THE UNIVERSITY OF CHICAGO

FORMALITY, PROCEDURE, AND JUDICIAL BEHAVIOR

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE LAW SCHOOL

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF JURISPRUDENCE

BY

ZHUANG LIU

CHICAGO, ILLINOIS

DECEMBER 2016

Table of Contents

List of Figures

List of Tables

Abstract

This JSD dissertation includes three essays that study formality, procedure and judicial behavior. Previous studies on judicial behavior show that judges are vulnerable to many undue influences (e.g., emotions, behavioral biases, and other manipulations). Studies have not addressed, however, the fact that judges are decision makers who are embedded in formalities and procedures that are devised to help them to improve decision quality. This dissertation suggests a set of debiasing formalities and procedures, and test their effectiveness against empirical and experimental evidences.

The Cautionary Effect of Formality:

An Experimental Investigation of Chinese Judges

Abstract: Previous studies suggest that one function of formality in law is that it makes people cautious before conducting a legally binding action. Translated into psychology, this means that formality activates the controlled and analytical cognitive processes, and induces people to deliberate. Theories predict that, under such circumstances, people ponder more carefully, and make less impulsive decisions. We tested this formality-cautionary effect in judicial decision-making. In a series of four experiments, we recruited Chinese judges (n= 184) to decide cases that evoked strong tensions between emotional instincts and deliberate decision outcome. Four treatment conditions were used as a proxy for formality. In study 1, we tried the structure of judicial opinion: judges in the treatment group were asked to write down formal reasons first, and then to give verdicts; judges in the control group did the opposite: they rendered verdicts first, and then gave reasons. In study 2, we used legalese as the stimulus: judges in the treatment group read legal arguments written in a more formalistic style than those presented to the control group. In study 3, we explored a priming effect from formalistic legal context: judges in the treatment group were asked to supply a complex legal definition before they made decisions. Yet the question simply framed a formal context, and was irrelevant to the case issue they were about to decide. In study 4, we tested the effect from publicity: judges in the treatment group were told their decisions were to be made public and discussed by law students. We found the formality-cautionary effect in studies 1 to 3, but not in study 4. The results provide empirical evidence of the cautionary function

of formality in law. Some types of formalities can be seen as a de-biasing mechanism that curbs intuition's adverse influence in the judicial process.

Introduction

Formality is pervasive in the law. Terminologies and legalese spread over legal documents and proceedings. Reasoning is required to be written down formally, according to certain protocols in most judicial decisions in civil law countries, and in some important cases in common law regimes. Formalities, such as the affixing and impressing of a wax wafer, are deliberately used by parties whose acts are to be judged by the law. Symbols, such as black robes, white wigs and gavels, are in the popular mind of legalism and weightiness. As an important feature of the law (Summers, 1992; Summers, 1997; Stinchcombe, 2001), the various effects of formality have been extensively studied. Formality functions as ceremony which deifies the law and impresses the litigants and the public (Posner, 1993, p. 127), as evidence that insures a satisfactory memorial of actions, and as a channel that offers a way for the legally effective expression of intention. (Fuller, 1941) Nevertheless, people sometimes criticize its overelaborate complexity and its inefficiency. (Posner,1996)

Among the analyses, a strand of literature suggests that an important function of formality is that it induces people to deliberate. Lon Fuller (1941) calls this the "cautionary function" – a formality may perform a cautionary or deterrent function, by acting as a check against inconsiderate action. It makes people think carefully before binding themselves to an agreement. By stating this, Fuller unintentionally points to the field of cognition and psychology: as a cautionary function, according to psychology, formality may activate the controlled cognitive process (compared to the unconscious process in the dual process theory), and change the subtle balance between intuition and deliberation. General psychology predicts that people under such circumstances ponder more

carefully, are more able to curb their emotional impulses, and as a result make more analytical and utilitarian decisions. (Greene et al., 2001, 2004, 2008; Valdesolo and DeSteno, 2006). Legal scholars have long invoked this psychological wisdom, but they have not examined it empirically. This paper tries to fill this blank and tests this theory. To concentrate on a single issue, we test it in the judicial context, in which formality is most prevalent. We propose that certain formalities make judges more prone to arriving at a more cautious (or less impulsive) result in deciding cases. We ran a series of four experiments to test the hypothesis and identify the channel through which formality makes an impact. In these experiments, incumbent Chinese judges (n= 184) were asked to decide cases that brought about strong tensions between emotional instincts and analytical deliberation. For example, in study 1's case scenario, an impoverished tortfeasor, riding a motorcycle to take his pregnant wife to the hospital, accidentally hit a man and caused him to stay in the hospital for 56 days for treatment. The loss of income amounted to an exceedingly large amount, larger than the tortfeasor could easily afford. According to current torts law, however, the tortfeasor was required to pay such an amount of money. The legal issue here is whether the judges should adjust the amount of compensation, given the tortfeasor's economic condition and the fact that his wife's water was broken and she was about to deliver. Here, sympathetic instinct warrants a lenient award of damage, whereas the law demands a utilitarian one. The other experiments used cases which shared similar controversies on the conflict of values and moralities. Notice that all these were close cases, which were different from the cases judges typically encounter. Nevertheless, close cases are the focus of professional and scholarly attention.

A battery of treatment conditions were used as a proxy for formality. The first was the structure of the judicial decision (study 1). In this setting, some judges (the control group) were asked to render a verdict first, and then give reasons; the other judges (the treatment group) wrote down their

reasons first, and then proceeded to the decision stage. Here, we hypothesize that reason-giving, a judicial protocol employed and required in many (but not all) decision settings, can activate the controlled cognitive processes, and thus change judicial decisions in close cases. The results confirmed our theory: judges in the treatment group, who wrote down their reasons prior to making their decisions, were more inclined to decide in a more law-abiding way.

Studies 2 to 4 extended Study 1 in two ways: first, the structure of reason-giving is one of many kinds of legal formality. We wanted to test the generalizability of the cautionary effect. So we explored different but interrelated psychological phenomena which derived from the dual process theory. Second, that writing reason makes judges deliberate more carefully has been noted by lawyers in previous researches, and it is perhaps not unanticipated to see judges decide differently when they deliberate more carefully by writing down their opinion. Nevertheless psychology has suggested some subtle effects that function on a more abstract level, and have not yet been discussed in legal studies. Study 2 to 4 further explore the issue along this direction.

In Study 2, we used legalese as the stimulus. We proposed that legalese, or "legal language", by alienating judges from a familiar scenario, and framing the scenario into a formalistic context, would curb emotional impulses, and induce more thoughtful decisions. To test this, for the treatment group, legal arguments were written in a more formalistic style than for the control group to elicit judges' changing mindsets and opinions. The result suggested that our prediction was correct. Judges made more cautious decisions with the legalese stimulus.

In Study 3, we explored the effect of a formal context of law on a more abstract level: we tried to use a priming technique from psychology to directly induce a change of decision. To set up a formal legal context, we asked a question about the definition of a legal terminology, ostensibly testing judges' legal knowledge, before the judges made any substantial decisions—the questions

were relatively difficult and required considerable effort to answer correctly; however, they were irrelevant to the legal issue the judges were about to decide. Consistent with the theory, we found a priming effect here. In the treatment group, judges made decisions in a less impulsive way than their peers in the controls, in which no priming stimulus was set.

Our fourth study tried to replicate the result in another domain: it used a different type of formality, the publicity of the judicial opinion, as stimulus. This study did not render any statistically significant finding. We nevertheless present it here for discussion. The treatment group was told that their decisions and reasoning will be anonymously presented to and discussed by law students in a legal writing class in Peking University Law School ("Publicity"). In the control group, none of this information was mentioned. Our result did not show any difference between the treatment group and the control. The no finding result in this setting could be due to a weak stimulus problem that arose from the experimental design; or it might suggest that anonymous publicity may not be as effective in framing judges' cognitive status as the other formalities.

Taken together, the results show that formality has a considerable impact. The findings provide an empirical response to the long-lasting question on the function of formality. Previous studies have suggested that judges are vulnerable to many cognitive and emotional influences when making decisions. The evidence from this paper further confirms such a conclusion. Besides, we show that a certain level of formality can serve as a de-biasing mechanism. This helps us better understand and improve our judicial processes.

Section I explains the research question and discusses relevant theories. Section II introduces the general experimental design, including summary statistics of the participants, the recruitment of judges, the randomization strategy, and other experimental settings. Section III to Section VI report

studies 1 to 4 in detail. Section VII discusses the validity and the implication of the experiments and concludes.

I. Formality and Related Theories

Formality

In legal studies, formality has a broad definition. Black's Law Dictionary defines it as "the conditions, in regard to method, order, arrangement, use of technical expressions, performance of specific acts, etc., which are required by the law in the making of contracts or conveyances, or in the taking of legal proceedings, to insure their validity and regularity."(Garner, 2004) In practice, the term is used to describe legalese, ("may it please the court", "fail not, at your peril"), terminologies ("promissory estoppel", "caveat emptor", and "bona fide"), the trappings of the judicial process (the high raising bench, the robes and wigs worn by magistrates on solemn occasions, etc.), the conditions which must be observed in making contracts ("consideration", or affixing and impressing a wax wafer), and some words that the law prescribes in order to render validity.

Lawyers and scholars often ask why formality is so pervasive. (Summers, 1992; Stinchcombe, 2001) Different kinds of formalities seem to stem from very different roots. For example, common terminologies allow people to communicate ideas more rapidly, with less need for lengthy explanations. For contract law, consideration is accepted as a proof of serious intent, and as an evidence to facilitate enforcement. (Fuller, 1941; Kull, 1992) The wigs and robes and other trappings of the court present a level of decorum and respect for the process. (Chase and Thong, 2012) As Judge Posner (1993, p. 127) puts it, "…the trappings of the judicial process—the impressive courtroom, with the judge sitting at an elevated bench and in uniform, and addressed

with honorifics unusual in a democratic society: all are a conscious effort to make the law more impressive than it would be if it were treated purely as a method of rational inquiry."

Apart from all these reasons, Lon Fuller, in explaining the function of consideration and other formalities in contract, provides an interesting angle. Summarizing previous scholars' view that "enforcement is denied gratuitous promises because such promises are often made impulsively and without proper deliberation", he suggests that a formality may perform a cautionary or deterrent function, by acting as a check against inconsiderate action. (Fuller, 1941) Despite its broad definition and diverse origins, some different types of formalities may serve a unified function: signaling a legal and formal context, a context that primes people into a more serious deliberation mindset. And by mentioning caution, Fuller unintentionally leads legal studies into the field of cognition and psychology. In psychological terms, a cautionary function suggests that formality may activate the deliberation center (System 2) of human cognitive process.

What is the consequence of caution? Psychology generates many predictions, yet one of the most important is that when people deliberate, their emotional impulses will be curbed, and their decisions become more analytical, or, in other words, more rational. This makes the issue potentially important to the law, especially in the setting of judicial decision-making, as judges very often need to deal with relevant questions.

Psychological theories

Work in psychology and neuroscience has revealed that human judgments are often mediated by two classes of brain processes ("dual-process theory", Greene et al., 2001, 2004; Haidt, 2001). One class consists of processes that automatically alter hedonic states in response to specific types of socially relevant stimuli, and this class probably arises from earlier evolutionary development. A second class consists of more effortful processes that underlie abstract reasoning, simulation, and

cognitive control. The automatic or intuitive processes resemble what Daniel Kahneman and Amos Tversky call System 1 of thinking, which is fast, instinctive, and emotional, while the deliberative processes are similar to System 2 in their definition, which is slower, more deliberative, and more logical (Kahneman, 2011).

As emotional responses and moral principles are usually consistent with goals that are evolutionary adaptive (aka, utilitarian, or rational), the intuitive and deliberative processes often work together to foster decisions which accord with the goals of both. However, there are some ethical decisions that induce competition between the two systems (for example, the trolley problem: whether to harm one person in order to save the lives of several people) (Greene et al., 2001; Greene et al., 2004). At the margin, the outcome of the competition depends on the relative activation levels between the two, and this subtle balance can be influenced by many factors. For example, psychological studies have found that when people are shaped into a positive state of mind before facing a moral dilemma, the happy feeling embedded in System 1 can suppress the negative feeling in making some decisions, which arises from the same system. Thus, happy people tend to make more utilitarian decisions. (Valdesolo and DeSteno, 2006) Another example is cognitive load. When decision makers encounter information in clearer form, they feel less burden in processing that information in System 2 (the deliberation center). Hence, more information fluency or less cognitive load is usually associated with more utilitarian decisions in moral judgements. (Greene, et al., 2008)

The dual process theory has a profound relationship with our studies in law and formality. As suggested by Fuller, formality makes people more cautious. Using the language of psychology, formality activates the deliberation center (or System 2) of the cognitive processes, in which an implicit (automatic) and unconscious process shifts partially to the explicit (controlled) and

conscious process. According to theories stated above, people under such circumstances make decision that are less instinctive but more deliberative in scenarios related to moral issues. (Greene et al., 2001, 2004, 2008; Valdesolo and DeSteno, 2006)

The "formality – controlled cognitive process" channel is particularly relevant to the study of law, especially of judicial behaviors. For one thing, formality pervades judicial processes. For another, judges have to exercise discretion in many cases. Most of those are close or hard cases, yet close cases are exactly the focus of professional attention. (Posner, 1993, p125) Take the prevailing "rule versus standard" problem for instance: terms such as "excessive self-defense" and "liquidated damages being too high to be enforceable" abound in the law. Judges must consider all related factors—factual, legal, and moral—to arrive at a prudent standard of "excessive," or to determine what "too high" is, on a case-by-case basis. The absence of rules and the pervasiveness of standards in so many areas of the law mean that the quality of the moral judgment exercised by judges has decisive importance in how they fulfill their role, and should thus be an essential factor in the understanding of judges, especially appellate court judges.

Hypothesis and why we should test it on judges

Drawing from the above theories, the hypothesis of this paper can be formalized as: judges make less instinctive and more cautious decisions when they are disciplined by or are exposed to a certain degree of formality.

General psychology has established, from experiments on lay people, a causal relation between the activation of the controlled process and cautious decisions. There are two problems to be solved in this paper: first, whether some certain kinds of formalities do serve a cautionary function; second, whether judges, who are already well-trained and experienced decision makers, can be influenced by this cautionary effect.

A strand of literature suggests that judges are no different than lay people in making decisions: judges maximize their utilities (Posner, 1993; Posner, 2010; Epstein et al., 2013) as lay people do. They are as subject as lay people to many behavioral biases (Guthrie et al., 2001; Guthrie et al., 2007), and they are generally vulnerable to the influence of extra-legal factors (e.g., Danziger and Avnaim-Pesso, 2011; Wistrich et al., 2014; Spamann and Klöhn, 2016). But as other legal realists, like Llewellyn (1940), have long pointed out, judges are not merely human – they are humans with a particular ability, training, and expertise. Judges are professionals who are highly trained, and specifically selected to interpret the legally relevant factors and to ignore the irrelevant ones (Spellman and Schauer, 2012; Kahan, et al., 2016). In particular, Kahan et al.'s study (2016) suggests that professional judgment imparted by legal training and experience confers resistance to biased cognition (but only for decisions that involve legal reasoning). The tension between these two literatures suggests that one should be agnostic about any particular psychological mechanism's influence on judges. As Llewellyn (1940) wrote seventy-six years ago, the real question is which other factors come into play, and how they influence the decision (e.g., Kennedy, 1998; Simon, 1998; Leiter, 2003; Epstein et al., 2013). Given that some factors might well have deep impacts on decisions as practiced in courts, we would be better off if we could clarify these potential influences, and carefully design de-biasing procedures (Guthrie et al., 2001; Guthrie et al., 2007; Thaler and Sunstein, 2008).

Proxies for formality

There are many types of formalities in law. Constrained by the feasibility of implementation, we could only use some specific types of formalities (proxies) to test a general effect. So what should we have chosen as proxies? We relied on theories to guide our selections. In study 1, we used the structure of judicial decisions as our first proxy. This selection was driven by both legal theories

and psychological studies. Legal studies have long noticed that reason-giving imposes a form of self-discipline, which improves the quality of decisions. (Cohen 2015) Giving written reasons before making decisions restrains instinct from dominating the decision process (the "it won't write" phenomenon) (Waits, 1983; Cohen, 2015), and makes judges deliberate (Merrill, 1981; Posner, 1995). At the same time, studies in psychology have also shown that the need to articulate reasons may reduce the impact of irrelevant factors (Lerner et al. 1998), although not always (Norton et al. 2004). These literatures, however, had not yet been combined, which gave us a chance to borrow wisdom from both academic traditions to study formality.

The standard applied for Studies 2 to 4 were similar: we searched for interactions between previous legal literature and psychological studies to help us pin down our proxies for formalities. We will defer the detailed illustrations to Section III to VI.

In total, we choose four proxies for formality, and designed Studies 1 to 4, which, though not a thorough account for all kinds of formalities, is sufficient to support (or invalidate) our general theory.

II. Experiment Settings

Experiment designs

127 judges participated in Studies 1 to 3. Our program randomly assigned each judge to two studies out of the three. The sequence of the cases was also randomized. Another 57 judges participated in study 4.

All case materials in the studies were designed to embody a conflict between a moral instinct and a deliberative decision. Each study covered a different group of treatment/control conditions. The experiment settings appear in Table 1.1.

Recruitment

We recruited judges in Zhejiang Province who were attending classes at Judges College in Hangzhou, in May 2016. Applicants for courses are judges from all divisions (criminal, civil, intellectual property, administrative) of the local, intermediate and high courts within the province. The College admits applicants based on court-specific quotas.

The 127 judges who participated in studies 1 to 3 came to attend academic writing class. The Chinese judiciary encourages judges to write academic articles. Every court above the intermediate level publishes its own journal, featuring the academic work of judges. Most articles are closely related to judicial practice, such as new problems or hard cases that the judges encountered. The journals also publish more theoretical and sometimes even descriptively statistical work. Judges contribute to these journals for fame and reputation, which can relate to promotion. A paper's quality is also seen as an indicator of a court's performance. The College offers two-day courses in this type of writing. Journal editors and judges with good publication records lecture about their writing techniques and research experience.

One hundred and twenty-one judges finished the studies and submitted their data. Among these judges, 55.8% of them were female; most of them (93.4%) were from local level courts; the majority (66.9%) were civil court judges; about 2/3 of them had a bachelor's degree, and 1/3 have a master's degree; and the mean age was 32 years old. This age seems young. But since Chinese judges, in general, start their career paths after they graduate from college (usually 22 years old) or receive a master degree (usually 24 years old), the mean age here represents a considerable level of judicial experience. And as some researchers suggest, the "frontier judges" – local judges who actually decide cases in practice (instead of becoming heads of departments and dealing mainly with administrative issues, or being promoted to higher level courts, or being transferred to other

government departments) are on average 35 years old. (Liu, 2003; Xiao and Chen, 2009) Our sample is close to this description. But, unfortunately, we do not have demographic information of the full sample of Chinese judges or judges in Zhejiang province to check against with our sample of participants.

It is worth noting that the judges here may not be a representative or random sample of the population. Judges who attended this program were academic oriented, which by itself suggests they can be systematically different from other judges. For example, it was possible that these judges were more ambitious than their peer judges who did not apply for such trainings, in the sense that they wanted to publish more academic papers in order to get promoted. It could also suggest that the judges had a better academic background, or writing technique, which is necessary for publishing academic papers. Nevertheless, we did not think a less representative sample would cripple our design, as in the experimental settings, results were drawn from the comparison between the treatment and the control, to which the judges were randomly assigned. A random assignment eliminates most potential biases and ensures internal validity.

The 57 judges who participated in study 4 attended a five-day judicial training program at Judge College. 52 participants finished the experiment. The demographics of the first and the second wave of judges were similar to each other.

Table 1.2 shows a statistical summary of the participants.

With the permission of the instructors, we invited judges to participate in our experiment during their class time. Most judges chose to participate. They used their laptops, iPads, or other mobile devices to enter the experiment on the Qualtrics platform. The experiment process was a standard one: judges read vignettes of case materials, and made decisions. On the last pages of studies 1 3,

we also asked how much confidence they had in their decisions. We did not do this for study 4, as we suspected it would confound our stimulus. This will be explained in detail in Section 6.

III. Study 1: Structure

Setup

Study 1 focused on the structure of judicial decision making. It tried to see whether the sequence of giving written reasons matters for decision outcome.

Reason-giving is expected in many contexts, including in judicial processes. The reason for reason-giving is multiplex, and most are well studied. (Cohen 2015) What concerns us is that reasoning is usually required to be formally written; and written reasons are usually put before a decision outcome. This practice, growing out of norms, customs, or legally binding protocols, is a common enough formality, yet it lacks studies from a cognitive perspective. Two issues are pertinent here: first, the requirement for written reasons; and, second, the sequence of reasoning and decision making. It is also worth noting that lawyers categorize deliberation and articulation of reasons as substance, while writing is a means to achieve the substance. This is why we think writing reason is a kind of formality. And the formality of writing reasons is not always required in legal procedures. In the U.S., for example, only a small proportion of judicial decisions are made accompanied by a written opinion. Judges need to balance the competing interests between efficiency (due to caseload and the need of moving the dockets) and decision quality. (Cohen, 2015)

Lawyers have noticed that writing reasons imposes a form of self-discipline, which improves the quality of decisions. (Cohen 2015) On the one hand, writing reason is described as restraining intuition from dominating the decision process. This is known as the “it won’t write” phenomenon. “Judges may frequently discover that the original opinion they originally had in mind ‘won’t

write' … The discipline of a written opinion … operates as an important control on judicial arbitrariness." (Waits, 1983; Cohen 2015) On the other hand, reasoning also makes judges deliberate more carefully. As former U.S. appellate judge Charles Merrill (1980) notes, "[T]he very act of writing opinions reinforces the decisional process. Misconceptions or oversights may come to light in the course of articulation." Judge Posner (1995) also writes about such a subtle process:

> "[R]easoning that seemed sound 'in the head' may seem half-baked when written down, especially since the written form of an argument encourages some degree of critical detachment in the writer, who in reading what he has written will be wondering how an audience would react. Many writers have the experience of not knowing except in a general sense what they are going to write until they start writing. A link is somehow forged between the unconscious and the pen. The link is lost to the judge who does not write."

Notice here he describes reason-writing as a trigger that switches the unconscious process to the critical reflection process.

The sense of lawyers has not been tested rigorously in the legal setting, but it is consistent with some psychologists' findings in experiments using lay people as research subjects. Studies find that the need to articulate reasons may reduce the impact of irrelevant factors (Lerner et al. 1998), although not always (Norton et al. 2004). Similarly, having time to reflect on the issues presented might eliminate hunches driven by legally irrelevant factors (analogous to Paxton et al. 2012), notwithstanding counter-examples (Schwitzgebel and Cushman, 2015). And growing out of a relevant mechanism, putting negative feelings into words, or, labeling negative emotions, can attenuate these negative feelings. (Lieberman et al., 2007)

Drawing from legal and psychological theories, we hypothesize that reason-giving, a judicial protocol employed and required in many (but not all) decision settings, can activate the controlled cognitive process and thus change judicial decisions in close cases.

Method

We used Study 1 to test this hypothesis. In the experiment, judges were asked to decide a legal case and write down their opinion. Some judges (the control group) were asked to decide the case first, and then give reasons; the other judges (the treatment group) wrote down their reasons first, and then proceeded to decide. To do this, the experimental website intentionally separated the reasoning page and the decision page. In the control group, judges were asked to make their verdicts and sentences first, and, proceeding to the next page, to give reasons. In the treatment group, a judge could only proceed to the decision stage after finishing the reasoning stage.

Case scenario

In the case scenario for study 1, an impoverished tortfeasor riding a motorcycle to take his pregnant wife to the hospital accidentally hit a man, causing him to stay in the hospital for 56 days to rehabilitate. His loss of income was exceedingly large, and more than what the tortfeasor could easily afford. According to current torts law, however, the tortfeasor should be required to pay such an amount of money. The legal issue here was whether the judges should adjust the compensation, given the tortfeasor's economic condition and the fact that his wife's water had broken and she was about to deliver. Here, sympathy and intuition warrants a lenient award of damage, whereas the law demands a utilitarian one and a full award of damage.

