



Legal Techniques for Rationalizing Biased Judicial Decisions: Evidence from Experiments with Real Judges

*John Zhuang Liu and Xueyao Li**

Judges rarely reveal their real reasoning in their opinions when they are influenced by factors that they know they should not consider. The natural next question is how, when a judge is improperly influenced, he or she reasons to justify a biased decision. In a set of experiments using incumbent Chinese judges, we first replicated the findings of previous studies that showed judges can be influenced by extra-legal factors. More importantly, we showed that judges may employ a range of legal techniques to rationalize decision biases: they interpret legal standards and legal concepts strategically, finesse the applicability of law, infer or deny causation and foreseeability, and draw different conclusions from facts. Our findings provide a more realistic understanding of how judges behave, and cast doubt on reasoned elaboration as a guarantee of judicial transparency and trustworthiness.

I. INTRODUCTION

Every judicial system consists of institutions designed to constrain judicial powers and ensure accountability. The centrality of reasoned elaboration in judicial decision making is one of these institutions. Reason-giving limits judicial discretion by ensuring that written decisions can be read and reviewed by higher courts (Shapiro 1987). It encourages public monitoring and participation in the judicial decision-making process (Eisenberg 1978; Fuller 1978). Giving reasons for decisions forms part and parcel of the authority of the judiciary, and the use of judicial power unsupported by reasons detracts from the trustworthiness of the judiciary (e.g., Rawls 1993:231–40; Schauer 1995; Sunstein 2007; Wechsler 1959). Henry Hart and Albert Sacks (1994:147–48) famously argue that “reasoned elaboration” distinguishes judicial decision making

*Address correspondence to Xueyao Li, Professor, KoGuan Law School, Shanghai Jiang Tong University, Koguan Law Bldg., 1954 Huashan Rd., Shanghai, China, 200030; email: lixueyao@sjtu.edu.cn. Liu is Assistant Professor of Law and Economics, School of Management and Economics, the Chinese University of Hong Kong, Shenzhen.

We thank Jianjun Bai, Omri Ben-Shahar, Yun-Chien Chang, Jinhua Cheng, Xin Dai, Yan Ge, William Hubbard, Holger Spamann, Yingmao Tang, Chao Xi, Wenming Xu, Qun Yang, and Wei Zhang for helpful comments and discussions. For excellent research assistance, we thank Yanyan Zhang. We are grateful to all of the participating judges in our studies.

from an exercise in “discretionary fiat.” In *Law’s Empire*, Ronald Dworkin (1986:94–96) describes reasoning as a source of integrity in law.

On the other hand, many commentators have noticed that judicial reasoning is habitually and unusually one-sided, and filled with certainty. Judge Posner (1988) states that most judicial opinions “imply that even the very toughest case has a right and a wrong answer and only a fool would doubt that the author of the opinion had hit on the right one.” He adds that judges “decide cases with greater confidence than the realities of judicial decision making permits, and speak and write with more confidence than they feel” (Posner 1988). This feature of judicial opinions is particularly perplexing, since many disputes are complex and fraught with conflict and uncertainty (e.g., Schauer 1988; Simon 1998). Karl Llewellyn, among many other jurists, was particularly skeptical of this kind of certainty. He stated that, for any but the easiest cases, legal certainty has never existed and never will. Most judicial reasoning is simply the “dressing up” of judicial opinions in the “garb of certainty” (Llewellyn 1989:73). In other words, reasoning is simply the clothing judges put on their decisions, rather than the basis on which they make those decisions.

Against this backdrop, there is an impressive collection of studies examining biases in judicial decision making. These studies examine mechanisms such as “hindsight bias,” “anchoring,” “probability neglect,” “emotional bias,” “identity protective cognition,” and the like (e.g., Guthrie et al. 2001, 2007; Kahan 2015), the effects of which bias both probabilistic and analytical reasoning. One strand of studies repeatedly confirms that irrelevant information, such as litigant characteristics, can unduly influence judicial decision making. Wistrich et al. (2015) use several experiments to reveal this general pattern. They provided judges with vignettes of cases in which they varied the defendants’ characteristics. Although the characteristics were irrelevant to the cases’ contexts and therefore should not have been considered when rendering a verdict and sentence, they strongly impacted the judges’ decisions. In an experiment resembling real-world judicial decision making (presenting judges with raw facts, precedents, and trial judgments and giving them about an hour to make decisions), Spamann and Klöhn (2016) also show that legally irrelevant characteristics of a defendant influence judges’ verdicts. Judges are more likely to convict an unrepentant defendant than one who is repentant.

More germane to our study, research also shows that judges rarely reveal the real reasons for their decisions when they have been influenced by factors they know they should not have taken into account. For example, Spamann and Klöhn (2016) required their experimental subjects to write reasons after they made decisions. Judges’ written reasons, however, never mention defendant characteristics at all. Liu (2018) found a similar pattern in his experiments with judges in China, showing that at least some judges are influenced by irrelevant defendant character evidence, but none of them mention it in their reasoning.

Our current study is a natural next step from this previous literature. Judges are expected, and often required, to provide reasons for their decisions. However, when judges are unduly influenced, how do they justify their biased decisions; in other words, what legal techniques do they employ to rationalize their biased decisions?

Theories and preliminary evidence suggest that judges may use a wide range of legal techniques to do so. For example, critics have long speculated that the judicial

interpretations that judges employ in their opinions are simply window dressing for their ideologies. When a judge uses originalism to interpret the Constitution, she does so not because she is a true adherent to that strategy and finds its result determinative, but because it is a means to a conservative end (Post & Siegal 2006; Chilton & Posner 2015). This speculation, however, can hardly be verified by rigorous empirical evidence, due to a conflation of legal and ideological preferences. In criminal law, Nadler and McDonnell (2011) found that actors' bad motives or bad moral character could not only increase perceived blame and responsibility, but also perceived causal influence and intentionality. This finding implies that laypeople may use causation and intention to rationalize their decisions, which have readily been influenced by irrelevant defendant characteristics. Sood and Darley (2012) found a similar pattern with respect to harm and criminal liability. To impose criminal responsibility that has been driven by irrelevant moral intuition, people impute harm to conduct where harm was not previously reported. Both Nadler and McDonnell (2011) and Sood and Darley (2012) focus on the psychological mechanism at work in certain legal contexts. They do not intend to document a more general pattern in the law.

In this article, we provide an empirical analysis lying at the intersection of two themes: biased judicial decisions and reasoned elaboration. We try to provide an outline of the legal techniques judges can employ to rationalize biased decisions. We report three experiments that studied this area, in the domains of contract law (Study 1), criminal law (Study 2), and torts (Study 3), and we studied an array of techniques, spanning both legal and factual matters. Our experiments used more than 100 real, incumbent Chinese judges as subjects, and employed a standard between-group design. Judges were randomly assigned to different conditions. They read cases vignettes that were adapted from real cases.

We first induced the judges to make biased decisions, using stimuli that are, under the law applicable to the precise question at hand, strictly irrelevant (e.g., the defendant's moral character in a determination of guilt). Study 1 was a lease contract case involving liquidated damages and the court's discretion in lowering such damages. In the treatment condition, the defendant was described as maintaining an extramarital relationship with a corrupt government official. In Study 2, a defendant who bred and sold parrots at home was charged with "illegal sale of precious and endangered wild animals and their related products." He was described as indulging in gambling in one experiment condition and as a good father in the other condition. In Study 3, the defendant had stored oxygen tanks in his courtyard, which led to a large fire, and he was being sued for tort liability by a plaintiff who voluntarily participated in firefighting but fell on the road and hurt himself. In one condition, the defendant was storing oxygen to produce crystal meth; in the other, he did so to better care for his mother's lung disease (this case was inspired by Nadler and McDonnell [2011]). Using these designs, we successfully replicated the findings of prior experimental studies showing that judges can be influenced by extra-legal factors (Wistrich et al. 2015; Spamann & Klöhn 2016; Liu 2018).

Second, as the focus of our study, we asked the judges to provide reasons for their decisions. A few reasons that are most relevant to the case issue were given to them in the form of options in a multiple-choice question, and the stimulus factor was also included

as one of the options; thus, judges could give reasons by choosing one or multiple options. We also gave judges an open-ended space to write their own answers and to subsequently code their written responses. The key for our design was its comparison among reasons given by the judges in different groups. In Study 1, we found that, in response to the stimulus, judges in the treatment group and the baseline group decided the case differently. In their reasoning, however, few judges, if any, identified the stimulus factor as their decision reason. Instead, they interpreted the legal standard concerning liquidated damages differently. In Study 2, we found a similar pattern. This time, the judges justified their decisions by interpreting a legal concept differently, applying different sources of law, and drawing different conclusions from facts. In Study 3, to rationalize their decisions, judges inferred causation and foreseeability differently from the same facts. Taken together, these studies show that judges can use a wide array of legal techniques to rationalize their biased decisions.

Third, we included the stimulus factor in the short list of possible reasons the judges could choose. Although it can be deduced, given the between-group design, that the stimulus was the exact factor that drove the judges to decide differently, hardly any judges identified the factor as their reason. It is worth analyzing whether the judges ignored the stimulus factor intentionally or unintentionally. One possibility is that the judges hid their real decision reasons intentionally and that they also deliberately employed legal techniques to do so. In other words, the judges did not want to acknowledge their decision bias, even when they clearly knew they were biased. Another possibility is that the judges were unconsciously influenced by irrelevant factors and reached decisions that were implicitly shaped by their moral instincts, and they chose decision reasons corresponding to their initial instincts. In other words, the judges were not able to acknowledge their decision bias, even when the biasing factor was very salient. We think both explanations are plausible, and both of the behavior patterns are problematic to find in the practice of judicial justice (see also discussion in Section VI).

The participants in our experiments were real judges. This feature is crucial. As scholars such as Llewellyn (1940) have long emphasized, judges are different from laypeople in many respects: they are professionals with particular abilities, training, experience, and expertise. In particular, they are decisionmakers who are familiar with the task of giving reasons. They are also embedded in environments devised to improve the quality of their decisions (Kahan et al. 2016; Liu 2018). The use of real judges as subjects brought considerably greater realism and ecological validity to our study.

As mentioned earlier, prior experimental studies showed that judges can be influenced by extra-legal factors (Wistrich et al. 2015; Spamann & Klöhn 2016). These studies used judges in the United States as their study subjects (except Liu 2018). It is worth mentioning that a part of our study replicated the findings of prior studies in the context of China. Given the replication crisis in social sciences, this replication *per se* provides valuable confirmation to the literature. More importantly, there are many notable differences between judges in China and their peers in the United States. The U.S. judges assume important roles in their political system, and they also enjoy strong discretionary power when deciding cases. On the other hand, judges in China, residing on the periphery of China's centralized government system, are supposed to confine

themselves strictly to the role of applying laws (Liu et al. forthcoming). In this regard, judges in China should be less capable of manipulating decision outcomes—yet we still found they did so, and they did so with a range of legal techniques. This suggests that the finding in our studies can very likely be applied to judges in other countries.

It is commonly believed that the requirement to give reasons promotes better judicial decision making by keeping the judges' discretionary powers in check. Thus, judges are expected to provide the facts and arguments that support their decisions. However, the aim of constraining judicial power is compromised by the possibility that judges can use seemingly justifiable reasoning to rationalize their decision biases, which prevents rather than facilitates accountability. We discuss the implications of our findings in Section VI.

The rest of the article proceeds as follows. In Section II, we outline some high-level design choices before discussing details of the design in the subsequent subsections. Sections III to V explain the design of Studies 1 to 3 in detail, and analyze the experiment results. Section VI discusses the externality validity and implications of the experiments. Section VII concludes.

II. EXPERIMENT DESIGN

We conducted our experiments on five batches of judges attending short-term training classes at Shanghai Jiao Tong University, China. The judges were from five local or intermediate courts (in the Chinese judicial system, intermediate courts handle important local cases in the first instance and appeals cases) in different provinces (two northeastern provinces and three eastern provinces; we prefer not to reveal the exact courts the judges work in because our studies revealed some arguably undesirable features of some judges). One of our authors (Li) gave a two-hour lecture on legal professionalism as a part of every training. After the end of each lecture, he invited the judges to spend about 15 minutes participating in an experiment. The judges used their laptops, iPads, and other mobile devices to participate in the online experiment, administered via the Qualtrics platform.

The experiment took place on five separate days in the spring of 2018. We recruited subjects for Studies 1 and 2 at multiple classes because the class sizes were insufficient for our desired number of observations.¹ We usually asked each judge to participate in one experiment. However, judges attending one of the classes ($N = 38$) participated in two experiments (Studies 2 and 3), spending about 30 minutes. When judges were involved in two experiments, the sequence of the two experiments was randomized.² We

¹The exact number of the participating judges is as follows: Study 1—14 judges in Class 4; 15 judges in Class 5; Study 2—15 judges in Class 1; 18 judges in Class 2; 35 judges in Class 3; four judges in Class 5; Study 3—38 judges in Class 3.

