

NBER WORKING PAPER SERIES

JUDGING THE JUDGES:
INDEXING OF COMPLEX INFORMATION REDUCES INJUSTICE

Andrew Caplin
Andrei Gomberg
Joyce Sadka

Working Paper 32587
<http://www.nber.org/papers/w32587>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2024

Caplin thanks the Alfred P. Sloan Foundation for Grant G-2023-19660 supporting research on Cognitive Economics at Work. Gomberg and Sadka thank Asociacion Mexicana de Cultura, International Development Research Center and Interamerican Development Bank for research support. We thank Lorenzo Aldeco, Roberto Gonzalez, Antonio Carbonell and numerous other research assistants who made this work possible. Finally, we thank Devesh Raval and Eric Spurlino of the Federal Trade Commission for valuable comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2024 by Andrew Caplin, Andrei Gomberg, and Joyce Sadka. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Judging the Judges: Indexing of Complex Information Reduces Injustice

Andrew Caplin, Andrei Gomberg, and Joyce Sadka

NBER Working Paper No. 32587

June 2024

JEL No. D8

ABSTRACT

We demonstrate that cognitive constraints produce injustice in its most paradigmatic form: incorrect verdicts in judicial trials. Responding to a request for help, we conducted a field study in Mexican labor courts in which time pressure on the judges leads to their making mistakes, many of which are appealed. All we did was to index pages to random case files to help judges quickly find relevant information. We made no other changes whatsoever. Our treatment greatly reduced successful appeals, fully 50% in complex cases, and resulted in judicial opinions that were shorter and more on point. These findings have already informed recent reforms in Mexico's labor court procedures and have broader applicability.

Andrew Caplin
Department of Economics
New York University
19 W. 4th Street, 6th Floor
New York, NY 10012
and NBER
andrew.caplin@nyu.edu

Joyce Sadka
Economics and Law Departments
Instituto Tecnológico Autónomo de México
Mexico City
Mexico
jsadka@itam.mx

Andrei Gomberg
Department of Economics and CIE,
ITAM
gomberga@gmail.com

Judging the Judges: Indexing Complex Information Reduces Injustice*

Andrew Caplin[†], Andrei Gomberg[‡], Joyce Sadka[§]

June 12, 2024

Abstract

We demonstrate that cognitive constraints produce injustice in its most paradigmatic form: incorrect verdicts in judicial trials. Responding to a request for help, we conducted a field study in Mexican labor courts in which time pressure on the judges leads to their making mistakes, many of which are appealed. All we did was to add **index pages** to random case files to help judges quickly find relevant information. **We made no other changes whatsoever.** Our treatment greatly reduced successful appeals, fully 50% in complex cases, and resulted in judicial opinions that were shorter and more on point. These findings have already informed recent reforms in Mexico’s labor court procedures and have broader applicability.

1 Introduction

Preventing judicial mistakes is a central task of a well-functioning justice system, as stressed by Posner (1995). The key role of cognitive constraints in contributing to errors is stressed by Tullock (1971), Stephenson (2010), Shavell (1995), and Cameron and Kornhauser (2006). A number of empirical papers provide suggestive evidence of the importance of such constraints in practice. An early laboratory study by Kelman et al. (1996) demonstrated susceptibility of legal environments to context effects. Guthrie et al. (2000), Wistrich et al. (2004) and Rachlinsky et al. (2007) recruited sitting judges to participate in testing for psychological biases documented for the general

*Caplin thanks the Alfred P. Sloan Foundation for Grant G-2023-19660 supporting research on Cognitive Economics at Work. Gomberg and Sadka thank the Asociacion Mexicana de Cultura, International Development Research Center and Interamerican Development Bank for research support. We thank Lorenzo Aldeco, Roberto Gonzalez, Antonio Carbonell and numerous other research assistants who made this work possible. Finally, we thank Devesh Raval and Eric Spurlino of the Federal Trade Commission for valuable comments.

[†]Department of Economics, New York University

[‡]Department of Economics and CIE, ITAM

[§]Department of Law, Department of Economics, and CIE, ITAM

population. Danziger *et al.* (2011) and Chen *et al.* (2016) provided observational evidences that such biases may manifest themselves in actual court decisions. Ash and McLeod (2015), using the variation in electoral incentives of judges, and Clark *et al.* (2018), using variation on the distractions affecting value of leisure, provide evidence that judges rationally respond to incentives in allocating their effort to the study of the cases they have to decide on. None of these papers directly measured the prevalence of actual errors in field settings; nor were they designed to test means of reducing this prevalence in practice.