Result

The experiment confirmed our prediction.

In the control group, 31.8% (14 out of 44) judges decided that the defendant should fully compensate the plaintiff -- a law abiding and utilitarian decision given the context. In the treatment group, the number increased to 64.8% (24/37). This difference is statistically significant (χ^2 = 8.813; Pr=0.0030), as shown in Table 1.3.

The amount of payment the judges granted was also different in the control and the treatment groups, as shown in Table 1.4. In the control group, judges on average granted 138,402 yuan compensation, but 183,490 yuan in the treatment group, a 45,088 yuan's increase. The difference is significant at .05 level (t=2.260).

These results suggest that the sequence of reasoning matters. As predicted by our theory, when judges make decisions directly, their intuition is more likely to prevail. When judges deliberate carefully by means of writing down their reasons, they become more deliberative and utilitarian.

To further analyze the result and make the estimates more precise, we used a logit model and an OLS model to control for judges' demographic information, including age, gender, education level, specialized field, and court level. The results are significant even after controlling for these variables (at .01 level), as shown in settings (1) and (2) in Table 1.5.

To better understand the mechanism underlying this effect, after the judges decided their cases, we asked them how confident they were in their decisions. Here, the reason-giving beforehand setting (treatment condition) significantly decreased the judges' confidence in their decisions by about 13%. (56.25% confidence in the treatment group, and 69.38% in the control group, t-value= -2.672, p= 0.0091. Table 1.12 setting (1)) Notice that judges in both conditions gave reasons. The difference lies in the sequence of reasoning. The fact that those giving reasons afterwards were more confident in their decisions suggests a motivated reasoning effect, where the judges in the control group were motivated by their decision results, and convinced themselves about their prior

opinions. This motivated reasoning pattern is consistent with previous findings in psychology and law. (Kunda, 1990; Kahan, 2012) The results on confidence also suggest that judges were more cautious about their decision outcome when they were asked to write down their reasons first.

IV. Study 2: Legalese

We found a significant effect from the structure of judicial decision making. Nevertheless, it is merely one of the many types of legal formalities. Studies 2 to 4 extended Study 1 in two ways: first, we explored different but interrelated psychological phenomena which derived from the dual process theory to test the generalizability of formality's effect. Second, that writing reason makes judges deliberate more carefully has been noted by lawyers in previous researches, and it is perhaps not unanticipated to see judges decide differently when they deliberate more carefully by writing down their opinion. Nevertheless, psychology has suggested some subtle effects that function on a more abstract level, and have not yet been discussed in legal studies. Study 2 to 4 explore the issue along this direction.

In Study 2, we propose that the use of legalese and terminology alienates judges from daily scenario, curbs emotional impulse, and induces more thoughtful decision.

Setup

Nelson Mandela once praised the magical power of the mother tongue, "[I]f you talk to a man in a language he understands, that goes to his head. If you talk to him in his language, that goes to his heart." A cordial language creates a congenial ambience, which easily disseminates happiness, compassion, or hatred. In some circumstances, however, people want to curb the power of emotion, and let reason dominate. Regarding this, cognitive studies have identified a foreign language effect: using a foreign language reduces cognitive biases in evaluating risks and uncertainties (Keysar et al., 2012). And people who think in a foreign tongue make substantially more utilitarian

decisions when faced with moral dilemmas. (Costa et al., 2014) Such phenomena are believed to stem from the reduced emotional response elicited by the foreign language, which consequently reduces the impact of intuitive emotional concerns. (Keysar et al., 2012) The reduction of emotion in a foreign language context increases psychological distance, and this leads people to construe situations in more abstract terms, which often aligns with less impulsive decision making. (Liberman and Trope, 2008; Trope and Liberman, 2010) The effect derives from the dual process theory, in which a foreign language curbs the unconscious/instinctive process, and shifts the balance to the controlled cognitive process.

We suspect that legalese, or "legal language", serves a similar function. It creates distance between the case scenario and the decision-maker by using abstract legal concepts, and hence channels the words into one's head rather than one's heart: it makes individuals deliberate, rather than follow their emotion and instinct. In doing so, it changes the subtle balance between the two cognitive processes. We predict that the use of legalese would induce less impuilsive judicial decisions.

Method

Study 2 tests this hypothesis. In its treatment group, legal arguments were written in a more formalistic style (by using more legalese) than in the control group, to lead judges into a legal context and induce them to deliberate. We chose the wordings of the legalese carefully to insure it did not affect the substance of the arguments, and to generate a treatment condition that provided zero new information compared to the control.

Case scenario

In the case scenario for Study 2, a businessman and his nephew were kidnapped for ransom, and committed a cruel murder of an innocent person to save themselves. The businessman was prosecuted for homicide. The issue is whether the judge should be lenient to this businessman on

the grounds that he killed a person to save his nephew and himself; and if so, then, to what extent. The case was adapted from a real case in China. And it also follows the trolley problem in philosophy (Foot, 1978) and cognitive psychology (a summary, see Greene, 2009): whether a utilitarian decision -- killing one to save several-- is morally correct (or, morally obligatory).

The legal issue embodied in this case has been widely discussed. (Thomson, 1985; Fischer, 1992) A featuring case is R v Dudley and Stephens (1884), which set up the doctrine that necessity is not a defense to murder. Current Chinese law governing this issue is undetermined, however. The general term of criminal law (article 232) states: "[W]hoever intentionally commits homicide shall be sentenced to death, life imprisonment or fixed-term imprisonment of not less than 10 years; if the circumstances are relatively minor, he shall be sentenced to fixed-term imprisonment of not less than three years but not more than 10 years." Nevertheless, there are exceptions. Certain legal theories suggest that the lack of "anticipated possibility" is a rightful defence. (Li and Liu, 1999; Qu, 2005; Zhang, 2009) The term "anticipated possibility" ("qi dai ke neng xing" in Chinese) is transplanted from the German law term "zumutbarkeit", which means "fair expectation that the offender will be able to avoid committing the particular offense". (Fletcher, 2007, p66)

Stimulus

Study 2 used legalese as stimulus. For its treatment group, legal arguments were written in more formalistic style than for the control group. The difference between the control and the treatment questions can be seen below.

This case is now presented to your court. There seems no specific rule in the law that governs this specific situation, but there are some theoretical debates in criminal law principle. **Control & [Treatment]:** Some people think that, ["qi dai ke neng xing" (anticipated possibility) is an element of crime,] when the suspect cannot be fairly expected to act otherwise under the circumstances, he should not be held liable for his act.

But [, as the theory of "qi dai ke neng xing" (anticipated possibility) is not completely recognized as a lawful defence,] other people think this should not be an excuse to acquit murder.

If you are the judge in charge of this case, will you convict the defendant murder?

Yes

No

For the control group, the "anticipated possibility" theory in criminal law was stated in plain language: "when the suspect cannot be fairly expected to act otherwise under the circumstances, he should not be held liable for his act." For the treatment group, the term "anticipated possibility" ("qi dai ke neng xing") appeared among the plain language explanations. We hypothesized that this formalistic context would increase the psychological distance between the judge and the case scenario, and thus activate the controlled cognitive process. (Costa et al., 2014)

Result

We hypothesized that killing an innocent person to make oneself survive would arouse a negative feeling from the judges, whereas the judges need a second thought to further consider the fact that

the killing had saved two persons, which presented a mitigating factor in the circumstance. Here, a less impulsive decision meant calculating the numbers and rendering a lenient verdict.

The results confirm our prediction. As shown in Table 1.6, in the control group, 69.2% (27 out of 39) judges chose verdict A, a decision that convicts the defendant murder. The rest (30.8%) chose verdict B, which acquits the defendant, a cautious decision given the context of the scenario. In the treatment group, the proportion of judges who chose to acquit increased to 55.8%. The difference is statistically significant at .05 level in a chi-test. And even after controlling for demographic variables of the participants in a logit regression, the effect is also significant at .05 level. (See setting (3) of Table 1.5.)

The sentences the judges gave were marginally different (at .1 level) between the control and the treatment groups. The means are 1.988 years and 3.378 years respectively. ($t = -1.860$, $p = 0.0667$. Table 1.7) An OLS regression further controlling for demographic variables shows that the effect is significant at .05 level (setting (4) in Table 1.5). This means that, keeping the other factors constant, judges in the treatment group were more lenient to the defendant who killed an innocent to make his nephew and himself survive.

We also recorded the confidence the judges had in their decisions. Unlike in Study 1, we do not find any significant difference between judges' confidence levels in Study 2. The confidence level is 62.30% in the treatment group, and 56.25% in the control group. ($t = 1.234$, $p = 0.2205$. Table 1.12 setting (2))

Confounding factor

There is, however, a confounding factor that could be an alternative explanation of our results. We appended a formal jargon ("anticipated possibility") in the treatment condition, and expected it adding no additional information to the plain language used in both the control and the treatment

condition. Nevertheless, though the legalese did not change the substance, it apparently increased the number of the arguments and the length of wordings in favor of the defendant. This might have increased the strength of the arguments, and shifted some judges' decision.

One fact that alleviates such a concern is that we do not see any significant increment of judges' confidence level in the treatment condition, contrary to what we would expect if the strength of the arguments was the factor that drove the shifting result. But this evidence does not conclusively settle the confounding problem, as we did not directly observe judges' psychological process. This problem leads us to Study 3, which tests a formal context on more abstract level.

V. Study 3: Priming

Study 2 employed legalese to decorate legal arguments and make them more formalistic. A formal context created by the use of legalese and terminologies alleviated judges from their emotional responses, and hence changed their opinions. In Study 3, we explored the effect of a formal context at a more abstract/cognitive level: we used a priming technique to directly induce a change of decisions. We propose that a formal context could prime judges into a deliberation mindset, activate the controlled process of reasoning, and bring about different decision results.

Setup

Study 3 tried to prime judges into a formal context by having them write down their answer to a legal question: in the treatment condition, a legal definition question ostensibly testing judges' legal knowledge was asked before judges made any substantial decisions—the questions was relatively difficult, and required considerable effort to correctly answer; it was irrelevant to the legal issue they were about to decide, however. We hypothesized that a formal context can activate the controlled cognitive process.

Psychological studies have suggested that judgments and attitude formation directly correlate with "the ease in which instances or associations could be brought to mind". (Tversky & Kahneman, 1973) Recent researchers have found that stimuli in the environment prime or activate content in memory, making related constructs more accessible and doing so even outside conscious awareness. For example, there is evidence that people who were assigned to vote in schools were more likely to support a school funding initiative. (Berger et al., 2008) The priming method is widely used in psychology to prime people into a different cognitive process. For example, when subjects are primed into a happy (Greene and Haidt, 2002; Valdesolo and DeSteno, 2006), cognitive fluent (Greene, 2008), reflective (Paxton et al., 2011), or less stressful (Youssef, et al., 2012) state of mind, they tend to make less impulsive/ more utilitarian decisions in moral dilemmas. In our priming setting, by making legal issues more salient in judges' mind (Hastie and Park, 1986; Berger et al., 2008), the treatment may shift their considerations from their moral intuitions to legal techniques in the subsequent decision making section. Moreover, a legal setting, which is supposed to be a rational inquiry, may elicit less intense emotional reactions relative to a daily scenario. Coming back to our general theory, a priming effect is a special case of the dual process/two systems theory (Greene et al., 2004; Kahneman, 2002), in which the priming treatment functions as a tool to activate the controlled and conscious process, and to curb the automatic and unconscious one, leads to more cautious decisions.

Case scenario

In case 3, a husband bequeathed his estate to his mistress rather than his wife, despite the devotion and care that his wife had given to him during his illness. The wife protested this result, and the judges were to decide on the validity of the will with regards to this bequest. The social norm in China would denounce the mistress, and the circumstances of the wife could cause the judges to

render a sympathetic decision; but as autonomy is one of the fundamental principles of law, the law in general respects people's free will, which facilitates the freedom of exchange and the disposal of property. The current Chinese family law is unsettled regarding this issue. On the one hand, the Law of Succession of China stipulates: "[A] citizen may, by making a will, donate his personal property to the state or a collective, or bequeath it to persons other than the statutory successors." (article 16) On the other, the law also authorizes judges to invalidate a will that conflicts with "social ethics" (General Principles of the Civil Law of China, article 7).

Method

We tried to see whether immersion into a legal context changes judges' decision patterns. In the treatment condition, judges were asked to explain a sophisticated legal term before they decided the case. The term was irrelevant to the case, and it was framed as a question testing their "legal knowledge". It reads:

Before proceeding to decision, let's do some exercise and test your legal knowledge: What is "wu quan xing wei wu yin xing"? please explain.

The phrase "wu quan xing wei wu yin xing" is presented in Chinese in the experiment. The terminology means "abstraction principle of real (property) conveyance", which denotes a German law principle of "abstraktionsprinzip". In general, it refers to the rule that a defect in an obligatory contract will not invalidate the contract on ownership transfer. (But what is an "obligatory contract", and what is a "contract on ownership transfer"? These issues relate to the complex

German civil law edifice. For a brief introduction, see Foster and Sule 2010, pp 414-416). It is often noted as "one of the most complicated and most important principles of German private law". (Foster and Sule, 2010, p414) Transplanting its legal scholarship and legal system from German law, China teaches this concept in legal education, although the principle is not stipulated in law, nor does it apply to practice in many circumstances. (Ge, 2007) Yet Chinese scholars and students often use the terminology, and see its underlying intricate legal principles as an example of "the beauty and precision of German civil law." (Wang, 1997; Wang, 1998; Sun, 2001)

The prediction for this experiment was that as legal context makes the judges deliberate, the activation of the controlled cognitive process would influence the subsequent decision making section, and thus induce less impulsive decisions.

Result

The results were consistent with the theory. As shown in Table 1.8, in the control group, 45.9% (17 out of 37) judges chose verdict A, a decision that respected the free will of the testator. The other judges chose verdict B or verdict C, which restrained the plaintiff (a "mistress" usually denounced by Chinese social norms) from getting her payment. In the treatment group, the proportion of judges that chose to respect the will increases to 67.5% (27/40). This difference is marginally significant at .1 level in a chi-test. And after controlling for demographic variables of the participants in a logit regression, the effect is significant at .05 level. (See setting (5) of Table 1.5.)

Shown in Table 1.9, the amounts of payment that the judges granted the plaintiff were also marginally different (at .1 level) in the control group and the treatment group. The means were 20270 Yuan and 28250 Yuan respectively. ($t=1.8835$, $p = 0.0635$) An OLS regression (setting (6)

in Table 1.5), further controlling for demographic variables, shows that the effect is significant at .05 level.

We also recorded the confidence the judges had in their decisions, but we did not find any significant difference between judges' confidence levels. The confidence level is 66.45% in the treatment, and 66.41% in the control. (t= 1.234, p = 0.2205)

VI. Study 4: Publicity

Setup

Our fourth experiment (Study 4) tried to replicate our results in another domain: it used a different kind of formality, the publicity of the judicial opinion, as stimulus. The hypothesis was that judges would deliberate more carefully when their decisions were to be made public. The underlying theory of this treatment condition is well documented. The publication of judicial opinions provides an opportunity to evaluate and scrutinize a decision. Judges who make their decisions public have greater incentives to take seriously the issues that confront them in a given controversy. There is antidotal evidence stating that unpublished opinions in general have a lower quality. (Pether, 2004)

Although the theory that publication leads to better decisions has been will articulated, the proposition had not yet been tested empirically. We used Study 4 to examine this theory. In the treatment condition, judges were told that their decisions and reasoning will be anonymously presented to, and discussed by, law students in legal writing class at Peking University Law School. For the control group, none of this information was mentioned and the judges believed that they were participating a study about their personal opinion on a legal issue. We predicted that the treatment stimuli would make judges believe their decisions were to be published, and hence discussed and evaluated by certain law students. The treatment conditions should thus make them

deliberate more carefully, and accordingly change their verdicts to a utilitarian direction. It is worth noting that due to the need of privacy protection, the judges were told that their decisions would only be published anonymously, and we did not record any information that could fully reveal their identities. This means the only thing at stake for them here, if any, was their pride as a group to be known as "judges from Zhejiang province". This was certainly not a perfect design for "publicity", and may have significantly reduced the strength of the stimuli.

A last note about the experimental setup is that we did not record judges' confidence in their decisions in this experiment, as we thought it was inappropriate to ask their confidence level when they were told the decisions would go public. We also worried that judges would overstate their confidence level, which could render inaccurate results.

Case scenario

In the case 4 scenario, a good Samaritan risked his own safety to save a drowning boy. The boy was rescued, but this good Samaritan unfortunately drowned .The father of the dead man sued the insurance company for compensation. The issue before the judge falls into a dilemma created by public policy and the law of contract. On the one hand, the boy would have been dead without the altruistic behavior of this good Samaritan. On the other hand, however, the insurance contract clearly exempted the insurance company where the insured's death was caused by self-induced danger. Is the insurance company liable for this accident? A sympathetic impulse may let the judges levy more obligation on the insurance company to compensate the loss of the victim, while a cautious decision would stick strictly to the contract (in the sense that enforcing the freedom of contract can maximize social utility).

Result

The experiment did not go as we had expected. We did not find any influence rendered by publicity. Judges in the control and the treatment groups granted similar verdicts (Table 1.10), and similar damage awards across groups (Table 1.11). (The sample number is larger in the verdict than in the sentence section, since four participants only gave verdict but did not render a certain amount of damage award.)

We are not sure what factors have influenced our results. The lack of finding in this setting might be due to a weak stimulus problem. As stated above, the stimulus, devised with some innate defects, may not have provided enough incentive for the judges to deliberate carefully. If the treatment condition did provide enough strength, however, the result suggests that publicity may not be as effective in framing judges' cognitive status as the other formalities. Nevertheless, we cannot draw any conclusion from the current results. More studies need to be done to ascertain the effect of publicity.

VII. Discussion

We discuss the internal and external validity of the experiments, and their implications.

In our study, we employed standard cognitive psychological research methods, which have proven successful in identifying decision-making patterns that people use in making real decisions. We do not claim that formality is mainly devised to make people cautious. Rather, we contend that some types of formalities share a similar cautionary function. We used the structure of judicial decision making, legalese and terminology, formalistic legal context, and the publicity of legal opinion as proxies for the broader concept of formality. This is not a perfect strategy, as formality includes a wide range of procedures, symbols, and forms which are far more complicated than we could cover in a single study. Nevertheless, we do think our experiments captured some

characteristics of formality in law. For one thing, the formalities we tested are common if not pervasive in the law, especially in the judicial decision-making setting. For another, the dual process theory that underlies the experimental designs demonstrates a strong predictive ability, which makes us believe that its explanatory power can go beyond this paper and extend to other scenarios with other types of formalities.

Our study solely focuses on the judicial setting. This is primarily because the judicial process is the place where formality is most prevalent. Nevertheless, it is easy to find formalities in other circumstances. And even in the judicial setting, there are other participants who are subject to the influence of formalities, e.g., litigants, lawyers, jurors, and the public. Yet we have good reasons to start our study with judges: first, judges are pivotal actors in the judicial process, which makes them practically important and become the focus of scholarly attention. Second, judges are professionals who are immersed in formalistic contexts every day. They are familiar with all kinds of formalities, they are used to dealing with them, and they should be the least affected by them. On a methodological level, if formalities have an impact on professional judges, we can well suspect that formalities also have a cautionary effect on other parties, who are more alien to the scenarios.

Our study shows the impact of formality in these experimental settings, but it does not conclusively demonstrate that this effect actually influences judges in the courtroom. Judges' actual decisions differ from those they made in response to our experiments in several important ways. In the courtroom, judges have a greater incentive to make accurate judgements, which likely makes them consider actual cases more carefully. In their jobs, judges consider more detailed and complicated factual materials, and have more time to deliberate before reaching a conclusion. These factors all contribute toward making them prudent and inducing them to deliberate. Moreover, not all legal

cases turn on issues regarding strong moral conflicts as do the cases in our experiments. The majority of cases are straightforward, and formality should not influence their outcomes. These differences suggest that we should be cautious in interpreting our results.

Nevertheless, counter-arguments show our results hold outside the laboratory, and have real world implications for judicial decision making. First, the experimental setting should suggest a bias against our findings. The artificiality of our experiment should have made it easier to ignore emotionally powerful but legally irrelevant material. A real-world judgment, on the other hand, has enormous impact on the litigants, and often on political and other important matters, which goes beyond the instant case. In the experiment, little was at stake. There were no actual litigants, and the judges risked nothing except, perhaps, their pride. Moreover, since the judges knew they were being observed, that should have made them particularly legalistic.

Second, our study focuses on close cases, which seems to narrow the applicability of our findings. We do not, however, need a theory to explain everything. Having certain explanatory power over judicial decisions in close cases is good enough as a theoretical contribution. Moreover, close cases are exactly where professional focus lies. We see this from the long debate between legal realism and legal formalism, and from today's empirical study of judicial behaviors. In fact, major scholarly attention is mostly attracted to close case scenarios, although these cases only occupy a small proportion of all disputes.

As for academics, our study provides a theoretical framework which can account for real world factors. By showing the balance between the two cognitive processes/systems, and how formality changes this balance, it offers a model for us to better understand other characteristic features of the judicial process. From our study, we can reasonably predict that some other procedural features would also lead to deliberative decisions, including, for instance, panel and chamber discussions,

and the adversarial system. Aside from this, our study contributes to two strands of literature in legal studies. First, legal scholars have long invoked the idea that formality makes pertinent parties more cautious, but they have never tested it seriously. This study sheds new light on this persistent question, with empirical evidence. Second, previous studies have shown that moral decisions are associated with judges' ideology and cost-benefit weighing. Our study provides a new angle, showing that cognitive factors can also shape judges' reasoning, and, more importantly, that the direction of the influence is predictable.

We hesitate to draw policy conclusions until more studies have replicated and refined our results. Taken at face value, our results would argue that formality can function as a de-biasing mechanism in judicial processes, as it curbs emotional impulses and promotes more thoughtful judgments. A potential application lies in the jury decision setting. It is well known that jury verdicts are less consistent and more emotional than judges' decisions. According to our study, requiring jurors to write down their reasons before they discuss case issues can be a helpful modification to trial by jury. And consistent with common practices, these written reasons should not be revealed in any event; the mere requirement of reason-writing is sufficient to accomplish our goal of making jurors deliberate.

A caveat here is that formality has its costs. It usually slows down a process. And it may conflict with other social values. For example, the requirement of consideration increases transaction costs. Reason-giving in judicial decisions can slow down dockets, induce insincere opinions, and evoke criticism and distrust. (Cohen, 2015) The use of legalese increases the difficulty of learning the law, makes lay people less able to litigate by themselves, and increases the costs of intermediaries (lawyers). The trappings of the court makes the judiciary formidable rather than agreeable, which

is contrary to many democratic values. These all should be taken into account when we evaluate the desirability of formality.

Conclusion

Formalities are a special feature of the law that has long attracted scholars' attention. As early as 1941, Lon Fuller has proposed that there is cautionary function of formalities. And by mentioning caution, he unintentionally guides legal studies into the field of cognition and psychology. In today's psychological terms, a cautionary function suggests that formality may activate the deliberation center (System 2) of human cognitive process, which can lead to a set of interesting effects. What we have done in this study is to combine these two separate but relevant literatures, develop Fuller's insight with more elaborated psychological theories, and test it with more rigorous methodologies. We find that formalities do have an impact. Even for well-trained professional judges, some certain formalities can still curb their emotional impulse, and make them decide in a more analytical way. In specific, we find such an effect from the sequence of reasoning, the use of legalese, and the priming of a legal context. Our study joins the growing literature on judicial behaviors. Similar to many other studies that employ experiments on judges, we also find that judges can be affected and manipulated. But formalities in our study are better to be seen as procedures that help judges restrain their cognitive defects and achieve unbiased judgments.

