SW 15: Judicial Decision-Making: Integrating Empirical and Theoretical Perspectives

Book of Abstracts

Convenors:
Klaus Mathis (University of Lucerne)
Piotr Bystranowski (Jagiellonian University)
Bartosz Janik (Jagiellonian University)
Maciej Próchnicki (Jagiellonian University)

Contents

Workshop schedule

Monday 8.07.19
Room 3.B48

14h00 - 16h00
Jeffrey Rachlinski, Andrew Wistrich - Judging Autonomous Vehicles
***

Gustavo Cevolani, Vincenzo Crupi, Roberto Festa - The Whole Truth About Dina: Judicial Reasoning and the Conjunction Fallacy
***

Christoph Winter - Exploring the Challenges of Artificial Judicial Decision-Making
***

Brian Sheppard - The Automation of the Reasonably Prudent Person Test
***

16h30 - 18h30
Michal Ovádek - Lawyers' Aversion to Politics: Choosing a Legal Basis of EU Legislation under Experimental Conditions
***

Jernej Letnar Černič - Measuring social dimension of judicial ideology at the Constitutional Court of Slovenia
***

Cláudia Toledo - Judicial Activism – the Need for Parameters: Analysis of Legal Reasoning in Judicial Review
***

Wessel Wijtvliet, Arthur Dyevre - Measuring Judicial Ideology in Economic Cases: An Expert Crowdsourcing Design

Tuesday 9.07.19
Room 3.B57

9h30 - 10h30
Marion Vorms - The Story-model of Jurors' Decision-making in the Light of an Analysis of Evidential Reasoning — Conceptual and Empirical Questions
***

Iris van Domselaar - Virtue-jurisprudence and Empirical Accounts of Legal Decision-making: Mutually Reinforcing?
***

11h00 - 12h00
Moa Lidén - "Guilty, No Doubt": Detention Provoking Confirmation Bias in Judges' Guilt Assessments and Debiasing Techniques
***

Andrew Higgins, Inbar Levy - Judicial Policy, Public Perception, and the Science of Decision Making: A New Framework for the Law of Apprehended Bias
***

14h00 - 16h00
Teneille Brown, Markus Kneer - Inaction, Negligence and the Duty to Warn: Empirical Data
***

Noel Struchiner - An Experimental Guide to Vehicles in the Park: The Psychological Fault Lines of the Hart Fuller Debate
***
Ivar Hannikainen - Cross-Linguistic Evidence of Essentialist Beliefs about the Law
***
Kevin P. Tobia - Testing Ordinary Meaning
***
16h30 - 18h30
Camillia Kong, John Coggon, Michael Dunn, Penny Cooper - Judging Values and Participation in Mental Capacity Law
***
Piotr Bystranowski, Bartosz Janik, Maciej Próchnicki - Debiasing and the Limits of Discretion: The Case of Numerical Legal Decision-making
***
Mariusz J. Golecki, Kamil Joński - Fast Adjudication and Bounded Rationality in Polish Administrative Courts-Foreseeing Judicial Response to Caseload Growth
***
Christoph Engel, Keren Weinshall - Manna from Heaven for Judges: Judges' Reaction to a Quasi-Random Reduction in Caseload
***

SW 15: Judicial decision-making: integrating empirical and theoretical perspectives

Convenors:
Klaus Mathis (University of Lucerne)
Piotr Bystranowski (Jagiellonian University)
Bartosz Janik (Jagiellonian University)
Maciej Próchnicki (Jagiellonian University)

Over last decades, the empirical research on judicial decision-making has bloomed, borrowing methods from a wide range of social sciences, including, most notably, cognitive psychology and behavioral economics. The experimental studies have focused, first and foremost, on a multitude of biases, which may affect the outcome of legal proceedings. More recently, ideas from experimental philosophy have started to blend with legal research, trying to shed new light on the philosophical aspects of legal intuition. What is more, sociology of law, criminology, and political science have also delivered some empirical results that could radically change the conventional picture of how judges decide cases.

Despite all these developments in empirical studies, their influence on traditional, conceptual theory of judicial decision-making has been mostly superficial. The impact of empirical findings on concepts such as judicial discretion, formalism, judicial rationality, legal interpretation, or rules vs. standards debate is yet to be determined. The aim of this interdisciplinary workshop is to combine perspectives of psychology, sociology, economics, criminology, neuroscience and other behavioral sciences, with the legal-theoretical approach to judicial decision-making.

We invite all contributions presenting philosophically important results of experimental and other empirical studies on judicial decision-making, with data collected from professional (judges or other professional decision-makers) or lay subjects. Papers showing legal-philosophical implications of existing research or trying to integrate it with traditional theories of legal reasoning, or the doctrinal approach in various branches of law, are also much welcome. Possible research questions may include for example:

1. How does intuition shape the results of legal rulings?

2. What is the role of moral intuitions in legal decision-making?

3. What is the ecological and external validity of the existing experimental research in the field?

4. Which cognitive biases affect the actual process of judicial decision-making?

5. How do identified judicial biases interfere with important legal values, such as rule of law?

6. What are effective and theoretically acceptable ways of debiasing judges and other legal decision-makers?

7. Should identified examples of judicial bias lead to a revision of basic principles of procedural or substantive law?

8. In what way differences in legal culture and institutional settings and the organization of judiciary between legal systems affect the process of judicial decision-making?

9. Are the results of neuroscientific research on decision-making and reasoning already relevant to the field?

Teneille Brown (University of Utah)
Markus Kneer* (University of Zurich)

Inaction, Negligence and the Duty to Warn: Empirical Data

Introduction

In the U.S., the common law of negligence expects people to conduct themselves in reasonably careful ways to avoid liability. However, the law typically did not hold defendants liable for negligence due to inaction. Even where a defendant realizes or should realize that action on his part is necessary for another's protection, this does not of itself impose upon him a legal duty to take such action. The U.S. is unusual in this regard. Most European countries impose some duty to rescue by statute. Only three states in the U.S. have imposed, by statute, an affirmative duty to rescue strangers.

Over the last forty years, this sharp distinction between careless action and inactionalso (known as misfeasance and nonfeasance) has been recognized as less legally significant. Where certain policy factors are present, or the plaintiff stands in some "special relationship" with the defendant—typically a fiduciary relationship where the plaintiff puts her safety in the hands of the defendant—most state courts recognize that a defendant may now be considered negligent for failing to act. A landmark California case provided a list of the factors judges are expected to weigh, when determining whether to impose a legal obligation to warn or rescue a stranger. Specifically, judges must weigh the a) foreseeability of serious harm, b) the ease of objectively measuring a tangible harm c) how likely it was that defendant's inaction caused plaintiff's injury, d) the ability to prevent this kind of injury by imposing a legal duty, e) the moral blameworthiness of the defendant's inaction, and f) the ability for defendants to afford this new duty or spread the cost of it through insurance. Judges in every state have followed this precedent when deciding whether to impose new, affirmative duties to warn. However, as is typical with fact-specific legal standards such as these, courts have been given no guidance as to how they ought to evaluate these criteria. Courts are free to put different weights on the different factors, or ignore some factors all-together. This has led to high unpredictability in decisions, and frustration among practitioners.

We conducted the first systematic series of experiments which explores different factors that might influence folk decisions in duty-to-warn cases. The purpose of this endeavor is to make juror dispositions in such cases explicit and more predictable, and to provide recommendations to courts that help establish reasonable and unbiased judgment guidelines.

Experiments

We will present novel data from an extensive series of experiments conducted with over 1400 subjects. The experiments drew on a variety of case-based scenarios covering several legal domains. In a full factorial design, we systematically varied

1. the extent to which a harmful consequence was foreseeable,

2. the difficulty to prevent such an event, and

3. the severity of harm that arose due to a failure to warn.

The experiments tested a variety of dependent variables, which were organized into three distinct types:

---

*Corresponding author: markus.kneer@gmail.com

1. mens rea (knowledge, recklessness, negligence),

2. judgments regarding a potential duty to warn,

3. and moral judgments (blame, punishment).

   We will present the data of the experiments at the workshop, discuss it with reference to previous decisions in similar cases, and try to tackle the complex question what kinds of factors and interactions do and do not constitute a bias from a legal point of view.

Piotr Bystranowski* (Jagiellonian University)
Bartosz Janik (Jagiellonian University)
Maciej Próchnicki (Jagiellonian University)

Debiasing and the limits of discretion: the case of numerical legal decision-making

The existing literature provides much evidence that legal decision-makers might issue their verdicts under the influence of fallible heuristics and biases, which often rely on non-legal factors. However, the proposed remedies – including the so-called debiasing techniques, which aim to reduce particular biases – are still highly controversial, not only from the viewpoint of their effectiveness, but also in their legal applicability.

In the following paper, we focus on numerical judgments (e.g., the amount of damages, the length of a prison term), which – as a result of reliance on heuristics and biases – are oftentimes issued in a highly inconsistent manner. It is widely acknowledged in the literature that deficiencies of numerical legal decision-making might result in serious unfairness (deciding similar cases differently) and substantial social costs (e.g., increasing the costs of legal uncertainty). This makes it worthwhile to consider possible ways of improving the quality of decision processes in such contexts. Is there a way to debias judicial numerical reasoning?

