Title
On Informing Jurors of Potential Sanctions

Permalink
https://escholarship.org/uc/item/0sb3c4z0

Author
Teitcher, Jennifer

Publication Date
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

On Informing Jurors of Potential Sanctions

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF ARTS

in Social Ecology

by

Jennifer Esther Feldstein Teitcher

Thesis Committee:
Professor Nicholas Scurich, Chair
Professor Elizabeth Loftus
Professor Mona Lynch

2017

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

ABSTRACT OF THE THESIS

On Informing Jurors of Potential Sanctions

By

Jennifer Esther Feldstein Teitcher

Master of Arts in Social Ecology

University of California, Irvine, 2017

Professor Nicholas Scurich, Chair

With few exceptions, jurors in criminal trials exclusively determine whether the defendant is guilty; they do not determine what the sanction is or even recommend what it should be. However, anecdotal evidence suggests that jurors make assumptions regarding the potential punishment, and that these assumptions inform their verdicts. This is rational behavior according to Decision Theory. Thus, several legal scholars have argued that jurors ought to be informed of the possible punishment that would follow a guilty verdict, in order to disabuse incorrect assumptions and make an informed decision. The present experiments tested: a.) whether jurors do make assumptions about the potential punishment that would follow from a guilty verdict; b.) whether those assumptions influence jurors' implicit threshold for reasonable doubt; and c.) whether informing jurors of the potential punishment additionally influences their implicit threshold. Experiment 1 manipulated the alleged crime (Grand Theft vs. Manslaughter) holding all other factors constant, and found that mock jurors (n=102, recruited via Amazon Mechanical Turk) had different expectations about the relative punishments but that these expectations did

not affect their implicit threshold for reasonable doubt. Experiment 2 manipulated the alleged crime as well as a judicial description of the potential punishment (e.g., term of incarceration of 2-6 vs 7-20 years). Again, mock jurors (n=297) were sensitive to the relative punishments, yet their implicit threshold did not differ on this basis. These findings call into question whether jurors should be informed of the potential punishment if the defendant is convicted.

INTRODUCTION

On March 30th 2016, Stanford University swimmer Brock Turner was convicted of three felony charges: (1) assault with the intent to commit rape of an unconscious person; (2) sexual penetration of an unconscious person; and (3) sexual penetration of an intoxicated person. Despite the deputy probation officer's sentencing report recommending three years in prison and three years of probation, Judge Aaron Persky sentenced Turner to only six months in jail and three years of probation. Turner was released in September, 2016, after only serving three months of his sentence (Grinberg, 2016; Kadvany, 2016; Rocha & Winton, 2016).

The short sentence drew criticism nationally and globally (Rocha & Winton, 2016). One person in particular, a juror on the case, wrote an anonymous letter to the judge expressing his frustration with the decision. He wrote that he expected that the judge would treat these convictions seriously and apply an appropriate sentence not only so that the defendant would suffer the consequences of assaulting an unconscious person, but also to serve as a deterrent for all on-campus rapes. However, the juror stated that the six-month sentence effectively ignored the convictions, and that despite the jurors' strong efforts, justice was not served because "the punishment did not fit the crime" (Kadvany, 2016, para. 1).

In most jurisdictions, members of juries, like those in the Brock Turner case, are not told what the punishment will be if they find the defendant guilty. Yet there has been considerable anecdotal evidence suggesting that, when deciding whether to convict a criminal defendant, jurors consider what they think the punishment would be (e.g., Freedman, Krismer, MacDonald, & Cunningham, 1994; Kerr, 1978). For instance, jurors might be unwilling to convict a criminal defendant even when there is a bevy of incriminating evidence because they feel that the

uncertainty of the defendant's guilt is too large to tolerate considering the severity of the punishment that will follow from a guilty verdict.

As described below, this is rational behavior according to Decision Theory in which rational decisions depend on the potential consequences. Thus, some legal scholars have argued that jurors ought to be apprised of the potential punishment that turns on their verdict because jurors might harbor mistaken and unrealistic assumptions about the potential punishments (e.g., J. Kaplan, 1967; Kozinski, 2015; Stoffelmayr & Diamond, 2000). This paper examines whether jurors do make inappropriate assumptions about the potential punishment and whether these assumptions affect their conviction threshold. First, however, it is necessary to understand the potential rationality of informing jurors of the punishment that would be occasioned by a guilty verdict.