References

Berger, J., Meredith, M. and Wheeler, S.C., 2008. Contextual priming: Where people vote affects how they vote. Proceedings of the National Academy of Sciences, 105(26), pp.8846-8849.

Chase, O.G. & Thong, J. 2012, Judging Judges: The Effect of Courtroom Ceremony on Participant Evaluation of Process Fairness-related Factors, Yale Journal of Law & the Humanities, vol. 24, no. 1, pp. 221.

Cohen, M., 2015. When Judges Have Reasons Not to Give Reasons: A Comparative Law Approach. Washington and Lee Law Review, 72.

Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J. & Keysar, B. 2014, Your Morals Depend on Language, PloS one, vol. 9, no. 4, pp. e94842.

Danziger, S., Levav, J., Avnaim-Pesso, L. & Kahneman, D. 2011, Extraneous Factors in Judicial Decisions, Proceedings of the National Academy of Sciences of the United States of America, vol. 108, no. 17, pp. 6889-6892.

Epstein, L., Landes, W.M. and Posner, R.A., 2013. The behavior of federal judges. Harvard University Press.

Fischer, J.M. 1992, The Trolley and the Sorites, Yale Journal of Law & the Humanities, vol. 4, no. 1, pp. 105.

Fletcher, G.P., 2007. The Grammar of Criminal Law: American, Comparative, and International, vol.1: Foundations. Oxford University Press.

Foot P.,1985. The Problem of Abortion and the Doctrine of Double Effect, Oxford Review, 19-33.

Foster, N.G. and Sule, S., 2010. German legal system and laws. Oxford University Press, USA.

Fuller, L., 1941. Consideration and Form. Columbia Law Review, vol.41, no.5, pp. 799-824.

Garner, B.A., 2014, Black's Law Dictionary, 10th ed, West Publishing, St.Paul.

Ge, Yunsong. 2007, Phoenix in Legend: A Discuss on the Transferring Mode of Real Right in Chinese Property Law (Wu Quan Xing Wei: Chuan Shuo Zhong De Bu Si Niao). Journal of East China University of Political Science and Law (Chinese), no. 6, pp. 102-118.

Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M. and Cohen, J.D., 2001. An fMRI investigation of emotional engagement in moral judgment. Science, 293(5537), pp.2105-2108.

Greene, J.D., & Haidt, J. 2002, How (and where) does moral judgment work?, Elsevier Ltd, LONDON.

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M. and Cohen, J.D., 2004. The neural bases of cognitive conflict and control in moral judgment.Neuron, 44(2), pp.389-400.

Greene, J.D., Morelli, S.A., Lowenberg, K., Nystrom, L.E. & Cohen, J.D. 2008, Cognitive Load Selectively Interferes with Utilitarian Moral Judgment, Cognition, vol. 107, no. 3, pp. 1144-1154.

Greene, J.D., 2009. The Cognitive Neuroscience of Moral Judgment. The Cognitive Neurosciences, 4, pp.1-48.

Hastie, R. and Park, B., 1986. The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. Psychological review, 93(3), p.258.

Haidt, J., 2001. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychological review, 108(4), p.814.

Kahan, D.M. 2013, Ideology, Motivated Reasoning, and Cognitive Reflection, Judgment and Decision Making, vol. 8, no. 4, pp. 407-424.

Kahan, D.M., Hoffman, D.A., Evans, D., Devins, N., Lucci, E.A. and Cheng, K., 2016. 'Ideology' or 'Situation Sense'? An Experimental Investigation of Motivated Reasoning and Professional Judgment. University of Pennsylvania Law Review, 64, pp.2015-26.

Kahneman, D., 2011. Thinking, fast and slow. Macmillan.

Keysar, B., Hayakawa, S.L. & An, S.G. 2012, The Foreign-Language Effect: Thinking in a Foreign Tongue Reduces Decision Biases, Psychological Science, vol. 23, no. 6, pp. 661-668.

Kennedy, Duncan. 1998. A Critique of Adjudication (fin de siècle). Harvard University Press.

Kull, A. 1992, Reconsidering Gratuitous Promises, The Journal of Legal Studies, vol. 21, no. 1, pp. 39-65.

Kunda, Z. 1990, The Case for Motivated Reasoning, Psychological Bulletin, vol. 108, no. 3, pp. 480-498.

Leiter, Brian. 2003. American Legal Realism. In W. Edmundson and M. Golding eds., The Blackwell Guide to Philosophy of Law and Legal Theory. Oxford: Blackwell.

Lerner, J.S., Goldberg, J.H. & Tetlock, P.E. 1998, Sober Second Thought: The Effects of Accountability, Anger, and Authoritarianism on Attributions of Responsibility, Personality and Social Psychology Bulletin, vol. 24, no. 6, pp. 563-574.

Li, Lizhong & Liu, Daihua, 1999, A Study on Anticipated Possibility (Qi Dai Ke Neng Xing Yan Jiu), Peking University Law Journal (Chinese), vol.1, pp.31-39.

Liberman, N. & Trope, Y. 2008, The Psychology of Transcending the Here and Now, Science, vol. 322, no. 5905, pp. 1201-1205.

Lieberman, M.D., Eisenberger, N.I., Crockett, M.J., Tom, S.M., Pfeifer, J.H. & Way, B.M. 2007, Putting Feelings into Words: Affect Labeling Disrupts Amygdala Activity in Response to Affective Stimuli, Psychological Science, vol. 18, no. 5, pp. 421-428.

Liu, Zhong, 2003, A Study on the Ages of Judges Coming into Office (Guan Yu Fa Guan De Xuan Ren Nian Ling), Journal of Comparative Law (Chinese), no.3, pp.90-97.

Llewellyn, Karl N. 1940. On Reading and Using the Newer Jurisprudence. Columbia Law Review 40:581-614.

Merrill, C.M. 1981, Could Judges Deliver More Justice If They Wrote Fewer Opinions.?, Judicature, vol. 64, no. 10, pp. 435.

Norton, M.I., Vandello, J.A. & Darley, J.M. 2004, Casuistry and Social Category Bias, Journal of Personality and Social Psychology, vol. 87, no. 6, pp. 817-831.

Paxton, J., Ungar, L. & Greene, J. 2012, Reflection and Reasoning in Moral Judgment, COGNITIVE SCIENCE, vol. 36, no. 1, pp. 163-177.

Pether, P. 2004, Inequitable Injunctions: The Scandal of Private Judging in the U.S. Courts, Stanford Law Review, vol. 56, no. 6, pp. 1435-1579.

Posner, E.A. 1996, Norms, Formalities, and the Statute of Frauds: A Comment, University of Pennsylvania Law Review,vol. 144, no. 5, pp. 1971-1986.

Posner, R.A. 1993, What Do Judges and Justices Maximize? (The Same Thing Everybody Else Does), Supreme Court Economic Review, vol. 3, pp. 1-41.

Posner, R.A. 1995, Judges' Writing Styles (And Do They Matter?), The University of Chicago Law Review, vol. 62, no. 4, pp. 1421-1449.

Posner, R.A., 2010. How judges think. Harvard University Press.

Qu, Xuewu, 2005, Capital Offense, Death Penalty and Anticipated Possibilities: A Legal Analysis on Women Who Killed Their Abusive Husbands out of Domestic Violence (Si Zui, Si Xing Yu Qi Dai Ke Neng Xing: Ji Yu Shou Nue Nv Xing Sha Ren Ming An De Fa Li Fen Xi), Global Law Review (Chinese), vol.1, pp.58-70.

Schwitzgebel, Eric, and Fiery Cushman. 2015. Philosophers' biased judgments persist despite training, expertise and reflection. Cognition 141:127-137.

Simon, Dan. 1998. A Psychological Model of Judicial Decision Making. Rutgers Law Review 30:1-142.

Spamann, Holger, and Lars Klöhn. 2016. Justice is Less Blind, and Less Legalistic, Than We Thought: Evidence From an Experiment with Real Judges. Working paper, Harvard University.

Spellman, Barbara, and Frederick Schauer. 2012. Legal Reasoning, in K. J. Holyoak & R. G. Morrison ed., The Oxford Handbook of Thinking and Reasoning, 2nd ed. Oxford University Press.
Stinchcombe, A.L. 2001, When Formality Works: Authority and Abstraction in Law and Organizations, University of Chicago Press, pp. ix-ix.

Summer, R.S. 1992, The Formal Character of Law, The Cambridge Law Journal, vol. 51, no. 2, pp. 242-262.

Summers, R.S. 1997, How Law Is Formal and Why It Matters, Cornell Law Review, vol. 82, no. 5, pp. 1165.

Sun, Xianzhong, 2001, More on the Theories of Juristic Act of Real Right (Zai Tan Wu Quan Xing Wei Li Lun), Social Sciences in China (Chinese), vol.5, pp.125-135.

Thomson, J.J., 1985. Double effect, triple effect and the trolley problem: Squaring the circle in looping cases. Yale Law Journal, 94(6), pp.1395-1415.

Trope, Y. & Liberman, N. 2010, Construal-Level Theory of Psychological Distance, Psychological Review, vol. 117, no. 2, pp. 440-463.

Tversky, A. and Kahneman, D., 1973. Availability: A heuristic for judging frequency and probability. Cognitive psychology, 5(2), pp.207-232.

Valdesolo, P. & DeSteno, D. 2006, Manipulations of Emotional Context Shape Moral Judgment, Psychological Science, vol. 17, no. 6, pp. 476-477.

Waits, K. 1983, Values, Intuitions, and Opinion writing: The Judicial Process and State Court Jurisdiction, University of Illinois Law Review, vol. 1983, no. 4, pp. 917.

Wang, Liming, 1997, Several Problems on Juristic Act of Real Rights (Wu Quan Xing Wei Ruo Gan Wen Ti Tan Tao), Chinese Legal Science (Chinese), no.3, pp.58-70.

Wang, Zejian, 1998, A Study on the Abstraction Principle of Real (Property) Conveyance (Wu Quan Xing Wei Wu Yin Xing Li Lun Zhi Tan Tao), Civil Law Theories and Case Studies (Min Fa Xue Shuo Yu Pan Li Yan Jiu), vol.1, pp.256-257.

Wistrich, A.J., Rachlinski, J.J. and Guthrie, C., 2014. Heart versus head: Do judges follow the law of follow their feelings. Tex. L. Rev., 93, p.855.

Xiao, Qiming & Chen Enze.2009, Facts on the Abilities of Present Judges Dissolving Disputes and Its Countermeasures (Dang Qian Fa Guan Jiu Fen Hua Jie Neng Li De Xian Zhuang Ji Dui Ce), Law Review, no. 2, pp. 138-143.

Youssef, F.F., Dookeeram, K., Basdeo, V., Francis, E., Doman, M., Mamed, D., Maloo, S., Degannes, J., Dobo, L., Ditshotlo, P. & Legall, G. 2012, Stress Alters Personal Moral Decision Making, Psychoneuroendocrinology, vol. 37, no. 4, pp. 491-498.

Zhang, Mingkai, 2009, A Summary of the Theory of Anticipated Possibility (Qi Dai Ke Neng Xing Li Lun De Shu Li), Chinese Journal of Law (Chinese), no. 1, pp. 60-77.

Case cited

R v Dudley and Stephens, 14 QBD 273 DC (1884).

Appendix 1.1: Case Materials

Case 1-Structure: Impoverished Tortfeasor

The plaintiff: Yue Feng, born on August 9, 1974, a senior associate of the Beijing office of xxx LLP, a renowned international law firm with its headquarter based in New York.

The defendant: Wang Zhiqiang, born on February 2, 1982, an administrative staff of xxx company Zhang Jiayi was Wang Zhiqiang's wife and was pregnant. At 17:30 pm on May 18, 2014, Zhang Jiayi's water unexpectedly broke. Wang Zhiqiang immediately took Zhang Jiayi to the nearby hospital with his motorcycle. On his way to the hospital, Wang Zhiqiang accidentally hit Yue Feng and injured him. The police officer came and issued a report of road traffic accident where Wang Zhiqiang was found at fault and therefore fully liable for the accident. Yue Feng's left arm, shoulders and ribs suffered varying degrees of injuries and he stayed in hospital for 56 days for medical treatment. The testimony of Yue Feng's doctor stated that Yue Feng was not fully recovered yet at the time he left the hospital but he insisted on leaving to attend to his work.

The plaintiff Yue Feng sued Wang Zhiqiang for medical expense and loss in income. The medical expense was 15,658.6 yuan and undisputed. According to the employment contract that Yue Feng had with the law firm and other relevant evidences he provided, his annual salary was 230,000 U.S. dollars. Thus, Yue Feng claimed a loss in income of 229,369 yuan. (calculated by 56 days, 230,000/365*56*6.5)

Wang Zhiqiang claimed that, although he was fully liable for this accident, he has the following excuses . It was because Wang Zhiqiang was so anxious to take his wife to hospital that he accidentally hit Yue Feng. Wang Zhiqiang has already apologized to Yue Feng and promised to make his best efforts to compensate Yue Feng's medical expense. However, the amount of loss of income that Yue Feng claimed is unreasonably large and the court should make a reduction on the

claimed amount as it considers appropriate. Wang Zhiqiang also submitted evidences proving that his monthly income is 6,200 yuan and has no other valuable assets besides the apartment he is currently living in, for which he has to pay a monthly mortgage for the next 6 years. Besides, Wang Zhiqiang has to support his parents who live in countryside and have no source of income. Now he has a son to raise after Zhang Jiayi delivered the baby. Zhang Jiayi is working in the marketing department of a local travel agency and has a monthly income of 4,500 yuan. Wang Zhiqiang alleged that the court should consider the economic situation of his family and reduce the amount of plaintiff's loss in income claim as it deems appropriate.

[Treatment:] Before writing down your decision, could you first write down your reasoning about the case on this page? Your decision should be written on the next page.

1. What is the legal basis of the plaintiff's claim?

2. What is the legal basis of the defendant's claim?

[Next page]

If you were the judge of this case, how would you decide?

A. Holding for the plaintiff, the defendant shall pay a compensation of 229,369 Yuan to the plaintiff.

B. Holding for the defendant. Adjust the compensation to _____ Yuan.

C. Other decision:_____

[Next page]

[Control:] Please write down your reasoning here.

Case 2- Legalese: Kill to Survive

Fan A is the chairman of the board of directors of a enterprise group in Nanning. In late 2014, Fan A and his nephew, Fan B were kidnapped by Liu, Yue, Chen, who were equipped with guns and knives, and were subsequently taken to a residence building in Cuiping District in the Town of Zhao Chang.

Liu, Yue and Chen threatened Fan A and Fan B with knives and guns and made Fan A agree to pay a ransom of 100 million yuan by March, 2016. To prevent Fan from reporting the abduction to the local police and to make sure he will pay the ransom on time, Liu, Yue and Chen beat Fan heavily and forced him, together with Liu, to strangle Jia Gu to death. Jia Gu was a villager living in town who was also kidnapped by Liu and Yue. Liu recorded the entire process of murdering by video. After Fan A was freed, he went to the local police station and reported the crime

Fan A was prosecuted of homicide by the local prosecutor.

[Next page]

This case is now presented to your court. There seems no specific rule in the law that governs this specific situation, but there are some theoretical debates in criminal law principle. **Control & [Treatment]:** Some people think that, ["qi dai ke neng xing" is an element of crime,] when the suspect cannot be fairly expected to act otherwise under the circumstances, he should not be held liable for his act.

But [,as the theory of "qi dai ke neng xing" is not completely recognized as a lawful defense,] other people think this should not be an excuse to acquit murder.

If you are the judge in charge of this case, will you convict the defendant murder?

Yes/No

[**Next page, asking**:] What sentence do you give? _____

[**Next page**]

Please write down your reasoning here:

Case 3- Priming: Mistress

Huang Yongbin and Jiang Xiaofang married in 1969 and later separated due to lack of affection. In 2001, Huang Yongbin met Zhang Xueyin and has been living with her since then. Beginning 2006, Huang Yongbin was hospitalized due to liver cancer. On April 18, 2007, Huang made a will where he bequeathed his share of the community property to Zhang Xueyin, which amounts to 40,000 yuan of total 80,000 deposits in bank account, and had this will duly executed. Huang died in June, 2007. After Huang's death, Zhang Xueying required Jiang Xiaofang to transfer the 40,000 yuan to her but was refused. Zhang brought this litigation against Jiang.

The defendant, Jiang Xiaofang claimed that she has been taking care of Huang Yongbin since his hospitalization and till his death. Zhang had extramarital affairs with Huang and cohabited with him during the existence of his marriage with Jiang, and therefore Huang's bequeathing of his estate of 40,000 yuan to Zhang is invalid for violating the principle of public policy and good customs.

The plaintiff, Zhang Xueying alleged that extramarital affairs and cohabitation does not necessarily cause the will invalid. The nature of extramarital relationship and the validity of bequest conduct are subject to two different areas of law. The actual intent of the testator should be respected and the will is valid.

[Next page]

[Treatment] Before proceeding to decision, let's do some exercise and test your legal knowledge: What is "wu quan xing wei wu yin xing", please explain.

[Next page]

If you are the judge in charge of this case, what is your decision?

A. Holding for the plaintiff and ordering the defendant, Jiang Xiaofang to transfer 40,000 yuan to Zhang Xueying

B. Holding for the plaintiff and ordering the defendant, Jiang Xiaofang to transfer ( ) yuan to Zhang Xueying

C. Dismissing the case

D. Other decision:_____

Case 4- Publicity: Reckless Good Man

The plaintiff: Zhang Zhenzhong, born on September 13, 1958, father of the insured

The insured: Zhang Quan, born on October 22, 1986, a clerk in the local branch of China Merchants Bank

The defendant: XX Insurance Company, a local insurance company in good standing

In the afternoon of December 13, 2014, Zhang Quan was on his way home. While he was walking along xx river, a woman (Zhang Guilan, born in 1982, a villager living near by) ran to Zhang Quan, trying to stop him, and claimed that her son (Jiang xx, born in 2005) just fell into the river and begged Zhang Quan to save him. Zhang Quan immediately jumped into water to rescue the boy. After a tough battle, Zhang Quan managed to push the boy to shore but was swept away by rushing water. The local police found Zhang Quan's body the next day and determined that his death was caused by drowning.

The defendant insurance company provided evidences of relevant terms of the insurance contract: (a) Article 4 of the insurance contract stipulates the exemptions of liability, which provides that (i) the insurance company shall not be liable for the death or disability of the insured where such death or disability is caused by the following reasons: (1) the insured intentionally performed certain conducts that caused unnecessary risks or danger to his own health or safety. (b) Article 3 of the insurance contract stipulates the scope of the liability of the insurance company, which provides that within the term of this contract, the insurance company shall compensate the insured according to following terms where any none self-induced accident caused the death or disability of the insured.

The plaintiff Zhang Zhenzhong claimed that Zhang Quan's death fell into the circumstances of Article 4 of the insurance contract where unintentional accident causes death or disability and

therefore the insurance company should compensate the dead according to the agreed provisions. Besides, Zhang Quan's scarifying his own life to save others demonstrated the traditional Chinese virtues of helpfulness and selflessness, and should be strongly encouraged. Zhang Quan's heroic behavior has been confirmed by the government authority as a Good Samaritan act. According to the doctrine of public policy and good customs in Chinese civil law, making compensations to Zhang Quan, which does not violate any mandatory law or contradict public policy, will reflect the values of fairness and justice. Finally, the provision of the insurance contract that excludes the Good Samaritan acts is a boilerplate clause which intended to eliminate the liabilities of the insurance company and increase the obligations of the insured. Such exculpatory clause is void and unenforceable according to contract law.

The insurance company claimed that, for the following reasons, Zhang Quan's death does not fall into the scope of compensable situations provided by the insurance contract and therefore the insurance company is not liable for his death. (i) At the time of the incident, Zhang Quan had a relatively accurate judgment on whether he would be able to save the child and swim back to shore safely. On the day the tragedy happens, the river has risen significantly due to the continuous rainfall in the past few days; the waves were huge and the whether was cold and windy. There was also a warning sign on the scene of the accident which reads "deep water, no swimming". Knowing that his conduct would create a significant risk to his life and safety, Zhang Quan still jumped into the river. In other words, Zhang Quan's conduct wass intentional and voluntary and as a result put his life in grave danger. The possibility of drowning caused by such reckless conduct was well foreseeable to Zhang Quan. Thus, this incident does not fall into the definition of "unintentional accident" stipulated by the insurance contract. (ii) The insurance company only compensates for unforeseeable and unintentional accident, rather than dangerous conducts that are intentionally

performed by the insured. An exclusion of the insured's intentional dangerous behaviors from the compensable circumstances is the essence of risk control of the insurance company and the forgoing of which will defect the purpose of the insurance contract, undermine the function of commercial insurance and impose unreasonably heavy liability on the insurance company. In fact, besides the intentional dangerous behaviors, the exemption clause in the insurance contract also includes the death or disability caused by diving, parachuting, rock climbing, adventure activities, martial arts game, wrestling, stunts, racing horses, racing and other high-risk activities. (iii) As a result, that Zhang Quang's voluntarily jumped into the river with the knowledge of the great danger of such behavior caused unnecessary risks to his life safety and fell into the exemption clause of Article 4 of the insurance contract. In sum, neither the cause nor the consequence of this incident was an "unintentional accident" to Zhang Quan. Objectively, what Zhang Quan has done, though heroic, was a reckless and highly dangerous self-injurious behavior. Subjectively, Zhang Quan has foreseen, or should have foreseen the risk of his conduct. Hence, the insurance company shall not be liable for Zhang Quan's death.

[Treatment: Your decisions and reasoning in this case will be PRESENTED to and discussed by students from Peking University Law School who attend the Legal Writing class, but, of course, ANONYMOUSLY.]

If you were the judge of this case, how would you decide?

A. Holding for the plaintiff, the insurance company shall fully compensate the defendant.

B. Adjust the compensation to _____ Yuan.

C. Holding for the defendant insurance company. The insurance company has no liability to compensate the dead.

D. Other decision:_____

[Next page]

Please write down your reasoning here.

Appendix 1.2: Tables

Table 1.1: Experiment Designs

Study/Case Number	Case Name	Stimulus	Treatment & Control	Number of Participants
First round (N=127)				
Study 1 /Case 1	Impoverished Tortfeasor	Structure	Treatment: Judges reason first, and then make a decision	37
			Control: Judges decide the case first, and then reason	44
Study 2 /Case 2	Kill to survive	Legalese	Treatment: Some legalese providing zero new information than wordings in the control group is used in the treatment group's case materials	41
			Control: plain language	37
Study 3 /Case 3	Mistress	Priming	Treatment: An irrelevant legal question is asked before judges make decisions, so as to prime judges into using a legal reasoning mindset (an activation of System 2)	40
			Control: no priming question	37
Second round (N=57)				
Study 4 /Case 4	Reckless Good Man	Publicity	Treatment (publicity): Judges are told that their decisions and reasoning will be reported by a national media; their decisions and reasoning will be shown on a famous website	25
			Control: no publicity stimulus	27

Note: The sum of participants in the treatment group and the control group is smaller than the total participation number in Case 1 to Case 3 because: 1. each judge was presented with and asked to decide two cases that were randomly drawn to them from Case 1 to Case 3; 2. not all participants finished the experiment and submitted their data; 3. among those who submitted their data, some participants skipped some of the questions, so we are only able to record parts of their information.

Table 1.2: Summary Statistics of Judges' Demographics

Variable	Obs	Mean	Std.Dev.	Min	Max
age (first round)	121	32.34	2.482	26	43
age (second round)	47	32.21	2.797	29	41

	First round		Second round	
	Obs	Percentage	Obs	Percentage
education				
High school	1	0.8%		
Bachelor's	85	70.3%	30	63.8%
Master's	34	28.1%	17	36.2%
Ph.D.	1	0.8%		
gender				
Female	67	55.8%	27	57.4%
Male	53	44.2%	20	42.6%
field				
Civil court judges	81	66.9%	29	61.7%
Criminal court judges	26	21.5%	4	8.5%
Administrative court judges	6	5.0%	3	6.4%
Other	8	6.6%	11	23.4%
court level				
Local	113	93.4%	39	83.0%
Intermediate	5	4.1%	6	12.8%
High	3	2.5%	2	4.2%

Note: in the second round experiment, 5 judges (out of 52) did not fully submit their personal information. So in summary statistics, the observation number is 47.