One of such psychological effects, potentially biasing the results of legal proceedings, to be studied in more detail in this paper, is the anchoring effect, i.e. the tendency to over-rely on numerical values present in the decision context, even if they are arbitrary and unrelated to the target problem. As research in general cognitive psychology indicates, the anchoring effect is notorious for its robustness and resistance to countermeasures. We will survey techniques suggested in the literature to diminish the negative outcome of cognitive fallacies, which may be useful to decrease judicial propensity to anchoring and other biases in numeric decision-making. These techniques range from making judges aware of the existence of such biases, to changes in institutional settings, to explicitly limiting the options the judge can choose from. Because of the last issue, the answer to the question about desirable modes of reasoning requires agreeing on the right trade-off between judicial discretion and accuracy.

In general psychological literature there is widespread consensus that debiasing requires intervention, i.e. that individuals cannot debias themselves. One of the factors contributing to this outcome is that people might not realise that their decision making process is flawed due to the poor feedback. In the case of judges this might be corrected via institutional feedback, but the usage of the decision outcomes to evaluate own decisions might be suboptimal in case of a specific decision environment.

Other kinds of remedies include other institutional solutions, such as modifying the judicial training or workstyle. Judges could be taught about the influence of their intuition and their susceptibility to cognitive effects such as anchoring. Another one is to limit the judicial workload: if judges have more time to decide, they may be more likely to rely on deliberative reasoning instead of being affected by irrelevant anchors. A somewhat opposite solution would be to increase the judicial workload by requiring judges to personally justify their rulings more frequently, thus increasing their accountability and forcing them to prove reasonableness of numeric values they came up with. Increasing the peer assessment of a judge's rulings by their more experienced colleagues, or a procedural division

---

*Corresponding author: piotr.bystranowski@emle.eu

of labour (one judge deciding the merits of a case, the other deciding about the numbers such as the amount of damages) may also be useful. Another procedural solution would be to encourage the parties to come up with a larger number of relevant anchors, or, to the contrary, organize legal proceedings in a way that minimizes the number of numeric values that a judge encounters.

However, if such debiasing proved impossible from the psychological or institutional reasons, maybe more invasive methods, like introducing specific mathematical formulae or tables, or enacting more or less binding official guidelines could be an answer to the problem? Each of these solutions pose both opportunities and problems, which should be assessed from the theoretical and practical viewpoints, yet the main issue here is to balance judicial discretion with debiasing policy.

Gustavo Cevolani* (IMT School for Advanced Studies Lucca)
Vincenzo Crupi (University of Turin)
Roberto Festa (University of Trieste)

The whole truth about Dina: Judicial reasoning and the conjunction fallacy

In the last decades, cognitive psychologists and behavioral economists have greatly deepened our understanding of how both experts and laymen reason, choose, and make decisions in a variety of contexts. Researchers have highlighted a number of "heuristics", or cognitive strategies routinely employed by human reasoners, as well as a number of "biases", or cognitive pitfalls, that can lead them astray.

In recent years, judicial decision-making has received renewed attention, leading to increased empirical investigation of judges' reasoning strategies and cognitive errors (cf. Rachlinski and Wistrich 2017). Results available up to now suggest that judges do not differ too much from other professionals (like physicians or managers or policy makers), especially as far as the biases they can fall prey of are concerned. This invites for an assessment of judicial reasoning with respect to both traditional accounts of legal reasoning and more abstract theories of human rationality.

The "conjunction fallacy" (also known as the "conjunction effect") is a widespread phenomenon showing that naïve reasoning systematically violates the rules of the probability calculus (Tversky and Kahneman 1983). To illustrate, in a study (Guthrie, Rachlinski and Wistrich 2009) 103 administrative law judges evaluated an alleged case of discrimination against a Muslim employee named Dina. After reading a short description of the case, most judges ranked the probability that the employer both "actively recruited a diverse workforce" (A) and also "unlawfully discriminated against Dina based on her Islamic religious beliefs" (B) as higher than the probability of either of the two events in isolation. This pattern of judgments is puzzling in that it conflicts with a basic and uncontroversial principle of probability theory, known as the "conjunction rule", prescribing that a conjunction A&B can not be more probable than any of its conjuncts A and B.

Since Tversky and Kahneman's seminal study, the conjunction fallacy has been a central issue in the analysis of human reasoning and decision making under uncertainty. Interestingly, it has attracted the attention not only of psychologists, but also of philosophers working in so-called formal epistemology (e.g., Levi 1985; Bovens and Hartmann 2003; Hintikka 2004; Crupi, Fitelson, Tentori 2008; Cevolani and Crupi 2015). In any case, the attempt of providing a satisfactory account of the phenomenon has proved rather challenging.

In this paper, we present two accounts of the conjunction fallacy, respectively based on the notions of verisimilitude (or truthlikeness) and confirmation (or inductive support) as explored within the philosophy of science. In both accounts, the participants' preference for A&B over A and/or B is normatively justified under specific circumstances. The common idea is that properly taking into account the role of information and truth as relevant factors in cognitive decision-making may explain why participants systematically deviate from sound probabilistic reasoning in cases like Dina's. We explore the conceptual relationships between these two accounts and discuss how they deal with the available behavioral data on the conjunction fallacy in judicial reasoning (Guthrie, Rachlinski and Wistrich 2009; Wojciechowski and Pothos 2018). Finally, we study the implications of the two accounts for competing models of legal reasoning and decision-making: the "story

---

*Corresponding author: gustavo.cevolani@imtlucca.it

model" (Pennington and Hastie 1993; Teichman and Zamir 2014), coherence-based models (Simon 1998), and models based on Bayesian probability.

References

Bovens, L., and S. Hartmann (2003). Bayesian epistemology. Oxford: Oxford University Press.

Cevolani, G., and V. Crupi. 2015. "Subtleties of Naive Reasoning. Probability, Confirmation, and Verisimilitude in the Linda Paradox." In Epistemology of Ordinary Knowledge, edited by M. Bianca and P. Piccari. Cambridge Scholars.

Crupi, V., B. Fitelson, and K. Tentori (2008). "Probability, confirmation and the conjunction fallacy". Thinking and Reasoning, 14, pp. 182–199.

Guthrie, C., J. J. Rachlinski, and A. J. Wistrich. 2009. "The Hidden 'Judiciary': An Empirical Examination of Executive Branch Justice." Duke Law J. 58(7):1477–530.

Hintikka, J. (2004). "A fallacious fallacy?". Synthese, 140, pp. 25–35.

Levi, I. (1985). "Illusions about uncertainty." British Journal for the Philosophy of Science, 36, pp. 331–340.

Pennington, N., and R. Hastie. 1993. "The Story Model for Juror Decision Making." In Inside the Juror, edited by Reid Hastie, 192–222. Cambridge: Cambridge University Press.

Rachlinski, J. J., and A. J. Wistrich. 2017. "Judging the Judiciary by the Numbers: Empirical Research on Judges." Annual Review of Law and Social Science 13 (1): 203–29.

Simon, D. 1998. "A Psychological Model of Judicial Decision Making." Rutgers Law Journal 30: 1, pp. 1-142.

Teichman, D., and E. Zamir. 2014. "Judicial Decision-Making: A Behavioral Perspective." The Oxford Handbook of Behavioral Economics and the Law. Oxford: Oxford University Press, pp. 664-702.

Jernej Letnar Černič* (Nova univerza)

Measuring social dimension of judicial ideology at the Constitutional Court of Slovenia

The Constitutional Court of Slovenia has in the past decaded delivered several seminal decisions relating to the values of pluralism, tolerance and broadmindness. However, not much if anything has been published as to the reasons and judicial ideology that triggered such judgements. What triggers constitutional judges to protect in some cases the rights of minority and in other the interests of majority? This paper stems from the research project on Ideology in the Courts: the Influence of Judges' Worldview on their Decisions (Slovenian Research Agency, 2017-2020), where project group has been empirically measuring the presence of three-fold judicial ideology at the Constitutional Court of Slovenia. The research group has in the course of the project developed three-fold methodological and theoretical model aimed at measuring judicial ideology, which has been measured within its economic, social and authoritorian dimensions. In doing so, the research group has empirically measured decisions and seperate opinions from selected periods of the Constitutional Court based on the results of the model. The objective of this paper is to present result of measuring ideological profiles of the Court and its individual judges relating to the social dimension of judicial ideology. In this way, three-fold judicial ideology model has assisted explains how judges to decide in the horizontal relationships and disputes between minority and majority. The empirical results will help the research group to develop guidelines for improving judicial-making at the Constitutional Court based on the values of pluralism, tolerance and broadmindedness in order to ensure impartial and independent functioning of the judicial system and individual judges.

---

*Corresponding author: jernej.letnar@gmail.com

Iris van Domselaar* (University of Amsterdam)

Virtue-jurisprudence and empirical accounts of legal decision-making: mutually reinforcing?

In the last decades, a wide range of literature has focused on the empirical dimension of morality, thereby also putting into question basic assumptions of rationalistic strands within normative moral philosophy. In this literature, often simply by passing, virtue theory not seldom received the compliment that it is the psychologically most sound approach to morality. Because of their emphasis on the agent as Haidt and Joseph (2004) have put it " [..] such theories fit more neatly with what we know about moral development, judgements and behavior than do theories that focus on moral reasoning or on the acceptance of moral high level principles such as justice."

Against this background, this paper examines the relation between virtue-jurisprudence, a philosophical approach to legal decision-making which draws upon the insights of virtue theory, and empirical research on judicial decision-making. More specifically, it examines to what extent indeed virtue-jurisprudence proves more apt to incorporate the (main) findings of these empirical approaches, compared to more traditional, rationalistic accounts of legal decision-making.