DECISION THEORY, BLACKSTONE RATIOS, AND STANDARDS OF PROOF

Many decisions are made under conditions of uncertainty. The decision to carry an umbrella to work depends in part on the likelihood that it will rain later that day. The decision to invest in a particular stock depends in part on the likelihood that the price will increase. Of course, these decisions involve more than simply the likelihood of some event occurring; they also involve the relative benefits/costs of a particular action. It might cost virtually nothing to carry the umbrella, so if it rains you are protected and if it doesn't you are inconvenienced, albeit meagerly. On the other hand, an investor can lose a significant sum of money on a poor stock choice; this is obviously not desirable but it is downright unreasonable if the potential returns were not great from the outset.

The process of making decisions under uncertainty can be structured with Decision Theory. Decision Theory takes into account the potential benefits of a 'correct' decision (e.g., carrying an umbrella when it rains) and the potential costs of an 'incorrect' decision (e.g., carrying an umbrella when it does not rain), and weighs these against the probability of a positive event occurring (i.e., the probability of rain). It has been proven mathematically that following the dictates of Decision Theory will, in the long run, lead to more optimal outcomes than not following the dictates (Keeney & Raiffa, 1976). Thus, "rational behavior" is defined by economists and others as adhering to Decision Theory. A variety of organizations use Decision Theory to make better-informed decisions.

In the late 1960s, a Stanford law professor—John Kaplan—pondered whether the then-fledgling Decision Theory could be used to structure the decisions that are made by jurors. Given that a juror is hardly ever 100% certain that a criminal defendant is guilty, Kaplan suggested that a juror needs to weigh the individual and/or societal costs ("disutility" in Decision Theory parlance) of possibly convicting an innocent person (a false positive), against the disutility of acquitting a guilty person (a false negative) when deciding if the certainty regarding the guilt of the defendant is sufficient to return a guilty verdict (1967). Kaplan, along with other legal commentators (Lempert, 1977), made the simplifying assumption that true negatives and positives are equally desirable in the context of adjudication.

The disutility tradeoff Kapan adverted to is widely-known as Blackstone's Ratio. William Blackstone, an 18th century jurist, famously said, "it is better that ten guilty persons escape than that one innocent person suffer." In other words, a false positive is ten times worse than a false negative. Historically, a false positive has been seen to be worse than a false negative for a number of reasons, such as promoting an unjust legal system or unwarranted monetary costs of

punishment (Risinger, 1997). However, as some have pointed out, those convicted in the 18th century were most often punished by death, and it is not clear whether the same 10:1 ratio of Blackstone would apply in a non-capital case (see, e.g., Lillquist, 2004). Furthermore, it appears that the general public does not necessarily agree that the proper ratio is 10:1. Undergraduates, for example, do not find one error worse than the other, with a ratio resembling 1:1 (Arkes & Mellers, 2002), and similar findings have been reported from a sample of jury-eligible individuals (Scurich, 2015).

Although Blackstone ratios have generated voluminous commentary over the centuries, J. Kaplan (1968) was the first to reify the implications of a Blackstone ratio for jury decision making. Specifically, he demonstrated that a particular Blackstone ratio implied a minimum level of certainty necessary to return a guilty verdict, and that ratio was determined by weighing the disutility of a false positive (E_{fp}) against the disutility of a false negative (E_{fn}). This ratio can then be converted into a probability to determine the individual threshold (T_c) to convict (see also Ceci & Friedman, 2000):

$$T_c = \frac{E_{fp}}{E_{fn}} \tag{1}$$

A juror should convict if, and only if, her certainty (expressed as a probability) in the guilt of the defendant is greater than the threshold; otherwise, the juror should vote to acquit. For example, assuming that the cost of a false positive is 10 times greater than a false negative, then T_c =10:1, which can be converted to a probability threshold of 0.91 (To convert from odds to probability, take the first term (i.e., 10) and divide it by the sum of the first and second terms (i.e. 10+1): 10/11 = 0.91). Thus, a juror should vote to convict if, and only if, she is more than 91% confident of the defendant's guilt after hearing the admissible evidence; if she is less than 91% confident, she should vote to acquit.