We present the first field study that directly confirms the impact of cognitive constraints on achievement of justice and of simple cognitively-informed interventions to raise its quality. Our study does just this in the context of a Mexican labor court, which provides a uniquely suitable environment for a study of this type. The court itself deals with a large number of legally similar cases; importantly, at the time of our study, judges were not present at oral hearings, but rather had to decide the cases from voluminous compilations of legal documents (*case files*), which implied that we were able to obtain the same view of each case as the judges did at the time of producing their decisions. Those decisions were often appealed to higher courts, and such appeals were frequently granted. This error-correction system, however, was costly and time-consuming, often taking too long to make a real difference for the parties involved, while also creating substantial additional work for the original court's judges. This motivated the court to seek the research team's help in improving their procedures. The result of this was a simple intervention, designed in close consultation with the judges themselves. All we did was to append index pages to case files in order to help judges find relevant information. We made no other changes, and no one other than the judge in question and ourselves was aware of which cases had such indices added.

The results of the intervention are now out and they are clear-cut. The rate of successful appeals was reduced by over a third (from 16% to under 11% of all cases). This reduction was concentrated on legally more complex cases (in a sense clearly defined in the applicable law), where the treatment cut the number of successful appeals in half. Furthermore, judges' opinions in the same legally more complex cases were on average shorter, fitting with the idea that the index allowed the judges to better concentrate on the most important aspects of the case.

This research has already had a substantial impact on judicial reform in Mexico. In fact, one of us (Joyce Sadka) has been actively involved in redesigning procedural rules for the entire system of Mexican labor courts. Based, in substantial part, on the lessons of this intervention, the new procedures stress clarity in the arrangement of evidence. New electronic filing systems will also make follow-up studies of this nature much easier to implement: something that we are presently working on. A major next step is to explore the broader economic benefits of improved procedures for the resolution of labor disputes.¹

The implications of our study go far beyond the confines of a courtroom. What we demonstrate is that improving presentation can in practice sharply improve the

¹See the recent work of Chemin *et al.* (2023) on the impact of procedures for the resolution of contract disputes in Kenya.

quality of decision-making. The lessons of this study may have impact in other similar environments. Effective indexing may be crucial in many areas in which options are complex and multi-faceted, not only in judicial contexts, but also in contractual and regulatory settings, and in consumer choice more broadly.

The remainder of the paper is organized as follows. Institutional details and the actual information treatment are discussed in section 2. We present the basic findings in section 3, and build a simple search model to rationalize our treatment effects in section 4. In section 5 we conclude by outlining the impact this study is having on institutional design in Mexico, and its broader ramifications.

2 Environment: legal, institutional, experimental

Environments suitable for empirical field study of judicial mistakes would ideally need to satisfy a number of conditions. First, the empirical researcher would need to be able to access detailed data on all the evidence presented to the judge in a large number of cases, all evidence considered by that judge, as well as on the final judicial disposition of the case. Second, she would likely want to choose a court that has to deal with a many similar and legally simple cases, the correct disposition of which does not normally depend on a complicated legal analysis and can be determined based on a relatively limited number of observable variables. Third, the researcher should be able to observe sufficiently frequent overturns on appeal, the main error correction mechanism available in the legal system. We were fortunate to conduct our field study in a setting with all these characteristics.

2.1 Field Setting

Our setting possesses features that made it uniquely suitable for a study of this type. For one, this court did consider a large number of similar cases which were frequently overturned on appeal, allowing us to collect sufficient data on legal errors. Crucially, at the time of our study the court did not conduct oral trials; written evidence admitted into the record (which we shall call the *case file*) was the sole basis for the decision produced by court officials, who were not themselves present in any oral hearings or depositions. We obtained access to the case file itself, allowing us to access all the information the judge herself had access to. In addition, we had access to the full texts of court decisions, which allowed us to match the evidence admitted with the evidence mentioned by the judge and the reasoning provided in her decision. Even without our intervention, the richness of the data is unusual. Our direct intervention in presentation of the case before the judge was even aware of it, is, to the best of our knowledge, unprecedented. We use all these unique features of our setting in this study.

This field experiment is part of a research agenda aimed at identifying causes of poor delivery of justice services in labor courts in Mexico, and testing policies that can improve both quality and access to justice, ensuring more efficient and equitable transitions in labor markets. While access to data, including case files, is backed by a strong public transparency law in Mexico, obtaining access to live court cases and carrying

out field interventions was only possible through constructing a strong relationship with labor courts, creating trust, and demonstrating that policies we proposed and tested had positive impact on outcomes the courts cared about, such as the rate of appeals that were granted against the decisions of the court.