²We do not find that the order of experiments has any significant effect on the decisions of the judges, that is, we do not find judges decide same cases differently when the cases were presented to them in different order.

Table 1: Summary Statistics of Judges' Demographics

<i>Variables</i>	<i>Study 1</i>	<i>Study 2</i>	<i>Study 3</i>
Age	34.0 (6.6)	36.9 (6.6)	37.2 (7.4)
Gender			
Female	47.8%	46.9%	56.3%
Male	52.2%	53.1%	43.8%
Education			
High school			
Bachelor's degree	87.0%	62.5%	46.9%
Master's degree	8.7%	37.5%	53.1%
PhD	4.3%		
Court Division			
Civil	82.6%	31.3%	40.6%
Criminal	13.0%	45.3%	56.3%
Administrative		12.5%	3.1%
Other	4.3%	10.9%	
Court Level			
Local	95.7%	6.3%	
Intermediate		68.8%	100.0%
Not reported	4.3%	25.0%	
<i>N</i>	23	64	32

NOTES: The statistics are based on the judges who submitted their demographic information. Standard deviations are in parentheses.

summarize the demographic information of the participants in Table 1, sorted by study. Some judges chose not to reveal their demographic information after they finished the experiment. The summary statistics in Table 1 reflect only the judges who submitted their information.

Our experiments employ a standard between-group design. The participating judges were randomly assigned to different experimental groups. They read cases vignettes adapted from real cases in China. The case materials were identical, with the exception of the stimuli. We tried to induce the judges in different conditions to make different decisions using stimuli of irrelevant factors, mostly defendant characteristics. In Study 1, we set up a baseline condition that did not include a stimulus, and a treatment condition that included a moral character. In Studies 2 and 3, we included moral characters in the opposite directions (good moral character vs. bad moral character) in the two experimental groups to induce contrary judicial decisions. We provided these two different types of experimental designs to improve the generalizability of our studies.

By way of background, in the Chinese judicial system, judges are responsible for deciding both factual and legal issues. The rules of evidence in China are stipulated separately by the Criminal Procedure Law, the Civil Procedure Law, the Administrative Procedure Law, and their related judicial interpretations. In principle, these rules require evidence to be objective, relevant, and legal. There are certain exclusionary rules for illegally obtained evidence. Yet, no specific rule covers relevance. This is partially because evidence laws in China regard truth-finding as the foremost value for litigation, and the method of appraising evidence takes the free evaluation of evidence through inner

conviction as a leading principle (Zhang & Walton 2010). In the Chinese system, judges can review any sort of evidence they deem necessary. The scope of inadmissible evidence is very limited.

The rules do not specify whether character evidence can be used to determine the facts of a case. Some legal scholars suggest judges should not consider character evidence of the defendant in a criminal trial; or, character evidence should be seen as irrelevant (Zhang & Liu 2017). Yet, no rule of admissibility specifically excludes character evidence. In practice, it is common for prosecutors in a criminal trial, and parties in a civil trial, to invoke arguments and evidence regarding a litigant's moral character. Anecdotes and empirical evidence also suggest that the moral character of a litigant unduly influence judges when making decisions (Liu 2018). More germane to our study, it is safe to say that no evidence rule in China specifically excludes the moral character evidence in the experiment stimuli, although, from the substantive law's perspective, these moral characters are legally irrelevant.

For the purposes of this study, the key aspect of our design was that we asked the judges to choose a reason (or reasons) for their decisions. We provided them with a multiple-choice question in each study, with a few decision reasons that are most relevant to the case issue as options. There were four to seven options in the questions, which covered two to three legal and factual issues debated in the case materials (and the irrelevant stimuli; see below), and the judges could select all the reasons that they thought should apply. Since our case materials were concise, and focused on one or two debated legal or factual issues, the reasons we provided in the options basically cover all case issues that are relevant. Yet, judges could also contribute their own reasons. That is, we also gave respondents an open-ended space to write their own answers and to subsequently code their written responses.³ Our design allowed us to see whether the judges in different groups gave similar or different reasons. That is, we determined whether their reasoning was in response to the stimuli.

Note that an alternative design would be to list the questions debated in the cases separately and ask the judges to choose yes or no as a simple answer. For example, in Study 1, we could first ask "Do you think the liquidated damages in this case are excessively higher than the actual loss?" and then ask the judges to decide yes or no. This design neglects the fact that, in practice, judges can choose not to mention a specific issue in their reasoning. Thus, it would induce judges to respond to a question they could avoid answering in actuality. Our current design reduces this problem in the sense that judges can select the reason or reasons that they think most relevant, instead of having to indicate yes or no to every reason. (We did, in fact, see some judges avoid mentioning certain issues in the options: e.g., in Study 1, some chose neither B nor D, avoiding the factual issue. See Section III.)

³Only 10 judges wrote their own reasons: three in Study 2, and seven in Study 3. We did not include these written reasons in the results presented below because coding the reasons requires much discretion, and we believe we should primarily report the objective results. That said, we also read the written reasons, and manually coded them into reasons for and against the defendant. Our results on reasoning are the same, or even stronger, after taking these written reasons into account.

Another key feature of our design is that we included the stimuli (the irrelevant factors) as an option in the multiple-choice questions. This design allowed us to see whether the judges were concealing or revealing these factors: given the experimental design, we knew in advance that these were the factors that would have driven them to decide differently in different conditions.

In total, we conducted three experiments. This was to ensure that our studies would cover a variety of areas of law and an array of legal techniques. Admittedly, we could not cover all areas of law or legal techniques. Yet we believe these experiments provide considerable generalizability to areas that we did not study directly.

III. INTERPRETATION OF A LEGAL STANDARD (STUDY 1)

We began by addressing an intuitive method that judges may employ: rationalizing their decisions by strategically interpreting a legal standard. The distinction between rule and standard has received substantial attention from legal scholars (e.g., Ehrlich & Posner 1973; Kaplow 1992; Fon & Parisi 2007). Laws are of necessity incomplete. Legislators cannot foresee all circumstances to which their laws could apply. They leave room for judges to decide cases according to specific facts and evidence. To capture these features, the definitions of rules and standards emphasize the distinction between whether the law is given content *ex ante* or *ex post* (e.g., Kaplow 1992). A rule may entail an advance determination of what conduct is permissible. A standard, usually more general and abstract, leaves the question of what conduct is permissible to the adjudicator. By definition, legal standards provide greater discretionary power to judges. We suspected that judges would use this power to manipulate decision outcomes. Study 1 was designed to document this phenomenon.

A. Case Scenario and Setup

In Study 1, a lessee (the plaintiff) and a lessor (the defendant) signed a lease for two commercial units. The parties (an individual and a firm) agreed on a set of terms, including a specific day for the delivery of occupancy and liquidated damages at RMB 300,000 yuan (about U.S. \$46,000). The rent was RMB 200,000 yuan per month. Then, two days before the specified move-in day, the lessor informed the lessee that the units could not be handed over on time because the last lessee had not yet moved. She asked for a delay of 10 days. The plaintiff claimed liquidated damages because it had made substantial preparation for moving, and as a private equity firm, the postponed opening of business would incur customer distrust, negatively influence the image of the company, and substantially affect its business.

The defendant claimed that the actual loss of the plaintiff included only potential compensation arising from its breach of contract with the moving company, totaling about RMB 20,000 yuan, as well as the loss of business opportunity due to the postponed move. However, the defendant claimed, the influence on the plaintiff's business was not substantial because the postponement was minor, and the plaintiff could still manage its business at the site of its previous office. Article 114, Paragraph 2 of the Contract Law

stipulates that where the amount of liquidated damages agreed on is excessively greater than the damages incurred, a party may petition the People's Court or an arbitration institution to make an appropriate reduction. The defendant argued that the liquidated damages of RMB 300,000 yuan stipulated in the contract were "excessively higher than the actual loss" of the plaintiff. She requested the court to lower the liquidated damages to RMB 20,000 yuan and was willing to compensate the lessee for 10 days' rent, about RMB 67,000 yuan.

In the treatment condition, the defendant was described as a person with bad moral character: she maintained an extramarital relationship with a corrupt government official. This information would be in a file that a judge sees since, in practice, it is a litigation strategy for litigants to provide evidence on the adverse moral character of their opponents, although such moral character is legally irrelevant and should not be taken into account by the court. The treatment also explicitly states that the extramarital relationship was irrelevant to the commercial units that were in dispute in this case. The stimulus was given at the end of the case vignette, in a separate paragraph.

[Treatment] The defendant Yang Li in this case did not appear at court. It was reported that the defendant maintained an extramarital relationship with the deputy director of the local land and resource bureau, Zhang. Zhang may have accepted bribes and possessed extensive assets with unclear sources. Zhang was being investigated and prosecuted by the procuratorate. Yang Li, who may have been involved in Zhang's crime, was suspected of accepting and soliciting bribes for Zhang. This case will be tried separately. The commercial units in this case were Yang Li's own, obtained through inheritance, and they were not related to Zhang's bribery case.

We included the sentence "The defendant Yang Li in this case did not appear at court" in the stimulus to make the description more coherent, because we later stated that the defendant was tried separately in a criminal trial. Chinese judges would understand that she was in custody and unable to appear at this civil trial. While the sentence helps consolidate our description, it inevitably brings in a confounding factor to the treatment. That is, the stimulus in the treatment includes both the defendant's moral character, and the fact that she was absent from trial. Thus, even if we find judges decide differently in response to the treatment, we cannot readily identify which factor has caused the result.

That said, we do not think this weakness in design would undermine our main findings. This is first because both the moral character and the appearance of the defendant are legally irrelevant factors that judges should not take into account when making a decision. More importantly, the focus of our design is to see whether judges use legal techniques to rationalize their readily biased decisions. We are less concerned with whether moral character or absence from court is the source of such biased decisions. In other words, as long as moral character and the other factors are irrelevant to the reasons judges gave, our main findings about legal techniques should be valid.⁴

⁴Another potential problem of our design in Study 1 is that we did not provide a balanced description of the defendant's moral character in the baseline condition. It is possible that the mere appearance of a legally irrelevant factor would have made the judges decide differently, regardless of the content of that factor (though we believe the chance is small). To address this problem, in Studies 2 and 3, we provide more balanced descriptions of moral character in different experimental groups.

The judges were asked to decide the case and provide their reasoning by choosing one of the reasons given below:

Your decision reasons are (select all that apply):

- A. The amount of liquidated damages agreed upon, RMB 300,000 yuan, is excessively higher than the actual damages incurred.
- B. The defendant did not have a substantial influence on the business of the plaintiff through the breach of the contract.
- C. The amount of liquidated damages agreed upon, RMB 300,000 yuan, is not excessively higher than the damages incurred.
- D. The defendant had a substantial influence on the business of the plaintiff through the breach of the contract.
- E. [Only for the treatment group.] The defendant maintained an extramarital relationship with the deputy director of the local land and resource bureau, Zhang, and was implicated in Zhang's bribery trial.
- E/F. Other reasons (please specify below).

B. Results

The stimulus in the treatment condition was designed to induce the judges to decide against the defendant, although it was technically irrelevant to the current case. The results showed that this design was successful. In the treatment group, 38.5 percent of judges decided that the plaintiff should win the case and the defendant should fully compensate the liquidated damages RMB 300,000 yuan, whereas none of the judges in the baseline condition made this decision. The difference was found to be statistically significant at the 0.05 level in the Fisher's exact test (Table 2, Column 1). We also asked the judges to give a specific compensation award when they did not support the awarding of the entire liquidated damage. The average damages granted by the judges in the treatment condition were about 185,000 yuan, whereas this number in the baseline condition was about 85,000 yuan.⁵ The difference was significant at the 0.01 level (Table 2, Column 1). These results successfully replicated the findings of previous studies that have employed a similar strategy by using extra-legal factors to induce a different decision (e.g., Nadler & McDonnell 2011; Wistrich et al. 2015; Spamann & Klöhn 2016; Liu 2018).

For our analysis, the key was the reason each judge chose for his or her decision. Chinese contract law authorizes judges to reduce liquidated damages they believe to be excessively greater than the actual loss. Lawyers in China suggest that they frequently encounter use of this doctrine in practice (e.g., Luo 2016). In response to the elements of this law, we generated two reasons that have been debated in the case materials and

⁵Where a judge supported the liquidated damages, the award was considered to be RMB 300,000 yuan.