This example illustrates that logic the undergirds supporting the beyond a reasonable doubt standard of proof that is required in all criminal cases (*In re Winship*, 1970). Implicit in beyond a reasonable doubt is the value judgment that the cost of convicting an innocent person is significantly greater than the cost of acquitting a guilty person; thus, the high level of certainty of a defendant's guilt that is theoretically required to return a guilty verdict (Kaplan, 1968). Whether jurors adhere to this ideal is an empirical question. A subsidiary issue concerns whether jurors modify their operationalization of beyond a reasonable doubt in light of the potential punishment that follows from a guilty verdict. Judge Jack Weinstein described this phenomenon eloquently:

> Should society be willing to risk 10 guilty defendants go free rather than one innocent person be convicted? Or is the proper ratio 100 to one? Should we be willing to accept lower risks in a spitting on the sidewalk case than in a capital homicide case? (Weinstein & Drewbury, 2006, p. 168-69)

STANDARDS OF PROOF AND JURORS

A number of empirical studies have examined how jurors respond to instructions on the beyond a reasonable doubt standard of proof. These studies have yielded mixed results. Some find that jurors are responsive to instructions on the standard of proof (Kagehiro & Stanton, 1985; Krauss & Scurich, 2014; Magnussen, Eilertsen, Teigen, & Wessel, 2013), while others do not (see Baguley, McKimmie, & Masser, in press). For example, Kagehiro and Stanton (1985) held constant the evidence in a trial but manipulated the standard of proof (e.g., clear and convincing evidence vs. beyond a reasonable doubt) and found that jurors were apparently sensitive to the different standards of proof (i.e., the conviction rates differed), but only when the

standard of proof was communicated in probabilistic terms. The rate of conviction did not differ when the standard of proof was communicated in conventional, legal terms.

Additionally, several studies have examined how the severity of punishment may impact conviction rates, and they, too, have yielded mixed results. For instance, Kerr (1978) systematically manipulated the severity of charge (Manslaughter, Second-Degree Murder; First-Degree Murder) and the severity of punishment (Mild (1-20 years), Severe (25-Life or Capital Punishment)), and found that the conviction rate decreased when the severity of the punishment was high. Specifically, Kerr (1978) reported a "marginally" significant (i.e., $p<.10$) decrease in the conviction rate when the penalty was severe for First-Degree Murder or Manslaughter. The pattern of results was slightly reversed for second-degree murder with the mild punishment yielding fewer convictions than when the punishment was severe. Freedman and colleagues (1994) failed to replicate these findings across nine separate studies, with over one thousand adult participants. Additionally, in the two studies that had jurors deliberate, the severity of punishment did not affect conviction rates (Davis et al., 1977; Nedermeier, Horowitz, & Kerr, 1999).

Aside from experimentally manipulating the punishment (and therefore explicitly telling participants what the punishment will be), no empirical study has examined participants' *assumptions* about the punishment. In most criminal cases, jurors are not told what the punishment will be if they find the defendant guilty. Although Freedman et al. (1994) assert that jurors have a general knowledge of what the punishment will be and appropriately assume that harsher crimes will yield harsher sanctions (as indicated by participants responding that harsher sentences for the less severe crimes were too severe in their study), it is not clear that their assumptions about the punishment are accurate. Thus, commentators such as Judge Alex

Kozinski (2015) and Stoffelmayr and Diamond (2000) argued that jurors could be speculating incorrectly about what the potential punishment is, and therefore making decisions on faulty assumptions. The solution may be that jurors ought to be explicitly informed of the potential punishment.

This argument is rational according to Decision Theory, since, as noted above, rational decisions depend on the consequences associated with false positive and false negative decisions. J. Kaplan (1967) specifically argues that withholding the potential punishment from jurors is "hard to defend" (p. 1075) from a Decision Theory perspective. But it is not clear that the prescriptions of the law are congruent with Decision Theory on this point. It is undeniable that criminal trials are governed by beyond a reasonable doubt; however, *In re Winship* (1970) did not hold that the beyond a reasonable doubt standard of proof applies with different force depending on the seriousness of the potential punishment. For instance, based on Decision Theory, we would assume that the threshold to convict the defendant would be different depending on the charged crime, because a false positive (convicting an innocent person) for a more severe crime (and therefore punishment) is presumably worse than a false positive for a less severe crime, making the threshold to convict *higher* for the more severe crime. However, *In re Winship* never explicitly stated that this ought to be true, and commentators are torn on whether they believe it should be (e.g., Stoffelmayr & Diamond, 2000; Lillquist, 2002; but see Laudan & Saunders, 2009). Although this debate is unlikely to be settled anytime soon, we believe that an empirical study could test some of the assumptions made by commentators to revive and bring some clarity to the inconclusive and dated literature. The present study endeavors to do just that.