Since we have access to the case files and judgments themselves, we can dig further into how these effects work. As noted above, of particular value in interpreting how mistakes are changed by clear presentation in the Mexican case is that judges are charged not only with *reaching* a verdict but also with *justifying* it based on identified *legal facts* extracted from the case file on which it is based. What allows a verdict to be successfully appealed are identifiable mistakes in this justification.

2.2 Data

We obtained access to detailed data from a state-level labor court in the metropolitan area of Mexico City.² This particular court has sole jurisdiction over labor disputes in seven large industrial and residential suburban municipalities. Between 2011 and 2016 we partnered with the court to conduct a number of interventions aimed at improving the the court’s performance. This paper reports results of a particular intervention undertaken during 2013-2014. We collected data from 891 case files which we further randomly divided into a treatment and control group for the purpose of the experimental intervention.

Our access to data from the court was complete: for each case we scanned and then coded the entire associated file, including all the evidence presented. We were particularly favored by the standard procedure in the court, under which a group of auxiliary judges would receive written evidence and run oral hearings, converting the entire submission into a written transcript, which is the authoritative document based on which the judgment would be made. Another set of auxiliary judges, who were not present in proceedings, would receive these case files and produce draft decisions. It is these draft-producing auxiliary judges that were the subjects of our study (from now on we shall simply refer to them as *judges*).³ At any given point during our experiment there were 6 to 8 of these judges active at the court. They handled all the case files assigned from 12 different courtrooms. The judges were allocated cases by a quasi-random assignment method predating our study.

It is particularly convenient for us that the actual authors of the decisions were never present at any of the hearings. Consequently, these decisions could not be influenced by any of the usual courtroom subtleties, such as the tone of voice or shiftiness of a witness, unless these were noted in the written record. Anything that is not recorded in the case file is therefore non-existent from the point of view of the judicial decision we study. This means that our judge subjects did not observe anything that we did not also observe in the case file.

² *Junta Local de Conciliación del Valle Cuautitlán Texcoco.*

³ Technically, all auxiliary judges work under the supervision of a head judge, who could request changes to a judgment before it is finalized and announced. In practice, this rarely happens in routine cases.

2.3 Legal Conflicts

The conflicts adjudicated by the court in our study are both simple and homogeneous. 95% of the cases dealt with by the court are unfair dismissal lawsuits, and we use only these cases for the study. Mexican labor law is quite favorable to workers, with the employee only having the burden of proving that an employment relationship existed; assuming the relationship did exist, the employer has to prove that unfair dismissal did not occur. Employers' defense in these suits normally follows one of three common feasible strategies: denying that the plaintiff was ever an employee in the first place; claiming that the plaintiff left the job or resigned voluntarily; and offering reinstatement to the worker. Submissions are also fairly standard, with more than 90% of the evidence ever presented in our study belonging to just 12 common types. In a typical case, a worker might allege being illegally dismissed after a given number of years of service, and provide a written contract or salary receipts. The firm for which he alleges to have worked might present its payroll accounts for the relevant period to demonstrate that it never paid anyone with that name. Alternatively, it might acknowledge the employment relationship and produce a resignation letter signed by the worker. Authenticity of all these documents could be objected to by the parties, with corresponding expert opinions being attached to the case file as additional evidence.

2.4 Appeals

Once cases were disposed by the labor court they could be appealed to a higher court under a legal procedure known as *writ of amparo*, alleging violation of a party's constitutional due process rights. We did not observe whether such an appeal was in fact filed, but we did observe if it was granted, since such cases would return to the same court for a revised disposition. In order to track this, in addition to the case files themselves, we obtained a list of all the case files that entered the decision-writing area of the court during the period from 2011 to 2019. The list contains the case file number, date of entry, and whether the case entered for the first time or was returned to the court by an appeals court finding that the initial judgment had to be modified. The list covers the period from two years before to six years after the period of our study. This allows us to establish if any of the cases resulted in successful appeals.⁴

In most courts successful appeals are relatively infrequent, and studying quality of first instance decisions using overturns on appeals is empirically challenging because of the sparseness of data. We are fortunate in that over 30% of the original decisions in this court are, in fact, appealed, and in the baseline we find that about half of these appeals are successful in full or in part.⁵ This may indicate relatively easy access to

⁴Since our access agreement was limited to the trial court, we were not able to observe the actual fact of appeal in case the latter was denied or was still pending in 2019. Nor do we know which side appealed the case in each case: while, obviously, decisions in favor of a worker would more likely be appealed by the employer, the worker could still appeal based on other details of the disposition of the case, such as the amount of an award.