Table 2: Decision Outcomes in Studies 1, 2, and 3

	<i>Study 1</i>	<i>Study 2</i>	<i>Study 3</i>
Plaintiff Wins/Conviction (%)			
(1) Bad character	38.5	64.9	23.8
(2) Baseline/good character	0.0	37.1	0.0
Fisher's exact	0.011	0.033	0.053
(1)/(2) (<i>N</i>)	13/16	37/35	21/17
Sentences			
(3) Bad character	18.5 (2.7)	1.5 (0.3)	2.3 (0.3)
(4) Baseline/good character	8.5 (1.0)	0.4 (0.1)	1.4 (0.3)
(3) - (4)	10.0***	1.0***	0.9*
<i>t</i> statistic	3.742	3.307	1.975
<i>p</i> value	0.001	0.002	0.056
(3)/(4) (<i>N</i>)	13/16	37/34	21/17

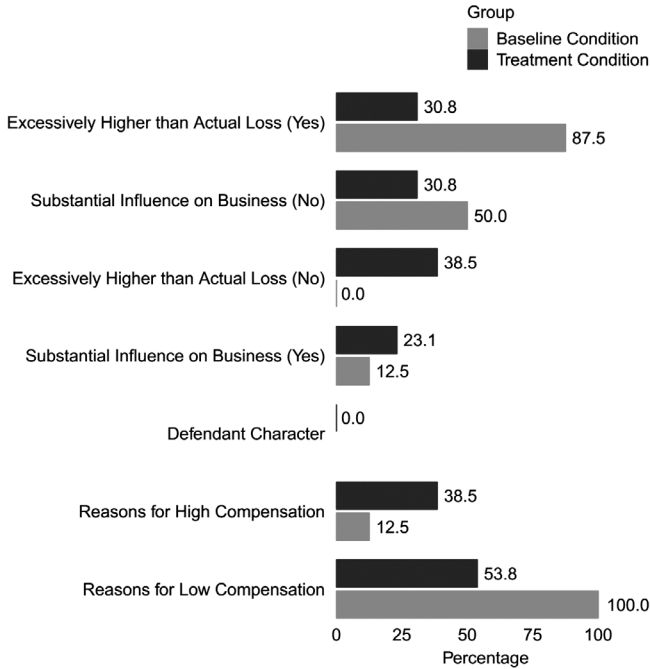
NOTES: In Study 1, judges in the treatment group received a stimulus designed to induce negative feeling about the defendant's moral character, while judges in the baseline group did not receive this stimulus. In Studies 2 and 3, judges in the bad character group received a stimulus designed to induce negative feeling about the defendant's moral character. Judges in the good character group received a positive stimulus; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

placed them in the options. One is whether the amount of liquidated damages is “excessively higher” than actual loss. The other is whether the late delivery had a “substantial influence” on the business of the plaintiff. Both reasons—excessively higher and substantial influence—are legal standards that are subject to interpretation, and they are correlated with each other. Excessiveness is a normative judgment that can be made only in relation to specific facts. The judges decided whether damages were excessive not only by looking at the monetary amount of the damages, they also were enjoined to examine the relationship between that amount and the facts of the case. Similarly, the issue of substantial influence—whether the actual loss incurred was substantial enough to justify the amount of liquidated damages—is relative to the amount of liquidated damages, and the relationship is defined by whether the amount is excessive. In the following analyses of the results, we first show the results for the two issues separately and then pool them together to show the aggregated results.

We observed whether the judges in the treatment and baseline conditions would interpret (choose) the legal standard in the case differently in response to the treatment. The judges were allowed to choose any option(s) that they thought should apply. This means that they could choose all options, even though some were mutually contradictory (e.g., Option A was the polar opposite of Option C), or they could choose any combination (including no options at all).

Figure 1 and Table 3 report the results. We found that judges chose different interpretations of the legal standard “excessively higher” in response to the stimulus. Judges in the treatment group were less likely to respond that the liquidated damages were excessively higher than the actual loss, a decision reason that would be in favor of the defendant (30.8 percent in the treatment vs. 87.5 percent in the baseline; the difference was statistically significant at the 0.01 level in the Fisher's exact test). Additionally, judges in the treatment group (38.5 percent) were more likely to choose Option C, “the liquidated

Figure 1: Decision reason in Study 1.



damages are not excessively higher than the actual loss,” a decision reason that would be unfavorable to the defendant. None of the 16 judges in the baseline condition chose this option. Note that about 30.7 percent of judges in the treatment group ($100\% - 30.8\% - 38.5\%$) and 12.5% ($100\% - 87.5\% - 0.0\%$) of judges in the baseline condition chose neither Option A nor Option C, avoiding mentioning this legal issue at all when giving reasons.⁶

We found a similar pattern in the second issue with respect to substantial influence. 30.8 percent of judges in the treatment condition selected the option that the defendant did not have substantial influence on the business of the plaintiff. This number was 50.0 percent in the baseline condition, but the difference was statistically insignificant ($p = 0.451$).⁷ The insignificance result is likely due to the small sample size we have in the experiment.

⁶We also conducted a three-way analysis using the Fisher’s exact test on: excessively higher (yes) (choosing Option A), excessively higher (no) (choosing Option C), and no opinion (choosing neither A nor C). The difference between the treatment and the baseline was statistically significant ($p = 0.003$). See Table A1 in the Appendix.

⁷We also conducted a three-way analysis using the Fisher’s exact test on: substantial influence (no) (choosing Option B), substantial influence (yes) (choosing Option D), and no opinion (choosing neither B nor D). The difference between the treatment and the baseline was statistically insignificant ($p = 0.692$). See Table A1 in the Appendix.

Table 3: Decision Reasons in Study 1

	(1) Excessively Higher than Actual Loss (Yes)	(2) Substantial Influence on Business (No)	(3) Excessively Higher than Actual Loss (No)	(4) Substantial Influence on Business (Yes)	(5) Defendant Character	(6) Reasons for High Compensation (3) (4)	(7) Reasons for Low Compensation (1) (2)
Treatment	30.8	30.8	38.5	23.1	0.0	38.5	53.8
Baseline	87.5	50.0	0.0	12.5	—	12.5	100.0
Treatment – Baseline	-56.7***	-19.2	38.5**	10.6	—	26.0	-46.2***
Fisher's exact	0.003	0.451	0.011	0.632	—	0.192	0.004
Treatment/Baseline(N)	13/16	13/16	13/16	13/16	13/16	13/16	13/16

NOTES: Judges in the treatment group received a stimulus designed to induce negative feeling about the defendant's moral character, while judges in the baseline group did not receive this stimulus; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

When we pooled the reasons for the same direction together, we found that in the treatment condition, 53.8 percent of judges selected the reasons that were in favor of the defendant (reasons for low compensation), that is, 53.8 percent of judges chose either Option A or Option B (or both), while 100 percent of judges in the baseline condition selected reasons that were in favor of the defendant. The difference was statistically significant at the 1 percent level. On the other hand, we do not find a similar pattern with respect to reasons for high compensation (Options C and D).⁸

An interesting pattern can be found in the option regarding the defendant's moral character. In the treatment condition, we provided Option E, which concerned the extramarital relationship of the defendant and her being implicated in a corruption case. We know from the between-group experimental design that this element is most likely the factor that drives the judges in the treatment group to make a harsher decision. However, no judge chose this option when asked to provide decision reasoning. A related question is whether the judges concealed their real decision reason and replaced it with the interpretation of a legal standard intentionally or unintentionally. We will defer the discussion of intentionality to Section VI.

IV. LEGAL CONCEPT AND APPLICABILITY OF LAW (STUDY 2)

Study 1 tested the use of a legal standard. Study 2 had two goals. The first was to replicate the findings of Study 1 in a criminal law case. The second goal was to document other techniques that judges may employ when rationalizing their decisions. Here, we focus on the interpretation of legal concepts and the applicability of certain rules, which are common techniques in practice.

Judges have certain discretionary power in interpreting legal concepts. For example, a judge can decide whether the concept of "human" includes the fetus, so that he can decide whether abortion is legal. Another judge can decide whether the declaration of "all *men* are created equal" is applied to all humanity, or is supposed to exclude women and children. Similarly, in many cases, judges also have discretionary power in deciding whether a specific rule or a specific precedent fits the case issue at hand, and whether they should apply this rule or precedent. In other words, both legal concept and applicability of law leave room for judges to interpret the law and reach their own conclusions. We suspect judges can easily use their techniques in interpreting legal concepts and deciding applicable law to rationalize their biased decisions. Study 2 is designed to provide evidence for this hypothesis.

⁸We also conducted a four-way analysis using the Fisher's exact test on: reasons for high compensation (choosing Options C or D, or both), reasons for low compensation (choosing Options A or B, or both), no opinion, and both (choosing from both reasons for high compensation and reasons for low compensation). The last combination is possible because one can rationally choose, for example, both excessively higher (yes) and substantial (yes); the former is a reason for low compensation, and the latter is a reason for high compensation. The difference between the treatment and the baseline was statistically significant ($p = 0.004$). See Table A1 in the Appendix.

A. Case Scenario

In Study 2, the defendant, Wang Peng, having always loved parrots, took home a female parrot found in the yard of the factory where he worked. He bought a male parrot online to mate with the female. The parrots reproduced very rapidly. One year later, Wang Peng already had 43 juvenile parrots. Later, he sold six parrots to his friend for RMB 2,000 yuan. According to the police, two of the six parrots were little golden parakeets, officially called the green-cheeked parakeet, a protected species, listed in the Appendix to the Convention on International Trade in Endangered Species of Wild Fauna and Flora. The police detained Wang Peng on suspicion of the crime of “illegal sale of precious and endangered wild animals and their related products.”

In court, the prosecution and the defense debated several issues. The first was whether domestically bred species were wild animals, as defined in criminal law. The prosecution cited Article 1 of the Interpretation of Several Issues Concerning the Specific Application of Law in the Trial of Criminal Cases Violating Wildlife Resources of the Supreme People’s Court (SPC), which explicitly included domestically bred species within the scope of protection of criminal law.

The defendant argued that, first, criminal law stipulated that the object of the crime is precious and endangered wild animals. Categorizing domestically bred animals as wild animals is self-contradictory. Second, in its interpretation, the SPC considered domestically bred animals to be wild animals. This interpretation goes far beyond the text of the criminal law. It violates the principle of *nulla poena sine lege* (no penalty without a law). Thus, the interpretation is not applicable.

Another issue is the factual one: Was the defendant aware of the illegality of his behavior? The Criminal Law of China defines two types of crime, differing according to the mental state of a defendant. An intentional crime is one committed as a result of clear knowledge that one’s own act will cause socially dangerous consequences, and with hope for or indifference to the occurrence of those consequences (Article 14). A negligent crime occurs when one should foresee that one’s act may cause socially dangerous consequences, but fails to do so (Article 15). Regarding the specific offense, when no mental state is specified in the statute, but seems to be an element of the offense, it should be assumed that the offense is stipulated as an intentional crime, in which “clear knowledge” and “with hope for” (or, “indifference”) are the required elements (Article 15).

The defendant claimed that he had raised many types of parrots. He did not know that he had bred protected ones, and he did by no means recognize the illegality of his behavior, which therefore should not be punished as a crime. By contrast, the prosecution argued that Wang Peng was a parrot lover who had knowledge of specific breeding and feeding techniques. He also frequently communicated his breeding experience to other parrot lovers online. It was unreasonable to believe that he did not know that the parrots he was raising were considered nationally protected animals. The current evidence suggests that Wang Peng knew or must have known of the illegality of his behavior. The prosecution suggested that the defendant should be sentenced to a five-year fixed-term imprisonment.

B. Setup

We used a between-subject design. Different descriptions of Wang Peng's motive for selling the parrots were included in the case vignettes. In the good moral character condition, the defendant was described as a good father, whereas he was described as a gambler in the bad moral character condition.⁹ According to the criminal law of China, judges should only consider the elements of the crime when making verdicts, and, in general, moral character is not an element of any crime. In other words, judges should not take the defendant's moral character into account when making a verdict. Applying this general rule to the current case, judges should not consider whether the defendant is a good father or a gambler when deciding whether the defendant committed the crime of "illegal sale of precious and endangered wild animals and their related products." The criminal law also prescribes the upper and lower bounds of a sentence for a given crime. Within these bounds, judges can consider aggravating factors such as repeat offenses, and mitigating factors such as voluntary surrender and first-time offender status. The law does not specially explain whether a defendant's moral character can be considered as an aggravating or mitigating factor. In practice, judges usually have the discretion to consider moral character evidence that is relevant to the case when giving sentences.

[Good Moral Character Condition] Wang Peng further explained that his daughter suffered from systematic lupus erythematosus, and she needed his care and company. He did not have time to take care of the juvenile parrots, so he sold them. He hoped that the sitting judge would not convict him.

[Bad Moral Character Condition] Wang Peng had been indulging in gambling online since 2015, was divorced from his wife, and had no time to take care of the parrots. He thus sold the juvenile parrots.

We included "He hoped that the sitting judge would not convict him" in the good moral character condition, but not in the bad moral character. This additional statement was made to avoid abruptness in the inclusion of the defendant's moral character, and to make the description of moral character coherent with the other parts of the case material. (The bad moral character was inserted as a part of the prosecution's charge, and already looked natural.) The additional statement may have created a confounding factor to our stimulus in the good moral character condition, but we believe this confounding factor was not substantial, and was worthwhile. Moreover, the statement was also legally irrelevant, and its inclusion would not have undermined our main design on reasoning.