Table 1.3: chi-test of verdicts (Structure-Case 1)

verdict	Control	Treatment	Total
Fully compensate (A)	14	24	38
Adjustment (B)	30	13	43
Total	44	37	81

Pearson $\chi 2$= 8.813; Pr=0.0030

Table 1.4: t-test of sentences (Structure-Case 1)

	Sentence Structure (yuan)
(1) Treatment	183,490
	(13,859)
(2) Control	138,402
	(14,308)
(1)- (2) diff	45,088**
	(19,951)
t-value	2.260
Pr	0.0269
Number (1)/(2)	36/38
Number	74

Table 1.5: Regression Results of Judges' Decisions (Case 1 to Case 4)

VARIABLES	(1) Logit Case 1 Verdict	(2) OLS Case 1 Sentence (Yuan)	(3) Logit Case 2 Verdict	(4) OLS Case 2 Sentence (Years)	(5) Logit Case 3 Verdict	(6) OLS Case 3 Sentence (Yuan)	(7) OLS Case 4 Sentence (Yuan)
Treatment (structure)	1.459***	66,926***					
	(0.539)	(23,054)					
Treatment (legalese)			-1.106**	-1.567**			
			(0.552)	(0.720)			
Treatment (priming)					1.363**	11,553**	
					(0.558)	(4,385)	
Treatment (publicity)							57,477
							(174,227)
constant	1.045	199,659	5.543	10.41	-2.057	7,377	1371,000
	(3.345)	(211,413)	(5.060)	(7.059)	(3.431)	(28,859)	(991,128)
controlling demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	76	80	80	81	74	77	47
R^2		0.167		0.226		0.209	0.092
Pseudo R^2	0.086		0.177		0.121		

Note: 1. Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1; 2. Demographic variables include age, gender, specialized field, court level, and education level; 3. Numbers of observation are smaller in the logit regressions as the models automatically delete observations that perfectly fail to predict.

Table 1.6: chi-test of verdicts (Legalese-Case 2)

verdict	Control	Treatment	Total
Conviction (A)	27	19	46
Acquittal (B)	12	24	36
Total	39	43	82

Pearson $\chi 2 = 5.209$; Pr = 0.022

Table 1.7: t-test of sentences (Legalese-Case 2)

	Sentence Legalese (year)
(1) Treatment	1.988
	(0.584)
(2) Control	3.378
	(0.476)
(1) - (2) diff	-1.391*
	(.747)
t-value	-1.860
Pr	0.0667
Number (1)/(2)	41/37
Number	78

Table 1.8: chi-test of verdicts (Priming-Case 3)

verdict	Control	Treatment	Total
Free will (A)	17	27	44
Social norm (B&C)	20	13	33
Total	37	40	77

Pearson $\chi 2 = 3.646$; Pr = 0.056

Table 1.9: t-test of sentences (Priming-Case 3)

	Sentence Priming (yuan)
(1) Treatment	28250
	(3159)
(2) Control	20270
	(2839)
(1) - (2) diff	7980*
	(4236)
t-value	1.8835
Pr	0.0635
Number (1)/(2)	40/37
Number	77

Table 1.10: chi-test of verdicts (Publicity-Case 4)

verdict	Control	Treatment	Total
Full Compensation (A)	18	16	34
Adjustment (B)	3	4	7
No Compensation (C)	8	5	13
Total	29	25	54

Pearson $\chi 2 = 0.6601$; Pr = 0.719

Table 1.11: t-test of sentences (Publicity-Case 4)

	Sentence Legalese (Yuan)
(1) Treatment	740909
	(93291)
(2) Control	678571
	(86121)
(1) - (2) diff	62337
	(127546)
t-value	0.2887
Pr	0.6272
Number (1)/(2)	22/28
Number	50

Table 1.12: t-tests of Judges' Confidence in Their Decisions

	Confidence Case 1- Structure (%)	Confidence Case 2- Legalese (%)	Confidence Case 3- Priming (%)
(1) Treatment	56.25	62.30	66.45
	(3.55)	(3.38)	(3.78)
(2) Control	69.38	56.25	66.41
	(3.39)	(3.55)	(3.24)
(1) - (2) diff	-13.13***	6.05	0.03
	(4.91)	(4.90)	(4.99)
t-value	-2.672	1.234	0.008
Pr	0.0091	0.2205	0.9937
Number (1)/(2)	40/44	42/40	40/39
Number	84	82	79

Note: 1. Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table 1.13: Regression Results of Judges' Confidence in Their Decisions

VARIABLES	(1) OLS Confidence (Case1)	(2) OLS Confidence (Case2)	(3) OLS Confidence (Case3)
Treatment (structure)	-12.43** (5.865)		
Treatment (priming)		2.199 (4.916)	
Treatment (legalese)			1.082 (5.144)
constant	-17.18 (53.79)	9.852 (32.36)	77.58 (50.42)
controlling demographics	Yes	Yes	Yes
Observations	80	77	81
R-squared	0.175	0.263	0.146

Note: 1. Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1;
2. Demographic variables include age, gender, specialized field, court level, and education level.

Reason Writing Reduces Emotional Judging:

Experimental Evidences from Judges in China

Abstract: Anecdotes and academic research show that emotion (judges' personal feelings) has undue influence on judging. Studies have not addressed, however, the fact that judges are decision makers who are embedded in procedures and formalities that are devised to help them to improve decision quality. In this paper, we suggest a simple debiasing procedure to curb the effect of emotion: requiring judges to write reasons before making a decision. We conducted a set of experiments on incumbent Chinese judges to test the effectiveness of this procedure. In the first study, with a 2*2 between subjects design, we explored the interaction between writing reasons and a stimulus that induces judges' negative feelings. We find that judges who were required to write reasons before they decided a case were significantly less affected by the emotional stimulus than those who directly entered the decision-making stage. At the same time, the confidence of the writing-reason-first judges in their decision were significantly lower than those who made their decisions first, according to their self-reports. This result suggests a higher degree of introspection in the judges who first wrote their reasons. A follow-up study and a placebo test also show that introspection and deliberation are channels through which the writing of reasons makes impacts.

Introduction

Judges are supposed to put aside their emotional reactions to litigants, whether they arise out of enmity or empathy. But both anecdotal evidence and academic research show otherwise. A well-known trope attributed to Judge Jerome Frank says, "Justice is what the judge ate for breakfast" (Kozinski, 1992), suggesting how trivial emotions can possibly influence a judge's decision. In a set of experiments using real judges as subjects, Wistrich et al. (2014) found that the defendants' characteristics, though irrelevant to the case context and which should not be taken into account when making a verdict or sentencing, do change judges' decisions by affecting judges' emotional status. In an experiment resembling real-world judicial decision-making, Spamann et al. (2016), also shows that legally irrelevant but emotionally salient characteristics of the defendants had an influence on judges' verdicts.

The finding that one set of rules applies to the sympathetic litigant and another to the unsympathetic litigant is surely problematic for the rule of law. Before jumping to a pessimistic conclusion and starting to search for solutions, however, it would be prudent to examine our current judicial system more carefully. Past studies have employed experiments to identify a causal relation between emotions and biased decisions in a laboratory setting. Their findings help lawyers to understand the cognitive process of judges: that is, judges and lay people are very similar in the way they make mistakes. One drawback of these studies is that they neglect that judges are decision makers who are embedded in procedures and formalities that were devised to help them improve decision quality. In the real world, some features of legal proceedings may well assist judges in correct their tendency toward errors. (Llewellyn, 1940; Kahan, 2015) In this paper, we suggest a simple, and already pervasive, debiasing procedure in the judicial setting: writing down reasons before making a decision.

As articulated by many judges and legal scholars, giving written reasons imposes a form of self-discipline on judges, which improves the quality of their decisions. (Posner, 1995; Wald, 1995; Oldfather, 2007; Cohen, 2015) Psychologists also find that giving reasons makes people decide more cautiously, as it activates system 2 of the human cognitive processes. (Paxton et al., 2012) There is evidence that shows that the need to articulate reasons effectively reduces the impact of irrelevant emotional factors (Lerner et al. 1998), and that putting negative feelings into words, or labeling negative emotions, attenuates these negative feelings. (Lieberman et al., 2007) Arising out of a similar two-system mechanism, Sieck and Yates (1997) found that subjects who engaged in a written exposition of their thought processes were less influenced by framing effects. Discussing opposite reasons helped experts reduce their anchoring biases when evaluating car prices. (Mussweiler et al., 2000)

The legal and the psychological studies both suggest that, as a procedure and formality in law, the requirement of giving reasons when making a decision urges judges to deliberate and thereby reduces their emotional biases. This theory, however, has not been tested empirically.

Our experiments are designed to test this theory. In Study 1, we use a 2*2 between subjects design to explore the interaction between reason writing and the influence of emotion. To do this, we first induce an emotional bias, using a standard treatment-control setting, where the defendant in the treatment group is described implicitly as a corrupt government official, though that corrupt behavior is irrelevant to the current case, which involves whether the defendant acted in self-defense or committed manslaughter. We randomly assign half of the judges in both the treatment and control groups to write formal reasons before they decide, and half of the judges in each group give their decisions directly. Our results show that the decisions of the judges in the two direct-decision groups diverge; the judges in the treatment group render significantly harsher decisions

on the defendant. In contrast, judges in the two writing-reason-first groups decide more consistently, which means that they are less influenced by irrelevant emotional factors. Taken together, the results suggest that giving written reasons before making a decision curbs the influence of emotion on judging.

To study the mechanism behind the results, we append another pair of treatment-control groups to Study 1, which extends Study 1 to a 2*3 design. For simplicity, we call this new pair of groups "Study 2". In the new treatment and control groups, we induce judges to think over the case before they are instructed to make a decision. To do this, we set a one minute mandatory cooling-off or deliberation period, before they can proceed to the decision page. We compare the decision result in Study 2 with that in Study 1. We find that compared to the two direct-decision groups in Study 1, the two deliberation-first groups in Study 2 decide in a more consistent manner and exhibit less emotional bias. Their decision pattern is similar to the two writing-reason-first groups in Study 1. This means that writing reasons and deliberating function similarly in reducing emotional bias. It also suggests that writing reasons has an impact because it promotes deliberation.

We further designed Study 3 as a placebo test to Studies 1 and 2. Study 3 uses a 2*3 design, which is equivalent to the combination of Studies 1 and 2. But the treatment in Study 3 is no longer an irrelevant emotion factor. Rather, it is an aggravating circumstance (a recidivist defendant) that judges should take into account according to the law. We find that reason writing and deliberation no longer reduce sentence divergences between the decisions of the treatment and the control groups, which suggests a correct application of law. The comparison of Studies 1, 2, and 3 suggests that reason writing only reduces the influence of irrelevant emotion factors, but not of legally applicable elements. It thereby confirms that reason writing increases the precision of judicial decisions.

Besides deliberation, by analyzing judges' confidence levels in their decisions, we identify introspection as another mechanism that leads to the above results. Legal theories suggest, though implicitly, that reason-giving may induce introspection, such that it restrains intuitions from dominating the decision process. This effect is known as the "it won't write" phenomenon: judges often change their initial decisions after writing down reasons. (Waits, 1983; Cohen 2015) We suspect that the confidence levels in their decisions would reflect these psychological processes: if judges become skeptical about their first instincts after introspection, or in other words, if their two cognitive processes (system 1 and system 2) are in conflict, then their confidence in their decisions should decrease. To test this hypothesis, we record judges' confidence level by asking "how much confidence do you have in your decision (0 to 100%)," after they make a decision.

Our data conforms to the prediction and shows that there is a decrease in confidence for the judges in the writing and deliberation conditions in Studies 1 and 2. This implies a conflict between intuition and reason inside the judicial mind.

The hypothesis is further confirmed by Study 3, in which judges' confidence remains unchanged in different decision procedures. While the decreasing confidence in Studies 1 and 2 arises from the fact that introspection over an erroneous intuition makes judges less certain about their decisions, there is no similarly biased emotion involved in Study 3, and the reception of the stimulus only induces a legally correct decision. Thus, as the conflict between system 1 and system 2 is absent, writing and deliberation does not change judges' confidence in their decisions.

Taken together, our experiments show that reason writing, by inducing introspection and deliberation, curbs judges' emotional biases when making decisions. These results contribute to previous experimental literature about judicial decisions. Past studies have successfully identified a battery of judges' decision biases, including the influence of emotional impulses, and many

behavioral effects (anchoring, framing, hindsight, heuristic, etc.). In this paper, however, we show that those studies may well have overlooked the complexity of real world judicial decision setting, especially those mechanisms that have been designed to curb biases. Another implication of our results points to the writing of reasons as a formal legal procedure per se. The results show a strong debiasing capability of the reason writing procedure. Legal scholars should take this idea into account when studying an adjudicate process. For example, only a small proportion of judicial decisions in the U.S. involves the writing down of formal reasons. This lack can become a source of judicial bias and arbitrariness, without judges even noticing it. Another application is in the jury setting: as emotional decision-making is a major problem with jurors, we may want to design a reason writing (but, of course, not publishing) procedure for jury trials.

The remainder of this paper is organized as follows. Section I reviews previous studies and raises our hypothesis. Section II introduces the main experiments and explains their results. Section III provides a placebo test to the main experiments. Section IV discusses the validity of our experiments and their implications.

I. Theory and Hypothesis

Judges are expected and required to put their feelings towards litigants aside when deciding a case, but research has shown that judges' emotional reactions to litigants significantly influence their decisions. Wistrich et al. (2014) used several experiments to reveal this general pattern. They provided judges with case vignettes, and varied the characteristics of the defendants. The characteristics were irrelevant to the case context and should not have been taken into account in rendering a verdict and sentencing. These characteristics did, however, have a strong impact on the judges' decisions. In an experiment resembling real-world judicial decision-making (by presenting the judges with raw materials of facts, precedents, and trial judgements), Spamann et

al. (2016) also show that legally irrelevant characteristics of a defendant can influence judges' verdicts. Judges were more likely to convict a defendant who is unsympathetic to his victim compared to a defendant who is sympathetic. In their experiment, they asked judges to give written reasons, and found their written reasons do not mention the defendant's characteristics at all, focusing instead on the precedent and other legal and policy considerations. It is worth noting that Spamann et al.'s study requested judges to write down reasons after reaching a verdict, so that judges were not able to revise their decisions. A closely relevant issue is what we want to explore: whether giving written reasons prior to making a decision reduces the influence of emotion on judging.

A. The function of giving written reason in law

Reason-giving is expected in many contexts, including in judicial processes. Lawyers have noticed that giving written reasons imposes a form of self-discipline on judges, which improves the quality of decisions. (Oldfather 2007; Cohen 2015) On the one hand, reason-giving is described as inducing introspection and restraining intuitions from dominating the decision process. This is known as the "it won't write" phenomenon. "Judges may frequently discover that the original opinion they originally had in mind 'won't write' … The discipline of a written opinion … operates as an important control on judicial arbitrariness." (Waits, 1983; Cohen 2015) This phenomenon occurs because "[T]he necessity for preparing a formal opinion assures some measure of thoughtful review of the facts in a case and of the law's bearing upon them. Snap judgments and lazy preferences for armchair theorizing . . . are somewhat minimized." (Leflar, 1961) And "[I]t is not so unusual to modulate, transfer, or even switch an originally intended rationale or result in midstream because 'it just won't write'. " (Wald 1995)

On the other hand, writing down reasons also forces and facilitates deliberation. As former U.S. appellate judge Charles Merrill (1980) noted, "[T]he very act of writing opinions reinforces the decisional process. Misconceptions or oversights may come to light in the course of articulation." Judge Posner (1995) also writes about such a subtle process:

> "[R]easoning that seemed sound 'in the head' may seem half-baked when written down, especially since the written form of an argument encourages some degree of critical detachment in the writer, who in reading what he has written will be wondering how an audience would react. Many writers have the experience of not knowing except in a general sense what they are going to write until they start writing. A link is somehow forged between the unconscious and the pen. The link is lost to the judge who does not write."

Notice here that he describes reason-writing as a trigger that switches the unconscious process to the critical reflection process.

As a result, scholars believe that reason-writing, by compelling judges to substantiate their thoughts based on deliberation and reason, improves the quality of judicial decision making, and reduces the effect of potential cognitive biases. (Guthrie et al., 2007)

B. Psychological theories of reasoning and writing

Though not yet tested rigorously, this sense of lawyers is consistent with psychological theory, and with findings in experiments using lay people as research subjects. Work in psychology and neuroscience has revealed that human judgments are often mediated by two classes of brain processes ("dual-process theory", Greene et al., 2001, 2004; Haidt, 2001). One class consists of processes that automatically alter hedonic states in response to specific types of socially relevant stimuli, and this class probably arises from earlier evolutionary development. A second class

consists of more effortful processes which underlie abstract reasoning, simulation, and cognitive control. The automatic or intuitive processes resemble what Daniel Kahneman and Amos Tversky call system 1 of thinking, which is fast, instinctive, and emotional, while the deliberative processes are similar to their system 2, which is slower, more deliberative, and more logical (Kahneman, 2011).

Psychological studies have suggested that writing reasons curbs system 1 and activates system 2. It restrains intuitions from making a judgment, and facilitates the use of the deliberative process of thinking. For example, the need to articulate reasonsfor a decision reduces the impact of irrelevant emotional factors (though the researchers attribute the effect to a stronger sense of accountability when subjects are asked to articulate reasons) (Lerner et al. 1998), but it does not always have this effect (Norton et al. 2004). Similarly, having time to reflect on the issues presented eliminates hunches driven by legally irrelevant factors (Paxton et al. 2012), although there are counter-examples (Schwitzgebel and Cushman, 2015). Moreover, putting negative feelings into words, or labeling them, can attenuate those negative feelings. (Lieberman et al., 2007) Taken together, this evidence indicates that reason writing can improve decision quality by reducing the adverse influence of emotional impulses.

Growing out of a similar two-system mechanism, writing reasons also reduces many behavioral biases. Sieck and Yates (1997) conducted experiments which found that subjects who engaged in written exposition of their thought processes were less influenced by framing effects. In another study, Mussweiler et al. (2000) found that car experts were less likely to be influenced by anchoring biases in evaluating car prices when they were induced to talk about opposing reasons. In explaining the mechanism, both studies suggested that writing encourages people to consider

more of the relevant information, which had previously been obscured by the frame, or biased by the anchor.

Although deliberation is usually beneficial to decisions involving analytical reasoning, it can be detrimental to decision making on an aesthetic task (McMackin and Slovic, 2000) and to self-satisfaction on the choice of a good (aka, realizing one's preference) (Wilson and Shooler, 1991; Wilson, et al., 1993). In these areas, intuition sometimes yields superior results, although not always. (e.g., Sieck and Yates, 1997) But as judicial decision-making is typically an analytical task, this potential weakness of deliberation is not directly relevant to the issue that concerns us in this paper.

C. Hypothesis: reason writing reduces judicial biases

Drawing from the above legal and psychological studies, we hypothesize that, as a procedure and formality in law, the requirement of giving written reasons when making a decision reduces judges' emotional biases, and thus makes judges decide in a more accurate and consistent manner. We use three studies on Chinese judges to test this hypothesis.

II. Experiments

We recruited 145 incumbent judges in Zhejiang Province who were attending short-term training classes at the Judges College in Hangzhou in September 2016. For summary statistics, see Table 2.1. Among these judges, 17 chose not to reveal their demographic information. So the summary statistics are based on the 128 judges who submitted their personal information. The average age of the judges is 33.93 years old. All of them have a bachelor's (59.4%) or a master's degree (40.6%). 45.3% are female. 51.7% come from the civil division, 44.5% are from the criminal division, and 3.8% are from the administrative or enforcement divisions. The majority (64.8%) are from local courts, 34.4% are from intermediate courts, and only one judge (0.8%) is from the high

court of Zhejiang Province. At the end of the study, we asked the judges, "in your practice, whether (and how often) your decision is influenced by your emotions, especially your feelings to the litigants". 12.5% of them chose "Never", 60.9% chose "Only in extreme circumstances", 25.8% chose "Occasionally", and 0.8% (one judge) chose "Often".

With the permission of the instructors, we invited the judges to participate in our experiment during their class time. They used their laptops, iPads, or other mobile devices to enter the experiment online on the Qualtrics platform. The experiment process was standard: judges read vignettes of case materials, and made decisions. We also asked how much confidence they had in their decisions.

All 145 judges decided a set of two cases, and the sequence of the cases was randomized. The combination of Study 1 and Study 2 is a 2*3 design (six groups in total), and they both used Case 1 as material. Study 3 is a 2*3 design (six groups) in itself, and it used Case 2 as material. Judges were randomly assigned to six groups in each case (that is, Studies 1 and 2's six groups, and Study 3's six groups).

A. Study 1

In Study 1, we use a 2*2 design to explore the interactions between the writing of reasons and the influence of emotion.

To do this, we first induce an emotional bias, using a standard treatment-control setting, where we describe the defendant in the treatment group as a corrupt government official, though her corrupt behavior is irrelevant to the current case, which concerns whether she acted in self-defense or committed manslaughter. We randomly assign half of the judges in both the treatment group and the control group to write formal reasons before they make their decisions, while half of the judges

give their decisions right away, but with no time limit imposed. We want to see whether judges who first write down their reasons are less subject to the influence of emotion.

Case Scenario

In Study 1's case scenario (see appendix Case 1), a lady in her car was robbed by three men. As the men fled on a motorcycle, the lady started her car and chased them to get her wallet back. In the chase, the lady negligently hit the motorcycle, and causing severe injuries to one man, who later died. The judges are to decide whether the lady legitimately acted in self-defense (which is a not guilty verdict in Chinese law), or whether it was an excessive act of self-defense (which is a manslaughter verdict in Chinese law). In the treatment group, the case material includes an additional piece of information, which states that "[I]t was later found that Ms. Long (the lady) was a government official, and that the 80,360 yuan she lost was from a bribe she had solicited the same day. The bribery issue will be tried at a separate trial." The material is adapted from several real cases which took place in China, in which judges decided differently from each other. The emotional stimulus is adapted from Wistrich et al., 2014, who tried to see whether judges' feelings about the litigant, evoked by irrelevant factors, influences their decisions.

Although the lady's corrupt behavior should not affect judges' decisions on the self-defense issue, we predict that judges who receive this information will be harsher in their verdict and sentencing. Nevertheless, we also predict that giving written reasons before making a decision will curb judges' emotional impulses, and hence reduce the decision divergence between the control group and the treatment (negative-emotion-induced) group.

Result

Our results confirm our predictions. First, the judges in the two direct-decision groups diverged in their decisions, in which judges in the treatment condition made significantly harsher decisions on the defendant.

The conviction rates (for manslaughter) in the treatment and the control are 70.8% and 55.6% respectively (see Table 2.2(a)). This difference is not statistically significant ($\chi 2$ =1.269, P = .260), however. This is most likely attributable to a small sample size problem, with only 24 judges in the treatment group, and 27 in the control group.

Although the divergence in conviction rates does not reach statistical significance, the sentences do show a noticeable difference. As shown in Figure 1 (A) and Table 2.3 (a), in the two direct-decision groups, judges in the treatment condition (with emotional stimulus) on average gave 2.04 years sentence to the defendant, whereas judges in the control condition (without emotional stimulus) on average gave a .74 years sentence. That difference is significant at the .01 level.

Second, compared to the two direct-decision groups, judges in the two writing-reason-first groups decided more consistently. As shown in Table 2.2 (b), the conviction rates in the treatment and the control were 57.1% and 60.0% respectively ($\chi 2$ =.038, P = .845). And the sentences they gave were .96 years and .62 years (t = .941, P = .352). It is worth noting that, although the difference in the sentences was not statistically significant, a gap between the sentences given by the treatment and control groups still exists.