⁵Although, as we explained, we did not have access to data on unsuccessful appeals at the case file level, we were provided by the court with the aggregate measure of 30% unsuccessful + successful

appeals, but also low quality of the first instance decisions.

2.5 The Field Study

We were able to implement a randomized intervention directly varying the presentation of the cases to the judges.⁶ As we obtained the original case files before these were presented to judges, we coded all relevant data, and produced for each case file a 2-page summary, which included party names and addresses, legal claims involved, facts alleged, and the first 5 items of evidence of each of three major types (documents, testimonies and depositions) admitted into the case file, with page numbers to indicate where each evidence item could be found.⁷

A summary of this form was created for all case files that entered the decision-writing division for the first time during the experimental period. These were only provided to the judges for a randomly selected subset of these. The rest were given to the judges in their original form, thus constituting a control group.⁸ We then were able to trace the files in the appeals list to determine whether or not the original decisions in both groups were later overturned on appeal. The consequent coding of the written opinions was likewise done for all the cases in both the treatment and the control groups.

Our main effect is shown for the full experimental data base, for which we were able to obtain information about granted appeals, which brought the file back to the draft decisions office of the court, with instructions from the higher court to correct the decision. Additionally, in order to investigate the mechanism behind the treatment effect, we coded, during 2017-2021, very detailed data from both the case file and the decision. To code a case file, we searched for it in a general archive to which concluded cases are sent (eventually) by the 12 courtrooms in the experiment, and scanned the entire file. For a variety of reasons, including the locations of the 12 courtrooms and their different policies about how soon to send a concluded case file to the general archive, this data set is somewhat smaller than the experimental data set, with 571 decisions coded. To account for the fact that the data in this smaller set is not randomly selected, we also present the treatment results on the reduced data set.

The experiment required close coordination with the draft decisions office of the appeals, at the beginning of our study.

⁶This experimental intervention emerged from an earlier attempt to improve the quality of the court decisions. As part of that earlier intervention, draft opinions would be read by outside monitors and any discrepancies between the drafts and the case files would be pointed out to the judges for their possible correction before the decisions would be announced. These suggestions were frequently ignored. Exit interviews with the judges indicated they were too overworked to revise completed drafts and preferred to take their chance with the appeals system. They also mentioned that case files themselves were often disorganized and badly presented. This suggested to us that this would be an ideal setting to study how case presentation may affect judicial errors.

⁷In order to do this we coded each case file on a large number of characteristics, and then used substantive and procedural legal rules derived from the structure of the applicable law to come up with the summary.

⁸The judges were aware that both the treatment and the control group files were part of the experiment and that the summaries existed for all case files. The experiment was explicitly described to them as part of a study designed to evaluate possible procedural improvements in the court.

court. We placed one full time research assistant in the judges’ office, and this RA was informed each time a group of case files arrived at the office. Before the case files were distributed to the judges, the RA scanned the case files and distributed them to additional research assistants who worked remotely, coding the basic information from each file in approximately two hours. We designed a simple program to convert the data coded by the research assistants into the two-page index. The summary sheets went through an additional rounds of quality control with different and more experienced research assistants, before being provided to the judges with the case file if they were selected for treatment, or saved but not provided to the judges if they were in the control group.⁹

Table 1 shows the balance between treated and untreated cases on some relevant statistics, including the number of items of evidence submitted by plaintiffs and defendants, the wage claimed by the plaintiff, the percentage of case files in the top quartile of items of evidence of plaintiffs and defendants. There is no statistically significant difference across treatment and control groups in any of these characteristics.

3 Empirical findings

The main treatment effect is shown in the first column of Table 2. Our intervention reduced overturn on appeal from 16% to 10.6%. The second column in Table 2 also shows the treatment effect on the smaller data set of 571 observations for which we were able to code detailed information from the case file and judge’s decision. The treatment effect in this group of case files is larger and equally significant.

Balance Table

Covariate	Control	Treatment	P
% Casefiles with # pl. proofs in top quartile	0.271	0.262	0.810
% Casefiles with # def. proofs in top quartile	0.274	0.315	0.279
# Defendant proofs	5.654	6.545	0.197
# Plaintiff proofs	11.490	12.025	0.626
Wage claimed by plaintiff (MXN\$)	565.514	486.104	0.510

Given that our intervention did not affect the content of the case file, but only the ease with which information in the case file could be accessed by the judge, we interpret this result as demonstrating that simplifying the judge’s attentional task by itself improves the quality of her decision, at least as this quality can be measured by the appeals process.