The judges were asked to decide the case and choose the reasons for their decision from among the following options. The defendant's moral character (whether the dependent is a good father or a gambler) is certainly irrelevant to these reasons.

⁹It may be worth explaining that we also indicated that the defendant divorced his wife as a consequence of gambling. Divorce, by itself, is not a sign of bad moral character in China.

Your decision reasons are (select all that apply):

- A. The species the defendant bred was a wild animal.
- B. The Interpretation of Several Issues Concerning the Specific Application of Law in the Trial of Criminal Cases Violating Wildlife Resources of the SPC was applicable in this case.
- C. Wang Peng was aware of the illegality of his behavior. That is, to a certain extent, he knew the animals he was raising and selling were protected animals.
- D. The species the defendant bred was not a wild animal.
- E. The Interpretation on Several Issues Concerning the Specific Application of Law in the Trial of Criminal Cases Violating Wildlife Resources of the SPC excessively expands the scope of wild animals, and it was not applicable in this case.
- F. Wang Peng was not aware of the illegality of his behavior. That is, he did not know the animals he raised and sold were protected animals.
- G. [For the good character condition] Wang Peng's daughter suffered from systematic lupus erythematosus. Wang Peng is sympathetic.
[For the bad character condition] Wang Peng gambled online. He did not have a good moral character.
- H. Other reasons (please specify below).

C. Results

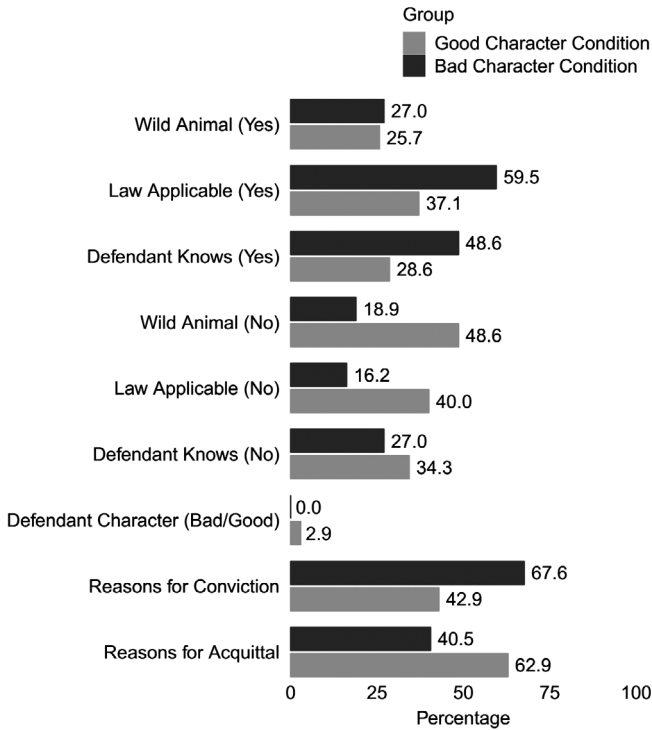
The experimental design was intended to induce judges to decide differently in response to the stimuli, based on the defendant's moral character, which should be irrelevant to the current case. The results of the judges' decisions are consistent with the design. In the good character condition, 37.1 percent of the judges decided that the defendant should be convicted, but 64.9 percent of judges made this decision in the bad character condition. The difference was found to be statistically significant at the 0.05 level in the Fisher's exact test (Table 2). We also asked the judges to give a specific sentence, if they did not acquit the defendant. The average sentence granted by the judges in the bad character condition was about 1.5 years, whereas the good character condition it was 0.4 years.¹⁰ The difference was significant at the 0.01 level (Table 2). These results show that the report of the irrelevant information induced a different judicial decision.¹¹

We covered three possible decision reasons in the options: the interpretation of a legal concept, that is, whether the parrots in this case should be judged wild animals; the

¹⁰If a judge acquitted the defendant, the sentence was counted as 0.

¹¹As stated above, criminal law in China prohibits judges from considering a defendant's moral character when rendering a verdict, but gives judges certain discretion in considering moral character in sentencing. On this view, the difference in verdicts between the experiment groups suggests decision biases. However, the differences in sentencing might not be considered as a bias.

Figure 2: Decision reason in Study 2.



applicability of a law, that is, whether the interpretation by the SPC was applicable; and the finding of fact, that is, whether the defendant was aware of the illegality of his behavior. Judicial interpretation is an official source of law in China that a court should follow when deciding a specific case. It is not precedent but, instead, abstract stipulations adopted by the SPC following a formal procedure outside a particular dispute (Liu 2016). Figure 2 and Table 4 summarize the results.

First, we found that the judges interpreted the concept of wild animals differently in response to the stimuli. In the bad moral character condition, it was found that 27.0 percent of judges responded that the parrots in this case were among the wild animals whose sale was proscribed in criminal law (Option A). In the good moral character condition, this option was chosen by 25.7 percent of judges. The difference was statistically indistinguishable ($p = 1.000$). On the other hand, 18.9 percent of judges in the bad character condition responded that the parrots in the case were not wild animals (Option D), but 48.6 percent of good-character-condition judges chose this option. The difference was significant at the 0.05 level ($p = 0.012$). This finding suggests that judges in the good character condition were more likely to reason that the concept of a wild animal in criminal law did not include domestically bred species.

Table 4: Decision Reasons in Study 2

	(1) Wild Animal (Yes)	(2) Law Applicable (Yes)	(3) Defendant Knows (Yes)	(4) Wild Animal (No)	(5) Law Applicable (No)	(6) Defendant Knows (No)	(7) Defendant Character (Bad/Good)	(8) Reasons for Conviction (1) (2) (3)	(9) Reasons for Acquittal (4) (5) (6)
Bad character condition	27.0	59.5	48.6	18.9	16.2	27.0	0.0	67.6	40.5
Good character condition	25.7	37.1	28.6	48.6	40.0	34.3	2.9	42.9	62.9
Bad – Good	1.3	22.3*	20.1*	-29.7**	-23.8**	-7.3	-2.9	24.7*	-22.3*
Fisher's exact	1.000	0.065	0.095	0.012	0.035	0.611	0.486	0.057	0.065
Bad/Good (N)	37/35	37/35	37/35	37/35	37/35	37/35	37/35	37/35	37/35

NOTES: Judges in the bad character group received a stimulus designed to induce negative feeling about the defendant's moral character. Judges in the good character group received a positive stimulus; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Further, the statistics suggest that the judges in the bad character group were less certain of whether the parrots in the case should be counted as wild animals for the purposes of criminal law. Only 45.9 percent (27.0% + 18.9%) of the judges referred to this issue (by choosing Options A or D). The rest avoided it. In the good character group, by contrast, the judges (25.7% + 48.6% = 74.3%) showed greater willingness to use this reason to acquit the defendant.

Judges also decided the applicability of the law differently. In the bad character group, 59.5 percent of judges decided that the interpretation issued by the SPC, a law that was unfavorable to the defendant, was applicable (Option B), whereas the number was 37.1 percent in the good character group ($p = 0.065$). In all, 16.2 percent of judges in the bad character group concluded that the interpretation of the SPC was not applicable (Option E), and this number was 40.0 percent for the good character condition ($p = 0.035$).

Third, we have some preliminary evidence that the judges addressed the same facts differently for the two conditions. In the bad character group, 48.6 percent of judges thought that the defendant was aware of the illegality of his behavior (Option C). This number was 28.6 percent for the good character group. The difference was marginally significant at the 0.10 level ($p = 0.095$). Yet we did not find a symmetrical pattern for Option F (that the defendant was unaware of the illegality of his behavior).¹²

When we pooled the reasons for the same direction together, we found that in the bad character condition, 67.6 percent of judges selected reasons for conviction, that is, 67.6 percent of judges chose either Option A, Option B, or Option C (or any combination of the three options), while only 42.9 percent of judges in the good character condition selected these reasons for conviction. The difference was statistically significant at the 0.10 level. We found a similar pattern with respect to reasons for acquittal (Options D, E, and F).¹³

Fourth, only one of the 72 judges chose the moral character of the defendant as part of his or her decision reason (Option G). Yet we know from the experimental design that his moral character is most likely the reason driving the judges to decide differently. The judge who decided based on character was in the good moral character condition. He or she used this reason to acquit the defendant.

¹²We also conducted three-way analyses (yes, no, and no opinion) using the Fisher's exact test on the three issues, respectively. The difference between the bad character condition and the good character condition is statistically significant with respect to the interpretation of wild animal ($p = 0.016$) and the applicability of the law ($p = 0.070$), but not significant with respect to the awareness of the defendant ($p = 0.214$). See Table A2 in the Appendix.

¹³We also conducted a four-way analysis using the Fisher's exact test on: reasons for conviction, reasons for acquittal, no opinion, and both (choosing from both reasons for conviction and reasons for acquittal). The last combination is possible because one can rationally choose, for example, both wild animal (yes) and law applicable (no); the former is a reason for conviction, and the latter is a reason for acquittal. The difference between conditions is statistically significant ($p = 0.083$). See Table A2 in the Appendix.

V. CAUSATION AND FORESEEABILITY (STUDY 3)

In the studies described above, we found that judges can use the interpretation of the law to rationalize their biased decisions. We suspected that finding of fact could function similarly. We used Study 3 to test this hypothesis. In particular, we focused on the inference of causation and foreseeability. Note that causation can be either a factual (e.g., cause in fact) or a legal issue (e.g., proximate cause). Yet this distinction is not incorporated in the formal legal system in China, despite the existence of many theoretical debates on this issue (e.g., Zhang 2013; Lao 2015).

Causation and foreseeability are pervasive concepts in the law: they are fundamental elements in tort, contract, insurance, and criminal law. Although fundamental, causation and foreseeability notoriously lack specific standards to guide outcomes. For example, inquiry into causation typically focuses on one or more of the following standards: whether there is a but-for relationship between an act and a result, the remoteness of the result, the extent to which the result seems accidental, the degree of dependence of the result on another person's act, and whether there is a concurrent actual cause, among others. These factors, however, are difficult to apply with any degree of precision. In practice, inquiry into causation may collapse into an intuition about whether it seems just to hold the actor liable for the result, given the circumstances of the case (Robinson 2008; Nadler & McDonnell 2011). We suspect that causation and foreseeability are convenient concepts for judges to manipulate.

A. *Case Scenario*

The defendant, Wang Shuqiang, living in a village, put two oxygen tanks near the wall in the courtyard of his house. In the middle of the night, they exploded, which led to a serious fire. According to the firefighting department, the explosion of the oxygen tanks was caused by a cigarette butt left by a passer-by outside the wall, but the specific person could not be found. The explosion of the oxygen tanks first set the attachments outside the wall on fire, which further expanded to the houses around.

The plaintiff, Zhao Xiaohua, living in the same village as the defendant, participated in the firefighting that night. It was dark as the plaintiff was carrying water to the site of the fire, and he fell into the roadside ditch. The fall caused several fractures in the bones of his thighs and wrist, and his thigh bone pierced his muscles and skin. The plaintiff was taken to a hospital.

The plaintiff claimed that his fall was the result of his contribution to the firefighting. There was a causal relationship between the defendant's storage of his oxygen tanks, the fire, and the plaintiff's injury. When the defendant stored the oxygen tanks, he should have foreseen the danger. Hence, the plaintiff asked the defendant to compensate him RMB 44,000 yuan, which would cover treatment costs and other expenses.

The defendant argued that there was no causal relationship between his oxygen tanks and the fire; it was also unreasonable to assume a causal link between the fire and the plaintiff's injury. In particular, some villagers had told him that the plaintiff had been at a celebration and had drunk some alcohol that night. Hence, his fall could have been

caused his drinking. Furthermore, the defendant argued that it was impossible for him to have foreseen that the plaintiff would fall and be injured in firefighting.

The law in China separates civil and criminal liability. A civil case should be tried according to civil procedure at the civil division of a court, and a criminal case should be tried according to criminal procedure at the criminal division of a court. The case in Study 3 is a civil case (torts), and the judges can understand this from the case materials and the questions.

B. Setup

We used a between-subject design. Different descriptions of the motive for the defendant to store the oxygen tanks were plugged into the case vignette. In one condition, the defendant was described as a son who was storing oxygen to better care for his mother's lung disease; in the other condition he did so to produce crystal meth.

[Good Moral Character Condition]

(In the first paragraph) Investigation found that the defendant's mother had suffered from lung disease for years and often needed oxygen at home. The defendant was storing the two oxygen tanks to take better care of his mother.

(In the last paragraph) The defendant further claimed that he had given up all opportunity to work outside the village to take better care of his mother. All his neighbors recognized that he was a good person. He hoped that the sitting judge would consider this factor.