THE PRESENT EXPERIMENTS

The present studies, approved by the University of California, Irvine Institutional Review Board (IRB), fundamentally address three questions: a.) do jurors make assumptions about the potential punishments that would follow from a guilty verdict; b.) do these assumptions influence their propensity to reach a guilty verdict; and c.) would informing jurors of the potential punishments affect their propensity to reach a guilty verdict such that punishment severity is inversely related to their propensity to convict.

Experiment 1 manipulates the crime (Grand Theft vs. Grand Theft and Involuntary Manslaughter) holding all other factors constant. Similar to most criminal trials, Experiment 1 does not explicitly state what the punishment will be if the defendant is convicted of the crime; rather, jurors will be asked ex post what they believe the punishment would be, should be, and what the judge's punishment will be. Consistent with Decision Theory, we hypothesize that jurors will operationalize the beyond a reasonable doubt standard of proof differently between the two crimes, specifically that they will use a more stringent conceptualization of beyond a reasonable doubt in the Manslaughter case since a stronger penalty will follow from a conviction for Manslaughter than for Grand Theft.

Experiment 2 again manipulates the crime (Grand Theft vs. Grand Theft and Involuntary Manslaughter), but it also manipulates and explicates the potential punishments that would follow if the defendant were convicted. These punishments were either at the upper end of the sentencing guidelines ("high punishment"), the low end ("low punishment"), or not specified. We hypothesize that explicitly providing the potential punishment would a.) affect jurors' assumptions about the potential punishments (namely, that the disclosed punishment would anchor participants' assumption within the range provided); and b.) that these modified

assumptions would be related to their propensity to convict. Essentially, explicitly disclosing a higher punishment will reduce jurors' propensity to convict.

It is important to note that we are interested in jurors' propensity to convict, not the rate of conviction per se. Previous studies that have focused on conviction rates and the likelihood of guilt have confounded the elements required to convict with the conviction itself (e.g., Kerr, 1978). They also do not directly speak to jurors' operationalization of beyond a reasonable doubt. Consistent with authoritative commentators (Kaye, Hans, Dann, Farley, & Albertson, 2007), we utilized an approach to estimate jurors' implicit threshold for the standard of proof to see if there is a difference in *thresholds*—not just conviction rates—across the various experimental conditions. A description of this approach appears below.

CALCULATING JURORS' IMPLICIT THRESHOLDS FOR REASONABLE DOUBT

The implicit threshold for reasonable doubt can be described as a threshold (t), which, if the likelihood of guilt ($\ell(x)$) exceeds, will result in a conviction (c). A conviction implies that $\ell(x) > t$. The willingness to convict is the conditional probability of a vote to convict (c = 1) given the subjective likelihood of guilt ($\ell(x)$), or p(c = 1 | $\ell(x)$). In plain English, this states the likelihood that a juror would convict conditional on her subjective likelihood of guilt. This conditional probability can be estimated from a logistic regression. Logistic regression is appropriate because the relation between c and $\ell(x)$ is not linear. Rather, the relation tends to follow the logistic curve where the portion of convictions is quite small when $\ell(x)$ is close to zero and quite large when $\ell(x)$ is close to one. The actual shape of the logistic curve is based on empirical data; specifically, on the distribution between subjective estimates of likelihood of

guilt and convictions. The curve is derived from conducting a logistic regression, which takes on the following form:

$$\ln\left(\frac{p}{1-p}\right) = \alpha + \beta(\ell(x)) \tag{2}$$

According to this equation, $0 \leq \ell(x) \leq 1$ and $0 < p < 1$ (since p is undefined at 0 or 1), and the natural logarithm of the odds of voting to convict is linear in $\ell(x)$. The logistic regression provides a maximum likelihood estimate of α and β.