In addition, by confronting virtue-jurisprudence and its philosophical background with empirical accounts of legal decision-making this paper also identifies certain grounds for criticism on (the methodological assumptions underlying) empirical research on legal decision-making.

---

*Corresponding author: I.vanDomselaar@uva.nl

Christoph Engel* (Max Planck Institute for Research on Collective Goods)
Keren Weinshall (Hebrew University of Jerusalem)

Manna from Heaven for Judges: Judges' Reaction to a Quasi-Random Reduction in Caseload

Judiciaries worldwide claim to be facing increasing caseloads. The perceived change is interpreted in very different ways. According to the judicial approach, the perceived crisis is mostly attributed to socio-legal developments that generate an increasing volume and complexity of cases, outpacing a smaller increase in the number of judges. This line of thought advocates raising the supply-side by appointing more judges, registrars and judicial staff)Richman & Reynolds 2012, Levy 2013, Adler 2014, Stras & Pettigrew 2010). On the other hand, the managerial approach claims that clogged courts are mainly the result of mismanagement and inefficacy, outdated legal procedures and judicial passivity (Mitsopoulos et al. 2010, Dalton et al. 2014, Castro & Guccio 2015, Moffett et al. 2016). In this interpretation, increasing courts' budgets or appointing new judges is not expected to be instrumental (Webber 2006, Agrast et al. 2011, Heaton & Helland 2011, Beenstock et al. 2004).

The managerial approach corresponds with theories of judicial behavior that view judges as individuals motivated in part by leisure preferences (Posner 1993). Thus, when more judges are appointed to a court with a given caseload, the judges who maximize personal utility are predicted to react by working less and taking more time for leisure. On the other hand, the judicial approach asserts that judges are predominantly motivated to do a good job (Engel & Zhurakhvoska 2017). If so, when judges have lighter caseloads they are expected to use the extra time for better resolving their remaining cases. This could result in an increase in court productivity, quantity and quality of judicial output. In this paper, we study this question empirically. Is increasing judicial staff effective in improving the quality of judicial services? To address this question, we exploit a natural, near-randomized experiment in the Israeli judiciary. In 2012, senior registrars were appointed for a pilot program to conduct afternoon hearings in two of the six Magistrate's Court districts. Considerations in assigning the registrars to the specific two districts pertained to physical court conditions, entirely exogenous to workloads or work procedures. These appointments constituted an increase of around 19% in judicial staff devoted to civil litigation in the two treated districts and a decrease of an estimated 15% of the weighted caseload for each civil judge active in the treated courts[1]. We test indicators for court productivity, judicial resources invested and decision-making outcomes before and after the exogenous shock on caseload, compared to those in untreated districts – before and after 2012 (diff in diff design).

The analysis is based on two sources of data: official court performance data for civil litigation in each district (annually published in 2007-2017) and an original random sample of 2085 small claims resolved in all six districts in the years 2011 (pre-treatment) and 2014 (post-treatment).

Initial results find more support for the judicial approach, although some claims of the managerial approach are also supported. After a short adjustment period in 2012 and 2013, the treated courts' performance does not improve and its efficiency – in terms

---

*Corresponding author: engel@coll.mpg.de
[1]Other magistrate's court districts did not experience changes in the size of their judicial staff or caseload during the period studied.

of number of closed and pending cases per judge, clearance rates and average length of proceedings – even decreases[2]. However, it appears that judges in the treated courts do not enjoy more leisure, but rather invest the additional time in better resolving their assigned cases. For example, after the reduction in caseload, judges in the treated courts use more laborious means of evidence (schedule more hearings, hear more witnesses); they are less likely to dismiss borderline cases on procedural grounds or write summary judgments; and are more likely to decide cases on the merits and write more elaborate opinions. Furthermore, caseloads seem to affect substantial outcomes: Treated judges tend to award petitioners with higher recoveries and trial costs. The significant difference in outcomes before and after treatment does not occur in the control group of judges from untreated courts, who did not benefit from a decrease in their workloads.

We discuss the implications of our findings on judicial management and judicial decision-making theories.

References

Adler, A.L. (2014). Extended vacancies, crushing caseloads, and emergency panels in the federal courts of appeals. Journal of Appellate Practice & Process, 15, 163-199.

Agrast, M.D., Botero, J.C. & Ponce, A. (2011). The Rule of Law Index 2011. Washington, DC: The World Justice Project.

Arlen, J. & Talley, E. (2008). Experimental law and economics. In J. Arlen & E. Talley (Eds.), Economic Approaches to Law Series: Experimental law and economics (pp. xv-lxi). Cheltenham, UK: Edward Elgar Publishing.

Beenstock, M. & Haitovsky, Y. (2004). Does the appointment of judges increase the output of the judiciary? International Review of Law and Economics, 24(3), 351-369.

Castro, M.F. & Guccio, C. (2015). Bottlenecks or inefficiency? An assessment of first instance Italian courts' performance. Review of Law & Economics, 11(2), 317-354.

Dalton, T. & Singer, J.M. (2014). Bigger isn't always better: An analysis of court efficiency using hierarchical linear modeling. Pace Law Review, 34(3), 1169-1189.

Engel, C., & Zhurakhovska, L. (2017). You are in charge: Experimentally testing the motivating power of holding a judicial office. The Journal of Legal Studies, 46(1), 1-50.

Heaton, P. & Helland, E. (2011). Judicial expenditures and litigation access: Evidence from auto injuries. Journal of Legal Studies, 40(2), 295-332. Huang, B.I. (2011). Lightened scrutiny. Harvard Law Review, 124(5), 1109-1152.

Iverson, B. (2017) Get in Line: Chapter 11 Restructuring in Crowded Bankruptcy Courts. Management Science. Published online in Articles in Advance 21 Aug 2017.

Levy, M.K. (2013). Judging the flood of litigation. University of Chicago Law Review, 80, 1007-1077.

Mitsopoulos, M. & Pelagidis, T. (2010). Greek appeals courts' quality analysis and performance. European Journal of Law and Economics, 30(1), 17-39. Posner, R. A. (1993). What Do Judges Maximize? (The Same Thing Everybody Else Does), 3 SUP. CT. Econ. Rev, 1, 3-4.

Richman, W.M. & Reynolds, W.L. (2012). Injustice on appeal: The United States courts of appeals in crisis. New York, NY: Oxford University Press.

Stras, D.R. & Pettigrew, S.M. (2010). Rising caseload in the fourth circuit: A statistical and institutional analysis. South Carolina Law Review, 61, 421-444.

Webber, D. (2006). Good budgeting, better justice: Modern budget practices for the ju-

---

[2]Note that before 2012, there is a common trend in performance indicators of all six courts, which remains constant after 2012 in the untreated courts

dicial sector (L&D Working Paper No. 3). Washington, DC: The World Bank.

Mariusz J. Golecki* (University of Lodz)
Kamil Joński (University of Lodz)

Fast Adjudication and Bounded Rationality in Polish Administrative Courts-Foreseeing Judicial Response to Caseload Growth

Judicial control of administration plays an important role in modern democratic states. It enables citizens with the possibility of challenging illegal administrative decisions, thus enhancing the rule of law and the control of government. Specialized administrative courts that handle disputes involving public institutions constitute important part of many continental judicial systems. However, such courts seem to be particularly prone to sudden increases in caseload, generated by changes in public policies or regulatory frameworks. That in turn could lower the level of fair trial standard, by (i) increasing backlog and (ii) decreasing quality due to human errors resulting from the overload.

In the paper we estimate production function (OLS, SFA) on the panel of 16 Polish Voivodship Administrative Courts (VACs) during 2010 – 2015 period to verify its ability to explain backlog accumulation. Specifically, it we examine whether production functions estimated on 2010 – 2012 data would be able to predict backlog accumulation during 2013 – 2015 period – when volume of cases grown by 16 percent – given correct caseload forecasts. To our best knowledge, it is first attempt to systematically test court's production function ability to explain backlog accumulation in such out of sample setting. This work attempts at integration of two strands of the literature: econometric modeling of court performance and studies of judicial decision making. The former, utilizes concept of production function in order to model court performance and explain sources of backlog. Given supply-demand remedies to the court congestion problem – as classified in seminal OECD paper (Palumbo, 2013) – our analysis focus on administrative courts productivity (i.e. ability to adjudicate incoming cases in a timely manner as required by article 6 of European Convention on Human Rights).

It is worth mentioning at least two approaches to modeling court performance. The first, popular among court administration practitioners (Lienhard, Kettiger, 2011) is based on the notion of weighted caseload. It attempts to recalculate raw volume of filed cases (caseload) to judicial workload, using case weights – empirically determined typical amount of time, required to resolve given type of case. Thus, workload can be interpreted as amount of judicial working time, required to handle incoming cases. That assumption have been criticized by Beenstock and Haitovsky (2004) who noted that despite growing caseloads, court systems managed to handle them without proportionate staffing increases. That observation inspired them to build 'rational judge' model, assuming that number of resolved cases is subject of choice between 'exerting more effort and thereby improving performance, or taking it easier, thereby risking the wrath of the court president'. Results convinced them, that 'planners of the judiciary might cynically conclude that they should let growing caseload pressure overwork judges even further, thereby increasing their productivity. Both, empirical strategy and results obtained by Beenstock and Haitovsky inspired strand of Law and Economics literature applying production functions to court performance (see summary by Voigt, 2016). We try to demonstrate that assumptions of weighted caseload methods can be incorporated to production function approach. At the end of the day, Cobb-Douglas production function, applied by most researchers implies that judge confronted with growing caseload would indefinitely increase number of reso-

---

*Corresponding author: mjgolecki76@gmail.com

lutions. Weighted caseload methods suggests, that at some point this increase would no longer be possible, so production function became flat (Joński, Mankowski, 2014). The results of our data analysis could be summarized in the following way. To handle issues of model uncertainty highlighted above, we estimated two models: (i) model I, involving Cobb Douglas production function popular in the Law & Economics literature and (ii) model 2, incorporating weighted caseload methods assumption. Specifically, model 1 is given by the equation (1), and model 2 by the following equation:

$$ln(res.cases/judges) = \begin{cases} \beta \, ln(caseload/judges) + \varepsilon, & caseload/judges < \Phi \\ \Delta + \varepsilon, & caseload/judges \geq \Phi \end{cases}$$

Where Δ represents maximal number of cases that judge can resolve during the year, and Φ is respective caseload. Together, these parameters define 'hockey sick' shape of such production function. Coefficient β is the slope of production function below threshold Φ (see fig. below).