Third, the above two points also show that the sentence divergence between the emotional stimulus group (treatment) and the no emotional stimulus group (control) becomes smaller in the writing-reason-first groups than it was in the direct-decision groups. This appears most straightforwardly in Figure 1 (A), which shows the sentences converged when the judges wrote down their reasons

first. A two-way ANOVA analysis of the interaction between emotional stimulus and writing-reason provides further support. The coefficient is -.96, which is marginally significant at the .1 level. (Table 2.3 (b))

Taken together, the above results confirm our theory. They show that giving written reasons before making a decision reduces the influence of emotion on judging. That influence, however, is not entirely eliminated by reasoning.

B. Study 2

Study 1 shows that reason-writing can reduce the influence of emotion on judicial decision making. We still want to know why this is the case, or, through what mechanism reason-writing has this impact. The theoretical analysis in Section I suggests that writing down reasons forces and facilitates deliberation in the judicial process, and that this induced deliberation restrains emotional impulses in judging. This theory has yet to be tested.

To open the black box of reason-writing and further study the mechanism behind the above result, we append another study using a pair of treatment-control groups, which expands Study 1 to a 2*3 design. For simplicity, we call this new pair of groups "Study 2". In the new treatment and control groups, we induced judges to deliberate over the case issue before they were instructed to make a decision. To do this, a one minute mandatory cooling-off or deliberation period was set before they could proceed to the decision page.

Comparing the decision result in Study 2 with that in Study 1, we find that the two deliberation-first groups in Study 2 decided in a more consistent manner and exhibited less emotional bias. e The sentences in the treatment and in the control are .98 year and .65 year respectively (Table 2.3 (c)), showing a much smaller gap than that of the direct-decision groups (Table 2.3 (a)). And their decision pattern is similar to the two writing-reason-first groups (Table 2.3 (b)). This result is also

shown in Figure 1 (A). A two-way ANOVA analysis of the interaction between emotional stimulus and deliberation partially confirms this result, with a -.97 coefficient, nearly marginally significant at the .1 level. (Table 2.3 (c)) The insignificance of the result is probably due to a small sample size problem, with only 41 judges in Study 2.

The above results show that reason writing and pure deliberation both reduce the effect of emotional bias. We may infer that reason writing has this effect through promoting deliberation. But the evidence is not conclusive. For example, some of the results have not reached the significant level, as we expected (e.g., the ANOVA test). Another example is in the verdicts. As shown in Table 2.2 (a), we see a 15.2% difference in conviction rates between the treatment and the control groups, when judges are asked to decide directly. When they deliberate first, the difference is 14.3% (Table 2.2 (c)). This means there is no apparent convergence of verdicts when judges are led to think more carefully. The fact that none of the above differences in verdicts is statistically significant may alleviate our concern. But it does not resolve the puzzle.

Taken together, it is prudent to say that the result in Study 2 provides prima facie evidence that reason-writing influences judges' decisions by forcing deliberation.

C. Judges' Decision Time Length and Confidence level

We further study the mechanism behind reason-writing by analyzing judges' decision time length and their confidence level in their decisions.

As suggested by theory, reason-giving induces introspection so that it restrains intuitions from dominating the decision process. This appears in the "it won't write" phenomenon: judges often change their initial decision after writing down their reasons. Similar to many other assertions by lawyers, this theory has yet to be tested rigorously. The difficulty of testing this mechanism is that we cannot directly observe what occurs in the judicial mind. But we suspect that these

psychological processes would be reflected in the confidence level in their decisions. The hypothesis is straightforward: if judges become skeptical about their first instinct after introspection, or in other words, if their two cognitive processes (system 1 and system 2) are in conflict, their confidence in the decision should decrease.

Our data comform to the prediction. We recorded judges' thinking time span before they made a decision (in the two writing-reason-first groups, this includes the time they spent on writing brief reasons). And after they made a decision, we also asked them "how much confidence do you have in your decision (0 to 100%)".

As shown in Table 2.4 (A), the time judges spent thinking were longer in the writing-reason-first groups (322 seconds) and the deliberation-first groups (284 seconds), than in the direct-decision groups (206 seconds), both significant at .01 level. However, theirconfidence in their decisions decreased when they spent more time thinking. In the two direct-decision groups, judges are on average 71.6% confident that their decisions were correct, while this number decreases to 64.4% in the writing-reason-first groups (the difference is -7.2%, significant at the .05 level), and 65.7% in the deliberation-first groups (the difference is -5.9%, marginally significant at the .1 level).

The data of judges' decision time lengths and confidence levels provides supplemental evidence about the introspective mechanism behind reason writing. The decrease in confidence when judges spend more time deciding the case suggests, though implicitly, a conflict between intuition and reason inside the judicial mind. Interestingly, when they are less confident in their decision, they are more likely to decide it correctly.

III. Placebo Test

Study 1 and Study 2 show writing reasons and deliberation reduces the divergence in decisions between judges who receive an emotional stimulus and those who do not. Study 3 is designed as

a placebo test to Studies 1 and 2, to confirm that reason writing and deliberation only reduces the influence of erroneous emotional factors, and not of emotionally charged, but legally applicable elements. In Study 3, the treatment is no longer an irrelevant emotion factor. Rather, it is an aggravating circumstance that judges should take into account according to the law.

To implement this test, Study 3 uses a 2*3 between subjects design, which is equivalent to the combination of Study 1 and Study 2. The "2" represents the treatment and control conditions, which vary depending on whether there is a circumstance of aggravation; the "3" represents the three decision procedures: direct-decision, writing-reason-first, and deliberation-first. 145 judges were randomly assign to the six (2*3) experimental groups.

Case Scenario

In the Study 3 case scenario, the defendant, who came from an impoverished family, stole 23,172 Yuan RMB in cash (a substantially "large amount" according to the current criminal law in China) from his co-workers, after his salary payment had been delayed for three months.

In the treatment condition, additional information states that the defendant had been imprisoned for ten months on a conviction for theft, and had just been released several months ago. This makes his current act a re-offense, which should be taken as an aggravating circumstance, for which there should be a sentence of longer imprisonment, according to current Chinese law.

As we believe that reason writing and deliberation only reduce the influence of erroneous emotional factors, and not of legally applicable elements, we predict that the judges' decision patterns are similar across the three decision procedures.

Result

The results conform to our prediction. We find that, in this scenario, reason writing and deliberation no longer reduces sentence divergences between the treatment and the control, which

suggests a correct application of law. This result is most obviously shown in Figure 1 (B). Detailed statistics are shown in Table 2.5: in the direct-decision groups, judges in the treatment condition (given the context of circumstance of aggravation) on average gave a 4.25 years' sentence to the defendant, while judges in the control condition (without circumstance of aggravation) on average gave a 3.20 years' sentence, with a 1.05 years' divergence ($t = 3.077$, $P = .004$, significant at .01 level). This pattern holds for the writing-reason-first groups and ithe deliberation-first groups. In the former groups, the numbers are 4.32 and 3.20, with a 1.01 years' divergence in sentences ($t = 2.816$, $P = .008$, significant at .01 level); in the latter groups, the number are 4.17 and 3.33, with a .85 year's divergence in sentences ($t = 2.364$, $P = .023$, significant at .05 level).

Two-way ANOVA tests further confirm that the divergence between the treatment and control groups remains unchanged across the decision procedures. The interaction between the aggravating circumstance and the writing-reason-first procedure has a coefficient of -.04, and the standard error is .50; the interaction between the aggravating circumstance and the deliberation-first procedure has a coefficient of -.20, with a .49 standard error. The results suggest that their trends/patterns are very much similar.

We also recorded the judges' decision time spans and their confidence in their decisions. As shown in Table 2.4 (B), judges in the writing-reason-first groups (239 seconds) and the deliberation-first groups (112 seconds) spent significantly longer time forming their decisions than judges in the direct-decision groups (81 seconds).But unlike in Studies 1 and 2, where judges' confidence decreased in the writing and deliberation conditions, judges' confidence held unchanged (or slightly increased, though not significantly) in the different decision procedures in Study 3. Table 2.4 (B) shows that judges in the writing-reason-first groups (76.3%) were more confident than judges in the direct-decision groups (69.8%), and the difference (6.4%) is marginally statistically

significant (at the .1 level). But there is no observable difference between the deliberation-first groups and the direct-decision groups.

In the above section, we propose that the decreasing confidence is due to a conflict between emotion and reason: a deliberation over an erroneous intuition makes judges become less certain about their decisions. Our data in Study 3 is consistent with this explanation. In Study 3, there is no biased emotion involved, and the reception of the stimulus only induces a legally correct decision. As the conflict between system 1 and system 2 is absent, deliberation does not change the judges' confidence in their decisions.

Study 3 further clarifies the function of reason-writing. Taken together, Studies 1, 2, and 3 suggest that reason writing and deliberation only reduces the influence of erroneous emotion factors, but not of legally applicable elements.

IV. Discussion

Our study employed standard cognitive psychological research methods, which provide considerable internal validity for our experiments. The real challenge to any experimental study, however, is its external applicability, which we discuss in detail here. Past studies employed experiments to find a causal relation between emotions and biased decision-making in a laboratory setting. These studies overlooked some important differences between the experimental and the real world setting. In the courtroom, judges consider more detailed and complicated factual materials, and have more time to deliberate before reaching a conclusion. Judges also have a greater incentive to make accurate judgements in the real world, which likely makes them consider actual cases more carefully. Moreover, judges are embedded in procedures and formalities which have been devised to help them to improve decision quality. These factors all contribute toward making them prudent and inducing them to deliberate.

While these studies do not capture all of these real world features, our research tries exactly to take some of these important factors into account. We show that reason writing, as a procedure that induces deliberation, does reduce the influence of emotion. So the problem of bias, found by previous studies, may not be as serious as we once thought. It is also worth noting that our results show that writing reasons does not eliminate the influence of emotion; it only reduces it. This means that judicial biases is still a topic that requires further attention and studies.

Our experiments solely tackle emotional bias (judges' personal feeling of the litigants), but the results of the experiments let us wonder whether the effects of reason writing extend to other cognitive biases that originate from a similar source: that is, the conflict and competition between system 1 and system 2. In particular, a strand of literature in legal studies has shown that a number of behavioral effects can influence judges, including anchoring, framing, hindsight bias, context dependence, etc., which undermine the accuracy, consistency and thus the legitimacy of judicial decisions. (Guthrie et al., 2001; Eisenberg et al., 2002; Leibovitch, 2016) These effects involve a spectrum of cognitive biases that judges may encounter: emotional bias is the most obvious one, which has been discussed for decades and is disputed in the recent empirical work. Judges are also self-aware of this issue. We predict that it is most likely to be curbed by reason writing, as professional judges are trained to control their emotions and to use reason when making decisions. Other biases are relatively subtle. Some are very hard to notice. For example, decision makers may barely realize that they are sometimes bewildered by choices and context (context dependence), or by how the question is framed (framing effect); and they may hardly notice that information that is completely irrelevant to the issue at hand can change their judgement (anchoring). As these effects are firmly embedded in cognitive processes, we believe that whether reason writing can reduce them is still an important question that has yet to be tested empirically.

Although our results show a debiasing capability from the procedure of reason writing, we hesitate to draw any policy conclusions at this point. We need more studies to replicate and refine our results. More importantly, writing reasons is not costless in the real world. An obvious cost of writing reasons is the time it requires judges to spend; time is important to both the judiciary and the litigants. The requirement of writing reasons can slow down procedures and become overly burdensome. Moreover, as suggested by Cohen (2015), reason-giving in judicial decisions can also induce insincere opinions, and evoke criticism and distrust of the judiciary.

Nevertheless, legal scholars should take the debiasing function of reason writing into account when studying a judiciary or an adjudicative process. For example, only a small proportion of judicial decisions in the U.S. involve the writing down of formal reasons. This is a potential source of judicial bias and arbitrariness, without judges' even noticing it. On the other hand, we may want to explore settings in which reason writing could promote procedural effectiveness. One potential scenario is the jury trial: as emotional decisions are a major problem with juries, we may want to design a reason writing procedure for jury trials, which would require jurors to write down their reasons before they discuss case issues. Consistent with common practices, these written reasons should not be revealed in any event; but the mere requirement of reason-writing would suffice to accomplish the goal of reducing jurors' emotional biases.

Conclusion

Legal realists and their successors today, the behavioral scholars in law, have long claimed that judges are merely human. Their major task is disenchantment: they use anecdotes, observations, data, and statistics to prove that judges can make self-serving or politically driven decisions (e.g., Posner 2010; Epstein, et al., 2013), and that judges are as vulnerable to emotional and behavioral biases as ordinary people (e.g., Guthrie, et al., 2000; Wistrich, et al., 2014; Leibovitch, 2016). This

paper joins this literature. We confirm that judges can be influenced by their emotions and feelings. But we also try to think about the issue from another direction. If judges are not reliable, why do we entrust our power with them? Why do we let them decide important issues? The design of the legal processes may be one of the reasons. In this paper, we identify a simple and already pervasive debiasing procedure in the judicial setting: writing down reasons before making a decision. We show that reason writing successfully reduces, but does not eliminate, the impact of judges' emotional biases in the experiments. We believe that there are more features of the legal processes in the real world that may well assist judges to avoid or correct their errors. And these features have yet to be studied to generate a more comprehensive model of judicial behavior.

References

Coffin, F.M., 1994. On appeal: courts, lawyering, and judging. WW Norton.

Cohen, M., 2015. When judges have reasons not to give reasons: a comparative law approach. Washington and Lee Law Review, 72.

Eisenberg, T., Rachlinski, J.J. and Wells, M.T., 2002. Reconciling experimental incoherence with real-world coherence in punitive damages. Stanford Law Review, pp.1239-1271.

Epstein, L., Landes, W.M. and Posner, R.A., 2013. The behavior of federal judges. Harvard University Press.

Frank, J., 1931. Are judges human? Part one: The effect on legal thinking of the assumption that judges behave like human beings. University of Pennsylvania Law Review and American Law Register, 80(1), pp.17-53.

Guthrie, C., Rachlinski, J.J. and Wistrich, A.J., 2000. Inside the judicial mind. Cornell Law Review, 86, p.777.

Guthrie, C., Rachlinski, J.J. and Wistrich, A.J., 2007. Blinking on the bench: How judges decide cases. Cornell Law Review, 93, p.1.

Kahan, D.M., 2012. Ideology, motivated reasoning, and cognitive reflection: an experimental study. Judgment and Decision Making, 8, pp.407-24.

Kahan, D.M., Hoffman, D.A., Evans, D., Devins, N., Lucci, E.A. and Cheng, K., 2016. 'Ideology' or 'Situation Sense'? An Experimental Investigation of Motivated Reasoning and Professional Judgment. University of Pennsylvania Law Review, 64, pp.2015-26.

Kahneman, D., 2011. Thinking, fast and slow. Macmillan.

Kelman, M., Rottenstreich, Y. and Tversky, A., 1996. Context-dependence in legal decision making. Journal of Legal Studies, pp.287-318.

Kozinski, A., 1992. What I ate for breakfast and other mysteries of judicial decision making. Loy. LAL Rev., 26, p.993.

Leflar, R.A., 1961. Some Observations Concerning Judicial Opinions. Columbia Law Review, 61(5), pp.810-820.

Leibovitch, A., Relative judgments, 2016, Journal of Legal Studies 45, forthcoming.

Lerner, J.S., Goldberg, J.H. and Tetlock, P.E., 1998. Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility. Personality and Social Psychology Bulletin, 24(6), pp.563-574.

Lieberman, M.D., Eisenberger, N.I., Crockett, M.J., Tom, S.M., Pfeifer, J.H. and Way, B.M., 2007. Putting feelings into words affect labeling disrupts amygdala activity in response to affective stimuli. Psychological science, 18(5), pp.421-428.

McMackin, J. and Slovic, P., 2000. When does explicit justification impair decision making?. Applied Cognitive Psychology.

Mussweiler, T., Strack, F. and Pfeiffer, T., 2000. Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. Personality and Social Psychology Bulletin, 26(9), pp.1142-1150.

Oldfather, C.M., 2007. Writing, Cognition, and the Nature of the Judicial Functions. Georgetown Law Journal, 96, p.1283.

Paxton, J.M., Ungar, L. and Greene, J.D., 2012. Reflection and reasoning in moral judgment. Cognitive Science, 36(1), pp.163-177.

Posner, R.A., 1995. Judges' writing styles (and do they matter?). The University of Chicago Law Review, 62(4), pp.1421-1449.

Posner, R.A., 2010. How judges think. Harvard University Press.

Rachlinski, J.J. and Jourden, F., 2002. The cognitive components of punishment. Cornell Law Review, 88, p.457.

Rodríguez, G. and Blanco, S., 2016. Contrast effect on the perception of the severity of a criminal offence. Anuario de Psicología Jurídica.

Sieck, W. and Yates, J.F., 1997. Exposition effects on decision making: Choice and confidence in choice. Organizational Behavior and Human Decision Processes, 70(3), pp.207-219.

Spamann, H. and Klöhn, L., 2016. Justice is less blind, and less legalistic, than we thought: evidence from an experiment with real judges. Journal of Legal Studies, Forthcoming.

Sunstein, C.R., Kahneman, D., Schkade, D. and Ritov, I., 2002. Predictably incoherent judgments. Stanford Law Review, pp.1153-1215.

Tversky, A. and Kahneman, D., 1975. Judgment under uncertainty: Heuristics and biases. In Utility, probability, and human decision making (pp. 141-162). Springer Netherlands.

Wald, P.M., 1995. The rhetoric of results and the results of rhetoric: Judicial writings. The University of Chicago Law Review, 62(4), pp.1371-1419.

Wilson, T.D. and Schooler, J.W., 1991. Thinking too much: introspection can reduce the quality of preferences and decisions. Journal of personality and social psychology, 60(2), p.181.

Wilson, T.D., Lisle, D., Schooler, J., Hodges, S.D., Klaaren, K.J. and LaFleur, S.J., 1993. Introspecting about reasons can reduce post-choice satisfaction. Personality and Social Psychology Bulletin, 19(3), pp.331-39.

Wistrich, A.J., Rachlinski, J.J. and Guthrie, C., 2014. Heart versus head: Do judges follow the law of follow their feelings. Texas Law Review, 93, p.855.

Appendix 2.1: Experiment Materials

I. Case 1 (Studies 1 & 2)

At 4:00 am, July 13, 2008, Zong Mo, Gui Pang, together with Tian Pang came to Ms. Long's garage located in Road A, District B, City C. Armed with white gloves and an iron drill each, Gui Pang and Tian Pang waited on both sides of the garage, while Zong Mo waited on his motorcycle nearby to help them to escape.

At around 5:30 am, seeing Ms. Long drove out of the garage, Tian Pang went to the driver's side and knocked down the car window with his iron drill. In the meantime, Gui Pang went to the other side of the car, broke down the car window, and robbed a handbag with 80,360 yuan and some other bills inside. **[Treatment:** It was later found that Ms. Long was a government official and the 80,360 yuan was a bribe she solicited the same day. The bribery issue will be tried on a separate trial.**]**

After they got the handbag, the two men ran to where Zong Mo waited and jumped onto his motorcycle. Zong Mo then started his motorcycle and tried to flee. Hoping to recover her bag and money, Ms. Long drove and chased the three men, and finally knocked the three men down when they ran to a piece of greenbelts on the north side of the residential zone. While Zong Mo and Gui Pang then stood up and managed to flee, Tian Pang was badly injured, and died on the way to hospital.

[Nest page. The case material appears on every page.]

[**Direct-decision**: we want you to think carefully about the case and then make decision. Options of verdict include "legitimate self-defense" and "excessive self-defense". You will decide the case on the next page.]

[**Writing-reason-first**: we want you to think carefully about the case and then make decision. Options of verdict include “legitimate self-defense” and “excessive self-defense”. We will ask you to decide the case on the next page. Before that, we want you to write about:

A. What legal and factual reasons may justify a “legitimate self-defense” (not guilty) verdict in this case?
B. What legal and factual reasons may justify an “excessive self-defense” (negligent homicide) verdict in this case?]

[**Deliberation-first**: we want you to think carefully about the case and then make decision. Options of verdict include “legitimate self-defense” and “excessive self-defense”. You will decide the case on the next page. This page has a mandatory one minute’s deliberation period. You can only proceed to the next page when the time passes. You can use the time to read the case again and consider carefully about a proper decision.]

[Nest page. The case material appears on every page.]

If you were the judge of this case, what is your verdict?

A. Legitimate self-defense

B. Excessive self-defense

[Nest page. The case material appears on every page.]

If you were the judge of this case, what sentence will you give?

As prescribed in Article 20, Clause 3 of PRC Criminal Law, criminal responsibility is not to be borne for a defensive act undertaken against ongoing physical assault, murder, robbery, rape,

kidnap, and other violent crimes that seriously endanger personal safety that causes injury or death to the unlawful infringer since such an act is not an excessive defense.

[Nest page. The case material appears on every page.]

How much confidence do you have in your decision? (0 to 100%)

[A percentage bar appears. The default point is at 50%]

2. Case 2 (Study 3)

The defendant Mr. Dou is from an impoverished family. **[Treatment: In October, 2014, he was imprisoned for ten months because of being convicted of theft. In April, 2015, he was released, and later]** He worked in a restaurant as a waiter. On February 11, 2016, he came to chief chef's dormitory to demand his unpaid salary that has been delayed for three months. The delayed payment of his salary had brought substantial difficulties to his living in the city. He couldn't find the chief chef in his dormitory. But he found that the roommates of the chef have fallen into sound sleep. He searched their belongings, and took 23,172 Yuan RMB cash from their drawers.

According to Article 264 of criminal law, those who steal relatively large amounts of public or private property and money or have committed several thefts are to be sentenced to three years or fewer in prison or put under criminal detention or surveillance, in addition to fines; or are to be fined. Those stealing large amounts of property and money or involving in other serious cases are to be sentenced to three to ten years in prison, in addition to fines.

[Nest page. The case material appears on every page.]

[**Direct-decision**: we want you to think carefully about the case and then make decision. We will ask you to give sentence on the next page.]

[**Writing-reason-first**: we want you to think carefully about the case and then make decision. We will ask you to give sentence on the next page. Before that, we want you to write about: why do you think a certain sentence (you will be giving) is appropriate.]

[**Deliberation-first**: we want you to think carefully about the case and then make decision. We will ask you to give sentence on the next page. This page has a mandatory one minute's

deliberation period. You can only proceed to the next page when the time passes. You can use the time to read the case again and consider carefully about a proper decision.]

[Nest page. The case material appears on every page.]

How many years' imprisonment will you give if you were the sitting judge?

[Nest page. The case material appears on every page.]