⁹While this, of course, required a substantial and costly effort on our part, an RA would spend only around 10% of the time spent by each judge on the same case; furthermore, as an RA typically was an upper level undergraduate law student or a very recent graduate, the overall cost of our intervention was only a very small part of the cost of even the initial disposition of the case.

3.1 Interaction of treatment effect and the legal controversy

The treatment effect is not evenly spread over all case files. In particular, we find that our treatment effect is highly concentrated in cases in which the legal controversy implies greater complexity. We use the legal Spanish-Latin term, *Litis*, to denote the type of legal controversy. This is determined by the interaction of the plaintiff's main cause of action and the defendant's main response.

Under Mexican labor law, if the defendant denies the existence of the labor relationship, the burden of proof belongs to the plaintiff, to show that this relationship existed. In such cases, evidence from the defendant is likely to be sparse. On the other hand, when the defendant recognizes the labor relationship, the burden is on the defendant to show that unfair dismissal did not take place. In this second category of cases, there is likely more evidence on the defendant's side, and there is also a larger number of legal theories that could justify the court's determination.

The cases in which the labor relationship is denied by every defendant will be classified as *Litis 1* and the rest as *Litis 2*. The last column in table 2 demonstrates that our treatment effect is entirely concentrated on *Litis 2*, that is, on the more complex cases, where the treatment reduces the overturns by over 50%. In stark contrast, there is no treatment impact on the *Litis 1* cases. The general moral is that clarification is of particular value for complex cases.

Further evidence that our experiment affected primarily the more complicated cases is provided in Table 3. Here, we control for case complexity by measuring the number of items of evidence presented by the defendant.¹⁰ In both *Litis 1* and *Litis 2*, cases with more defendant evidence are more likely to be overturned, though the effect is more significant when the defendant carries the burden of proof. However, only in *Litis 2* cases did our treatment reduce overturns on appeal, with the effect being stronger the more items of evidence are presented by the defendant.¹¹

3.2 What drives the improvement in quality?

Being able to compare the original case files and the written opinions of the judge allows us to further explore the impact of our treatment on the content of the opinion. Thus, we can find out whether the evidence presented at trial is actually mentioned in the judicial opinion. Our original conjecture was that our treatment would help judges to avoid missing evidence, which, we suspected, might have provided grounds for appeals. We therefore constructed a pair of variables called *Defendant proof omissions* and *Plaintiff proof omissions*, which counted the number of documents presented by each party and admitted into the case file, which were not mentioned in the opinion. The impact of

¹⁰The number of proofs presented by the plaintiff did not matter in any of our regressions, and therefore is omitted here.

¹¹We do not have a direct measure of either the time judges used to produce their decisions, or of the impact our treatment may have had on that time. Judges typically received a bunch of both treated and untreated cases together and returned their decisions in those cases at the same time. We believe the judges took about 2-3 days per case on average to come up with a decision, but we do not have statistics for individual cases.

Table 1: Treatment Effect on probability of Overturn

Dependent Variable: Model:	(1)	Overturn	
		(2)	(3)
<i>Variables</i>			
Treatment Main	-0.054** (0.023)	-0.071** (0.029)	0.037 (0.056)
Litis 2			0.081* (0.045)
Treatment Main $\times$ Litis 2			-0.145** (0.065)
Constant	0.160*** (0.017)	0.171*** (0.022)	0.112*** (0.035)
Sample	All	Opinions coded	Opinions coded
<i>Fit statistics</i>			
Dependent variable mean	0.13468	0.13660	0.13660
Observations	891	571	571
<i>Heteroskedasticity-robust standard-errors in parentheses</i>			
<i>Signif. Codes: ***: 0.01, **: 0.05, *: 0.1</i>			

Table 2: Treatment effect on Overturn interacted with # Defendant proofs by litis

Dependent Variable: Model:	Overturn	
	(Litis 1)	(Litis 2)
<i>Variables</i>		
Treatment Main	0.021 (0.057)	-0.046 (0.037)
# Defendant proofs	0.006* (0.003)	0.006*** (0.002)
Treatment Main $\times$ # Defendant proofs	0.0005 (0.005)	-0.005** (0.002)
Constant	0.054* (0.031)	0.121*** (0.029)
<i>Fit statistics</i>		
Dependent variable mean	0.12925	0.13915
Observations	147	424
<i>Heteroskedasticity-robust standard-errors in parentheses</i>		
<i>Signif. Codes: ***: 0.01, **: 0.05, *: 0.1</i>		