Fig. 4 Estimated production functions and data points, 2010 – 2015 sample
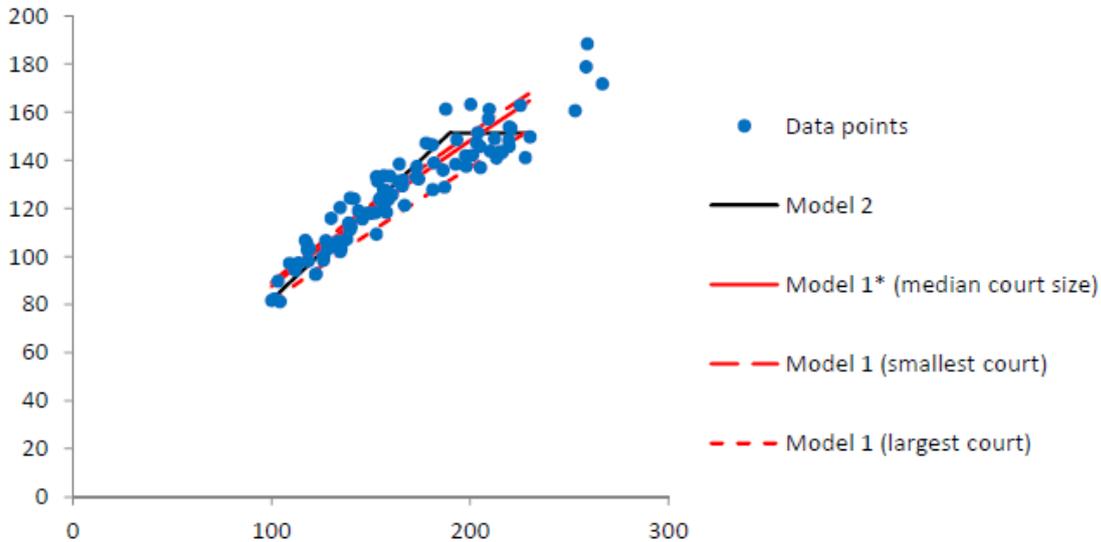
*- Since Cobb Douglas model allows for scale effects (estimation results suggests negative scale effects, in other words the larger court the lower number of resolved cases per judge), we plotted production function for median, as well as smallest and largest court.

To facilitate out of sample exercise, all models were estimated in three time windows: (i) 2010 – 2012 – before caseload growth, (ii) 2013 – 2015 – period of caseload growth and (iii) 2010 – 2015. To obtain first approximation of results, we estimated OLS regressions on pooled data. As a next step, we estimated model 1 using panel Stochastic Frontier Analysis (SFA). SFA is a log-linear production function, whose error term (ϵ) is modeled as a mixture of two components: (i) strictly nonnegative (halfnormal) one, interpreted as inefficiency and (ii) symmetrically distributed one, interpreted as random noise (detailed description in Kumbhakar, Lovell, 2000).

The second part of the paper is focused on the impact of increased caseload upon the quality of adjudication and the judicial strategy. Both judgments and decisions demonstrate systematic departures from the rational choice model (Golecki, 2011, 2015). This

finding refers both to legal and non legal contexts. It has been observed that judges are prone to both types of departures from the standard rational choice model (Vermule, 2006). This phenomenon is partly explained by the way in which actors apply the so called rules of thumb (Kahneman, Tversky, 1979). The research questions concentrate on the presence and frequency of particular judicial heuristics and biases and their impact upon the output of litigation. The underlying hypothesis is based on the assumption that heuristics and biases enable fast thinking and more effective, albeit unreliable processing of information and decisionmaking under time pressure. The paper thus explores the relationship between the caseload and heuristics/biases and the success rate of the citizen who brings the suite against illegal decisions. According to the main hypothesis if the number of cases exceeds the performance frontier (approximately 220 cases per judge), the success ratio drops dramatically. Moreover the quality of judgment deteriorates as well, which poses a serious threat to rule of law and the quality of judicial control.

References

Beenstock M., Haitovsky Y., (2004), Does the appointment of judges increase the output of the judiciary?, International Review of Law and Economics, Elsevier, vol. 24(3), pages 351-369, September.

Golecki M.J., 2011. Miedzy pewnoscia a efektywnoacia. (Between Certainty and Efficiency) Warszawa, Wolters- Kluwer.

Golecki M.J., 2015. New York Times v. Sullivan in European Context. The Optimal Liability for Libel from the Perspective of Behavioural Law and Economics [in:] K. Mathis (ed.), European Perspectives on Behavioural Law and Economics, Cham, Heidelberg, New York, Dordrecht, London, Springer, pp. 243-267.

Guthrie, C., J.J. Rachlinski, A.J. Wistrich,. 2000. Inside the judicial mind. Cornell Law Review 86: 777-892.

Jolls Ch., C.R Sunstein, R. Thaler, 1998, A Behavioral Approach to Law and Economics, 50 Stanford Law Review, vol. 50.

Jonski, K. and Mankowski, D., 2014. Is Sky the Limit? Revisiting 'Exogenous Productivity of Judges' Argument. International Journal for Court Administration, 6(2), pp.53–72.

Kahneman D., P. Slovic, A. Tversky, 1982. Judgment Under Uncertainty: Heuristics and Biases, New York.

Kahneman D. 2011, Thinking, Fast and Slow. London: Macmillan.

Lienhard A., Kettiger D., (2011), Research on the caseload management of courts: methodological questions, Utrecht Law Review, Volume 7, Issue 1, January 2011

A. de Oliveira Gomes, T. de Aquino Guimaraes, L. Akutsu (2016) The Relationship between Judicial Staff and Court Performance: Evidence from Brazilian State Courts, International Journal for Court Administration, Vol. 8 No. 1, October 2016

Palumbo G, Giupponi G., Nunziata L and Mora-Sanguinetti J.S, 2013, The Economics of Civil Justice: New Cross-Country Data And Empirics, Economics Department Working Papers No. 1060 OECD

Peer E., Gamliel E., 2014 Heuristics and Biases in Judicial Decisions, 114 Court Review – Vol. 49.

Petersen N. 2013. Avoiding the common-wisdom fallacy: The role of social sciences in constitutional adjudication, International Journal of Constitutional Law vol. 11.

Van der Torre A., Jonker J., van Tulder F., Steeman T., Paulides G., (2007), Rechtspraak: productiviteit in perspectief

Vermeule A. 2006 Judging under Uncertainty. An Institutional Theory of Legal Interpre-

tation, Harvard University Press.

Voigt S. (2016) Determinants of judicial efficiency: a survey, European Journal of Law and Economics, 2016, vol. 42, issue 2, No 1, 183-208

Ivar Hannikainen* (Pontifical Catholic University of Rio de Janeiro)

Cross-Linguistic Evidence of Essentialist Beliefs about the Law

A core question in legal philosophy concerns the necessary and sufficient conditions something must satisfy in order to count as a law or a legal system. Lon Fuller offered a major contribution to this discussion when he proposed a novel set of necessary conditions. Specifically, Fuller argued that a social arrangement is a legal system insofar as that arrangement satisfies eight principles that he collectively called "the inner morality of law." These principles include, for example the generality principle (that legal systems must have general rules of conduct) or the publicity principle (that legal rules of conduct must be made public for those regulated to learn of their rights and duties).

We conducted four experiments in order to probe the folk and expert concepts of the law. Our studies sought to evaluate the extent to which Fullerian principles garner support, and also to investigate whether beliefs about the nature of law are robust to manipulations of construal level and evaluation mode.

Our results revealed limited support for Fuller's procedural natural law theory: First, United States adults (n = 454) with no specific training or knowledge of the law tended to cast doubt on the truth of Fuller principles. Second, we replicated this result in a sample of bar association members (n = 91). Though some individual principles were widely endorsed, others were widely rejected. Thus, insofar as Fuller aimed to capture the predominant concept of law, his theory misses the mark at least in part.

We also found that beliefs about the nature of law were quite unstable: When we varied experimental conditions, from joint to separate evaluation, participants became more likely to view Fullerian ideals as necessary properties of law—and less likely to view them as empirically true. In fact, in separate evaluation, participants tended to exhibit a peculiar combination of beliefs: agreeing that 'there could be no retrospective laws' while also believing that 'there are laws that are retrospective'. Follow-up research has now shown that this effect emerges in various languages, including Polish, Spanish and Dutch.