[Bad Moral Character Condition]

(In the first paragraph) Investigation found that the defendant was illegally producing small amounts of crystal meth (methamphetamine) at home, and the two oxygen tanks were to be used for this production. This criminal case was being investigated separately by the police.

(In the last paragraph) The plaintiff further claimed the defendant's illegal production of crystal meth (methamphetamine) was responsible for his injury. The defendant should be punished, and he should compensate him for the costs.

The manipulations in Study 3 were first presented in the first paragraph of the case material from a neutral perspective. In the last paragraph, we reiterated the manipulation from the plaintiff's and the defendant's perspectives, with some variations, to make the statement more natural. Presenting the moral character from different perspectives may have created a confounding factor to our manipulations, but we believe it was not substantial. Moreover, the perspective of this presentation was also a legally irrelevant factor, the inclusion of which would not have undermined our main design on reasoning.

The judges were asked to decide the case and choose their decision reasons from the following options:

Your decision reasons are (select all that apply):

- A. There is a causal relationship between the defendant's act of oxygen storage and the plaintiff's injury.
- B. The defendant could have foreseen the possibility that someone (the plaintiff) would be injured.

- C. Although the plaintiff had drunk some alcohol, the defendant's act of oxygen storage was the major reason of the plaintiff's injury.
- D. There was no causal relationship between the defendant's act of oxygen storage and the plaintiff's injury.
- E. The defendant could not have foreseen any possibility that someone (the plaintiff) would be injured.
- F. The plaintiff's drinking of alcohol is the major reason for his injury.
- G. [For the bad character condition] The defendant produced crystal meth at home and had poor morals. His behavior should be punished.
[For the good character condition] The defendant was taking care of his mother. He had good morals. It would be unreasonable to ask him to make compensation.
- H. Other reasons (please specify below).

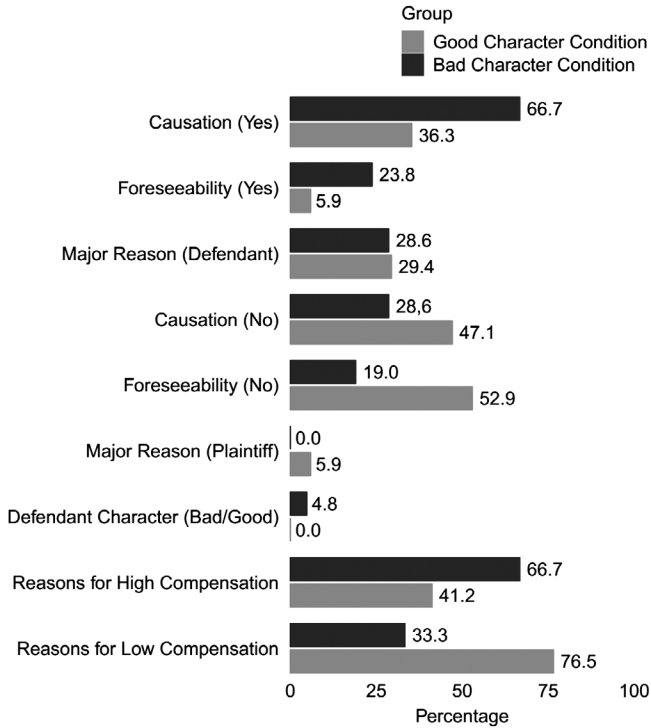
C. Results

The experimental design was intended to induce the judges to decide differently in response to the stimuli. The judges' decisions were consistent with our design. In the condition with a defendant of a good moral character, no judges determined that the plaintiff should win, that is, that the defendant should pay full damages (RMB 44,000 yuan) to the plaintiff. In the condition with a defendant of a poor moral character, however, 23.8 percent of the judges granted the plaintiff full compensation ($p = 0.053$). We also asked the judges to give a specific damage amount. In the bad character condition, the average damage amount was about 23,000 yuan, whereas in the good character condition, the average damage amount was about 14,000 yuan ($p = 0.056$). (The average calculation includes judges who awarded zero damages.)

Three issues were covered in the options for decision reasons: causation, that is, whether there was a causal relationship between the defendant's act of oxygen storage and the plaintiff's injury; foreseeability, that is, whether the defendant could have foreseen the possibility that someone would be injured because of a fire caused by the oxygen tanks; and a separate factual issue of whether the defendant's act (oxygen storage) or the plaintiff's (having drunk alcohol that night) was the major reason for the plaintiff's injury. Note that we did not frame this as an issue about concurrence cause. We used the ambiguous concept of major reason, frequently used in Chinese legal practice, to frame it as a general factual question. Figure 3 and Table 5 summarize the results.

First, we found some evidence that the judges interpreted causation differently in response to the stimuli. In the bad character condition, 66.7 percent of judges found a causal relationship between the defendant's act and the plaintiff's injury (Option A). In the good character condition, however, only 35.3 percent of judges chose this option. The p value of the difference was 0.101. On the other hand, 28.6 percent of the judges in the bad character condition thought that there was no causal relationship between the defendant's act and the plaintiff's injury (Option D). For the good character condition, this number was 47.1 percent. However, this difference was not statistically significant.

Figure 3: Decision reason in Study 3.



Second, the judges addressed foreseeability differently: 19.0 percent of the judges in the bad character condition chose Option E, that is, they thought that the defendant could not have foreseen the possibility that someone would be injured. Significantly more judges (52.9 percent) chose this option in the good character group ($p = 0.042$). Yet no significant difference regarding Option B (the defendant could have foreseen the possibility that the plaintiff would be injured) was found between the judges in the two experimental groups.

Third, we find no statistically distinguishable difference between the judges in different conditions regarding the factual issue of the “major reason.” Option C was chosen by 28.6 percent of the judges in the bad character group and by 29.4 percent of the judges in the good character group. Almost no judge chose Option F for either condition.¹⁴

¹⁴We also conducted three-way analyses (yes, no, and no opinion) using the Fisher’s exact test on the three issues, respectively. The difference between the bad character condition and the good character condition was statistically significant with respect to foreseeability ($p = 0.081$), but not significant with respect to causation ($p = 0.129$) and major reason ($p = 0.845$). See Table A3 in the Appendix.

Table 5: Decision Reasons in Study 3

	(1) Causation (Yes)	(2) Foreseeability (Yes)	(3) Major Reason (Defendant)	(4) Causation (No)	(5) Foreseeability (No)	(6) Major Reason (Plaintiff)	(7) Defendant Character (Bad/Good)	(8) Reasons for High Compensation (1) (2) (3)	(9) Reasons for Low Compensation (4) (5) (6)
Bad character condition	66.7	23.8	28.6	28.6	19.0	0.0	4.8	66.7	33.3
Good character condition	35.3	5.9	29.4	47.1	52.9	5.9	0.0	41.2	76.5
Bad - Good	31.4	17.9	-0.8	-18.5	-33.9**	-5.9	4.8	25.5	-43.1**
Fisher's exact	0.101	0.197	1.000	0.318	0.042	0.447	1.000	0.190	0.011
Bad/Good (N)	21/17	21/17	21/17	21/17	21/17	21/17	21/17	21/17	21/17

NOTES: Judges in the bad character group received a stimulus designed to induce negative feeling about the defendant's moral character. Judges in the good character group received a positive stimulus; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

When we pooled the reasons for the same direction together, we found that in the bad character condition, 66.7 percent of judges selected reasons supporting high compensation, that is, 66.7 percent of judges chose either Option A, Option B, or Option F (or any combination of the three options), while only 41.2 percent of judges in the good character condition selected these reasons. The difference was statistically insignificant. In the bad character condition, 33.3 percent of judges selected reasons for low compensation, that is, 33.3 percent of judges chose either Option C, Option D, or Option E (or any combination of the three options), while 76.5 percent of judges in the good character condition selected these reasons. The difference was statistically significant at the 0.05 level.¹⁵

Only one of the 38 judges cited the moral character of the defendant as a decision reason (Option G). Yet we know from the experimental design that moral character was most likely the reason that drove the judges to decide differently. The sole judge who chose Option G was in the bad character condition. It is likely that he or she was using this reason to make a harsher decision.

VI. DISCUSSION

A. External Validity

The studies presented employ randomized experiments that provided considerable internal validity. The challenge to these studies is their external applicability. In experimental studies, we inevitably omit certain potentially important differences between experiments and the real world. For example, in the courtroom, judges consider more detailed and complicated factual materials and have more time to deliberate before reaching a conclusion (Spamann and Klöhn [2016] took this into account in their experiment). Judges also have a greater incentive to make accurate judgments in the real world, which likely makes them consider actual cases more carefully. Moreover, they are subject to procedural requirements that are designed to limit their power and arbitrariness (Liu 2018). All these factors contribute to making judges prudent and to reducing their decision biases.

That said, we believe our studies still provide strong external applicability. Our case materials were adapted from real cases in China. We provided judges with a considerable time to decide the cases (about 15 minutes). Our experimental procedures allowed the judges to alter their decisions after they gave their reasons (i.e., after they deliberated). Moreover, many past studies have used experiments to find causal relations among emotions, impulses, heuristics, and biased decisions in laboratory settings (e.g., English et al.

¹⁵We also conducted a four-way analysis using the Fisher's exact test on reasons for high compensation, reasons for low compensation, no opinion, and both (choosing from both reasons for high compensation and reasons for low compensation). The last combination was possible because one can rationally choose, for example, both causation (yes) and major reason (defendant); the former is a reason for high compensation, and the latter is a reason for low compensation. The difference between conditions was statistically significant ($p = 0.033$). See Table A3 in the Appendix.

2006; Guthrie et al. 2007; Kahan 2010, 2013; Nadler & McDonnell 2011; Rachlinski et al. 2011; Sood & Darley 2012; Wistrich et al. 2015). Researchers using more realistic settings (Spamann & Klöhn 2016) and using real-world empirical evidence (Leibovitch 2016) have confirmed that the judicial decision patterns found in the laboratory are similar to those found in the real world. These findings reduce our concerns over external validity.

We used particular case scenarios and legal issues in the present experiments. A natural problem is whether the patterns found in these scenarios and regarding these legal issues can be extended to other cases and issues. We do not think this concern will jeopardize the applicability of our findings. One feature of experimental study (or any empirical study) is that it uses a specific scenario to test general theories. In our studies, we carefully designed three scenarios that cover questions of both civil and criminal law: a contract law case regarding liquidated damage, a criminal law case regarding illegal sales of protected animals, and a tort case regarding causation and foreseeability. We also designed our experiments to cover both the interpretation of law and findings of fact, which together account for the bulk of legal training and practice. We found similar patterns for all three of these cases. It is reasonable to think that these behavioral patterns have strong generalizability.

B. Hiding and Rationalizing: Intentionally or Unintentionally?

We found that the participant judges used a variety types of techniques to rationalize their arguably biased decisions, and they rarely mentioned the factors actually driving their decisions. It is worth discussing whether they did so intentionally or unintentionally.

One possibility is that the judges hid their real decision reasons intentionally and that they also deliberately employed legal techniques to do so. We provided the real decision reasons (moral character) among the options, but almost all judges avoided choosing these options. This behavioral pattern may suggest that judges were already aware that they should not be influenced by an irrelevant factor, but the knowledge did not alter their decisions. They still decided the cases in the direction of the irrelevant factor and then chose other reasons to cloak their real reason. For example, in Study 1, it may be that a judge would want to punish the defendant, who was the mistress of a corrupt government official, so he or she would make a harsher decision against her in an arguably irrelevant circumstance. Yet such a judge would also know that he or she should not take the identity of the defendant into account when deciding a case, so he justifies his decision by giving a reason that facially only depends on the interpretation of the law.

Another possibility is that the judges were unconsciously influenced by irrelevant factors and reached decisions that were implicitly shaped by their moral instincts, choosing decision reasons corresponding to their initial decisions and instinct. For example, in Study 2, a judge in the condition with a defendant of good character may have intuitively decided that the defendant was not guilty. His reasoning would then simply make his reason consistent with his instincts (Simon 2004; Kahneman 2011:20–26). This is also a typical phenomenon of motivated reasoning. Research shows that people's reasoning processes are biased when motivated by directional goals. Such goals may involve a "wish,

desire, or preference that concerns the outcome of a given reasoning task” (Kunda 1990:480). Such reasoning influences not only the conclusion that is reached but also the procedures, methodologies, and facts that underlie that judgment (e.g., Edwards & Smith 1996; Taber & Lodge 2006).

Given the current evidence, we are not able to pin down which is the actual mechanism, but it is worth noting that both mechanisms outlined above are problematic and challenges to confidence in judicial justice. It would undoubtedly be problematic for a judge to conceal his real decision procedure and replace it with a camouflage. Yet it would also be problematic if motivated reasoning has such a pervasive influence in the law. In theory at least, judges should decide cases after deliberating over a wide range of factual and legal issues rather than deciding them first and then employing factual and legal questions as tools to defend their intuition.