The log-odds indicate the willingness to convict—as a continuous variable. However, we are interested in a point estimate, specifically the point where a participant is equipoise between conviction and acquittal. When p = 0.5, the log-odds are zero, indicating that a conviction is equally likely as an acquittal. The relevant query is determining what subjective likelihood of guilt ($\ell(x)$) corresponds to this level of willingness. This can be determined by substituting p = 0.5 into the previous equation, which reduces the log-odds to zero, and solving for $\ell(x)$:

$$t = \ell(x)_{1/2} = -\frac{\alpha}{\beta} \tag{3}$$

The implicit threshold $\ell(x)_{\frac{1}{2}}$ indicates the subjective likelihood of guilt $\ell(x)$ at which a conviction is as equally likely as an acquittal. By conducting a logistic regression for each condition, we can identify what the implicit threshold is, and how disclosing the punishment impacts the threshold.

EXPERIMENT 1

Participants and Procedure

One hundred and two participants were recruited online through TurkPrime, a premium platform of Amazon's Mechanical Turk (see generally, Mason & Suri, 2012; Litman, Robinson & Abberbock, 2016). To simulate the makeup of a typical jury, participants were eligible if they were at least 18 years old, had no felony convictions, and a United States citizen at the time of the study. Upon completion of the study, participants received a small monetary reward as compensation. The median age of the sample was 32 (inter-quartile range [IQR] = 14.25). The sample was composed of 54% (n = 55) males; 50% (n = 51) of participants identified as politically liberal, 21% (n = 21) identified as politically conservative, 27% (n = 28) identified as Independent, and the rest indicated some "other" type of political affiliation. An attention check question was used to ensure participants were paying attention to the materials, though no participants failed the attention check embedded in this study.

Participants read a synopsis of a criminal case that has been used in previous research (D. Simon et al., 2004; D. Simon & Scurich, 2011). The defendant in the case was charged with either Grand Theft (n = 51) or Grand Theft and Involuntary Manslaughter (Manslaughter) (n = 51). Grand Theft was defined as it is in the California penal code, as "taking possession of property valued at over $2,500 that is owned by someone else," while Involuntary Manslaughter was defined as it is in the California penal code, as "causing the death of another person while committing a crime." Participants were randomly assigned to one of the two possible conditions.

The facts of the case were the same in each experimental condition except that in the Manslaughter case a security guard was found dead at the scene. In both conditions, the defendant allegedly stole $5,200 from the company's safe. Incriminating evidence included a confident eyewitness, video footage of a car similar to the defendant's leaving the parking lot around the time that the crime was committed, and possible motives such as a paid-off debt and a

delayed promotion. Evidence for the defense included an alibi witness who saw the defendant dressed in different clothing shortly after the crime and far away from the crime scene, as well as an explanation for the debt. Participants also received official judicial instructions on the standard of proof and the presumption of innocence before and after they read the case. The approximate length of the case materials was 900 words.

After participants read the case they were asked to render a verdict, state their subjective belief regarding the likelihood that the defendant committed the alleged act (1-100%), and answer the following questions about their opinion and expectations of what the punishment would be if the defendant were convicted: i) How long (in years) would you sentence the defendant to prison?; ii) How long (in years) do you think the judge will sentence the defendant to prison?; and iii) What do you think the prison sentence would be (in years) if convicted of this crime? The order of these questions was counterbalanced. Finally, participants provided their demographic information.

Results

Table 1 contains participants' verdicts, likelihood of guilt ratings, and estimates of the associated punishment, decomposed by experimental condition.

Table 1. *Experiment 1 Conviction Rates, Likelihood of Guilt Ratings, and Sentencing Beliefs Decomposed by Experimental Condition*

	Grand Theft *(n = 51)*		*Manslaughter* *(n = 51)*	
	Mean	*SD*	*Mean*	*SD*
Guilty Verdict	0.51	0.51	0.45	0.50
Likelihood of Guilt	67.80	25.14	68.03	23.38
Participant Prison Sentence**	3.48	4.92	11.80	11.50
Judge Prison Sentence**	4.96	7.63	11.62	10.09
General Prison Sentence***	7.00	14.25	16.22	8.67

* $p < .001$, comparing Grand Theft results to Manslaughter results

** Participants were asked what punishment they would give the defendant if they were the judge, and what punishment they think the judge would give the defendant.