How much confidence do you have in your decision? (0 to 100%)

[A percentage bar appears. The default point is at 50%]

Appendix 2.2: Figure and Tables

Figure 1: Differences in Sentences (Studies 1, 2, and 3)

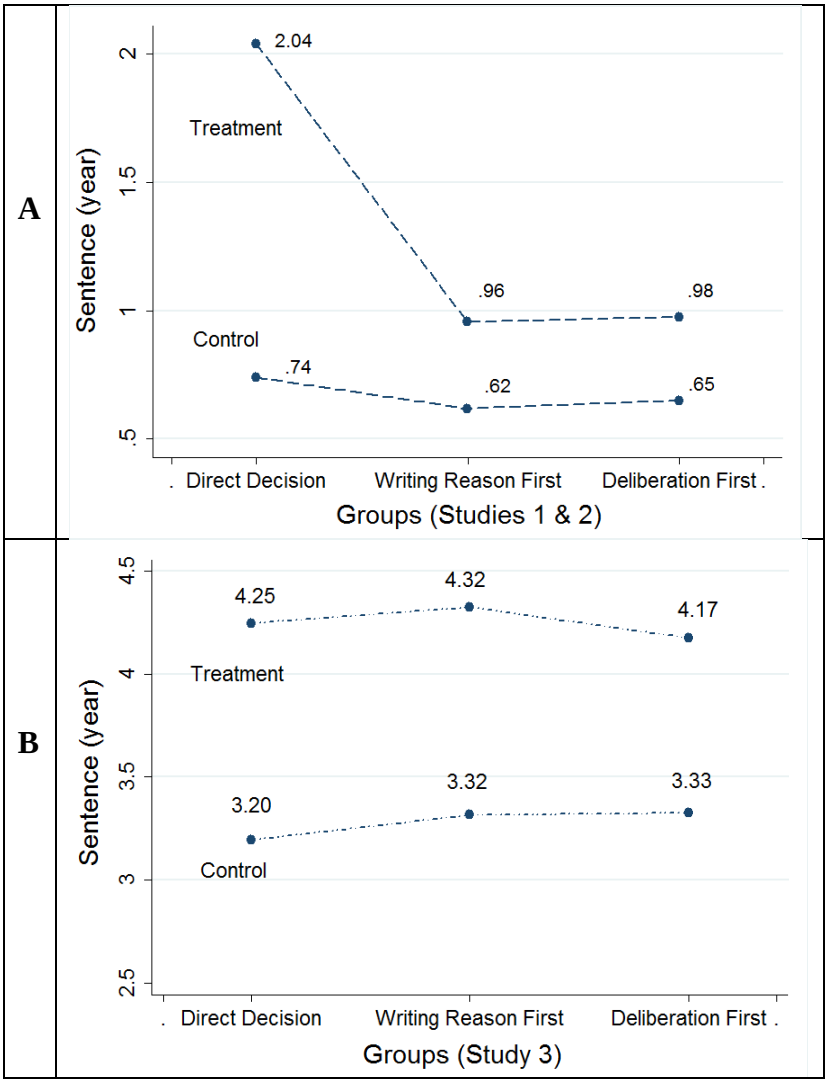

Table 2.1: Summary Statistics of Judges' Demographics

Variable	Mean/ Percentage	Variable	Mean/ Percentage
Age	33.93		
(Std. Dev.)	(6.06)		
Education		Gender	
Bachelor's	59.4%	Female	45.3%
Master's	40.6%	Male	54.7%
Field		Court level	
Civil	51.7%	Local	64.8%
Criminal	44.5%	Intermediate	34.4%
Other	3.8%	High	0.8%
Emotion's influence in practice*			
Never	12.5%		
Only in extreme circumstances	60.9%		
Occasionally	25.8%		
Often	0.8%		
Observations	128		

Note: 1. 145 judges participate in the studies. 17 of them choose not to reveal their demographics. The summary statistics is based on the 128 judges who submit their information. 2. * Before the end of the studies, we ask judges, "in your practice, whether (and how often) your decision is influenced by your emotion, especially your feeling to the litigants".

Table 2.2: Differences in Verdicts (Studies 1 & 2)

Manslaughter (proportion)	(a) Direct Decision	(b) Writing Reason First	(c) Deliberation First
(1)Emotional stimulus	70.8%	57.1%	61.9%
(2)No emotional stimulus	55.6%	60.0%	47.6%
Pearson $\chi 2$	1.269	.038	.865
Pr	.260	.845	.352
Number (1)/(2)	24/27	25/21	21/21
Number	51	46	42
Interaction (two-way ANOVA)		Emotion* Writing Reason First	Emotion* Deliberation First
		-.55	-.08
		(.84)	(.86)

Note: Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table 2.3: Differences in Sentences (Studies 1 & 2)

Sentence (year)	(a) Direct Decision	(b) Writing Reason First	(c) Deliberation First
(1)Emotional stimulus	2.04	.96	.98
	(.34)	(.27)	(.38)
(2)No emotional stimulus	.74	.62	.65
	(.21)	(.23)	(.23)
(1)- (2) diff	1.35***	.34	.32
t-value	3.372	.941	.731
Pr	.002	.352	.470
Number (1)/(2)	24/27	25/21	21/20
Number	51	46	41
Interaction (two-way ANOVA)		Emotion* Writing Reason First	Emotion* Deliberation First
		-.96*	-.97
		(.53)	(.59)

Note: Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table 2.4: Decision Time and Confidence

	(A) Studies 1 & 2			**(B) Study 3 (Placebo)**		
	(a) Direct Decision	(b) Writing Reason First	(c) Deliberation First	(a) Direct Decision	(b) Writing Reason First	(c) Deliberation First
Time spent before making decision (s)	206 (18)	322 (26)	284 (28)	81 (8)	239 (31)	112 (8)
Confidence in decision (%)	71.6 (2.1)	64.4 (2.8)	65.7 (2.9)	69.8 (2.2)	76.3 (2.6)	70.1 (3.3)
Time (s)						
(b)- (a)		116*** (31)			158*** (28)	
(c)- (a)			78** (32)			31*** (11)
Confidence (%)						
(b)- (a)		-7.2** (3.5)			6.4* (3.4)	
(c)- (a)			-5.9* (3.5)			0.2 (3.7)

Note: Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table 2.5: Differences in Sentences (Study 3 Placebo)

Sentence (year)	(a) Direct Decision	(b) Writing Reason First	(c) Deliberation First
(1) Circumstance of aggravation	4.25	4.32	4.17
	(.27)	(.28)	(.34)
(2) No aggravation	3.20	3.32	3.33
	(.19)	(.23)	(.13)
(1)- (2) diff	1.05***	1.01***	.85**
t-value	3.077	2.816	2.364
Pr	.004	.008	.023
Number (1)/(2)	26/23	17/19	23/23
Number	49	36	46
Interaction (two-way ANOVA)		Aggravation *Writing Reason First	Aggravation *Deliberation First
		-.04	-.20
		(.50)	(.49)

Note: Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

The Economic Structure of the Jury System

Abstract: We analyze the economic structure of the jury system, and argue that juries alleviates an information problem in judicial markets. Judicial decisions, a typical expert service, are a credence good, and consumers of the service (the general public) have difficulty ascertaining its quality. The jury thereby functions as an input monitoring scheme, and reduces judges' tendency to excessively collaborate with the government in convicting defendants. Our theory predicts that juries should promote public confidence in the judiciary. We tested this hypothesis against comparative data across 113 countries. We find that judiciaries in countries that conduct jury trials retain a higher level of public trust, controlling for legal origins, economic development, the level of judicial corruption, and the level of democracy. The data also shows that these judiciaries perform more effectively in the criminal adjudication system. Placebo tests show that the jury does not have an impact on creditor right protection cases or contract enforcement cases (in which the jury is certainly not involved in most countries), which suggests that omitted background variables are not driving the influence of the jury on criminal trial efficacy. The seemingly paradoxical patterns that are identified—that the jury, which is less proficient than professionals in deciding cases, in fact makes the judiciary more effective—show an intrinsic economic logic for the jury, besides a political institution and judicial body.

Introduction

The institution of the jury is ancient, and throughout its history it has evoked both praise and condemnation. In modern law and social science, trial by jury, as a judicial institution, is frequently criticized for its inaccuracy and redundancy. Conversely, it is also largely embraced as a political institution for its advocacy of direct democracy. (Tocqueville, 1898, p359) In spite of the debate, there seems to be a consensus that the jury is inefficient, and redeemed, if at all, by the noneconomic values that the system protects. We try to show that this conventional wisdom is wrong.

The institutional and doctrinal structure of the jury has a subtle, though intuitive economic logic, which stems from the information structure of the market. Judicial decisions, the product the judiciary provides, are credence goods by their very nature. The consumers of the product, other than the litigants, are mainly members of the general public who are concerned with social justice. For them, it is difficult and maybe impossible to ascertain the quality of a judicial decision. This is largely because the law is complex. A deeper reason is that the truth of the matter is rarely known in any particular case, and there would be little need for a trial if the truth were known. Moreover, a case is a past event, and a past event is always "constructed" rather than "ascertained", by memories, words, physical evidences, common sense, defendant's and witnesses' performances in the court room, and most importantly, by decision makers' beliefs. The special feature of fact-finding places the consumers (the public) and the producers (the judges) in an unbalanced position regarding information. Such an information asymmetry leads to many problems. For example, as a typical under-treatment problem taking place in credence good markets, judges can be effort averse (Posner, 2010; Epstein, et al., 2013), where they exert less effort than the socially optimal level in providing services, when the quality of their services is unascertainable.

In a credence good setting, when consumers have difficulties in determining the output quality of a firm, mechanisms that monitor the inputs often emerge to alleviate the problem and retain the market. This applies to judicial markets as well, and it explains the laws which surround judges, governing appointments, compensation, tenure, and conflicts of interest: they are all designed as input monitoring schemes and to reduce judges' incentive to provide a defective service. (Posner, 2009, p48-49) So is the jury system such a mechanism.

Lawyers have long noticed a most imperative problem with the judiciary in criminal trials: absent scrutiny of the quality of its decisions, the judiciary can excessively collaborate with the government. And judge can be beholden to the state, which politicizes justice. (Glaeser, et al., 2002) This comes naturally from the institutional structure: day after day, judges are in a repeated game with government officials (e.g., prosecutors and police), but typically in a one-time game with the defendant. Collaborating with the government costs judges less effort, while turning down government requests and protecting individuals can involve much more labor. At the same time, judges lack the power of either the purse or the sword. They count on the government to finance them and enforce their decisions. Judges' career paths, e.g., promotion to a higher position, are also decided by the government (e.g., the president (U.S.)[1], or the political party (China)). Judges can be easily captured by the powerful. In the worst case scenario, a judge would conspire with an oppressive government, and use the criminal procedure to suppress dissidents.

This is where the jury comes in. As direct representatives of the public, they monitor and participate in the production of judicial decisions. Jurors may outperform judges for two reasons. First, jurors are peer citizens. Thus, their initial preference is more likely to align with the

[1] US Federal Judges are appointed by the president with Senate confirmation. State judges are sometimes appointed, and in some cases elected.

defendant. Second, jurors are summoned (usually randomly) from a large pool of citizens, on an ad hoc basis. These features, the increased numbers of decision makers, the random selection process, and the one time game nature, make it much harder for them to be captured by the government, as compared to judge. Thus, they are less likely to render decisions that have been manipulated by prosecutors or other government agents.

Notice that the credence good nature of the judicial decision is the key to understanding the economic structure here. There are many scenarios in which the government can become sluggish or abusive, but when it is easy to ascertain the outcome quality, there is no need for any direct public participation to monitor the producing process. For example, the public does not need to participate in the procurement of a winter service vehicle so that they can clear their thoroughfares of snow, nor does the public need to concern itself with whether to grant a patent to a new technique or license a new drug. As their decision quality can eventually be ascertained in the market, these are issues that can be addressed by simply setting government accountability.

In the second part of the paper, we test the implications of the theory against data. A direct implication of the theory is that juries should improve public confidence in the judiciary, which is how the credence good theory is defined. We collected data for up to 113 countries from varied sources to analyze the impact of juries. The empirical evidence supports the view that the jury system consolidates people's trust in the judiciary. We show that there is a significantly positive association between the jury and trust. In general, we find that citizens trust the judiciary 6.2% to 7.3% more in countries that have adopted a jury system as opposed to those that have not, controlling for legal origins, economic development, level of judicial corruption, and level of democracy. The data on trust are from World Value Surveys.

It is worth noting that, although juries partially solve the information problem in the judicial market, the use of the juries is expensive. As previous literature has well pointed out, the jury is not an efficient system when seen narrowly. It imposes opportunity costs on all the jurors, and it often slows down the judicial process. Nor do juries always accurate find out the facts, especially when the issue before a jury is complicated. The jury trial is radically imperfect from a Utopian standpoint, which requires us to weigh its costs and benefits, and examine its overall effectiveness. We thereby test juries' overall impact on judicial performance. Our data provides preliminary evidence that shows that the overall influence of the jury is positive. We show that judiciaries with juries perform 7.0% to 10.8% more effectively in the criminal adjudication system. Placebo tests show that the jury does not have an impact on creditor right protection cases (in corporate insolvency) and contract enforcement cases, in which the jury is certainly not involved in most countries, which suggests that omitted background variables are not driving its influence on criminal trial efficacy. Taken together, the empirical findings provide seemingly paradoxical patterns: the jury, which is designed to constrain judicial power, ends up empowering the judiciary, and lay people, who are less proficient than professionals in deciding cases, in fact make the judiciary more effective. In contrast to previous theories which separated juries' judicial and political functions, the results of this paper indicate that the two functions are, by their very nature, symbiotic. Neither cheap nor highly accurate, the jury system is radically imperfect as an adjudicative body. But by maintaining public confidence, it has irreplaceable merits. And seen from a comparative perspective, the cost of juries is bearable given the benefits they bring to the system.

Section I of this paper reviews the literature concerning the jury, and formulates questions and hypotheses accordingly. Section II provides an economic theory of the jury system. Sections III,

IV, and V provide empirical evidence about the impacts of juries. Section VI discusses the implications of our findings.

I. Why Do We Need the Jury? Previous Literature

A. The Jury as a Political Institution

A strand of literature categorizes the jury mainly as a political institution which promotes direct democracy in the judicial process. Under the title "Trial by Jury in the United States Considered as a Political Institution," Tocqueville (1989, p361) wrote that "[the jury] places the real direction of society in the hands of the governed, or of a portion of the governed, and not in that of the government." Tocqueville was certainly not the first to support the idea of public participation in adjudication. Centuries earlier in Athens, the popular courts, or dicasteries, although a center of controversy, were defended as a bulwark of the democracy. (Morrow, 1940) Plato said that "[i]n the judgment of offenses against the state the people ought to participate, for when anyone wrongs the state they are all wronged and may reasonably complain if they are not allowed to share in the decision." (Plato, Benjamin Jowett transl., 1888, 767a-767b) These views were not held by Plato alone. Aristotle included among the essential elements of a constitution "the system of popular courts, composed of all citizens or of persons selected from all, and competent to decide all cases—or, at any rate, most of them, and those the greatest and most important." (Aristotle, Ernest Barker transl., 1899, 1317b.) Plato's rationale for lay participation in adjudication was straightforward: "He who is without the right of sitting with his fellows in the courts of law thinks that he is excluded from citizenship." (768b) Aristotle defined a citizen as "a man who shares in the administration of justice and in the holding of office."(1275a) Together, these observations indicate that the legitimacy of the judiciary comes directly from citizens' participation in the judicial process. Similarly, theorists today have emphasized the jury's role in representing the

community and reflecting citizens' values and judgements. (Vidmar & Hans, 2007) The jury is an example of a participatory institutional design that can catalyze civic responsibility and public awareness. (Dzur, 2012)

Because it shifts power away from the state, and thrusts it into the hands of citizens, the jury is also largely seen as a system which counterbalances judges and, behind them, the state's power, and which safeguards individual rights and freedom. (Delvin, 1956, p164) In the country that originated the modern jury system, Sir William Blackstone extolled the jury as "the glory of English law" and claimed that it placed "a strong … barrier between the liberties of the people and the prerogatives of the Crown." (Blackstone, 1791/1966, p379) The founders of the United States shared this view. Thomas Jefferson's Declaration of Independence denounced George III's abuse of power and improper interference with judicial administration, as well as '[f]or depriving us in many cases of the benefit of Trial by Jury." As a result, Article III of the United States Constitution specifically guarantees the role of the jury system in criminal justice: every defendant has the right to be judged by a jury. (James, 1962) The U.S. Supreme Court has further explained that the nation's founding fathers insisted on further resistance against arbitrary behaviors, besides creating an independent judiciary: "[P]roviding criminal suspects with the right to be judged by a jury composed of their companions has brought them immeasurable protection in order to confront corrupted or ultra procurators as well as biased and weird judges." (Duncan v. Louisiana, 391 U.S. 145 (1968))

B. The Jury as a Judicial Institution

Some studies on the capability of juries have a very different perspective than those who see the jury as a democratic body, however. Although jury research finds that juries generally are competent decision-making bodies (Vidmar & Hans, 2007), there are also valid challenges.

(Devine, 2012) These scholarly challenges, together with anecdotes involving high profile cases, e.g., the O.J. Simpson and Rodney King trials, are responsible for ambivalence and even distrust in the jury's overall impact on a society.

One strand of literature targets the competence of juries. Scholars disagree with one another, but some evidence clearly casts doubt on the ability of juries to make sound decisions in areas of special challenges, e.g., instructions, and technical evidence. As an extreme example, from studies of death penalty trials, found that in South Carolina 14% of jurors who served as jurors in capital cases believed that the death penalty was the only acceptable punishment for murder, 57% believed it was the only acceptable punishment for premeditated murder, and 48% believed it was the only acceptable punishment for the killing of a policeman or prison guard. (John Blume et al, in Garvey, ed., 2002) As another example of juror incompetence, jurors can confuse evidence in complex or combined criminal trials: a joining of offenses can lead to the confusion of facts from one offense to another, and it can also lead to unfavorable ratings of the defendant's character. Moreover, judicial instructions have been found to be ineffective in reducing these biases. (Tanford et al., 1985) This situation has been made worse by the rise of scientific evidence and expert testimony. (Vidmar & Hans, 2007, pp174-189) In fact, concerns about the possibility of juries' reliance on unsound science underlie a series of 1990s rulings on expert evidence by the U.S. Supreme Court. In these rulings, called the Daubert trilogy after one of the cases, the Court held that trial judges should serve a gatekeeper role, evaluating the scientific basis for expert testimony before permitting it to be introduced at trial. (Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993); General Electric v. Joiner, 522 U.S. 136 (1997); Kumho Tire v. Carmichael, 526 U.S. 137 (1999))

Another problem with the juries is that they tend to be inconsistent and prejudiced, which compromises a basic requirement of justice: settled laws that treat like cases alike. It is not uncommon for juries to decide similar cases differently, without good reason. (Abramson, 1994) Prejudice is a more serious problem, and its types vary, from barely noticeable examples—discrimination against foreigners (Vidmar & Hans, p152)—to the most notorious and pernicious ones—racism in criminal procedures and even in capital punishment decisions. Death penalty cases exhibit the most appalling patterns. This was famously shown in Furman v. Georgia, where U.S. Supreme Court justices cited the evident racism in the death penalty system, under which blacks were disproportionately more likely to be sentenced to death. In Georgia, cases with white victims were more than four times as likely to result in death sentences than comparable cases with black victims. (David C. Baldus, et al., in Dorothy K. Kagehiro, et al., ed., 1992) Racism in jury trials has not merely been historically rampant; it is currently pervasive. A recent rigorous empirical study, using a data set of felony trials in Florida between 2000 and 2010, found evidence that juries formed from all-white jury pools convicted black defendants significantly more often than white defendants (16% more often), and that this gap in conviction rates disappeared entirely when the jury pool included at least one black member. (Anwar et al., 2012)

C. The Theoretical and Empirical Questions

Jury trials are surely a way of obtaining public participation. But the question remains: why do we need any lay people sitting in the courtrooms to maintain the effectiveness of the judiciary? To emphasize the democratic value the jury provides presents a puzzle rather than an answer. After all, adjudication – finding the truth of a case and applying the law -- is an expert service, and in other practices the lay public are seldom placed in charge in settings that depend upon expertise. For example, we do not need public participation in the administrative process of granting a new

patent or licensing a drug; nor do we need lay participation in forming a currency policy. If these affairs can be left to experts, why do we need lay people in the courtroom? This question is further complicated by the fact that juries are at best as competent as judges in deciding a case, while they always involve higher costs, and slow down the trial process, etc. These questions require us to offer a different explanation.

II. The Economic Structure of the Jury System

We analyze the economic structure of the jury system, and provide a theory to explain the function of juries accordingly. The institution of the jury has a subtle, though intuitive and implicit economic logic, which has its root in the information structure of the market. As judicial decisions are credence goods and the market can be fragile, the jury system has an important function in curing the possible inefficiencies created by these features.

A. Judicial Decisions as Credence Goods

The judiciary is a special kind of service provider; the only product it provides is judicial decisions. Judicial decisions are credence goods.

A credence good is a good whose quality (utility impact) is difficult for the consumer to ascertain, not only before purchasing it, but also after using it. Expert service is a typical example: an expert always knows more about what a consumer needs than the consumer herself, which is exactly why a consumer needs an expert. This asymmetric information structure can lead to many problems. An expert may defraud a consumer, or she can exert less effort than is in the best interest of the customer. Commonplace advice regarding the services of car mechanics goes: “if a mechanic tells you that he has to replace a part in your car, don't forget to ask him to put the replaced part into the trunk of your car." (Dulleck, et al., 2006) On other occasions, an expert may provide excessive services. A survey conducted by the Department of Transportation in the U.S. estimated that more

than half of auto repairs are unnecessary. (Wolinsky, 1993) Other instances include taxi rides, where a stranger to a city cannot be sure whether her driver takes the proper route to her destination (Balafoutas, 2013), and medical treatments, where doctors may provide a suboptimal diagnosis to sell the most profitable treatment. (Emons, 1997) We categorize these problems as "moral hazard", a common consequence of information asymmetry, which induces unethical behaviors, and distorts the market away from an efficient one.

Judicial decisions are a unique kind of expert service. The problems with credence goods remain here. The consumers of a judicial decision include the direct participants of the services (the litigants). But litigants, especially defendants in a criminal trial, are not the only consumers. A more important consumer is the general public, because a judicial decision not only affects the parties to a case, by redistributing their wealth or restricting their freedom, but also influences the society at large by setting precedents and signaling norms and rules. Moreover, every citizen cares about law and order, and they also more or less care about justice. An unjust decision hurts consciences or evokes feeling of sympathy, which are externalities levied upon citizenry. This relation can also be seen from the supply side. The service of the judiciary is mainly paid for by tax payers rather than litigants, especially in criminal cases. This implies that the judiciary is a service provider to the public, rather than to the defendant. In fact, it is typically a "service" that the defendant would prefer to avoid.

Now we consider the information structure in the judicial market. A judicial decision is a credence good. This is first because the law is complex. But complexity per se is not the source of information asymmetry. The manufacture of a car is complex but there is no information problem in the auto market, for people know the overall quality of a car after driving it; information thus diffuses and accumulates in the market as a brand name, amazon review stars, and reputation.

(Ackerloff, 1970) The problem with the judiciary comes instead from the difficulty people may have in ascertaining decision quality in a trial, even after knowing all the law and evidence.

The substantial problem comes from fact finding. The truth of any particular case is rarely known. If it were, there would be little need for a trial. Discovery of facts is thereby a complicated issue, and its quality will be difficult for the public to ascertain.. Take the shooting of Michael Brown, for instance. Evidence in the case includes the sworn testimony of more than 60 witnesses, ballistics report, autopsies and other similar information. (DOJ Report on Shooting of Michael Brown, 2015) Working through all this information is time-consuming — and beyond the capacity of most of commentators, not to mention average people. Moreover, a deeper difficulty behind fact finding is that a case is a past event, and a past event is always "constructed" by memories, words, physical evidence, common sense, defendant's and witnesses' performances in the court room, and most importantly, by the decision makers' beliefs. People with different prior beliefs can hardly reach consensus on the "truth" of a case.

These together explain why, though we see fact-finding as the most important function of a trial, a trial is never solely about finding the truth, and no legal system spends unlimited resources on fully discovering a past event, no matter how important it might be. Instead, procedural rules and standards of evidence govern the trial. The burden of proof is never "finding the exact truth". Rather, it is to establish the truth "beyond reasonable doubt" or by "preponderance of the evidence".

Taken together, these special features of the judicial process decide that, even after knowing all the evidence in a case, people may still be uncertain about the decision quality of the fact finding, and the application of the law to the facts. This uncertainty puts the judiciary and the public in a credence goods market.

B. The Jury as an Input Monitoring Scheme

Information asymmetry often leads to inefficiencies. But what is the exact problem in the market for judicial decisions? What fraud can a (criminal court) judge, who is a legal expert and public servant, perpetrate on his consumers (the citizens)?