C. Salience and Decision Bias

We included the stimulus factor in the short list of possible reasons the judges could choose. The sheer bluntness of the inclusion of this irrelevant information makes the stimulus factor extremely salient. In other words, the judges must have consciously considered the stimulus factor. Yet judges still decided in a biased manner. This suggests that making the possibility of bias salient is not enough for judges to completely overcome it. For example, in Study 1, if salience has been able to eliminate bias, the decision outcomes in the treatment group should be similar to those in the baseline group, where the stimulus factor is absent.

Yet this finding is different from findings in some past studies, and needs some further explanations. Previous studies on race and implicit bias suggest that making racial bias salient helps reduce and even eliminate racial bias. Both black and white jurors exhibit a racial bias by being more likely to find defendants of a different race guilty than defendants who are of the same race. In a set of experiments, Sommers and Ellsworth (2000, 2001) found that making racial issues salient in a trial reduced white juror racial bias toward a black defendant (but not black juror bias toward a white defendant); Cohn et al. (2009) further found that the effect of salience is robust even for individuals who reported high levels of racism. In these studies, race salience was manipulated by presentation of testimony that frames the incident in question in racially charged terms (Sommers & Ellsworth 2000, 2001) or voir dire questionnaires including questions about jurors’ racial attitudes (Sommers 2006). Without the salience of racial issues, white jurors were significantly more likely to find a black defendant guilty, compared to a white defendant. However, when race was made salient, white juror verdicts did not differ in response to defendant race. These studies used aversive racism to explain the race-salience effect: when race is made salient, white jurors are reminded that their actions could be interpreted as racist and against certain social norms, so they tend to respond in a socially appropriate manner, being less likely to find a black defendant guilty. In contrast, in the absence of the salient factor, white jurors may not realize that their behavior could be interpreted as racist, and they are more likely to rely on negative stereotypes about blacks and to be more likely to find a black defendant guilty.

In contrast to the previous studies, we found that making the biasing factor salient does not eliminate bias. This is likely because previous studies deployed the salient stimulus early on, before respondents encountered the key dependent measure, while our salient factor was a part of the questions on reasoning, coming after the questions on decisions. Although judges can always go back and change their initial choice of decisions, this design may have provided a much weaker influence on their decision making. Moreover, the effects of awareness-generating processes are detail dependent and nuanced. Although we believe we have made the stimulus factor salient, and the judges must have consciously considered the stimulus factor, we do not have direct evidence to verify this. Thus, it is unclear how the salient factor interacted with judges' decision-making processes in our studies. It is also worth noting that our results do not suggest that salience does not reduce bias; rather, they simply show that salience does not help judges to completely overcome bias. It may well be that salience already reduced decision bias, without eliminating it. In other words, if we did not list the biasing factor in the short list of options, the disparity in the decisions and reasoning between the treatment group and the control group would have been even larger.¹⁶

D. Implications

It is easy to see why the practice of justifying biased decisions would be problematic for a legal system, regardless of whether judges do it consciously or unconsciously. The authority of the usually unelected judiciary to exert its powers in a legitimate manner is a perennial concern for any judicial system (Bickel 1986; Dworkin 1986; Marmor 2007; Breyer 2010).¹⁷ It is commonly believed that the requirement to give reasons promotes better judicial decision making by keeping judges' discretionary powers in check. Thus, judges are expected to disclose facts and arguments that shape their actions, and they are even usually held to a higher standard of transparency than representatives of other branches of government when they give reasons (Cohen 2010). However, the tactic of demanding reasoning from judicial powers can be compromised by the possibility that judges may present justifications as a means of preventing, rather than facilitating, accountability. Another problem associated with justifying biased decisions is the stability and predictability of the law. Reasoning in judicial opinions explains how judges understand and address specific rules, thus providing guidance to lower courts, future parties, and the public. As Frederick Schauer (1995:649) put it, "giving a reason creates a prima facie commitment on the part of the reason giver to decide subsequent cases in accordance with that reason." However, providing different reasons in response to undisclosed

¹⁶To answer the question of whether salience reduces (rather than eliminates) decision bias, we need to randomly assign the salience factor to judges, and compare their decision outcomes. This can be a future research topic.

¹⁷These commentators were motivated by constitutional questions, which are more fundamental than the more mundane cases we used in our experiments. Yet obviously, even in ordinary cases, there is need for constraining the discretionary power of the judges.

biasing factors can make reasoning elusive and misleading, and it can levy additional costs for individuals to ascertain and follow the law.¹⁸

Although the rationalization of biased decisions is problematic for the legitimacy of the judiciary, we hesitate to draw any policy implications at this point. This is partially because our goal in this article is simply to identify a behavioral pattern and provide a more realistic understanding of the judicial decision-making process in general. Moreover, no perfect solution to the current problem readily presents itself. In practice, there is no good method available for ascertaining whether a judge is biased and whether he provides justifications for the bias. Even if we were to remind judges about the possibility of bias (e.g., making the bias salient), they may neglect to take it into account. Commentators have proposed that it may sometimes be better for judges not to give a reason when such a requirement would result in boilerplate or insincere explanations (Cohen 2010, 2015). This suggestion is worth considering, on the one hand, because reasoning that rationalizes decision biases can sometimes be more troubling than no reasoning at all, as rationalizations tend to be misleading. On the other hand, however, asking judges to give reasons may nonetheless be a good idea because people will notice when judges cook up two different answers in response to two very similar cases.

Our results also suggest an intriguing question at the base of legal training. On the one hand, the purpose of legal training is to assist judges to identify the specific legally relevant factors at hand and dismiss the irrelevant ones (Spellman & Schauer 2012; Kahan 2015). So it is likely that legal training, expertise, and exposure to multiple cases buffers against the tendency to use biased reasoning. On the other hand, in our experiments, as they rationalize their judgments that are based on bias, judges put the exquisite legal skills they received from their formal training to use. In this regard, it is also likely that the better the training a judge receives, the better he or she is at decorating a biased decision. This raises problems for legal training and legal education in general: Does good legal education reduce or increase people's capability of providing biased reasoning? Or, is legal training irrelevant to decision biases, since bias is simply human nature? Previous studies have provided mixed evidence. On the one hand, studies find that individuals with better analytical reasoning skills are more likely to be associated with ideologically motivated cognition because they are better at fitting their beliefs to their ideology (Kahan 2013). On the other hand, Kahan et al. (2016) also find that judges are less likely than the general public to exhibit culture bias when analyzing statutory interpretation problems, suggesting that legal training and experience confer certain resistance to

¹⁸It is worth noting that the idea that judges must really believe what they say in their legal opinions is surprisingly controversial. Some influential judges and legal theorists explicitly reject the notion of judicial candor. For example, Calabresi and Bobbitt (1978) argue that a lack of judicial candor may be justifiable to avoid the destructive consequences of openly recognizing "tragic choices" between conflicting fundamental values. Other commentators suggest that sincerity can be sacrificed to promote the coherence of legal doctrine (Idleman 1995:1392-94), to obtain public compliance with controversial judgments in hard cases (Dan-Cohen 2002:28-32; Hirsch 1993), and to preserve collegiality in the courts (Posner 2003:343). Taken together, these arguments imply that the departure from sincerity is sometimes needed to maintain the legitimacy of the judiciary.

However, these concerns seem of little relevance to our context, in which judges are rationalizing their readily biased decisions.

identity-protective motivated reasoning in professional judgment. Yet it remains a question whether this finding can be extended to other domains, such as the biases driven by heuristics, emotions, and the moral character of the litigant. Given the current evidence, it is prudent to say the relation between legal training and decision bias is not completely clear, and has yet to be thoroughly examined.

Our findings contribute to the literature on the biases in judicial decision making. Previous studies have found a wide range of decision biases, induced by emotions, impulses, and heuristics (e.g., Guthrie et al. 2007; Kahan 2010, 2013; Rachlinski et al. 2011; Wistrich et al. 2015; Leibovitch 2016; Spamann & Klöhn 2016; Liu 2018). The fact that one set of rules may apply to a sympathetic litigant and another set to an unsympathetic litigant certainly poses a challenge to the rule of law. However, justifying a biased decision with elaborated reasons, as we find in this article, seems to further complicate the problem because it makes judicial decision biases hard, if not impossible, to detect, and it partially disables public monitoring in the judicial decision-making process. Moreover, where judges cover their biases intentionally, our findings also cast doubt on the integrity of this profession.

Considered more broadly, our findings have implications for the literature that studies how ideologies shape judicial decisions (e.g., Posner 2010; Epstein et al. 2013). Commentators have long speculated that judges use judicial interpretation and other legal techniques to cover their ideologies. For example, a judge uses originalism to interpret the Constitution not because she is a believer of that strategy and finds its result determinative but because she uses originalism as a means to a conservative end (Post & Siegal 2006; Chilton & Posner 2015). On the other hand, many legal scholars, more often than political scientists, attribute originalism to legal analysis instead of politics (Whittington 2011). We do not intend to join this debate. However, past studies have had difficulty separating the influence of the law and the ideologies, due to a conflation of legal and ideological preferences. Our study shows a potential method that can help disaggregate these effects. For example, it is possible to design a controlled experiment that triggers judges' ideological bias (similar to Kahan et al. [2016]), and see whether judges in the treatment and control groups provide different judicial interpretations of the same precedent or statute.

VII. CONCLUSION

Llewellyn (1931) critically argued that rule of law has never really existed because judges have always made law themselves, using rules to make their decisions seem plausible. Reasoned elaboration in the judicial decision-making process has been thought to constrain judicial arbitrariness and safeguard judicial transparency and trustworthiness (Fuller 1978; Hart & Sacks 1994; Waldron 2011). In this article, however, we find that judges can use reasons to incorporate their personal feelings. We provide evidence that they do so using a wide range of legal techniques and in many areas of the law: from interpreting legal standards to deciding applicability of law and from inferring causation to confirming mere facts. Our findings cast doubt on the effectiveness of reasoned elaboration

as a means to ensure judicial transparency, and provide a more realistic understanding of how judges behave. How to reduce bias and constrain judicial discretionary power is a direction in which future research may proceed.

REFERENCES

- Bickel, Alexander M. (1986) *The Least Dangerous Branch: The Supreme Court at the Bar of Politics*. New Haven, CT: Yale Univ. Press.
- Breyer, Stephen (2010) *Making Our Democracy Work: A Judge's View*. New York: Vintage.
- Calabresi, Guido, & Philip Bobbitt (1978) *Tragic Choices*. New York: W. W. Norton & Company.
- Chilton, Adam S., & Eric A. Posner (2015) "An Empirical Study of Political Bias in Legal Scholarship," 44 *J. of Legal Studies* 277.
- Cohen, Mathilde (2010) "Sincerity and Reason-Giving: When May Legal Decision Makers Lie," 59 *DePaul Law Rev.* 1091.
- (2015) "When Judges Have Reasons Not to Give Reasons: A Comparative Law Approach," 72 *Washington & Lee Law Rev.* 483.
- Cohn, Ellen S., Donald Bucolo, Misha Pride, & Samuel R. Sommers (2009) "Reducing White Juror Bias: The Role of Race Salience and Racial Attitudes," 39 *J. of Applied Social Psychology* 1953.
- Dan-Cohen, Meir (2002) *Harmful Thoughts: Essays on Law, Self, and Morality*. Princeton, NJ: Princeton Univ. Press.
- Dworkin, Ronald (1986) *Law's Empire*. Cambridge, MA: Harvard Univ. Press.
- Edwards, Kari, & Edward E. Smith (1996) "A Disconfirmation Bias in the Evaluation of Arguments," 71 *J. of Personality & Social Psychology* 5.
- Ehrlich, Isaac, & Richard A. Posner (1973) "An Economic Analysis of Legal Rulemaking," 3 *J. of Legal Studies* 257.
- Eisenberg, Melvin Aron (1978) "Participation, Responsiveness, and the Consultative Process: An Essay for Lon Fuller," 92 *Harvard Law Rev.* 410.
- Englich, Birte, Thomas Mussweiler, & Fritz Strack (2006) "Playing Dice with Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making," 32 *Personality & Social Psychology Bulletin* 188.
- Epstein, Lee, William M. Landes, & Richard A. Posner (2013) *The Behavior of Federal Judges*. Cambridge, MA: Harvard Univ. Press.
- Fon, Vincy, & Francesco Parisi (2007) "On the Optimal Specificity of Legal Rules," 3 *J. of Institutional Economics* 147.
- Fuller, Lon L. (1978) "The Forms and Limits of Adjudication," 92 *Harvard Law Rev.* 353.
- Guthrie, Chris, Jeffrey J. Rachlinski, & Andrew J. Wistrich (2001) "Inside the Judicial Mind," 86 *Cornell Law Rev.* 777.
- (2007) "Blinking on the Bench: How Judges Decide Cases," 93 *Cornell Law Rev.* 1.
- Hart, Henry M. Jr., & Albert M. Sacks (1994) *The Legal Process: Basic Problems in the Making and Application of Law*, ed. by W. N. Eskridge, Jr. & P. P. Frickey. Westbury, NY: Foundation Press.
- Hirsch, Alan (1993) "Candor and Prudence in Constitutional Adjudication," 61 *George Washington Law Rev.* 858.
- Idleman, Scott C. (1995) "A Prudential Theory of Judicial Candor," 73 *Texas Law Rev.* 1307.
- Kahan, Dan M. (2010) "Culture, Cognition, and Consent: Who Perceives What, and Why, in Acquaintance-Rape Cases," 158 *Univ. of Pennsylvania Law Rev.* 729.
- (2013) "Ideology, Motivated Reasoning, and Cognitive Reflection: An Experimental Study," 8 *Judgment & Decision Making* 407.
- (2015) "Laws of Cognition and the Cognition of Law," 135 *Cognition* 56.