*** Participants were asked what they thought the prison sentence would be for someone convicted of the charge, irrespective of the case they read.

As seen in Table 1, although there were slightly fewer convictions in the Manslaughter ($M = .45$, $SD = .50$) condition compared to Grand Theft ($M = .51$, $SD = .51$), the difference was not statistically significant ($\chi^2 (1) = .35$, $p = .55$). Additionally, the means for likelihood of guilt were almost identical for both conditions and were not statistically significant, $t(100) = -.05$, $p = .96$; Cohen's d (d) = .01 , 95% Confidence Interval (CI) [-.54, .56]. However, the means for all questions regarding prison sentence were significantly different between the conditions. In particular, participants would sentence the defendant to a longer sentence in the Manslaughter condition, $t(67.72) = -4.75$, $p < .001$; $d = .94$, 95% CI [.36, 1.52], they believed the judge would sentence the defendant to a longer prison sentence in the Manslaughter condition, $t(93.14) = -3.76$, $p < .001$; $d = .75$, 95% CI [.18, 1.31], and they believed someone convicted of

Manslaughter would be sent to prison for a longer term than if the person were convicted of Grand Theft, $t(100) = -3.95$, $p < .001$; $d = .78$, 95% CI [.21, 1.35]. These differences validate the experimental manipulation, and also suggest that jurors do have assumptions regarding the potential punishment and that these assumptions systematically vary by crime.

A logistic regression with participants' likelihood of guilt rating as the independent variable and their verdict as the dependent variable was conducted for each experimental condition. Both models revealed that the likelihood of guilt is a strong predictor of the propensity to convict, $\chi^2(5)_{Grand\ Theft} = 29.63$, $p < .001$; $Exp(B) = 1.09$, 95% CI [1.04, 1.14]; $\chi^2(5)_{Manslaughter} = 29.33$, $p < .001$; $Exp(B) = 1.14$, 95% CI [1.05, 1.22]. Figure 1 (below) shows a fitted logistic regression curve. The implicit threshold is estimated for Grand Theft at 69.40% ($SE = 4.63$; 95% CI [58.92, 78.38]), and for the Manslaughter condition the threshold is estimated at 74.72% ($SE = 3.07$, 95% CI [67.44, 79.93]) (See Wright, Strubler, & Vallano, 2011 on calculating the standard error and confidence interval). Since the confidence intervals for each condition overlap, the thresholds are not statistically different. Thus, although participants in the Manslaughter condition assumed that the prison term would be longer than participants in the Grand Theft condition, their implicit operationalization of the beyond a reasonable doubt standard of proof did not change between the two experimental conditions. In other words, participants' assumptions regarding the potential sanctions did not affect their propensity to convict.
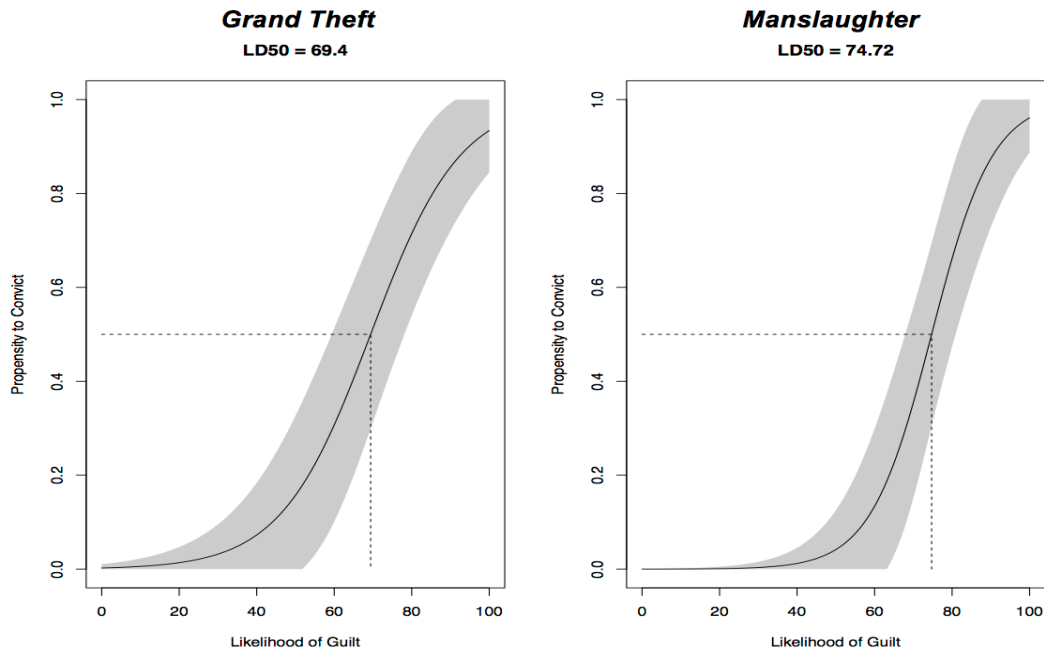
Figure 1. *Logistic regression predicting the probability of a guilty verdict from participants' likelihood of guilt rating. Note that shaded area is 95% Confidence Interval.*