Our results also speak to two common objections levied against folk psychological evidence on philosophical issues: the expertise defense and the reflection defense. Regarding the former, did experienced legal professionals reveal different intuitions? They did not; legal professionals were also divided with regard to the truth of Fuller principles and susceptible to the effect of construal level. Regarding the reflection defense, did conditions favoring more careful reflection influence beliefs about the inner morality law? Indeed, our evidence indicated that, when prompted to resolve the tension between their conflicting intuitions, individuals were more likely to conclude that Fuller principles are contingent, not necessary, properties of law. Together, these experiments suggest that laypeople and legal experts have both empirical and essentialist beliefs about the nature of law. Perhaps, naturalist and positivist views about law are differentially supported by these two ways of thinking about what laws are like. Essentialist beliefs emerged more clearly when participants reasoned about hypothetical legal systems instead of actual legal systems, and when they described the abstract essence of law instead of concrete instances of law.

---

*Corresponding author: ivar.hannikainen@gmail.com

Andrew Higgins (University of Oxford)
Inbar Levy* (University of Melbourne)

Judicial Policy, Public Perception, and the Science of Decision Making: A New Framework for the Law of Apprehended Bias

The right to an impartial and independent tribunal is fundamental to the administration of justice and public confidence in it. However the test for apprehended bias is not informed by psychological research on cognitive biases, and while courts purport to give effect to the views of a fair-minded and informed member of the public on the risk of bias, little attention has been given to what the public thinks in reality. Using doctrinal analysis and drawing on psychological literature, the paper argues that the law must be re-examined with a view to closing the gap between the case law on which factors give rise to a reasonable risk of bias, public attitudes, and psychological research on decision-making. The article proposes a new framework for the law, including a judicial code that identifies circumstances when judges should and should not sit based on legal policy considerations, measured public opinion, and relevant psychological studies, and new procedures and tests for courts dealing with cases that are not identified as automatic disqualification or non-disqualification scenarios under a code.

Camillia Kong* (Birkbeck College)
John Coggon (University of Bristol)
Michael Dunn (University of Oxford)
Penny Cooper (Birkbeck College)

Judging Values and Participation in Mental Capacity Law

Judges applying the Mental Capacity Act 2005 (MCA) are charged with making profound decisions that directly impact the lives of a very large number of individuals. This task is challenging because the weight accorded to the person's (P) values and testimony in judicial deliberation about his or her capacity and best interests remains heavily contested, with little consensus emerging in judicial practice. Part of the reason for this is that incommensurable values often collide in such cases, whether between those held by the judge, those held by P, or those that pre-dominate within caregiving institutions or across society more generally. In this paper, a strict and flexible interpretation of the MCA's values-based approach to making decisions about P's capacity and best interests is outlined. Within this outline, attention is drawn to different problematic implications that these two identified approaches have both for the status and normative force of P's values, and the participatory role that P ought to play in judicial deliberations about decisions concerning his or her care or treatment. Two main areas of concern are discussed: first, on the strict interpretation, the central requirement for P to participate in judicial proceedings is explored, with attention drawn to a false separation between ascertaining P's values and the intrinsic value of enabling P to participate in courtroom proceedings. Second, on the flexible interpretation, the discretion that judges possess to draw on values that they deem significant in ways that can legitimately override the expressed values of P in judicial decision-making is revealed and explored. The values that judges draw upon, which take the form of internal judicial values (values inherent to what the law is and is not supposed to do) and extra-legal judicial values (values drawn from additional social, political, and moral sources), are shown to be ambiguous, and in need of further clarification, particularly with regards to how they intersect with the values held by P him/herself.

This ambiguity demands a research agenda which undertakes empirical investigation into the role of values in judicial deliberation, so as to identify relevant trends in a legal system where precedent is a limited doctrine. The AHRC-funded Judging Values and Participation in Mental Capacity Law project will be engaged in such research, exploring crucial questions through interviews with judges and advocates, including:

- Which values are important for judges when making decisions under the MCA and why? Where do these values come from and how are they interpreted?

- Does the MCA itself incorporate certain legal, political, and moral values?

- What implicit assumptions / a priori values inform whether P's testimony/values are accorded evidentiary weight?

This empirical data will function as a grounding for the development of a normative framework to probe the criteria of reasoning, values, and legal procedures which enable judges to fulfil their obligations towards P in accordance with the empowering ethos

---

*Corresponding author: camillia.kong@bbk.ac.uk

of the MCA and Articles 12 and 13 of the United Nations Convention on the Rights of Persons with Disabilities. Probing the theoretical basis of broader ethical obligations owed to P will help resolve issues of what is owed to P at the level of effective and just participation in judicial reasoning, and feed into crucial professional development tools and policy guidelines around judicial practice and deliberation. Specifically, the empirical and normative analyses will inform recommendations around best practice within the MCA's statutory framework exploring the procedural and policy changes and training tools that are required to fulfil obligations to P as well as encourage greater transparency about the judge's own values in the deliberative process.

Moa Lidén* (Uppsala University)

"Guilty, No Doubt": Detention Provoking Confirmation Bias in Judges' Guilt Assessments and Debiasing Techniques

In 1989, the European Court of Human Rights (ECHR) decided a case called Hauschildt v. Denmark in which a judge who made repeated decisions to detain Hauschildt awaiting trial also decided about his guilt. The ECHR concluded that Hauschildt's right to a fair trial (Article 6) had been breached since the judge was partial when deciding about guilt. However, with reference to differences in Danish and Swedish standards of proof for detention, the Swedish legislator did not see any reason to change the Swedish Code of Judicial Procedure as a result of the ECHR's verdict. Thus, Swedish judges are allowed to decide about both detention and guilt, unless a claim of partiality is substantiated[1].

Subconscious biases such as confirmation bias are not a traditional basis for claims of partiality. However, the theory of confirmation bias emphasizes how easily individuals form hypotheses and the great difficulties they have in reasoning independently of their hypotheses[2]. This makes them remarkably perseverant and unresponsive to new input and that their beliefs persist long after they ought to have been abandoned or at least seriously questioned[3]. Although the more general topic of bias in the Court setting is well researched[4], confirmation bias in law-learned judges is not.

The present study examines whether judges' pretrial detention decisions trigger confirmation bias in their guilt assessments. It also tests two strategies to mitigate confirmation bias: (1) to have different judges decide about detention and guilt and (2) to reduce cognitive load by structuring the evaluation of evidence. In Experiment 1, Swedish judges (N = 64) read 8 scenarios in which they either decided themselves about detention or were informed about a colleague's decision. Then, participants rated the defendant's trustworthiness, the strength of each piece of evidence, the total evidence and decided about guilt. In Experiment 2, Law students (N = 80) either first rated each piece of evidence separately and then the total evidence (structured evaluation) or only the total evidence (unstructured evaluation), and then decided about guilt. Overall, detained defendants were considered less trustworthy and when participants themselves detained, they rated the guilt consistent and total evidence as stronger and were more likely convict, compared to when a colleague had detained. The total evidence was considered stronger after unstructured than structured evaluations of the evidence but the evaluation mode did not influence guilt decisions. This suggests that changing decision maker holds greater debiasing potential than structuring evidence evaluation.

---

*Corresponding author: moa.liden@jur.uu.se

[1]The Swedish Government Bill, Prop. 1992/93:25 p. 16.

[2]Greenwald, A. (1980). The totalitarian ego. American Psychologist, 35, pp. 603–618, Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. Journal of Personality and Social Psychology,37(11), pp. 2098–2109, Nickerson, R. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. Review of General

[3]Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm.

[4]See for instance Englich, B., & Mussweiler, T. (2001). Sentencing under uncertainty: Anchoring effects in the courtroom. Journal of Applied Social Psychology, 31, 1535–1551, Rachlinski, J.J. (2012). How Judges Make Decisions, in Judicial Decision Making in Civil Law: Determinants, Dynamics and Delusions. Russo, J. E., & Schoemaker, P. J. H. (1989). Decision traps: Ten barriers to brilliant decision-making and how to overcome them.

Although statistics are lacking regarding how often the same judge decides about both detention and guilt in real life criminal cases, estimations made by Chief Judges at the District Courts in Sweden suggest that across the different Courts this varies from 1% to 80% (M = 26.39%, SD = 24.40%) of the cases. It is also uncertain exactly how many individual cases this entails, but using official statistics from The National Court's Administration (2017), an estimation is that approximately 2847 (8.60% of 33 107) of the cases decided in 2017 included at least one detention hearing and in 656 (26.39% of 2487) of these, the same judge decided about both detention and guilt. The three most commonly mentioned factors that could influence whether the same or different judges decided were organizational factors such as rotation schedules for allocating cases (79.17%), work load (62.50%) and the type of case to be determined (33.33%).

The above described findings provide an impetus for future research to gain further knowledge of the extent of the problem as well as whether rotation schedules or similar can be used to avoid situations in which the same judges decide about both detention and guilt, using the Danish model as a reference. For the European countries that are required to incorporate the ECHR's judgments, the need for such research is evident from Hauschildt v. Denmark but also outside of Europe this is motivated by the defendant's right to a fair trial.

In sum, this research identifies a trigger of confirmation bias, the detention, that is both inherent and very common in criminal cases. The tested debiasing techniques, especially the change of decision maker, hold some potential in mitigating confirmation bias after a detention but there is an apparent need for a more sound and nuanced framework. Even so, it seems important that Courts take proactive organizational actions to avoid situations in which the same judge is in charge of both the pretrial detention and the main hearing. If this is not accomplished by the Courts it is indeed reasonable for defendants' to put forward claims of partiality and for the Courts to take such claims seriously.