Table 3: Treatment effect on defendant proof omissions litis 1 - And 2SLS - Omissions litis 2

Dependent Variables: IV stages Model:	Defendant Proof Omissions (Litis 1)	First (Litis 2)	Overturn Second (Litis 2)
<i>Variables</i>			
Treatment Main	-0.767 (0.847)	1.32* (0.684)	
Defendant Proof Omissions			-0.082 (0.050) ¹
Constant	1.74** (0.711)	2.35*** (0.453)	0.387** (0.156)
<i>Fit statistics</i>			
Dependent variable mean	1.3878	3.0142	0.13915
Observations	147	424	424
<i>Heteroskedasticity-robust standard-errors in parentheses</i>			
<i>Signif. Codes: ***: 0.01, **: 0.05, *: 0.1</i>			
<i>Note 1: P-value 0.100345</i>			

our treatment on the *Defendant Proof Omissions* is shown in Table 4.

Not only did the treatment not “help” the judges to avoid omissions: it made omissions more common in the same *Litis 2* cases in which we find the treatment effect. What we discovered is that, controlling for the number of items of evidence admitted in the case file, in *Litis 2* treated cases the judges in fact omitted more evidence, not less. Our results in fact suggest that it is the shorter opinions that may be driving the reduced overturn probabilities. The last two columns of Table 4 can be read together as a two-stage least-squares regression. While the first stage, in column 2, shows that our treatment is significant at the 10% level in increasing defendant proof omissions, the predicted defendant proof omissions barely miss significance at the 10% level as a predictor for the overturn on appeals. Our results may be indicative of the treatment effect working through shortening judges’ decisions, and in particular through them omitting the discussion of evidence that may not be relevant to their theory of the case.

4 A Simple Search-Based Rationalization

The first order impact of our intervention is that it makes it easier to read the court case in the appropriate order. A search model could represent this mechanism. We present a simple specification of how this might operate and produce the patterns observed in our data. There is no claim of generality: the goal of the model is purely illustrative.

Consider a world with two independent items of evidence $e_a, e_b \in \{0, 1\}$. Suppose

that for the employer to be guilty requires $e_a = e_b = 0$, otherwise they are innocent. Assume that the goal of the judge is to minimize the expected number of mistakes with utility losses of making a mistake normalized to 1. Suppose also that she must view at least one piece of evidence to be able to write up a verdict.

To keep the analysis simple, assume that *ex ante* guilt is the more likely verdict (for the purposes of this example, this reflects the general pro-worker bias of the Mexican labor law). This means that the optimal decision for the judge if they stop after seeing the first piece of evidence is to correctly announce innocence if that piece of evidence so indicates (if $e_{a/b} = 1$), and otherwise to announce guilt, while knowing that this might be wrong in the less likely case in which the unseen evidence establishes innocence. We define the probabilities of e_a, e_b respectively being 1 as β_a, β_b , with item a having the higher odds of establishing innocence:

$$0.5 > \beta_a > \beta_b > 0. \quad (1)$$

Note that these are precisely the probabilities with which seeing the additional piece of evidence will overturn the verdict and so define the value of that additional piece of information.

The only question then is whether or not to observe the second item of evidence after looking at the first. That obviously depends on a cost-benefit trade-off. Given the wide variation both in length of case files and in time pressure, we assume that there is a random cost of undertaking the second search. To keep things simple, we take this cost to be uniformly distributed over $[0, 1]$. We assume that this cost is known after reviewing the first item of evidence, and hence when deciding whether or not to observe the second piece of evidence. If realized costs are at the top end of this range, search is certainly not worthwhile. If they are at the bottom of this range, it certainly is worthwhile. The precise cutoff depends on the corresponding value β_a, β_b . To pin down costs and benefits as they depend on which item is seen first, we take the two possibilities in turn.

First, suppose that e_a is seen first. We now compute total expected losses, which depend both on how likely it is that there will be a mistaken verdict, and the costs of search that are expected to be incurred in the optimal strategy. To compute how likely is a mistake, note that this requires that the first piece of evidence be non-dispositive ($e_a = 0$), that the second piece of evidence is not observed due to high realized costs of learning, and that so observing would have established innocence. The corresponding probability is the product of three independent terms:

1. $1 - \beta_a$, the probability that $e_a = 0$;
2. $1 - \beta_b$, the probability that costs are above the critical level β_b , making further search worthwhile;
3. β_b , the probability that this piece of evidence would have established innocence.