For one thing, a judge can be effort averse, which is a typical under-treatment problem in a credence goods market, where the expert exerts less than socially optimal effort in providing services, when the quality of product is unascertainable. In fact, the problem of effort aversion should be more likely in the judiciary than in most other employments, because of life tenure and fixed salary (for Article III judges in the U.S.). Empirical evidence confirms these theoretical predictions. (e.g., Huang, 2011; Epstein, et al., 2013, pp36-37)

However, that judges might be lazy does not justify the use of juries, for there is no reason to believe that lay people will be more diligent, especially when they are one-timers in the trials, and have no career concerns and do not responsibility as the judges do.

A real concern with judges is their impartiality, namely, the judiciary can excessively collaborate with the government, and judges can be beholden to the state, which politicizes justice. (Glaeser, et al., 2002) This comes naturally from the institutional structure: day after day, judges are in a repeated game with government officials, e.g., the prosecutors and the police, but typically in a one-time game with the defendant. Collaborating with the government costs judges less effort, while turning down government requests and protecting individuals can involve much more labor. At the same time, judges lack the power of the purse and the sword. They count on the government to finance them and enforce their decisions. Judges' career paths, e.g., promotion to a higher position, are also decided by the government, e.g., the president (U.S.), or the political party (China). As a result, judges can be easily captured by the powerful. In the worst case scenario,

judges can conspire with an oppressive government and use the criminal procedure to suppress its dissidents.

Submitting to political pressure is a type of under-treatment problem, or effort aversion in a broader sense. The problem is still rooted in the nature of credence goods. If the public perfectly knew the facts of a case, they could easily ascertain a suspect's guilt or innocence, and there would be no place in the legal procedure for the government to step in and abuse individual rights, as the consumers/citizens concerned about social justice would know about the intervention and oppose it in this perfect information world. But when the cost of information is prohibitively high, so that the consumers are unaware of an unjust verdict, then oppression cannot be distinguished from crime, and no counter-measures can be launched.

There are, however, mechanisms that alleviate these effects. In a credence goods setting, when consumers have difficulties in determining the output quality of a firm, mechanisms that monitor the inputs often emerge to alleviate the problem and retain the market. (Posner, 2009, pp48-49) In expert services markets, examples of input monitoring include professional education and licensing requirements. Ethical training in medical and law schools, and the Hippocratic Oath of a physician and its counterpart in the law, are also ways to improve and monitor the quality of labor inputs. Economists have found many other examples. As Dulleck, et al (2006) noted,

> "[T]he separation of doctors and pharmacies is an institution to avoid overtreatment by disentangling the incentives to prescribe drugs from the profit made selling them. The fixed part of a taxi ride tariff provides incentives for the driver to serve many individual customers and therewith not to take longer routes than necessary. To avoid overcharging as well as overtreatment, in many repair industries the chamber of commerce issues

> standard worktimes for some repairs to allow customers to better check upon a workman's bill—by providing a comparison to usual hours for the ordered repair."

Monitoring inputs applies to judicial markets as well, and it explains the laws that surround judges governing appointments, compensation, tenure, and conflicts of interest: they are all designed as input monitoring schemes which can reduce judges' incentives to market a defective service. (Posner, 2009, pp48)

The jury is another input monitoring scheme. Just as the separation of doctors and pharmacies is an institution to avoid a credence goods problem, the separation of the jury verdict and the judge's sentencing also reduces the possibility for judges to "defraud" the consumers. But there are differences between the medical and judicial settings. The separation of diagnosis and treatment aims to solve an over-treatment problem, in which physicians tend to prescribe excessive drugs and treatments to obtain higher profits. The separation of the jury verdict and the sentencing aims to solve an under-treatment problem as suggested above. In the absence of juries, judges tend to save their efforts. They especially try to avoid extra cost brought by being afoul of the government. Regarding this problem, jurors outperform judges in two aspects. First, jurors are peer citizens. Their initial preference is more likely aligned with the defendant's, while judges are more likely to enforce the will of the state. This is clear from a review of American history. Juries played a crucial role in helping their peer American colonists to subvert the authority of royal governors. "[J]ury trials resulted in the acquittals of John Peter Zenger (tried for criminal libel against British interests), smugglers prosecuted under the Navigation Acts, rioters against the Stamp Act, and participants in the Boston Tea Party." As Mark Roe (2007) pointed out, this is why the British were ambivalent to transplant their jury system to their colonies.

Second, jurors are summoned (usually randomly) from a large pool of citizens, on an ad hoc basis. They are outsiders to the judicial process, and do not typically have ongoing experiences or relationships with other actors in the justice system or government sector, like the police or prosecutors. Thus, jurors' decisions are less likely to reflect either bind faith in the ability of these other actors or a need to get along with them. (Lempert, 1992) Moreover, the features of jury trials -- the increased numbers of decision makers, the random selection process, and the one time nature of the game — decide that they are harder for the government to capture as compared to judges, and thus the outcomes of jury cases are less likely to have been manipulated by prosecutors or other government agents.

A recent example comes from Russia. Jury trials were gradually reintroduced to Russia as part of the liberal reforms following the dissolution of the Soviet Union. In 2008, the State Duma decided to end jury trials for terrorism and treason, in their already limited areas. Officially, this decision was explained as necessary to improve counterterrorism enforcement. Many lawyers and experts pointed out, however, that the curtailment of jurors' powers is primarily in the interest of law enforcement agencies and special services. (Barinov, 2012) In an open letter to President Dmitry Medvedev, the Jurors' Association, well-known actors, human rights activists, prominent lawyers, and economists argued that,

> "[K]nowing the inner workings of the investigatory agencies and the courts, we are confident that trying cases of terrorism, mass rioting, espionage and other felonies without jurors will not make the prevention of such crimes more efficient. On the contrary, the result will be a mockery of crime prevention in which dozens of innocent people will be sent to prison." (Mishina, 2012)

As the letter points out, the core value of jury trials is its function in freeing the judiciary from political control. In fact, it is reported that, although the use of juries in Russia has been highly restricted since their introduction (only 0.01% of all criminal cases are tried before juries) (Barry, 2010), they have effectively counterbalanced the state's power. Juries granted many more acquittals, in 15% to 20% of cases, compared with fewer than 1% in cases decided by the bench. Jurors also gained a reputation for indenpendence as they refused to submit high-profile cases, such as the murder of Igor V. Izmestiev, to the state, even under pressure and threats from the government. (Ibid)

As the jury has more compassion, and as the procedures of trial by jury are set intentionally to avoid manipulation, the jury can be seen as a monitoring scheme in the producing of justice, and thus it prevents judges from being abusive. The presence of the jury thus signals a positive message, and hence should increase overall confidence in the judicial market.

Notice that the nature of the judicial decision as a credence good is the key to understanding the economic structure here. There are many scenarios in which the government can become sluggish or abusive, but when it is easy to ascertain the outcome quality, there is no need for direct public participation to monitor the producing process. For example, the public does not need to participate in the municipal procurement of a winter service vehicle to clear their thoroughfares of snow, nor does the public need to concern itself with whether to grant a patent to a new technique or license a new drug. As the quality of these decisions can eventually be ascertained in the market, these are the issues that can be addressed by simply setting government accountability.

C. Predictions

Our theory proposes that the jury partially solves the credence good problem in judicial markets. A prediction from the theory is that the use of juries will increase public confidence and trust in

the judiciary. This prediction runs afoul of the literature that criticizes the jury as an inaccurate and redundant judicial institution.This conflict requires an empirical examination. We use the following hypothesis to capture this question:

Hypothesis #1: Jury trials help the judiciary gain a higher degree of trust from the people.

The economic theory of the jury is consistent with the political theory that praises the jury as an advocate of direct democracy, by predicting that the use of juries will increase people's trust in the judiciary. (Gastil, et al., 2010, Chapter 7) But as we pointed out, the political theory per se can hardly explain why this is the case. Moreover, a question that follows the question of trust concerns judicial effectiveness and performance. On the one hand, despite the debate among legal scholars and political theorists, there seems to be a consensus that jury trials are inefficient, as is evident in the discussion of the competence of juries. On the other hand, theories of organizational trust suggest that trust is critical to organizational effectiveness. Trust enhances cooperation, improves communication, and facilitates citizenship behaviors (Tyler et al., 2002), in addition to improving group and organizational performance. (e.g., Davis, et al., 2000; Dirks, 2002) H.L.A. Hart argued that an effective legal system depends on the willing compliance of most people, perhaps for normative reasons. Their compliance makes it possible for the state to focus its limited resources for coercion on the few who are potentially disobedient. (Hart, 2012, p196) Following this line of thought, and answering the question of why people obey the law, Tom Tyler's studies show that trust in judicial procedures is a most important factor in promoting peoples' compliance with judicial decisions. (Tyler, 2006) Together, these observations indicate that, if a jury system increases trust in the judiciary, it should also make more people comply voluntarily, thus making it more effective and efficient. Accordingly, we propose hypothesis #2 to test these contested theories and to examine the jury's overall effect on judicial performance:

Hypothesis #2: Judiciaries that have jury trials perform more effectively.

III. Data and Basic Empirical Facts about the Jury System around the World

We use data across countries to test the above hypotheses. Our dataset combines several data sources about juries, public trust in the judiciary, and judicial performance, and it comprises up to 113 jurisdictions, including 39 jurisdictions that have adopted a jury system(see Table 3.1) (within the dataset, however, the scope of countries that we are able to analyze also depends on the availability of other variables in the models).

Using empirical evidence to comparatively analyze the impact of an institution follows a tradition in economics. (e.g., La Porta, et al., 2008) Prior to this paper, the only study to tackle a relevant topic—lay participation's impact from a comparative and empirical perspective — was conducted by Voigt (2009). He sent questionnaires to lawyers in 71 countries to collect data on their jury and lay assessor systems, but he did not find many statistically significant effects of lay participation on several outcome variables, including judicial corruption, judicial independence, and governance quality. He did find, however, that jury systems display a significant positive effect on government effectiveness, while lay assessor systems do not. This paper builds on Voigt's study. We use Voigt's data on whether a country employs a jury system, and supplement it with data from the U.S. Department of State's Human Rights Report (2011, 2013, and 2015). Nevertheless, we explore some different (but relevant) topics, including the impact of the jury system on people's trust in the judiciary, and in judicial performance. So our empirical questions are in general different from Voigt's.

The jury is largely seen as a common law institution. For example, in their canonical work on legal origins and financial development, La Porta, et al. (2008) and Glaeser, et al., (2003) see the jury as a decisive feature that separates common law origins from civil law origins. This might have

been true in the 13th century, but it is not the case for today. We see from Table 3.1 that about one-third of civil law countries use jury trials, while many common law countries operate their judicial systems without a jury. There is no statistically significant correlation between whether a country is a common law regime and whether it uses jury trial.

The data on people's trust in the judiciary are from the World Values Surveys. From 1989 to 2014, the World Values Survey has conducted six waves of surveys covering approximately 120 countries around the world. The data includes both a nation dimension and a year dimension, but these are not strictly "nation-year" panel data— although most countries have been surveyed more than once, no country is surveyed in every one of the six runs. The World Values Survey asks how much confidence participants have in their courts (on a scale of 1–4) and in the government in general (on a scale of 1–4). Political scientists have shown that people from different countries and cultures have very different understandings of these questions. East Asians are more inclined to report higher government support and confidence, but this does not necessarily mean that they trust their government more than people in the Western world. For example, the mean confidence in the government given by Chinese people is usually higher than that of U.S. citizens, but this is problematic since it does not capture the actual variations in people's attitudes across the two countries. The true reason for the divergence is that Chinese people tend to give relatively higher scores on questionnaires for cultural reasons, or because they are in the shadow of political pressure, while U.S. citizens express a more acritical attitude. (e.g., Delhey, et al., 2011) This phenomenon undercuts the validity of direct comparisons of numbers given by people from different countries. To solve this problem, we use a relative score, instead of an absolute score, to document people's trust in the judiciary. To construct the relative score for "trust in the judiciary," we subtract "confidence in the government in general" from "confidence in the court," which

cancels general tendencies driven by culture or political pressure. This also enables us to directly compare cross-national numbers.

Table 3.2 shows the sources for all data.

Some of the controlling variables are highly correlated with each other and thus may have a multicollinearity problem. In the empirical analysis, we control for different subsets of judicial independence, common law origin, judicial integrity, ln GDP, and ln GDP per capita, in separate regression settings. The results are generally similar when we control for different subsets of variables. See Table 3.3 for a Pearson correlation matrix of the main independent variables.

IV. The Jury System and Trust in the Judiciary

We first test hypothesis #1. We use an OLS model to estimate the association between whether a country has a jury system and the public trust in the judiciary. (Diagnostic tests are done for the OLS model by checking outliers, heteroscedasticity, and linearity.) The dependent variable in the regression is "trust in the judiciary." We also control for time fixed effects, which enables us to estimate different countries/time in one model. For example, the Survey covers China in 1999 and 2004, the U.S. in 1995, 1999, and 2004, and France In 1995 and 2004. The year fixed effects enable us to compare China and the U.S. in 1999 and 2004, China and France in 2004, and the U.S. and France in 1995 and 2004. To avoid an over-estimate problem, brought by using a country for multiple times, the standard errors are clustered by country.

The model takes the following form:

$$\mathrm{Trust}_{it} = \beta_0 + \beta_1 \mathrm{Jury}_i + \beta_2 \mathrm{CommonLaw}_i + \beta_3 \mathrm{Jury}_i * \mathrm{CommonLaw}_i + \beta_4 \mathrm{X}_{it} + \text{Year Fixed Effects} + \varepsilon_{it}$$

where Trust_{it} indicates people's trust in the judiciary in country i at time t, Jury_i is a dummy variable indicating whether country i has adopted a jury system, CommonLaw_i is a dummy variable indicating whether i is a common law origin country, and X_{it} is a vector of country-related characteristics, including ln GDP, ln GDP per capita, judicial integrity (absence of judicial corruption), level of judicial independence, and the level of democracy. The coefficients to be estimated are β_0, β_1, β_2, $\beta_3$and β_4; ε is an error term.

During the last 30 years, six countries have changed their judicial systems and adopted a jury system: Spain (starting from 1978), Russia (starting from 1992), Kazakhstan (2008), Japan (2009), South Korea (2012), and Argentina (2015). A possible design for empirical analysis would be to compare the trust levels in these countries before and after the introduction of the jury. Unfortunately, we do not have data for Spain and Russia before they adopted the jury. Nor do we have data for the rest after they adopted the jury. The limitations of the data prevent us from inferring any causation between the jury and public trust using a differences-in-differences design. This is also why we do not use a country fixed effect in the regression. Nevertheless, the regression still shows important patterns, as can be seen in Table 3.4.

The OLS regressions show a statistically significant correlation between the use of a jury system and public trust in the judiciary. Across different jurisdictions, the use of jury trials is associated with an increase in public trust in the judiciary of around 6.2% to 7.3% (the percentages differ across settings (1) to (5)).[2] The results hold across settings (1) to (5), in which different controlling variables are applied.

[2] The dependent variable "Trust in the judiciary" spans from -0.85 to 0.85, adding up to a 1.70 scale. The effect of using jury trials, the percentage of trust it improves, can be calculated as, for

An immediate question is whether we can infer any causal relation between the use of juries and trust from these statistical results. An omitted variable problem raises a major concern, namely, whether there is a factor that is responsible for both the existence of a jury system and higher public trust in the judiciary. For example, studies show that judicial performance is generally of a better quality in common law countries, so a common law origin (often associated with a jury system) rather than the jury per se could be a determining factor in increased trust. To address this problem, we control in the regressions for legal origin, GDP per capital, GDP, judicial corruption, judicial independence index, and the level of democracy.. This suggests that the effect of using juries is significant even after taking these factors into account. For example, controlling for legal origin (whether a country is a civil or common law country) implies that among common law countries (or among civil law countries), a jury system can still increase public trust in the judiciary.

Another concern is reverse causation. That is, the statistical correlation might also be explained by the possibility that people in a country who trust their judiciary relatively strongly will urge the judiciary to reform and adopt a jury system. However, this concern can be eliminated by looking closely at the history of jury systems across jurisdictions. For most countries, the jury is a relatively mature institution which was transplanted through colonization decades ago. (Ryan Park, 2010) This means that the adoption of the jury is largely exogenous. Spain and Russia are the only two countries in the data set (covering years from 1989 to 2008) that reformed to a jury system during

---

example, in setting (1), 0.124/1.70, which is about 7.3%. In setting (4), this number is 0.106/1.70, which is about 6.2%.

The scales of the Judicial Independence index and Judicial Integrity index are 0.93 and 0.95. We can compare the effects of the jury system, judicial independence and judicial integrity. Here, improving from a lowest judicial independence level (country) to a highest one can increase 28.1% of trust (0.93*0.514/1.70 in setting (5)); from a most corrupt country to a least one improves 26.6% of trust (0.95*0.476/1.70 in setting(5)). Given the adoption a jury system *per se* can improve trust around 7%, the effect size of the jury is considerable.

the last thirty years. The aim of their reforms was to improve trust in their judiciaries. In Spain, the right to a trial by jury was legislated by the Spanish Constitution of 1978, a furtherance of the country's transition to democracy. In Russia, the adoption of the jury system came with the political reforms that followed the dissolution of the Soviet Union. These observations suggest that the adoption of the jury system is an exogenous variable to trust, from which we infer that the use of juries improves public trust in the judiciary rather than the opposite.

The above analyses are further confirmed by an interesting pattern that comes from the interaction term between jury trials and a common law origin (Jury System * Common Law) in the regressions. The negative coefficient suggests that, compared to civil law countries, the jury system is less important in promoting public trust in common law countries; that is, when a civil law country conducts jury trials, people's confidence in the judiciary increases to a greater extent than when a common law country does so. This is consistent with our previous understanding of legal origins. The common law system was per se associated with stronger judicial independence and judicial power, and the jury was one of many institutions that protected individual rights, whereas civil law countries generally had more centralized and authoritarian elements in their rulings, with their judiciaries being instrumental in realizing state authority. (e.g., La Porta, et al., 2003) Hence, because the jury is an institution that counterbalances state power once it is adopted, it would have influenced the system in a civil law country to a larger extent and thus improved the people's trust in the judiciary more significantly.

V. The Jury System and Judicial Performance

Hypothesis #2 states that judiciaries with jury trials perform more effectively. As the jury is only used in criminal trials in most of the countries (except the U.S., and Canada), we thereby test juries'

influence on criminal adjudication effectiveness, and also explore a counterfactual relation --the impact of using juries on civil case performance-- as a placebo test.

A. The Jury's Impact on Criminal Justice

We use an OLS model to estimate the relation between whether a country has a jury system and its judiciary's performance in criminal justice. (Diagnostic tests are done for the OLS model by checking outliers, heteroscedasticity, and linearity.) The data for the two dependent variables in the regressions are from the World Justice Project the Rule of Law Index 2012-2013. The World Justice Project surveys lay people, lawyers and other legal experts in 97 countries to get their perceived measurement about over four hundred factors concerning rule of law in their country. These factors (variables) are then constructed into 48 indicators. (The World Justice Project, 2013. P1) The major indicator we use, "criminal adjudication system is timely and effective", refers specifically to the perception of judicial performance in criminal trials. Another indicator, "criminal justice system in general is effective", refers to the perception of the justice system in general, including the performance of prosecutors and the police. The OLS model takes the following form:

$$\mathrm{CriminalJustice_i} = \beta_0 + \beta_1 \mathrm{Jury_i} + \beta_2 \mathrm{CommonLaw_i} + \beta_3 \mathrm{Jury_i} * \mathrm{CommonLaw_i} + \beta_4 \mathrm{X_i} + \varepsilon_i$$

where $\mathrm{CriminalJustice_i}$ denotes the effectiveness of the criminal justice system in country i in 2012, $\mathrm{Jury_i}$ is a dummy variable showing whether country i has adopted a jury system, $\mathrm{CommonLaw_i}$ is a dummy variable that indicates whether i is a common law origin country, and $\mathrm{X_i}$ is a vector of country-related characteristics, including the level of judicial independence, ln GDP, ln GDP per capita and judicial integrity (absence of judicial corruption), etc. The coefficients to be estimated are β_0, β_1, β_2, $\beta_3$and β_4; ε is an error term.

The analysis is cross-sectional, comparing different countries' criminal justice systems at a given time (2012). The results of the regressions are shown in Table 3.5. The OLS regressions show a statistically significant correlation between the jury system and criminal adjudication performance. In settings (1) to (3), across different jurisdictions, the use of jury trials is associated with an increase in criminal adjudication timeliness and effectiveness of about 7.0% to 10.8%.[3] The estimated coefficients are different in settings (1) (2) and (3), where different controlling variables apply. Setting (6) also shows a positive correlation between the use of the jury and the effectiveness of the criminal justice system in general, though the results do not hold in settings (4) and (5), which suggests the jury's impact becomes weaker in criminal justice systems other than the judiciary.

As discussed above, the jury, as an institution that was adopted decades ago, is an exogenous variable to today's judicial performance in criminal adjudication. This implies a causal relation between them. The controlling variables indicate that the jury's effect is significant after taking these factors into account. A caveat here is that we should be especially careful when inferring any causation from mere correlations.

B. Placebo Test

[3] The dependent variable "criminal adjudication system is timely and effective" spans from 0.14 to 0.90, adding up to a 0.76 scale. Jury trials' effect, the effectiveness they improve, can be calculated as, in setting (2), 0.0534/0.76, which is about 7.0%. In setting (3), this number is 0.0822/0.76, which is about 10.8%.
The scale of the Judicial Independence index is 0.93. We can compare the effect of the jury system and judicial independence. Here, from a lowest judicial independence level (country) to a highest one, can increase 43.8% of effectiveness (0.93*0.358/0.76 in setting(3)). The adoption a jury system *per se* can improve effectiveness around 7.0% to 10.8%, which shows that the effect of the jury is considerable.

We conduct a placebo test to rule out an omitted variable problem of the above regressions, that is, it is not the jury but some other variables behind the jury that increase criminal adjudication performance. In the placebo test, we explore whether the jury also makes an impact on creditor right protection or contract enforcement cases, in which jury trials are certainly irrelevant in most countries.

The data on creditor protection and contract enforcement cases are from the World Bank Doing Business Index 2012. As reported in the code book, "[T]he creditor right protection indicator studies the time, cost, and outcomes of insolvency proceedings involving domestic entities. The data for indicators for resolving insolvency are derived from questionnaire responses by local insolvency practitioners and have been verified through a study of laws and regulations as well as public information on bankruptcy systems." (The World Bank, 2012B, p7) And "[I]ndicators for enforcing contracts measure the efficiency of the judicial system in resolving a commercial dispute. The data are built by following the step-by-step evolution of a commercial sale dispute before local courts and are collected through study of civil procedure codes and other court regulations as well as questionnaires completed by local litigation lawyers and judges." (The World Bank, 2012A, p8)

We use an OLS regression model. The results are shown in Table 3.6. As shown in the table, we do not find any significant correlation between the jury and judicial performance in civil cases. The factor that has a strong influence is ln GDP per capita, which suggests that economic development level is significantly associated with judicial capability in creditor right protection and contract enforcement cases.

The comparison between Table 3.5 and Table 3.6 shows that the jury's influence is constrained in the criminal adjudication setting. This provides further evidence that shows that the significant

correlations between the jury and criminal adjudication effectiveness is not very likely driven by omitted variables rather than by the jury system per se.

A caveat here is that the regressions in the placebo test cover only 65 countries, because the Doing Business Index only surveys these countries. The insignificant results can be due the small sample, which does not show the full landscape of the issue.

VI. Discussion

During the last decade, the study of the jury gained special practical importance, as many countries tried to reform their judicial processes by introducing jury trials in their judicial systems. Russia and Spain, for example, incorporated jury trials into their criminal procedures in the 1990s. (Spain stipulated a right to trial by jury in the 1978 Constitution, but the first jury trial was held in 1997). (Thaman, 1999) South Korea (Park, 2010) introduced a reform to accommodate a jury system in 2012, and Argentina (Llenas, 2015) in 2015,.Promoting public participation through the establishment of a jury system has become a heated topic and has raised serious discussions in China. (e.g., Liao, 2012) While many new democracies welcome the jury system, controversies have arisen from jury verdicts in high-profile cases in the United Kingdom, the United States, and other countries, (see e.g., Bower, 1992; Hans, 2005, pxi) and some, including Britain, the supposed mother of trial by jury, have called for the reform of existing jury systems. (The Economist, 2009) For these legal reformers, the findings of this paper may provide a better understanding of the potential benefits of the jury. For readers in countries that have adopted jury trials, a comparative examination of the jury system further clarifies the importance of public participation in court.