EXPERIMENT 2

Experiment 1 manipulated the crime presented to participants, but it did not inform them of the potential punishment that would follow from a conviction. Experiment 2 not only manipulates the crime, it also manipulates the potential punishment by explicitly disclosing what the potential punishment would be if the defendant were convicted.

Participants and Procedure

Three hundred and three participants were recruited through TurkPrime. Consistent with current practice (Oppenheimer et al., 2009), six participants were excluded from analysis for failing the attention check (n=4), being younger than 18 years old (n=1), or not a U.S. citizen (n=1), resulting in a total of 297 participants who were at least 18 years old and a U.S. citizen that remained in the analysis. The median age for the sample was 39.36 years old (*IQR* = 22).

The sample was composed of 43.8% males ($n = 130$); 46.5% identified as politically liberal ($n = 138$); 32.6% ($n = 97$) identified as politically conservative, 17.5% ($n = 52$) identified as Independent, and the rest indicated some "other" type of political affiliation.

In this study, we attempted to experimentally manipulate participants' threshold by having the judge inform participants before they render a verdict that, if convicted, the defendant will receive a 2-6 year prison term ("low punishment"), a 7-20 year prison term ("high punishment"), or those in the control group were not told a sentence length. Again, we manipulated the crime of which the defendant was charged. The experimental design was a 2 (crime: Grand Theft or Manslaughter) X 3 (punishment disclosed: low or high or none (control)) between-participants fully-crossed factorial design. Thus, participants were randomly assigned to one of six possible cells.

All the facts in the case were the same as Experiment 1. The question "What do you think the prison sentence would be (in years) if the defendant is convicted of this crime?" was removed for this study, as most participants were told what the punishment would be if convicted.

Results

Table 2 contains participants' verdicts, likelihood of guilt ratings, and estimates of the associated punishment, decomposed by experimental conditions.

Table 2. *Percentage of Guilty Verdicts, Likelihood of Guilt Ratings, and Sentencing Opinions Decomposed by Experimental Condition*

| | Grand Theft | | | | | | Manslaughter | | | | | |
| | Low Punishment Disclosed (n = 48) | | High Punishment Disclosed (n = 59) | | No Punishment Disclosed (n = 43) | | Low Punishment Disclosed (n = 50) | | High Punishment Disclosed (n = 41) | | No Punishment Disclosed (n = 56) | |
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Guilty Verdict	0.35	0.48	0.39	0.49	0.56	0.50	0.40	0.50	0.54	0.51	0.46	0.50
Likelihood of Guilt Rating	64.73	22.97	68.05	25.46	63.95	30.38	65.34	26.37	76.29	19.11	65.88	25.65
Participant Prison Sentence	1.89	2.03	6.00	14.53	3.71	6.12	5.40	14.93	11.44	6.87	14.41	16.95
Judge Prison Sentence	2.47	1.81	5.69	4.64	4.23	4.46	3.84	3.48	11.66	6.02	13.57	12.21