Michal Ovádek* (KU Leuven)

The Apolitical Lawyer: Experimental Evidence of a Framing Effect

The integration of cognitive psychology into law and economics has sparked a fruitful (Jolls et al., 1998; Korobkin and Ulen, 2000) and disruptive (Rachlinski, 2011; Sheffrin, 2017) research agenda. A growing body of evidence emerging over the past three decades has demonstrated how decision-making of legal practitioners – judges, lawyers, clerks – falls prey to various heuristics and biases (Gigerenzer and Engel, 2006; Guthrie et al., 2001). A great deal of this scholarship has been produced in the common law context – notably the United States – while focusing on distortions in cognitive abilities to make quantitative judgments (Robbennolt, 1999; Guthrie et al., 2001; Rachlinski et al., 2015). The interest in an individual's numerical skills has not only methodological appeal but also a long-standing pedigree in behavioural research (Tversky and Kahneman, 1974, 1981).

The present article aims to contribute to behavioural legal research by developing a novel hypothesis combining the potency of framing and legal attachment to (perceived) objectivity and impartiality. The 'apolitical hypothesis' rests on two important behavioural findings. The 'political' treatment is experimentally applied through the medium of framing (Tversky and Kahneman, 1981, 1986; Keren, 2011). Labelling an argument as 'political' or 'politically motivated' is hypothesized to constitute a negative frame that will sharply reduce the attractiveness of the associated legal option.

The experimental setup constructed for the purpose of testing this hypothesis engaged law students in 'traditional' – in continental European terms – legal analysis consisting of the application of doctrinal rules and principles to two factual scenarios in order to make a motivated choice between two legal options. Uniquely in the experimental legal literature, European Union (EU) law was used as the substantive area of law for the experiment. The experiment revealed that legal argumentation treated with a 'political' frame made law students 12% to 24% more likely – varying with legal indeterminacy of the scenario – to select the untreated, 'apolitical' legal option. This conclusion is paradoxical: in trying to be (or appear) impartial, lawyers become influenced by legally irrelevant information.

The experiment took place in a class concerning EU law at a traditional Belgian law school in November 2018 with 88 students. A particular issue of EU law to which the class was devoted – the choice of legal basis of EU legislation – was selected as the substantive area for the experiment and two factual scenarios with differing degrees of legal ambiguity were drafted. Scenario A was designed to be as indeterminate as possible, while scenario B was intended to have a relatively clear-cut legal solution. In each scenario, subjects had to choose a legal basis from two possibilities. The students had some prior knowledge of EU law, varying individually and had already studied at the law school for over three years at the time of the experiment.

The table below reports the frequencies of subjects' legal choices for both scenarios. In the aggregate, 15 more subjects chose the 'apolitical' legal basis – counted as a 'success' x – in the treatment group than the control group.

Scenario	$n_{control}$	$n_{treatment}$	$x_{control}$	$x_{treatment}$
A	45	42	18	22
B	42	45	2	13
A+B	87	87	20	35

With the experimentally generated data we can construct loglikelihood functions to visualize the differences in the proportion of successes for each group. The loglikelihood function takes the sample size n and the number of successes x to derive the unknown parameter θ. What we are particularly interested in is the value of θ at which each function is maximized. The maximum (log)likelihood for a binomial distribution with iid random variables is simply $\theta = x / n$.
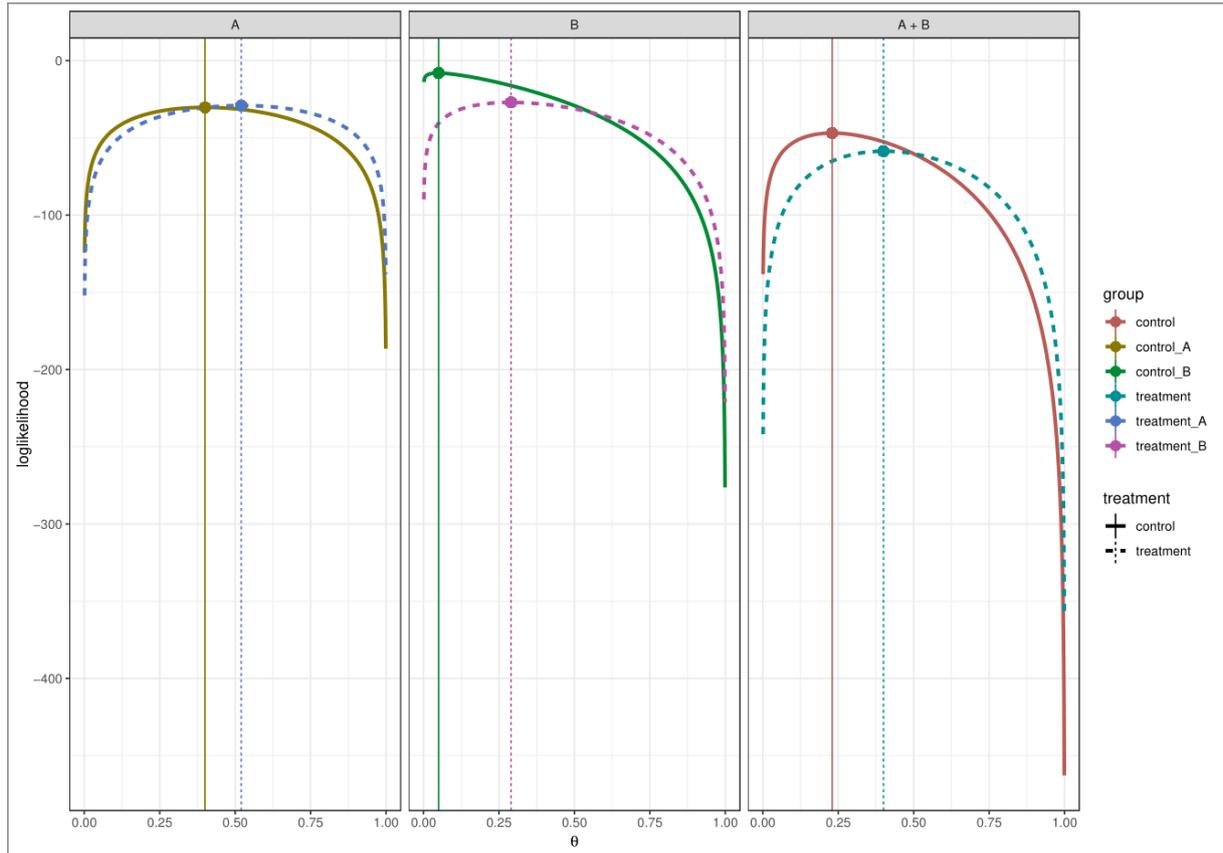
The $\Delta\theta$ difference between the control and treatment function under each scenario represents the effect size of the treatment. We can see strong evidence for the alternative hypothesis, because in all scenarios the maximum likelihood point of the treatment function lies to the right of the maximum likelihood of the control function. In other words, both at the disaggregated and aggregated level the subjects were more likely (positive effect) to select the 'apolitical' legal basis under treatment conditions. The table below summarizes the effect sizes, expressed as difference in proportions, loglikelihood ratio (LR) and Cohen's h, under each scenario.

Scenario	$\Delta\theta$	LR	Cohen's h
A	0.12	0.96	0.24
B	0.24	3.37	0.69
A+B	0.17	1.25	0.37

In terms of statistical significance, all test statistics tell the same story: the between-

group difference is statistically significant at p-value < 0.01 for the aggregate data and scenario B but not scenario A. The p-values mirror the ordering of the effect sizes for the different experimental conditions: hB > hAB > hA. While formally we must conclude that data for scenario A does not demonstrate a statistically significant difference in the treatment group, h = 0.24 represents a 'small' effect size in the standardized vocabulary of Cohen. In light of the significant and larger effect sizes of the aggregate and scenario B data, it seems plausible that the hypothesized apolitical effect does exist but the design of scenario A – with a great degree of legal ambiguity and therefore stochasticity – was not sufficiently sensitive (unlike scenario B) to detect the effect at n = 87.

Nevertheless, the most important finding from the theoretical perspective is the more than medium-sized and clearly statistically significant effect identified under scenario B. Despite the extremely low threshold of a success in the control group proving the low ambiguity of the legal scenario, the treatment induced subjects to be more than three times as likely to opt for an apolitical legal basis. Scenario B was devised as a hard case – pushing the apolitical hypothesis to its limits – and the results point unambiguously towards the existence of a sizeable and significant effect (h = 0.69, p-value ¡ 0.005). Political treatment can thus make lawyers turn away from even the most unambiguously clear answers to legal questions.

Jeffrey Rachlinski* (Cornell University)
Andrew Wistrich

Judging Autonomous Vehicles

The emergence of self-driving cars will present numerous challenges to the legal system. Although regulatory agencies will determine when and how such cars may be placed into service, judges will likely play a central role in defining the standards for liability for such vehicles. Will judges be biased against such cars? People disfavor injuries arising from artificial sources, expressing a "naturalness bias." In this paper we find that judges express a similar bias against self-driving cars. They both assign more liability to self-driving cars than they would to a human-driven car and treat accidents as more serious when caused by a self-driving car.