To set against these verdict-based losses are expected costs of learning, which are incurred with probability β_b and when they are incurred average $\frac{\beta_b}{2}$. Hence the total

expected utility loss L_a to the judge of seeing e_a first can be computed by adding the probability of a mistake to the costs of search,

$$L_a = \beta_b(1 - \beta_a)(1 - \beta_b) + \frac{\beta_b^2}{2} \quad (2)$$

$$= \beta_b(1 - \beta_a - \frac{\beta_b}{2} + \beta_a\beta_b). \quad (3)$$

The logic if β_b is seen first is entirely analogous and gives rise therefore to corresponding expected losses with the roles of β_a and β_b reversed,

$$L_b = \beta_a(1 - \beta_b - \frac{\beta_a}{2} + \beta_a\beta_b) \quad (4)$$

In terms of expectation, the lack of an index produces expected losses that are the average of L_a, L_b . With an index, the minimum of these losses is chosen by picking which item to search first. It is clearly optimal to search e_a rather than e_b first. This reduces mistake-related losses since, with $\beta_a > \beta_b$,

$$\beta_b(1 - \beta_a)(1 - \beta_b) < \beta_a(1 - \beta_a)(1 - \beta_b). \quad (5)$$

It also reduces expected search costs from $\frac{\beta_a^2}{2}$ to $\frac{\beta_b^2}{2}$. Finally, note that it shortens the expected number of items listed, since only 1 piece of evidence is listed if the first is dispositive or it is not but search costs are too high. If e_a is examined first, the total probability is therefore $\beta_a + (1 - \beta_a)(1 - \beta_b)$ as opposed to $\beta_b + (1 - \beta_a)(1 - \beta_b)$. The higher probability of only one search directly implies that the opinion length is on average shorter. This simple model, therefore, while staying close to the nature of our experiment (providing a summary to the judges to guide search) generates the two major features that we observe in the data: decisions in treated cases are both shorter and less likely to be overturned.

5 Broader Implications

The experiment we report in this paper has had an important real-life impact on labor institutions. In 2017, sparked by internal debate and international pressure, Mexico passed a constitutional amendment, starting a major reform of its labor justice institutions. Legislative debate to reform the Federal Labor Law ensued, and due to the research team's strong ties to the labor authorities, we were invited to present the results of our field interventions, including this experiment. The convincing evidence that our intervention increased the quality of justice and lowered costs incurred by the appeals process had an impact. It was also noted that this improved quality of justice not only benefits those who appeal when they believe the first-instance judgment is erroneous, but has further indirect benefits. Many others may not appeal an erroneous judgment for reasons such as lack of funds, and for these parties, improved quality of the lower court's decisions implies access to better justice.

The bottom line is that we were able to propose parts of the new labor lawsuit procedures. The new federal labor law, passed in 2019, makes major changes to the lawsuit process. Plaintiffs now must provide, as part of their initial filing, a list of items of evidence they will present at trial, with a justification of each. In the defendants' answer they must also provide any rebuttal evidence and its justification. Trials are now oral and presided over directly by the same judge who is to write the decision. Finally, courts are obliged to have electronic case management systems, including an electronic mailbox in which parties can see all the dispositions in their case in real time and can receive formal notifications from the court. In the future, case summaries like the ones we implemented will be easily produced by the courts' case management systems, and it will be possible to measure the impact of the presentation of the case information to the judge on the final decision.

While our experiment was conducted in the specific setting of Mexican labor courts, we believe it addresses an issue of general importance to administration of justice. Judicial errors are not a feature of just this court, as evidence of appeals elsewhere amply attests. Judges everywhere face time pressures that constrain their attention, and courts everywhere have to design procedures that would minimize the wedge between the private incentives of judges and the public interest in achieving legally correct case disposition and avoiding costly appeals processes. Our study provides direct evidence that these problems are both empirically relevant and potentially remedied through simple cognitive interventions.

Cognitive constraints are increasingly important in other legal settings, particularly related to contractual obligations. In many countries the law makes it the duty of those who sign contracts to understand them. The application of this duty is especially controversial in the context of consumer contracts, which consumers generally do not read and are, in fact, very hard to understand, as detailed in McDonald and Cranor (2008) and Benoliel and Becher (2019). Developing further methods of altering presentations to raise choice quality could have high returns. Our results indicate that such simple measures that ease the search for key contract clauses may have much to add. The same applies to disclosure statements, which can best be viewed as efforts to make key contract clauses easier to understand. Much more research is needed to test out appropriate methods of presenting such clauses to simplify search and raise comprehension.