For practitioners, the size the effects of the jury is another important question. Skeptics of the jury's efficacy certainly overlook its value shown above, but given that juries have a real cost of implementation, what is the relative weight of the jury's impact? According to the findings of this

paper, the adoption of a jury system can increase people's trust in the judiciary by about 6.2% to 7.3%. For comparison, improving from a lowest judicial independence level (country) to a highest one increases trust by 28.1%; improving from a most corrupted level (country) to a least one increases trust by 26.6%, ceteris paribus. Comparing the jury's number with these more important factors' shows that the size of the effect of the jury on trust is considerable. The jury's effect on judicial performance in criminal adjudications is also strong. For example, the jury increases the effectiveness of criminal justice systems in general by about 7.0% to 10.8%. For comparison, improving from a lowest judicial independence level (country) to a highest one increases effectiveness by 43.8%. This analysis does not offer a thorough costs and benefits evaluation of the jury, as we have no way to translate the jury's effects into a monetary return, and compare it with the implementation costs it may incur. Nevertheless, the results show that the benefits of the jury are substantial, especially when we take into account that improving judicial independence and fighting corruption are usually not an easy task for transitional societies. (See China's example, liebman, 2007)

Our theory and the empirical results show a positive association between the use of the jury and people's trust in the judiciary. This echoes the history in which people see the jury as a natural safeguard against an oppressive government. It is no coincidence that jury systems, which had spread throughout continental Europe after the French revolution, disappeared in countries like Spain, Germany, and Russia when these countries came under authoritarian control. Spanish history has seen the jury emerge several times under liberal regimes, only to disappear under authoritarian ones. (Gleadow, 2000) In France, Napoleon advocated the jury system for the republic and then crippled it after his coronation. (Donovan, 1999) In Japan, the jury was most extensively used in the late 1920s, a pre-war period when individual freedom seemed to be

increasing. But trials by jury diminished as militaristic elements came to dominate the Japanese government, and were suspended during the Second World War. (Lempert, 1992)

The data of this paper extends the historical narrative to a broader, empirical regularity. From a comparative standpoint, the judiciaries in countries that use jury trials are more trustworthy in the eyes of their people. Moreover, the results also show that jury trials are more important in civil law countries. They improve public trust in the judiciary to a greater extent than in common law countries. This is consistent with the fact that people have less leverage to defend themselves through the law in civil law (usually more centralized) regimes.

Conclusion

Nearly two centuries ago, Alexis de Tocqueville contended that the jury served a larger civic function and brought greater benefits outside the courtroom. What Tocqueville provides is a paradoxical pattern. He wrote: "[T]he jury… which seems to restrict the rights of the judiciary, does in reality consolidate its power; and in no country are the judges so powerful as where the people share their privileges." (Tocqueville, 1898, p367)

Our paper provides an economic theory to explain the underlying mechanism of this paradox, and tests its implications against data. We stress that the nature of judicial decisions as a credence good is the key to understanding the structure of judicial markets, and that, besides a judicial and political body, the jury's major function is as an inputs monitoring scheme to amend a market failure and improve people's confidence in the judiciary.

References

Abramson, J.B., 2000. We, the jury: the jury system and the ideal of democracy. Harvard University Press.

Ackerloff, G., 1970. The market for lemons: Quality uncertainty and the market mechanism. Quarterly Journal of Economics, 84(3), pp.488-500.

Anwar, S., Bayer, P. and Hjalmarsson, R., 2012. The impact of jury race in criminal trials. Quarterly Journal of Economics, 127(2).

Aristotle, Ernest Barker translated, 1899. Politics.

Balafoutas, L., Beck, A., Kerschbamer, R. and Sutter, M., 2013. What drives taxi drivers? A field experiment on fraud in a market for credence goods. The Review of Economic Studies, 80(3), pp.876-891.

Baldus, D.C., Woodworth, G. and Pulaski Jr, C.A., 1992. Law and statistics in conflict: Reflections on McCleskey v. Kemp. In Handbook of psychology and law (pp. 251-271). Springer New York.

Barinov, A., 2012. Problems with trial by jury in Russia, at official website of Russian Legal Information Agency, available at http://rapsinews.com/legislation_publication/20120911/264633744.html.

Barry, E., 2010. In Russia, Jury Is Something to Work Around. New York Times. November 15th. Available at http://www.nytimes.com/2010/11/16/world/europe/16jury.html?pagewanted=all&_r=0

Botero, J.C. and Ponce, A., 2011. Measuring the rule of law. Available at SSRN 1966257.

Blackstone, W., 1791/1966. Commentaries on the laws of England, Dawsons of Pall Mall.

Blume, J., Eisenberg, T. and Garvey, S., 2003. Lessons from the capital jury project. In Beyond repair, pp.144-177.

Caldeira, G.A., 1986. Neither the purse nor the sword: Dynamics of public confidence in the Supreme Court. American Political Science Review, 80(04), pp.1209-1226.

Caldeira, G.A. and Gibson, J.L., 1992. The etiology of public support for the Supreme Court. American Journal of Political Science, 635(36), pp.635-664.

Davis, J.H., Schoorman, F.D., Mayer, R.C. and Tan, H.H., 2000. The trusted general manager and business unit performance: Empirical evidence of a competitive advantage. Strategic Management Journal, 21(5), pp.563-576.

Delhey, J., Newton, K. and Welzel, C., 2011. How general is trust in “most people”? Solving the radius of trust problem. American Sociological Review, 76(5), pp.786-807.

U.S. Department of Justice, 2015, Department of justice report regarding the criminal investigation into the shooting death of Michael Brown by Ferguson, Missouri Police Officer Darren Wilson, at https://www.justice.gov/sites/default/files/opa/press-releases/attachments/2015/03/04/doj_report_on_shooting_of_michael_brown_1.pdf

Devlin, B.P., 1966. Trial by jury. Stevens.

Devine, D.J., 2012. Jury decision making: The state of the science. NYU Press.

Dirks, K.T. and Ferrin, D.L., 2002. Trust in leadership: meta-analytic findings and implications for research and practice. Journal of applied psychology, 87(4), p.611.

Donovan, J.M., 1999. Magistrates and Juries in France, 1791-1952. French Historical Studies, pp.379-420.

Dulleck, U. and Kerschbamer, R., 2006. On doctors, mechanics, and computer specialists: The economics of credence goods. Journal of Economic literature, 44(1), pp.5-42.

Dunn, J., 1984. The concept of trust in the politics of John Locke. Philosophy in history: Essays on the historiography of philosophy, pp.279-301.

Dwyer, W.L., 2002. In the hands of the people: The trial jury's origins, triumphs, troubles, and future in American democracy. MacMillan.

Dzur, A.W., 2012. Punishment, participatory democracy, and the jury. Oxford University Press.

Emons, W., 1997. Credence goods and fraudulent experts. The RAND Journal of Economics, pp.107-119.

Epstein, L., Landes, W.M. and Posner, R.A., 2013. The behavior of federal judges. Harvard University Press.

Fukuyama, F., 1995. Trust: The social virtues and the creation of prosperity. New York: Free press.

Garoupa, N.M. and Ginsburg, T., 2010. Reputation, Information and the Organization of the Judiciary. Journal of Comparative Law, 4(2).

Gastil, J., Deess, E.P., Weiser, P.J. and Simmons, C., 2010. The jury and democracy: How jury deliberation promotes civic engagement and political participation. Oxford University Press.

Glaeser, E.L. and Shleifer, A., 2002. Legal origins. Quarterly Journal of Economics, pp.1193-1229.

Gleadow, C., 2000. History of trial by jury in the Spanish legal system (No. 6). Edwin Mellen Press.

Hans, V.P. ed., 2006. The jury system: Contemporary scholarship. Ashgate.

Hardin, R., 1998. Trust in government. Trust and governance, 1, pp.9-27.

Hart, H.L.A., 2012. The concept of law. Oxford University Press.

Huang, B.I., 2011. Lightened Scrutiny. Harvard Law Review, pp.1109-1152.

James, F., 1963. Right to a Jury Trial in Civil Actions. The Yale Law Journal, 72(4), pp.655-693.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A. and Vishny, R.W., 1998. Law and finance. Journal of political economy, 106.

La Porta, R., Lopez-de-Silane, F., Pop-Eleches, C. and Shleifer, A., 2003. Judicial checks and balances. Journal of Political Economy 112(2).

La Porta, R., Lopez-de-Silanes, F. and Shleifer, A., 2008. The economic consequences of legal origins. Journal of economic literature, 46(2), pp.285-332.

Lee, S.H., 2008. South Korea struggles to incorporate a young jury system, New York Times, July 7th.

Locke, J., 2016. Second treatise of government and a letter concerning toleration. Oxford University Press.

Lempert, R., 1992. A Jury for Japan?. The American Journal of Comparative Law, 40(1), pp.37-71.

Llenas, B., 2015. First jury in history of Buenos Aires, Argentina finds defendant 'Not Guilty', FOX NEWS LATINO, March 15th, available at http://latino.foxnews.com/latino/news/2015/03/13/first-jury-in-history-buenos-aires-finds-defendant-not-guilty/

Liao, Y., 2012. The reform of lay assessor system in a transitional society (She hui zhuan xing bei jing xia ren min pei shen yuan zhi du ge lu jing tan xi), Chinese Journal of Legal Studies (Zhong Guo Fa Xue), 3.

Mishina, E., 2012. Trial by jury in Russia: Revival and survival, at Institute of Modern Russia, http://imrussia.org/en/rule-of-law/186-trial-by-jury-in-russia-revival-and-survival.

Morrow, G.R., 1940, January. Plato and the rule of law. In Proceedings and Addresses of the American Philosophical Association (Vol. 14, pp. 105-126). American Philosophical Association.

Park, R., 2010. The globalizing jury trial: Lessons and insights from Korea. American journal of comparative law, 58(3), pp.525-582.

Plato, Benjamin Jowett translated, 1888. Laws.

Posner, R.A., 1993. What do judges and justices maximize? (The same thing everybody else does). Supreme Court Economic Review 41(1).

Posner, R.A., 2009. Public intellectuals: A study of decline. Harvard University Press.

Putnam, R., 1993. Making democracy work: civic traditions in modern Italy. Princeton University Oress.

Ríos-Figueroa, J. and Staton, J.K., 2014. An evaluation of cross-national measures of judicial independence. Journal of Law, Economics, and Organization, 30(1), pp.104-137.

Roe, M.J., 2007. Juries and the political economy of legal origin. Journal of Comparative Economics, 35(2), pp.294-308.

Shapiro, M., 2013. Courts: a comparative and political analysis. University of Chicago Press.

Tanford, S., Penrod, S. and Collins, R., 1985. Decision making in joined criminal trials: The influence of charge similarity, evidence similarity, and limiting instructions. Law and Human Behavior, 9(4), p.319.

Thaman, S.C., 1999. Europe's new jury systems: The cases of Spain and Russia. Law and Contemporary Problems, 62(2), pp.233-259.

The Economist, 2009. The jury is out. Feb 12th, 2009, available at http://www.economist.com/node/13109647.

The World Bank, 2012A. Doing Business 2012: Enforcing Contract, available at http://www.doingbusiness.org/reports/global-reports/~/media/GIAWB/Doing%20Business/Documents/Annual-Reports/English/DB12-Chapters/Enforcing-Contracts.pdf

The World Bank, 2012B. Doing Business 2012: Resolving Insolvency, available at http://www.doingbusiness.org/reports/global-reports/~/media/GIAWB/Doing%20Business/Documents/Annual-Reports/English/DB12-Chapters/Resolving-Insolvency.pdf

The World Justice Project, 2013. Rule of Law Index, 2012 – 2013, available at http://worldjusticeproject.org/sites/default/files/WJP_Index_Report_2012.pdf

Tocqueville, A.H.C.C., 1862. Democracy in America, tr. by H. Reeve.

Tyler, T.R. and Huo, Y., 2002. Trust in the law: Encouraging public cooperation with the police and courts through. Russell Sage Foundation.

Tyler, T.R., 2006. Why people obey the law. Princeton University Press.

Vidmar, N. and Hans, V.P., 2007. American juries: The verdict. Prometheus Books.

Vile, M.J.C., 2012. Constitutionalism and the Separation of Powers. Liberty Fund.

Voigt, S., 2009. The effects of lay participation in courts—A cross-country analysis. European journal of political economy, 25(3), pp.327-339.

Wolinsky, A., 1993. Competition in a market for informed experts' services. The RAND Journal of Economics, pp.380-398.

Case cited

Baker v. Carr, 369 U.S. 186 (1962).

Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579(1993).

Duncan v. Louisiana, 391 U.S. 145 (1968).

General Electric v. Joiner, 522 U.S. 136 (1997).

Kumho Tire v. Carmichael 526 U.S. 137 (1999).

Powers v. Ohio, 499 U.S. 400, 407 (1991).

Appendix 3: Tables

Table 3.1: Legal Origins and Jury Trial

Common Law Jurisdictions		**Non-Common Law Jurisdictions**		
Jury Trial		**Jury Trial**		
Yes	**No**	**Yes**	**No**	
Australia	Bangladesh	Algeria	Albania	Kuwait
Canada	Cyprus	Austria	Argentina (Before 2015)	Kyrgyzstan
Ghana	Kenya	Belgium	Armenia	Latvia
Jamaica	India	Brazil	Azerbaijan	Lebanon
Liberia	Israel	Burkina Faso	Belarus	Lithuania
Ireland	Malaysia	Cote d'Ivoire	Bosnia	Macedonia
Malawi	Nepal	Denmark	Bulgaria	Mexico
New Zealand	Nigeria	Dominican Republic	Cambodia	Moldova
Sierra Leone	Pakistan	El Salvador	Cameroon	Mongolia
Sri Lanka	Singapore	Finland	Chile	Morocco
Trinidad and Tobago	South Africa	France	China	Netherlands
United Kingdom	Tanzania	Georgia	Colombia	Peru
United States	Thailand	Greece	Croatia	Philippines
	Uganda	Libya	Czech	Poland
	Zambia	Madagascar	Ecuador	Romania
	Zimbabwe	Mali	Egypt	Rwanda
		Nicaragua	Estonia	Senegal
		Norway	Ethiopia	Slovakia
		Panama	Germany	Slovenia
		Portugal	Guatemala	South Korea (before 2012)
		Qatar	Hungary	Switzerland
		Russia	Indonesia	Taiwan
		Spain	Iran	Turkey
		Sweden	Iraq	Uruguay
		Tunisia	Italy	Uzbekistan
		Ukraine	Japan (Before 2009)	Venezuela
			Jordan	Vietnam
			Kazakhstan (Before 2008)	Yemen

Note: The data source is Voigt (2009) and the U.S. Department of State's Human Rights Report (2011, 2013, and 2015).

Table 3.2: Sources of Data

Factors	Variables	Explanations	Sources
Trust	Public trust in the judiciary	Confidence in the courts. Answers to "[H]ow much confidence do you have in the courts (in your country)?"	The World Value Survey
	Public trust in the government in general	Confidence in the government. Answers to "[H]ow much confidence do you have in the government in general (in your country)?"	The World Value Survey
Politics	Judicial constraints on government power	Government powers are effectively limited by the judiciary.	The World Justice Project
	Judicial integrity (absence of judicial corruption)	Government officials in the judicial branch do not use public office for private gain.	The World Justice Project
	Judicial independence index	An index combining both de facto and de jure judicial independence.	Ríos-Figueroa and Staton (2014)
	Democracy index	The latest version of the Polity index.	The Polity IV Project
Judicial performance	Criminal adjudication system is timely and effective	An indicator showing the effectiveness of the criminal adjudication system in a country in 2012. The indicator is constructed from a bunch of variables drawn from two original sources of data collected from independent sources in each country: a General Population Poll and a series of Qualified Respondents' Questionnaires (experts' questionnaires).	The World Justice Project the Rule of Law Index 2012-2013
	Criminal justice system in general is effective	An indicator showing the effectiveness of the criminal justice system in general in a country in 2012.	The World Justice Project the Rule of Law Index 2012-2013

Table 3.2: Sources of Data (continued)

	Contract enforcement	"Indicators for enforcing contracts measure the efficiency of the judicial system in resolving a commercial dispute. The data are built by following the step-by-step evolution of a commercial sale dispute before local courts. The data are collected through study of civil procedure codes and other court regulations as well as questionnaires completed by local litigation lawyers and judges."	The World Bank Doing Business Index 2012 (The World Bank, 2012B, p7)
	Creditor protection (resolving insolvency)	"Doing Business studies the time, cost, and outcome of insolvency proceedings involving domestic entities. The data for the "resolving insolvency" indicators are derived from questionnaire responses by local insolvency practitioners and verified through a study of laws and regulations as well as public information on bankruptcy systems."	The World Bank Doing Business Index 2012 (The World Bank, 2012B, p8)
Institutions	Jury system	Whether a country has adopted jury trials in its criminal procedure.	Voigt (2009); The U.S. Department of State's Human Rights Report (2011, 2013, and 2015).
	Legal origins	Divided into Common Law countries and non-Common Law countries for the purposes of this paper.	La Porta et al. (2008)
Economy	GDP GDP per capita		The World Bank

Table 3.3: Pearson Correlation of the Main Controlling Variables

	Jury Trial	Judicial Independence	Democracy	Ln GDP per capita	Common Law Origin	Judicial Integrity
Jury Trial	1					
Judicial Independence	.224**	1				
Democracy	.201**	.820***	1			
Ln GDP per capita	.194**	.504***	.243***	1		
Common Law Origin					1	
Judicial Integrity		.299***		.257**	0.315***	1

Note: Time = 2012; Pearson correlations. Significance levels: *** p<0.01, ** p<0.05

Table 3.4: The Jury System and Public Trust in the Judiciary

VARIABLES	(1) OLS Trust in the Judiciary	(2) OLS Trust in the Judiciary	(3) OLS Trust in the Judiciary	(4) OLS Trust in the Judiciary	(5) OLS Trust in the Judiciary
Jury System	0.124**	0.122**	0.110*	0.106**	0.107**
	(0.0492)	(0.0500)	(0.0568)	(0.0499)	(0.0508)
Common Law	0.148*	0.154*	0.165*	0.143*	0.140
	(0.0825)	(0.0917)	(0.0927)	(0.0841)	(0.0931)
Jury System *	-0.263***	-0.267**	-0.266**	-0.261**	-0.259**
Common Law	(0.0992)	(0.103)	(0.107)	(0.101)	(0.102)
Judicial Independence	0.385***	0.369***	0.499***	0.503***	0.514***
	(0.0935)	(0.124)	(0.163)	(0.119)	(0.166)
Judicial Integrity	0.464*	0.460*	0.434*	0.474*	0.476*
	(0.244)	(0.240)	(0.246)	(0.245)	(0.241)
Democracy			-0.00655	-0.00574	-0.00593
			(0.00898)	(0.00903)	(0.00955)
Ln GDP	0.0200	0.0186		0.0193	0.0199
	(0.0137)	(0.0161)		(0.0136)	(0.0168)
Ln GDP per capita		0.00545	0.0175		-0.00272
		(0.0238)	(0.0197)		(0.0249)
Constant	-0.529	-0.527	-0.177	-0.545	-0.545
	(0.377)	(0.382)	(0.160)	(0.383)	(0.385)
Year Fixed Effect	YES	YES	YES	YES	YES
Observations	131	131	128	128	128
R-squared	0.516	0.517	0.528	0.538	0.538

Notes: 1. Robust Standard Errors in parentheses, clustered by the country. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.
2. Settings (1) (2) each covers 68 countries in the regressions; settings (3) (4) (5) each covers 65 countries. The difference is due to availability of the "Democracy" variable.
3. Sources of data see Table 2.

Table 3.5: The Jury System and Criminal Adjudication System Effectiveness

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	OLS	OLS
VARIABLES	Criminal Adjudication System (timely and effective)			Criminal Justice System in General		
Jury System	0.0578**	0.0534*	0.0822**	0.0364	0.0332	0.0620**
	(0.0283)	(0.0283)	(0.0327)	(0.0227)	(0.0227)	(0.0261)
Common Law		0.0506	0.0939**		0.0365	0.0797**
		(0.0334)	(0.0419)		(0.0269)	(0.0333)
Jury System*			-0.103*			-0.103**
Common Law			(0.0609)			(0.0485)
Judicial	0.333***	0.303***	0.358***	0.384***	0.362***	0.417***
Independence	(0.113)	(0.114)	(0.117)	(0.0903)	(0.0913)	(0.0931)
Judicial	-0.121	-0.180	-0.189	0.0240	-0.0189	-0.0274
Integrity	(0.112)	(0.118)	(0.117)	(0.0899)	(0.0948)	(0.0930)
Democracy	-0.0179***	-0.0176***	-0.0196***	-0.0127***	-0.0125***	-0.0144***
	(0.00460)	(0.00457)	(0.00466)	(0.00369)	(0.00367)	(0.00371)
Ln GDP	-0.00964	-0.0123	-0.0145	-0.00697	-0.00887	-0.0111
	(0.00970)	(0.00978)	(0.00977)	(0.00778)	(0.00787)	(0.00778)
Ln GDP per	0.0464***	0.0573***	0.0568***	0.0523***	0.0602***	0.0597***
capita	(0.0165)	(0.0179)	(0.0177)	(0.0132)	(0.0144)	(0.0141)
Constant	0.258	0.234	0.266	0.0883	0.0709	0.102
	(0.188)	(0.187)	(0.186)	(0.151)	(0.150)	(0.148)
Observations	91	91	91	91	91	91
R-squared	0.398	0.414	0.434	0.627	0.636	0.654

Note: 1. Standard Errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.
2. The regressions cover 91 countries.
3. Sources of data see Table 3.2.

Table 3.6: Placebo Test

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
VARIABLES	Creditor Right Protection			Contract Enforcement		
Jury System	-2.123	0.0442	0.578	1.686	1.832	6.808
	(5.159)	(4.667)	(5.814)	(3.382)	(3.433)	(4.125)
Common Law		19.19***	19.56***		1.296	4.809
		(4.967)	(5.559)		(3.654)	(3.944)
Jury System*			-1.545			-14.40**
Common Law			(9.873)			(7.005)
Judicial	-3.380	-6.051	-4.778	6.535	6.354	18.22
Independence	(21.65)	(19.45)	(21.24)	(14.19)	(14.31)	(15.07)
Judicial Integrity	-5.627	-38.08**	-37.86**	-13.83	-16.02	-13.95
	(16.42)	(16.97)	(17.18)	(10.77)	(12.48)	(12.19)
Democracy	0.139	-0.0522	-0.0929	-0.861*	-0.874*	-1.253**
	(0.739)	(0.666)	(0.720)	(0.485)	(0.490)	(0.511)
Ln GDP	-0.718	-0.961	-1.012	-1.886*	-1.903*	-2.382**
	(1.542)	(1.386)	(1.436)	(1.011)	(1.019)	(1.019)
Ln GDP	7.604***	11.13***	11.11***	8.976***	9.214***	9.076***
per capita	(2.399)	(2.340)	(2.362)	(1.573)	(1.721)	(1.676)
Constant	-8.841	-35.74	-34.85	31.31	29.50	37.75*
	(33.03)	(30.47)	(31.25)	(21.65)	(22.41)	(22.17)
Observations	65	65	65	65	65	65
R-squared	0.207	0.371	0.372	0.435	0.437	0.476

Notes: 1. Robust Standard Errors in parentheses, clustered by country. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.
2. The regressions cover 65 countries.
3. Sources of data see Table 3.2.