*Corresponding author: jjr7@cornell.edu

Brian Sheppard* (Seton Hall University)

The Automation of the Reasonably Prudent Person Test

The reasonably prudent person standard is the crutch upon which much of the American tort system balances. The variety of circumstances that give rise to injury-producing carelessness is so large, that legislators have struggled to articulate clear and satisfactory conduct rules for the minimization of injury. The reasonableness standard is a low-cost way to provide some clue regarding permitted or prohibited conduct, but it does not provide those who are subject to it (or those applying it) with much guidance about how to behave in a given situation. Moreover, it can serve as a smokescreen to mask ideological or self-serving decision-making: by invoking the standard, a result can be made to appear as if it was dictated by law even if it was dictated by the whims of the interpreter. So while the versatility and context-sensitivity of the test are hailed as virtues, its murkiness and lack of constraining power are criticized as its vices. Rules, by contrast, provide the inverse.

This trade off begs for a technological solution. For a long time, we have had the computing capacity to automate decisions that are powered by simple rule structures. If we accepted a quantitative interpretation of reasonable conduct (such as the average speed of a driver for the area in question) then it would not be challenging to automate the application of the reasonableness test to a situation in which we had the appropriate numerical data (such as the speed the defendant was traveling). This does not solve the context-sensitivity problem, however, because we might want different rules for different circumstances (such as rules that changed depending on the road, weather, time of day, etc.). Enter algorithmic decision-making. Scholars like Casey and Niblett have argued that emerging technology, such as machine learning, can provide context-responsive directives with clarity and constraining power. By building variables into an algorithm and collecting data in real time, we might be able to determine the parameters of, say, reasonable conduct—a finding that can give rise to machine-written bright-line directives that apply to a specific place and time.

The question I plan to analyze here is whether technology like this is capable of providing the virtues of the reasonably prudent person test in American tort law without its vices. I will focus on the different ways that the technology could be designed and highlight the weaknesses of these approaches. In particular, I explain how certain approaches, such as those that use prior jury determinations to train the machine, risk stagnation that does not sully the jury-driven reasonably prudent person test.

---

*Corresponding author: brian.sheppard@shu.edu

Noel Struchiner* (Pontifical Catholic University of Rio de Jareiro)

An Experimental Guide to Vehicles in the Park: The Psychological Fault Lines of the Hart Fuller Debate

Jurisprudence, or general philosophy of law, in as much as it attempts to uncover the nature of law and is interested in questions such as what it means for law to exist and what makes law efficacious, must describe how law affects the practical reasoning of the full range of its subjects. Gerald Postema, in his "Positivism, I Presume?... Comments on Schauer's "Rules and the Rule of Law" (Harvard Journal of Law and Public Policy 14, 1991, 797–822) surmises it this way:

> In one way or another the law plays a role in the practical reasoning of everyone in society, and in reasonably well functioning societies, law works as an internal guide to (nearly) everyone in society, and not just to appellate judges. It is to say that a general jurisprudential theory would be radically incomplete and seriously misleading, if it failed to give some account of the place of law in the practical reasoning of officials, lawyers, and lay citizens alike.

Even though we agree with Postema that jurisprudence is an enterprise in practical philosophy, we don't think he went far enough. If jurisprudence is to live out the true meaning of its creed, then it does not suffice to say that "one way or another" the law plays a role in practical reasoning. We have to strive for precision and uncover in exactly what way or ways law plays a role in practical reasoning. It is also not enough to speculate from the armchair how law could make a practical difference (i.e. particularism, rule-sensitive particularism, presumptive positivism, hard core formalism). Philosophical speculation is not worthless. It is certainly food for thought and a starting point to obtain hypotheses that must be tested empirically and experimentally. To describe the nature of law is not (or not only) to describe what law could possibly do in terms of interference with practical reasoning, but what law actually does to practical reasoning (of officials, lawyers, and lay citizens alike).

One device that law uses to make a difference in practical reasoning and thus provide internal guidance is to communicate with rules. Hardly anyone would deny this. There is, however, a long tradition of philosophical disagreement regarding both what a rule is and what a rule does. In the jurisprudential field, the Hart x Fuller debate epitomizes this disagreement. While for Hart a rule is not to be confused with its purpose, and can have a core meaning given solely by the words and corresponding concepts used in its formulation, Fuller insisted that the language of the rule itself is never enough to ascertain the meaning of the rule, i.e., rule and purpose can never be teased apart. They dwelled about the meaning of the "no vehicles in the park rule" most prominently, but each had his share of examples. Younger generations have jumped into the debate taking sides (e.g. Schauer sustains the idea that rules are generalizations that have semantic autonomy towards their background justifications – purposes – and offer resistance to them), but, until now, no one has empirically verified their intuitions.

But what is the ordinary (folk and expert – non-philosophical) concept of a rule? Do legal professionals and ordinary people think alike when it comes to rules? We strive to answer these questions through a set of experiments and to uncover which philosophical

---

*Corresponding author: struchiner@gmail.com

theory is vindicated by the empirical data. Our results help to understand the psychological underpinnings present in a longstanding jurisprudential debate about the nature of rules.

Kevin Tobia* (Yale University, ETH Zurich)

Testing Original Public Meaning

Influential theories of legal interpretation recommend using dictionaries or corpus linguistics to provide evidence about the "original public meaning" of legal texts. For example, Public Meaning Originalism might recommend using dictionaries or corpus linguistics to uncover the original public meaning of "bear arms" in the U.S. Constitution's Second Amendment. Such an interpretive inquiry is typically understood as an empirical one, aiming to discover a fact about how ordinary people actually understood the text at the time it became law. However, a central methodological question remains unanswered: Do these methods of corpus linguistics and dictionary-use accurately reflect original public meaning?

This paper develops a novel method of "experimental jurisprudence" to assess this question. Experimental studies reveal systematic divergences among the verdicts delivered by ordinary concept use, dictionary use, and corpus linguistics use. For example, the way in which people today apply the concept of a vehicle is systematically different from the way in which people apply the modern dictionary definition of a "vehicle" or interpret the modern corpus linguistics data concerning vehicles. Strikingly similar results arise across levels of legal expertise, including "elite-university" law students (e.g. from Harvard Law) and United States judges. These findings provide evidence about the accuracy of dictionaries and corpus linguistics in estimating modern public meaning. I argue that these studies also provide evidence about the accuracy of similar (originalist) methodologies in historical times and that these findings shift the argumentative burden to Public Meaning Originalism to demonstrate that their methods reliably track public meaning.

---

*Corresponding author: kevin.tobia@gmail.com

Cláudia Toledo* (Federal University of Juiz de Fora)

Judicial Activism – The need for parameters: analysis of legal reasoning in judicial review

In Democratic Legal States, judicial decisions must be objectively and rationally grounded in statutes and precedents. Due to the system of checks and balances, Judicial Power must control the acts and omissions of the other Powers. However, the suitability and intensity of this control is often questioned with the frequent criticism of judicial activism, understood as undue judicial interference with the competence of Legislative and Executive Powers.

This paper presents some partial results of a two-year research project currently underway, that aims (i) to identify objective parameters for the assessment of judicial decisions justification (bibliographic research), and (ii) analyze the kind of arguments that ground judicial review, applying those parameters in an international comparative study of the constitutional case law of Brazil, Argentina, Mexico and Germany, and in a national comparative study of the infra-constitutional case law of different Brazilian states (empirical research). On the basis of these data, it is then possible to demonstrate the activist performance (or not) of Judicial Power.

For the analysis of the arguments used in the legal reasoning of judicial review, relevant contributions can be found in theories of legal argumentation. Robert Alexy's theory – together with the Neil MacCormick's one – is internationally acknowledged as the standard theory of legal argumentation. According to Alexy, there are two types of discourse, empirical discourse and general practical discourse. The empirical discourse describes reality, either by past, present and future concrete facts, or by scientific data from natural and social sciences. The general practical discourse is a normative speech, in which the claim to correctness is raised. According to Jürgen Habermas (whose discourse theory is the main bases of Alexy's work), the general practical discourse is formed by pragmatic, ethical, and moral arguments. Pragmatic arguments are related to the choice of techniques and strategies of action, based especially on the criteria of efficiency and utility. Ethical arguments are related to the tradition that gives identity (cultural and political self-understanding) to a certain society. Moral arguments refer to the "symmetrical interest of all", having the semantic form of categorical imperatives.

Finally, according to Alexy, legal discourse is a special case of general practical discourse. It is composed of institutional arguments - statutes, precedents and (ruling) doctrine. Institutional arguments are defined by authoritative issuance and social efficacy. They are the most suitable arguments to ground judicial decisions, which are also institutional acts. Although often not sufficient to justify decisions, institutional arguments are necessary for such justification, so that they are objective, rational, controllable acts. Some conclusions can also be drawn from this:

- the greater the number and relevance of institutional arguments in judicial review, the greater the chance that Judicial Power is acting within its competence – thus, the lower the probability of judicial activism;

- the greater the number and relevance of non-institutional arguments in the ratio decidendi of judicial review, the greater the chance that Judicial Power is acting beyond its competence – thus, the higher the probability of judicial activism;

---

*Corresponding author: toledo.claudia@direito.ufjf.br

- the greater the approach to the issue at stake by institutional arguments, the greater the scope of judicial review and the greater the chance that Judicial Power is acting within its competence – thus, the lower the probability of judicial activism.

Fundamental social rights claimed before the court are a paradigmatic issue in giving rise to the critiscism of judicial activism. These are fundamental rights to positive State provision of goods, services or financial benefits. Therefore, the judicial order for their compliance means, for example, the court order for the State to provide medical treatment to a person or a place in public school for a child, situations that raise the criticism of undue judicial interference with the implementation of public policy.