- Kahan, Dan M., David Hoffman, Daniel Evans, Neal Devins, Eugene Lucci, & Katherine Cheng (2016) "Ideology" or "Situation Sense"? An Experimental Investigation of Motivated Reasoning and Professional Judgment," 64 *Univ. of Pennsylvania Law Rev.* 2015.
- Kahneman, Daniel (2011) *Thinking, Fast and Slow*. New York: Farrar, Straus & Giroux.
- Kaplow, Louis (1992) "Rules Versus Standards: An Economic Analysis," 42 *Duke Law J.* 557.
- Kunda, Ziva (1990) "The Case for Motivated Reasoning," 108 *Psychological Bulletin* 480.
- Lao, Dongyan (2015) "Causation in Fact and Consequence-Based Liability in Criminal Law" ["Shishi Yinguo yu Xingfa Zhongde Jieguo Guize"], 2 *China Legal Science [Zhongguo Faxue]* 131.
- Leibovitch, Adi (2016) "Relative Judgments," 45 *J. of Legal Studies* 281.
- Liu, Fengjing (2016) "Definition and Implementation of the Authority of Judicial Interpretation" ["Sifa Jieshi Quanxian de Jieding yu Xingshi"], 3 *China Legal Science [Zhongguo Faxue]* 207.
- Liu, John Zhuang, Lars Klöhn, & Holger Spamann (forthcoming) "Precedent and Chinese Judges: An Experiment," *American J. of Comparative Law*.
- Liu, Zhuang (2018) "Does Reason Writing Reduce Decision Bias? Experimental Evidence from Judges in China," 47 *J. of Legal Studies* 83.
- Llewellyn, Karl N. (1931) "Some Realism About Realism: Responding to Dean Pound," 44 *Harvard Law Rev.* 1222.
- Llewellyn, Karl N. (1940) "On Reading and Using the Newer Jurisprudence," 40 *Columbia Law Rev.* 581.
- Llewellyn, Karl N. (1989) *The Case Law System in America*. Chicago, IL: Univ. of Chicago Press.
- Luo, Kun (2016) "On the Judicial Discretionary Reduction of Liquidated Damages" ["Woguo Weiyuejin Sifa Zhuojian de Xianzhi yu Paichu"], 2 *Legal Science [Falv Kexue]* 115.
- Marmor, Andrei (2007) *Law in the Age of Pluralism*. New York: Oxford Univ. Press.
- Nadler, Janice, & Mary-Hunter McDonnell (2011) "Moral Character, Motive, and the Psychology of Blame," 97 *Cornell Law Rev.* 255.
- Posner, Richard A. (1988) "The Jurisprudence of Skepticism," 86 *Michigan Law Rev.* 827.
- (2003) *Law, Pragmatism, and Democracy*. Cambridge, MA: Harvard Univ. Press.
- (2010) *How Judges Think*. Cambridge, MA: Harvard Univ. Press.
- Post, Robert, & Reva Siegel (2006) "Originalism as a Political Practice: The Right's Living Constitution," 75 *Fordham Law Rev.* 545.
- Rachlinski, Jeffrey J., Andrew J. Wistrich, & Chris Guthrie (2011) "Probable Cause, Probability, and Hindsight," 8 *J. of Empirical Legal Studies* 72.
- Rawls, John (1993) *Political Liberalism. John Dewey Essays in Philosophy*. New York: Columbia Univ. Press.
- Robinson, Paul H. (2008) *Criminal Law: Case Studies and Controversies*, 2d ed. New York: Wolters Kluwer Law & Business.
- Schauer, Frederick (1988) "Judging in a Corner of the Law," 61 *Southern California Law Rev.* 1717.
- (1995) "Giving Reasons," 47 *Stanford Law Rev.* 633.
- Shapiro, David L. (1987) "In Defense of Judicial Candor," 100 *Harvard Law Rev.* 731.
- Sommers, Samuel R. (2006) "Racial Diversity and Group Decision Making: Identifying Multiple Effects of Racial Composition on Jury Deliberations," 90 *J. of Personality & Social Psychology* 597.
- Sommers, Samuel R., & Phoebe C. Ellsworth (2000) "Race in the Courtroom: Perceptions of Guilt and Dispositional Attributions," 26 *Personality & Social Psychology Bulletin* 1367.
- (2001) "White Jurors' Bias: An Investigation of Prejudice Against Black Defendants in the American Courtroom," 7 *Psychology, Public Policy, & Law* 201.
- Simon, Dan (1998) "A Psychological Model of Judicial Decision Making," 30 *Rutgers Law J.* 1.
- (2004) "A Third View of the Black Box: Cognitive Coherence in Legal Decision Making," 71 *Univ. of Chicago Law Rev.* 511.
- Sood, Avani M., & John M. Darley (2012) "The Plasticity of Harm in the Service of Criminalization Goals," 100 *California Law Rev.* 1313.
- Spamann, Holger, & Lars Klöhn (2016) "Justice Is Less Blind, and Less Legalistic, Than We Thought: Evidence from an Experiment with Real Judges," 45 *J. of Legal Studies* 255.

- Spellman, Barbara, & Frederick Schauer (2012) "Legal Reasoning," in K. J. Holyoak & R. G. Morrison, eds., *The Oxford Handbook of Thinking and Reasoning*, 2d ed. Oxford: Oxford Univ. Press.
- Sunstein, Cass R. (2007) "If People Would Be Outraged by Their Rulings, Should Judges Care?" 60 *Stanford Law Rev.* 155.
- Taber, Charles S., & Milton Lodge (2006) "Motivated Skepticism in the Evaluation of Political Beliefs," 50 *American J. of Political Science* 755.
- Wechsler, Herbert (1959) "Toward Neutral Principles of Constitutional Law," 73 *Harvard Law Rev.* 1.
- Whittington, Keith E. (2011) "Is Originalism Too Conservative?" 34 *Harvard J. of Law & Public Policy* 29.
- Wistrich, Andrew J., Jeffrey J. Rachlinski, & Chris Guthrie (2015) "Heart Versus Head: Do Judges Follow the Law or Follow Their Feelings," 93 *Texas Law Rev.* 855.
- Zhang, Jikun, & Libing Liu (2017) "Relevance of Evidence in Rape Cases" ["Qiangjian Anjian Zhong Zhengju Xiangguanxing Panduan yu Shiyong"], 3 *People's Procuratorial Semimonthly [Renmin Jiangcha]* 74.
- Zhang, Mingkai (2013) "On the Theory of Objective Attribution of Liability" ["Yetan Keguan Guize Lilun"], 2 *Peking Univ. Law J. [Zhongwai Faxue]* 300.
- Zhang, Nanning, & Douglas Walton (2010) "Recent Trends in Evidence Law in China and the New Evidence Scholarship," 9 *Law, Probability & Risk* 103.

APPENDIX

Table A1: Three-Way and Four-Way Tests for Study 1

Percentage	(1)		(2)		(3)	
	<i>Excessively Higher than Actual Loss</i>		<i>Substantial Influence on Business</i>		<i>Aggregated Reasons</i>	
	<i>Treatment (Bad)</i>	<i>Baseline (NA)</i>	<i>Treatment (Bad)</i>	<i>Baseline (NA)</i>	<i>Treatment (Bad)</i>	<i>Baseline (NA)</i>
Positive	30.8	87.5	23.1	12.5	53.8	87.5
Negative	38.5	0.0	30.8	50.0	38.5	0.0
No opinion	30.8	12.5	46.2	37.5	7.7	0.0
Both positive and negative	—	—	—	—	0.0	12.5
Fisher's exact	0.003		0.692		0.004	
Treatment/Baseline (N)	13/16		13/16		13/16	

NOTES: Judges in the bad character group received a stimulus designed to induce negative feeling about the defendant's moral character, while judges in the baseline group did not receive this stimulus. "Positive" suggests a judge chose only positive reason/reasons, without choosing any negative reason/reasons; "Negative" suggests a judge chose only negative reason/reasons, without choosing any negative reason/reasons; "No opinion" suggests a judge did not choose any option as a reason; and "Both positive and negative" suggests a judge chose both positive and negative reasons. Two judges in the baseline group chose from both reasons for high compensation and reasons for low compensation.

Table A2: Three-Way and Four-Way Tests for Study 2

Percentage	(1) <i>Wild Animal</i>		(2) <i>Law Applicable</i>		(3) <i>Defendant Knows</i>		(4) <i>Aggregated Reasons</i>	
	<i>Bad</i>	<i>Good</i>	<i>Bad</i>	<i>Good</i>	<i>Bad</i>	<i>Good</i>	<i>Bad</i>	<i>Good</i>
	<i>Character</i>	<i>Character</i>	<i>Character</i>	<i>Character</i>	<i>Character</i>	<i>Character</i>	<i>Character</i>	<i>Character</i>
Positive	27.0	25.7	59.5	37.1	48.6	28.6	56.8	37.1
Negative	18.9	48.6	16.2	40.0	27.0	34.3	29.7	57.1
No opinion	54.1	25.7	24.3	22.9	24.3	37.1	2.7	0.0
Both positive and negative	—	—	—	—	—	—	10.8	5.7
Fisher's exact	0.016		0.070		0.214		0.083	
Bad/Good (N)	37/35		37/35		37/35		37/35	

NOTES: Judges in the bad character group received a stimulus designed to induce negative feeling about the defendant's moral character. Judges in the good character group received a positive stimulus. "Positive" suggests a judge chose only positive reason/reasons, without choosing any negative reason/reasons; "Negative" suggests a judge chose only negative reason/reasons, without choosing any negative reason/reasons; "No opinion" suggests a judge did not choose any option as a reason; and "Both positive and negative" suggests a judge chose both positive and negative reasons. Four judges in the bad character group and two judges in the good character group chose from both reasons for conviction and reasons for acquittal.

Table A3: Three-Way and Four-Way Tests for Study 3

Percentage	(1) <i>Causation</i>		(2) <i>Foreseeability</i>		(3) <i>Major Reason</i>		(4) <i>Aggregated Reasons</i>	
	<i>Bad</i>	<i>Good</i>	<i>Bad</i>	<i>Good</i>	<i>Bad</i>	<i>Good</i>	<i>Bad</i>	<i>Good</i>
	<i>Character</i>	<i>Character</i>	<i>Character</i>	<i>Character</i>	<i>Character</i>	<i>Character</i>	<i>Character</i>	<i>Character</i>
Positive	66.7	35.3	23.8	5.9	28.6	29.4	61.9	23.5
Negative	28.6	47.1	19.0	52.9	0.0	5.9	28.6	58.8
No opinion	4.8	17.6	57.1	41.2	71.4	64.7	4.8	0.0
Both positive and negative	—	—	—	—	—	—	4.8	17.6
Fisher's exact	0.129		0.081		0.845		0.033	
Bad/Good (N)	21/17		21/17		21/17		21/17	

NOTES: Judges in the bad character group received a stimulus designed to induce negative feeling about the defendant's moral character. Judges in the good character group received a positive stimulus. "Positive" suggests a judge chose only positive reason/reasons, without choosing any negative reason/reasons; "Negative" suggests a judge chose only negative reason/reasons, without choosing any negative reason/reasons; "No opinion" suggests a judge did not choose any option as a reason; and "Both positive and negative" suggests a judge chose both positive and negative reasons. One judge in the bad character group and three judges in the good character group chose from both reasons for high compensation and reasons for low compensation.

A. Experiment Materials

a. Study 1

On March 15, 2017, the plaintiff Shenzhen Kaiyuan Investment Co., Ltd (the lessee) and the defendant Yang Li (the lessor) signed a lease. The two parties agreed on that from May 18, 2017 to May 17, 2019, the defendant would lease two commercial units to the

plaintiff: Rooms 1611 and 1612 in the No. 2 Building of Zhuoyue Century Center, located at the intersection of Fuhua 3rd Road and Jintian Road, Futian CBD, Shenzhen. The rent was RMB 200,000 yuan per month. The two parties also agreed that the lessee could move in on May 18, 2017, and the lessor would allow the lessee to occupy the units on that date and finish the relevant delivery of occupancy procedures. According to the contract, if either party broke the contract, he or she should pay liquidated damages of RMB 300,000 yuan.

