape and control the law, then she will want to write a high-quality opinion in order to maximize her policy gains. Thus, ensuring that
judges have an appropriate role to play in the policy-making process, combined with the judiciary’s intrinsic incentives to do good work, might best address the accountability-independence tradeoff.

**Institutional possibilities.** Our analysis shows that, consistent with Posner’s (1993) conjecture, under at least one degree of insulation—here, the insulation experienced by judges at the US courts of appeals—judges behave as all workers appear to when their product is either unobservable or less observable than other features of their workflow. In many ways our findings are consistent with both conventional wisdom and with past studies of the effects of significant distraction on workplace performance. However, what sets this analysis apart is a particular set of features that characterize these judges. Unlike salaried workers in most firms or academics and students (e.g., Lozano 2011, Clotfelter 2012), the unavoidable flow of cases faced by federal court of appeals judges makes it difficult to adjust their work schedules. Judges in the United States largely lack explicit performance incentives. Moreover, the importance of their work—resolving federal cases—instills delays and quality deterioration with deep normative implications that are not as significant in the context of a slowdown in the production of, say, cars or title insurance contracts. Indeed, the normative consequences of delay and deteriorating quality in the US federal courts has been a matter of grave concern for some time (Kaufman 1990, Bannon 2014).\(^\text{13}\)

Of course, judges are not alone in the dynamics we study, and there are reasons to suspect some of the behaviors we identify will have parallels in other settings. For example, many career bureaucrats do not face obvious performance incentives other than the possibility of being able to affect public policy (e.g., Gailmard and Patty 2007). Similarly, legislators have incentives to credit claim and position take, but it is fairly easy for them to avoid blame for performance failures, as they can cite congressional dysfunction, collective action problems, and policy gridlock when campaigning Arnold (1990). Thus, our analysis has more general implications for institutional design, forcing us to confront inherent tensions among the various desiderata that underlie the construction of effective political institutions.

\(^{13}\)Importantly, our research design proposes one way of testing the causal impact of a marginal increase in a judge’s leisure utility. However, it does not allow us to evaluate changes in both judges’ utility and the institutional setting. As a consequence, we cannot identify what the consequences of various institutional reforms might be, and so we are not in a position to evaluate the impacts of alternative institutional arrangements. What we can do, however, is suggest a framework for evaluating policy proposals for judicial institutional design.
References


URL: https://hbr.org/2012/03/the-march-madness-really-destr/


**URL:** [http://CRAN.R-project.org/package=XML](http://CRAN.R-project.org/package=XML)


**URL:** [http://www.jstatsoft.org/v59/i05/](http://www.jstatsoft.org/v59/i05/)