Our results have broader ramifications for the emerging field of cognitive economics (Caplin 2024). Handel and Schwartzstein (2018) review an extensive literature documenting consumer mistakes due to limited understanding of available information. Bhargava et. al. (2017), Ericson and Starc (2016) and Hastings and Tejeca-Ashton (2008) are among many studies showing that the manner in which information is presented significantly affects choice. Our field study is in this tradition. It involves a tightly constrained change in the manner of presentation. We were not allowed to change the content of case files in any respect, but rather added an index up front to guide the reader in how to assign their cognitive effort. This makes it crystal clear that the main channel of impact on decisions is attentional. The beneficial impact of our intervention not only for the judges in question, but for the functioning of the legal system, is equally clear.

Bibliography

1. Ash, Elliott, and W. Bentley MacLeod. “Intrinsic motivation in public service: Theory and evidence from state supreme courts.” *The Journal of Law and Economics* 58.4 (2015): 863-913.
2. Benoliel, Uri, and Shmuel I. Becher. “The duty to read the unreadable.” *BCL Rev.* 60 (2019): 2255.
3. Bhargava, Saurabh, George Loewenstein, and Justin Sydnor. “Choose to lose: Health plan choices from a menu with dominated option.” *The Quarterly Journal of Economics* 132.3 (2017): 1319-1372
4. Cameron, Charles M., and Lewis A. Kornhauser. “Decision rules in a judicial hierarchy.” *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift Für Die Gesamte Staatswissenschaft* (2005): 264-292.
5. Cameron, Charles M., and Lewis A. Kornhauser. “Appeals mechanisms, litigant selection, and the structure of judicial hierarchies.” *Institutional Games and the US Supreme Court* 173 (2006): 204.
6. Caplin, Andrew. “Data Engineering for Cognitive Economics.” *Journal of Economic Literature*, forthcoming (2024)
7. Chemin, Matthieu, et al. “Data Science for Justice: The Short-Term Effects of a Randomized Judicial Reform in Kenya.” (2023): 63-p.
8. Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue. “Decision making under the gambler’s fallacy: Evidence from asylum judges, loan officers, and baseball umpires.” *The Quarterly Journal of Economics* 131.3 (2016): 1181-1242.
9. Clark, Tom S., Benjamin G. Engst, and Jeffrey K. Staton. “Estimating the effect of leisure on judicial performance.” *The Journal of Legal Studies* 47.2 (2018): 349-390.
10. Danziger, Shai, Jonathan Levav, and Liora Avnaim-Pesso. “Extraneous factors in judicial decisions.” *Proceedings of the National Academy of Sciences* 108.17 (2011): 6889-6892.
11. Ericson, Keith M. Marzilli, and Amanda Starc. “How product standardization affects choice: Evidence from the Massachusetts Health Insurance Exchange.” *Journal of Health Economics* 50 (2016): 71-85.
12. Guthrie, Chris, Jeffrey J. Rachlinski, and Andrew J. Wistrich. “Inside the judicial mind.” *Cornell L. Rev.* 86 (2000): 777.

13. Handel, Benjamin, and Joshua Schwartzstein. "Frictions or mental gaps: what's behind the information we (don't) use and when do we care?." *Journal of Economic Perspectives* 32.1 (2018): 155-178.
14. Hastings, Justine S., and Lydia Tejada-Ashton. Financial literacy, information, and demand elasticity: Survey and experimental evidence from Mexico. No. w14538. National Bureau of Economic Research, 2008.
15. Kelman, Mark, Yuval Rottenstreich, and Amos Tversky. "Context-dependence in legal decision making." *The Journal of Legal Studies* 25.2 (1996): 287-318.
16. McDonald, Aleecia M., and Lorrie Faith Cranor. "The cost of reading privacy policies." *Isjlp* 4 (2008): 543.
17. Posner, Richard A. *Overcoming law*. Harvard University Press, 1995.
18. Rachlinski, Jeffrey J., Chris Guthrie, and Andrew J. Wistrich. "Heuristics and biases in bankruptcy judges." *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft* (2007): 167-186.
19. Shavell, Steven. "The appeals process as a means of error correction." *The Journal of Legal Studies* 24.2 (1995): 379-426.
20. Stephenson, Matthew C. "Information acquisition and institutional design." *Harv. L. Rev.* 124 (2010): 1422.
21. Tullock, Gordon. "Public decisions as public goods." *Journal of political economy* 79, no. 4 (1971): 913-918.
22. Wistrich, Andrew J., Chris Guthrie, and Jeffrey J. Rachlinski. "Can judges ignore inadmissible information? The difficulty of deliberately disregarding." *University of Pennsylvania Law Review* (2005): 1251-1345.