On May 16, 2017, the lessor, Yang Li, informed the lessee that she could not hand over the rooms to the lessee on time on May 18 because the last lessee had not yet moved, and she wanted to defer delivery by about 10 days. The lessor and the lessee did not reach an agreement on deferred delivery. During the negotiation, the lessor Yang Li expressed a negative attitude and asserted that she had a personal relationship with government officials and was not afraid of being sued. The lessee brought suit against the lessor and asked for RMB 300,000 yuan as liquidated damages.

The plaintiff claimed that the company had been preparing for removal and had already closed its business and packed all furniture and office equipment. Preparing for a move on May 18, the company had signed a contract with a moving company. If they breached the contract, they would suffer losses. Moreover, the company was a private equity investment firm. Customers came and visited their office every day. The postponed opening date would incite customer distrust and cast a negative light on the company image, substantially affecting the company's long-term business.

The defendant Yang Li's agent ad litem claimed that the actual loss of the plaintiff included only potential compensation arising from breaching the contract with the moving company, about RMB 20,000 yuan (which was accurate, according to the court's investigation), as well as the loss of business opportunities due to the postponement of moving. In fact, the plaintiff was still able to manage its business on the site of its previous office, so there would be no loss of business. Furthermore, the lessor did not defer the lessee's moving in date willfully. This occurred because the last lessee did not move out on time. Therefore, the lessor was not at fault. The liquidated damages RMB 300,000 yuan stipulated in the contract was significantly higher than the actual loss of the plaintiff. Article 114, paragraph 2 of the Contract Law stipulates that where the amount of liquidated damages agreed upon are excessively higher than the damages incurred, a party may petition the People's Court or an arbitration institution to make an appropriate reduction. The defendant thus requested the court to lower the liquidated damages to RMB 20,000 yuan, in addition to which, she was willing to compensate the lessee for 10 days of rent, about RMB 67,000 yuan.

[(*Treatment Condition: Bad Moral Character*)] The defendant Yang Li in this case did not appear at court. It was reported that the defendant maintained an extramarital relationship with the deputy director of the local land and resource bureau, Zhang. Zhang may have accepted bribes and possessed extensive assets with unclear sources. Zhang was being investigated and prosecuted by the procuratorate. Yang Li, who may have been involved in Zhang's crime, was suspected of accepting and soliciting bribes for Zhang. This case will be tried separately. The commercial units in

this case were Yang Li's own, obtained through inheritance, and they were not related to Zhang's bribery case.]

The laws and regulations that may be involved in this case:

Article 114, paragraph 2 of the Contract Law stipulates that where the amount of liquidated damages agreed upon are excessively higher than the damages incurred, a party may petition the People's Court or an arbitration institution to make an appropriate reduction.

1. If you were the judge in this case, what decision would you make?
 - A. The plaintiff wins. The defendant should pay the plaintiff the liquidated damages of RMB 300,000 yuan.
 - B. The defendant should pay the plaintiff his loss arising from breaching the contract with the moving firm (RMB 20,000 yuan) and compensate him for 10 days rent (about RMB 67,000 yuan). The compensation should be RMB 87,000 yuan in total.
 - C. Adjust the liquidated damages. The defendant should pay the plaintiff RMB _____ yuan in total.
 - D. Other (Please write here) _____
2. Your decision reasons are (select all that apply):
 - A. The amount of liquidated damages agreed upon, RMB 300,000 yuan, is excessively higher than the actual damages incurred.
 - B. The defendant did not have a substantial influence on the business of the plaintiff through the breach of the contract.
 - C. The amount of liquidated damages agreed upon, RMB 300,000 yuan, is not excessively higher than the damages incurred.
 - D. The defendant had a substantial influence on the business of the plaintiff through the breach of the contract.
 - E. [Only for the treatment group.] The defendant maintained an extramarital relationship with the deputy director of the local land and resource bureau, Zhang, and was implicated in Zhang's bribery trial.
 - E/F. Other reasons (please specify below).

b. Study 2

The defendant Wang Peng, born in 1983 in Jiujiang County, Jiangxi Province, was working at Diya Computational & Numerical Control Factory in Shenzhen at the time of the case. In April 2014, Wang Peng took home a female parrot from the yard of the factory to his dormitory. He had always been interested in parrots. In May 2014, Peng Wang bought a male parrot online to mate with the female one. The two parrots reproduced rapidly. He had 43 juvenile parrots only one year later. At the beginning of April 2016, Wang Peng sold six parrots to his friend for RMB 2000 yuan. According to the police, the six included four cockatiels and two golden parakeets, officially named the green-cheeked parakeet, a protected species listed in the Appendix of the Convention on International Trade in Endangered Species of Wild Fauna and Flora. On May 18, 2016, the

Forest Branch of the Shenzhen Public Security Bureau detained Wang Peng on suspicion of the crime “illegal sale of precious and endangered wild animals and their related products.”

The prosecution and the defense debated the following issues:

First, whether domestically breeding species should be categorized as wild animals.

The prosecution claimed that the parrots sold by Wang Peng belonged to a protected species. At the same time, Article 1 of the Interpretation on Several Issues Concerning the Specific Application of Law in the Trial of Criminal Cases Violating Wildlife Resources of the SPC has explicitly included domestically breeding species within the scope of criminal law protection.

The defendant claimed that, first, criminal law stipulates that the object of crime for the “illegal sale of precious and endangered wild animals” is precious and endangered wild animals. The meaning is clear, that is, the species involved must be wild animals. Categorizing domestically breeding animals as wild animals is self-contradictory. Second, in the Interpretation of Several Issues Concerning the Specific Application of Law in the Trial of Criminal Cases Violating Wildlife Resources, the SPC considered domestically breeding animals as wild animals. This is far beyond the text of the criminal law. It constitutes an extensive interpretation and violates the principle of *nulla poena sine lege* (no penalty without a law). It also contradicts the legislative intention of the criminal law. The interpretation is not applicable.

Second, whether an actor who is unaware of the illegality of his action should be held guilty.

The defense claimed that Wang Peng raised a large number of parrots, including a variety of species. He did not know that he was actually breeding protected ones. He by no means realized that his behavior was illegal. Thus, his actions should not be punished as a crime.

The prosecution argued that Wang Peng was a parrot lover and he had specific breeding and feeding techniques. He also frequently communicated his breeding experience with other parrot lovers online. It was unreasonable to believe that Wang Peng did not know that the parrots he was raising were nationally protected animals. Thus, Wang Peng must have realized the illegality of his behavior.

[(*Bad Moral Character Condition*) Wang Peng had been indulging in gambling online since 2015, was divorced his wife, and had no time to take care of the parrots. He thus sold the juvenile parrots.] The prosecution suggested that the defendant should be sentenced to five years of fixed-term imprisonment for “illegal sale of precious and endangered wild animals.”

[(*Good Moral Character Condition*) Wang Peng further explained that his daughter suffered from systematic lupus erythematosus, and she needed his care and company. He did not have time to take care of the juvenile parrots, so he sold them. He hoped that the sitting judge would not convict him.]

The laws and regulations that may be involved in this case:

Article 341, Paragraph 1 of the Criminal Law: Those who illegally hunt and kill rare and endangered wild animals that are accounted for in the state key production plan or illegally purchase, transport, or sell rare and endangered wild animals and/or their manufactured

products are to be sentenced to not more than five years of fixed-term imprisonment or criminal detention and may in addition be sentenced to a fine. In serious cases, those law offenders are to be sentenced to not less than five years and not more than 10 years of fixed-term imprisonment and may in addition be sentenced to a fine. In especially serious cases, such offenders are to be sentenced to more than 10 years of fixed-term imprisonment and in addition be sentenced to a fine and confiscation of their property.

1. If you were the judge in this case, would you convict the defendant of the crime illegal sale of precious and endangered wild animals?
 - A. Yes
 - B. No
 - C. Neither yes nor no, but _____
2. Your decision reasons are (select all that apply):
 - A. The species the defendant bred was a wild animal.
 - B. The Interpretation of Several Issues Concerning the Specific Application of Law in the Trial of Criminal Cases Violating Wildlife Resources of the SPC was applicable in this case.
 - C. Wang Peng was aware of the illegality of his behavior. That is, to a certain extent, he knew the animals he was raising and selling were protected animals.
 - D. The species the defendant bred was not a wild animal.
 - E. The Interpretation on Several Issues Concerning the Specific Application of Law in the Trial of Criminal Cases Violating Wildlife Resources of the SPC excessively expands the scope of wild animals, and it was not applicable in this case.
 - F. Wang Peng was not aware of the illegality of his behavior. That is, he did not know the animals he raised and sold were protected animals.
 - G. [For the good character condition] Wang Peng's daughter suffered from systematic lupus erythematosus. Wang Peng is sympathetic.
[For the bad character condition] Wang Peng gambled online. He did not have a good moral character.]
 - H. Other reasons (please specify below).
3. The sentence you would give is:
 - A. Imprisonment for _____ year(s)
 - B. Acquittal and no penalty
 - C. Other _____

c. Study 3

The defendant Wang Shuqiang lives in 16 Meitan village, Sanhe Town, Dapu County, Chaozhou City, Guangdong Province. The defendant put two oxygen tanks near a wall in the courtyard of his house. At 11 p.m., August 12, 2017, the two oxygen tanks exploded, and a large fire ensued. According to the firefighting department, the explosion of the oxygen tanks was caused by a cigarette butt left by a passer-by outside the wall, but they

could not find the specific person. The explosion of the oxygen tanks first set attachments outside the wall on fire, and the fire further expanded to the houses around.

[(*Good Moral Character Condition*) Investigation found that the defendant's mother had suffered from lung disease for years and often needed oxygen at home. The defendant was storing the two oxygen tanks to take better care of his mother.]

[(*Bad Moral Character Condition*) Investigation found that the defendant was illegally producing small amounts of crystal meth (methamphetamine) at home, and the two oxygen tanks were to be used for this production. This criminal case was being investigated separately by the police.]

The plaintiff Zhao Xiahua, living in the same village as the defendant, participated in the firefighting at that night. It was midnight, and the road was dark when the plaintiff was carrying water to the fire site but fell into a ditch by the roadside. The fall caused several fractures of the bones of his thigh and wrist, and his broken thigh pierced his muscle and skin. The plaintiff was sent to a hospital. The injury cost him RMB 44,000 yuan, including treatment costs and other expenses.

The plaintiff claimed that his fall was the result of his help with the firefighting. There was a causal relationship between the defendant's oxygen storage, the fire, and his injury. Further, when the defendant stored the oxygen, he should have foreseen the potential danger. Hence, the plaintiff asked the defendant to compensate him RMB 44,000 yuan, including treatment costs and other expenses.

The defendant argued that there was no causal relationship between his oxygen tanks and the fire; it was also unreasonable to assume a causal link between the fire and the plaintiff's injury. In particular, he was informed that the plaintiff had been at a celebration and had drunk alcohol that night. Hence, the fall of the plaintiff could have been caused by drinking. Besides, the defendant argued that it was impossible for him to have foreseen that the plaintiff would fall and be injured because of firefighting.

[(*Good Moral Character Condition*) The defendant further claimed that he had given up all opportunity to work outside the village to take better care of his mother. All his neighbors recognized that he was a good person. He hoped that the sitting judge would consider this factor.]

[(*Bad Moral Character Condition*) The plaintiff further claimed the defendant's illegal production of crystal meth (methamphetamine) was responsible for his injury. The defendant should be punished, and he should compensate him for the costs.]

1. If you were the judge in this case, what decision would you make?
 - A. The plaintiff wins, and the defendant pays the plaintiff RMB 440,000 yuan.
 - B. Adjust the amount of compensation to RMB _____ yuan.
 - C. The plaintiff loses, and the defendant does not need to make any compensation.
 - D. Other (Please write here) _____
2. Your decision reasons are (select all that apply):
 - A. There is a causal relationship between the defendant's act of oxygen storage and the plaintiff's injury.

- B. The defendant could have foreseen the possibility that someone (the plaintiff) would be injured.
- C. Although the plaintiff had drunk some alcohol, the defendant's act of oxygen storage was the major reason of the plaintiff's injury.
- D. There was no causal relationship between the defendant's act of oxygen storage and the plaintiff's injury.
- E. The defendant could not have foreseen any possibility that someone (the plaintiff) would be injured.
- F. The plaintiff's drinking of alcohol is the major reason for his injury.
- G. [For the bad character condition] The defendant produced crystal meth at home and had poor morals. His behavior should be punished.
[For the good character condition] The defendant was taking care of his mother. He had good morals. It would be unreasonable to ask him to make compensation.
- H. Other reasons (please specify below).