Among fundamental social rights, the right to the existential minimum is the one whose claim raises the most criticism of judicial activism, since existential minimum is the only definitive right (object of a rule) among all fundamental rights (prima facie rights, object of principles). Therefore, "existencial minimum" was the search expression used in the empirical research on the Courts websites.

1. all plenary decisions that refer to the search expression were surveyed. From 2004 (the first literal reference to "existential minimum" by the Constitucional Court) until 2017 (the beginning of the research project), there were 12 judicial decisions;

2. there is a clear growing appeal to Judicial Power for the guarantee of the right to existential minimum. Whereas the first reference to this right in 2004 was made in a monocratic decision, in 2008 there was 1 plenary decision on this issue and in 2017 there were 7 plenary decisions on it;

3. in all 12 decisions surveyed, institutional arguments were used. Statutes and precedents were mentioned in 100% of the decisions, while reference was made to doctrine in 92% of them (i.e., in 11 decisions);

4. all non-institutional arguments were also used, considering the 12 decisions in total. Most decisions presented more than one non-institutional argument. Among the general practical arguments, the pragmatic ones are the most used (in 6 decisions – 50%), followed by the ethical (5 decisions – 42%) and the moral ones (4 decisions – 33%). Among the empirical arguments, there were arguments related to concrete cases in 6 decisions (50%) and to scientific data in 5 decisions (42%).

Judicial activism is not a phenomenom identified according to a binary code "yes" or "no". It presents a gradual structure, i.e., undue judicial interference may be lighter or more serious depending on the conditions of the concrete case. The research project is not over yet but seems to lead to the conclusion that, at least in relation to the highly controversial issue surveyed, there are not enough arguments that rationally and objetively justify serious judicial activism in Brazilian Constitutional Court.

Marion Vorms* (University Paris 1, IHPST, CNRS, Paris)

The story-model of jurors' decision-making in the light of an analysis of evidential reasoning — conceptual and empirical questions

The story-model advocated by Pennington and Hastie (1986, 1988, 1992, 1993) is among the most influential views of jurors' decision-making. According to it, jurors in criminal trials make sense of the evidence through the construction of a mental representation of the events, rather than through the estimation and combination of probabilities. Stories consist in causal explanatory scenarios of the crime, and they are composed of sub-stories called "episodes", which are made of causal links between physical and mental events (goals, intentions, beliefs, etc.). In addition to governing the interpretation of the evidence, stories drive the choice of a verdict (hence govern jurors' decision).

The goal is of this paper is twofold. First, we propose a critical clarification of some aspects of the story-model in the light of an analysis of evidential reasoning, drawing from categories proposed by Schum (2004), and in the line of Lagnado's work on legal reasoning (e.g. Lagnado & Harvey 2008, Lagnado 2011). Second, we present two experimental studies we have run (results are currently under analysis), which aim at testing two implications of the story-model (which are problematic from the perspective of our criticism).

According to Pennington and Hastie, jurors spontaneously impose a story-structure to the evidence, before making any decision regarding the verdict (this is not a post hoc rationalisation). The very process of making sense of evidence consists in the construction of a causal, explanatory, mental model of the events. This leads Pennington and Hastie to make a strong claim, which is at odds with the basic requirement of evidential reasoning: the organisation, interpretation, and processing of evidence is mediated through story construction. In constructing a story, 'the juror chooses to believe only some of these assertions and will make inferences that other events occurred' (1988, 524). Hence, the evaluation of evidence credibility, relevance, and probative force, is conditional on story construction, and on the various criteria that make a particular story 'good' (coherence, completeness, coverage). As we will argue, this leaves open the question of how jurors come up with a story in the first place. This question is all the more difficult to answer that the very notion of 'story' lacks clear definition.

This conceptual criticism leads us to empirical explore two implications of the story-model.

First, the more narratively the evidence is presented to jurors at trial, the more confident they should be in reaching their verdict. Since forensic evidence is less narrative than lay testimony, which is more readily accommodated within a story, it is to be expected that testimonial evidence should weigh relatively more in jurors' decision than forensic evidence. Even more, the relative credibility and inferential force of either should not matter much in decision-making. Our first experimental study is aimed at testing this implication of the model, by testing the relative weight of good story with weak evidential basis / strong piece of (forensic) evidence that does not fit well in a story on jurors' reasoning and decision. Does strong evidence without scenario trump a good explanatory scenario with weak evidence (or vice versa)?

Our second experimental study is aimed at clarifying the role of stories in jurors' decision-making through the testing of the 'uniqueness' hypothesis, according to which one good story is better than two.What happens when subjects are presented with two

---

*Corresponding author: mvorms@gmail.com

incompatible stories, which are however both compatible with (and explanatory of) the same evidence, and which go in the same direction (inculpatory)? Normatively, having a second story added to a first one should not lessen our degree of belief in guilt (if they are both incriminating). Empirically, the story-model predicts that a second story makes it harder to accept one or the other, and sheds some doubt on the whole thing, thus lessening belief in guilt. We will present both quantitative results, and qualitative answers given by the participants to explain their decision.

Wessel Wijtvliet* (KU Leuven)
Arthur Dyevre (KU Leuven)

Measuring Judicial Ideology in Economic Cases: An Expert Crowdsourcing Design

Whether and to what extent judicial outcomes reflect the influence of the sitting judges' personal attributes, notably their ideological leaning, is a question that is central to judicial behaviour research. With the US literature serving as methodological reference for empirical legal research in the rest of the world, scholars have undertaken to apply these techniques to courts in Latin America, Canada, Australia and Europe. Yet in many settings, such as the EU judiciary, their application is made difficult or impossible due to the unavailability of voting records, the low saliency of judicial appointments (newspapers ignore them) and the lack of transparency of the appointment process. Recent research has seen interesting attempts to overcome these obstacles. These alternative measurement strategies, though, present serious limitations.

We present an alternative measurement approach designed to leverage the knowledge of legal experts. We ask EU competition law experts to rate General Court judges on a number of dimensions and then use these ratings to construct judicial scores for Europhilia, competition law expertise and business-friendliness. We show that, unlike Europhilia and competition law expertise, the pro-business score of the panel median is a significant predictor of General Court decisions in competition and state aid cases.

Methodologically, our analysis suggests that expert ratings constitute a viable and promising alternative in settings where other measurement methods are unavailable. Substantively, we show that while the ratings indicate that General Court judges differ in their knowledge of competition law as well as in their attitude towards business and European integration, business-friendliness appears to be the main ideological divide when it comes to agreeing on the disposition of competition and state aid disputes. Specifically, we find that the Court is more likely to rule in favour of the Commission when the panel median is less pro-business on our expert crowd-sourced measure of judicial attitudes. These results contradict the notion that EU judges diverge primarily in their degree of Europhilia/Euroscepticism. Yet they are consonant with the US literature on US federal courts, which emphasise economic policy preferences.

---

*Corresponding author: wessel.wijtvliet@kuleuven.be

Christoph Winter* (Harvard University)

Exploring the challenges of artificial judicial decision-making
The application of artificial intelligence (AI) to judicial decision-making has already started. Significant progress has not only been made in the United States, most prominently with regard to bail decisions,[1] but also in Russia.[2] China placed more than 100 robots in courts offering legal advice to the public,[3] and a recently developed algorithm was able to predict 79% of the European Court of Human Rights (ECtHR) judgments concerning cases of torture, fair trail and privacy.[4] Against this background, it seems plausible to assume that future advances in AI will revolutionize the judicial sector. Issues of discrimination have emerged in the context of bail decisions leading to intensive debates around the criminal justice system and demands for what has been coined "algorithmic fairness". Yet, while it seems possible to overcome respective ethical and technical problems,[5] further challenges that arise from the shift in authority from human intuitions to artificial intelligence within the judicial system remain unclear and neglected. Until this day, there is no systematic analysis of the risks of implementing AI into judicial decision-making.

This project therefore sets out to identify and engage with the challenges of artificial judicial decision-making. In this way, it offers a first attempt to its systematization and qualification of solvability from an interdisciplinary perspective. While a comparative point of view between natural (human) intuitions and artificial intelligence will be applied throughout the analysis, special emphasis will be given to ethical issues arising from a change in data and training information taken into account by humans and AI respectively. Furthermore, the potential effects in both civil and common law jurisdictions with their different understandings of the judiciary's role within the separation of powers will be analyzed from a social-psychological and political perspective. For instance, judgments of the European Court of Justice or the United States Supreme Court are arguably perceived as having a much stronger legislative character than those of the German Federal Constitutional Court further challenging democratic principles. Lastly, I will identify and engage with the most relevant legal issues thereby focusing on the demands of a fair trial as recognized by both national and international law.

---

*Corresponding author: christoph_winter@fas.harvard.edu

[1]Angwin et al, ProPublica (23 May 2016) https://www.propublica.org/article/machine-bias-riskassessments- in-criminal-sentencing.

[2]Russia Beyond (13 February 2018) https://www.rbth.com/science-and-tech/327585-free-legal-advicerobolawyer.

[3]World Government Summit (11 July 2018) https://www.worldgovernmentsummit.org/Observer/list/couldan-ai-ever-replace-a-judge-in-court.

[4]Aletras et al, 2016, Predicting Judicial Decisions of the European Court of Human Rights: a Natural Language Processing Perspective, PeerJ Computer Science, https://doi.org/10.7717/peerj-cs.93

[5]See most notably Kleinberg et al., Human Decisions and Machine Predictions, Q J Econ 237 (2017); see also Sunstein, Algorithms, Correcting Biases, Social Research (forthcoming).