A logistic regression with the experimental factors as independent variables and participant's verdicts (guilty or not guilty) revealed that neither crime, punishment disclosed, nor the interaction significantly predicted verdicts, χ^2 (5) = 6.49, p = .26. Additionally, the results from a two-way ANOVA with participants' likelihood of guilt ratings as the dependent variable indicated no significant main effect for crime, $F(1, 291)$ = 1.47, p = .23, partial η^2 = .01, no significant main effect for punishment disclosed, $F(2, 291)$ = 2.62, p = .074, partial η^2 = .02, nor for an interaction, $F(2, 291)$ = .63, p = .53, partial η^2 = .004. However, a two-way ANOVA with participants' prison sentence (i.e., how long the participants would sentence the defendant to prison if convicted) yielded a significant main effect for crime, $F(1, 291)$ = 21.39, p < .001; partial η^2 = .07, and for punishment, $F(2, 291)$ = 6.14, p = .002, partial η^2 = .04, but no significant interaction, F(2, 291) = 2.30, p = .10, partial η^2 = .02. Participants would impose longer prison sentences when the crime was Manslaughter (marginal overall M = 10.41; SE = 1.01) compared to Grand Theft (marginal M = 3.87, SE = .98), p < .001, d = 6.57, 95% CI [6.00, 7.15]. With respect to punishment disclosure, participants would impose longer prison sentences

when high punishment was disclosed (marginal $M = 8.72$, $SE = 1.23$) or no punishment was disclosed (marginal $M = 9.01$, $SE = 1.23$) compared to when low punishment was disclosed (marginal $M = 3.64$, $SE = 1.22$), (High vs. Low: $p = .004$, $d = -8.13$; CI [-9.18, -7.45]; No vs. Low: $p = .004$, $d = -8.71$; CI [-9.62, -7.81]).

With respect to how long participants thought the judge would send the defendant to prison, there was a significant main effect for crime, $F(1, 291) = 52.58$, $p < .001$, partial $\eta^2 = .15$, punishment, $F(2, 291) = 24.13$, $p < .001$, partial $\eta^2 = .14$, and the interaction, $F(2, 291) = 9.10$, $p < .001$, partial $\eta^2 = .06$. Participants assumed that the judge would impose a longer sentence when high ($M = 13.57$, $SE = .86$) or no punishment ($M = 11.66$, $SE = 1.02$) was disclosed for the Manslaughter conditions versus the respective Grand Theft conditions ($M = 5.69$, SE = .85; $M = 4.23$, $SE = 1.00$, respectively), ps $< .001$ (High: $d = -9.23$, 95% CI [-10.57, -7.89]; No: $d = -7.35$, 95% CI [-8.45, -6.25]), but there was no difference in the judge's prison sentence between Manslaughter and Grand Theft when low punishment was disclosed ($M = 3.84$, $SE = .93$; $M = 2.47$, $SE = .95$, respectively), $p = .30$, $d = -1.45$, 95% CI [-1.90, -1.01]. This latter finding (i.e., no difference between Grand Theft and Manslaughter) could be due to a restriction of range in the potential punishment, which spanned from 2-6 years. Nonetheless, this general pattern of results is consistent with Experiment 1 in that participants appear to assume that the judge will impose a longer sentence in the Manslaughter case as would the participants themselves, but these different assumptions about the possible punishment did not affect the overall conviction rate. Put differently, the experimental manipulations affected participants' assumptions about the potential punishment but not their verdicts.

Again, to calculate the implicit threshold, a logistic regression for each of the six experimental conditions was conducted with the likelihood of guilt as the independent variable

and verdict as the dependent variable. There were no differences in the implicit thresholds across the experimental conditions (all of the 95% confidence intervals overlap). As seen in Figure 2 (below), in the Grand Theft condition, the threshold to convict when a low punishment was disclosed was estimated at 76.05 (*SE* = 3.65, 95% *CI* [69.75, 84.55]), the threshold for a high punishment disclosure was estimated at 80.14 (*SE* = 3.22, 95% *CI* [73.46, 86.22]), and the threshold when no punishment was disclosed was 58.91 (*SE* = 8.50, 95% *CI* [36.37, 73.62]). In the Manslaughter condition, the threshold to convict when a low punishment was disclosed was estimated at 76.74 (*SE* = 3.46, 95% *CI* [68.83, 83.40]), the threshold for a high punishment disclosure was estimated at 76.10 (*SE* = 3.50, 95% *CI* [68.91, 83.03]), and the threshold when no punishment was disclosed was 70.06 (*SE* = 4.49, 95% *CI* [60.47, 79.07]).
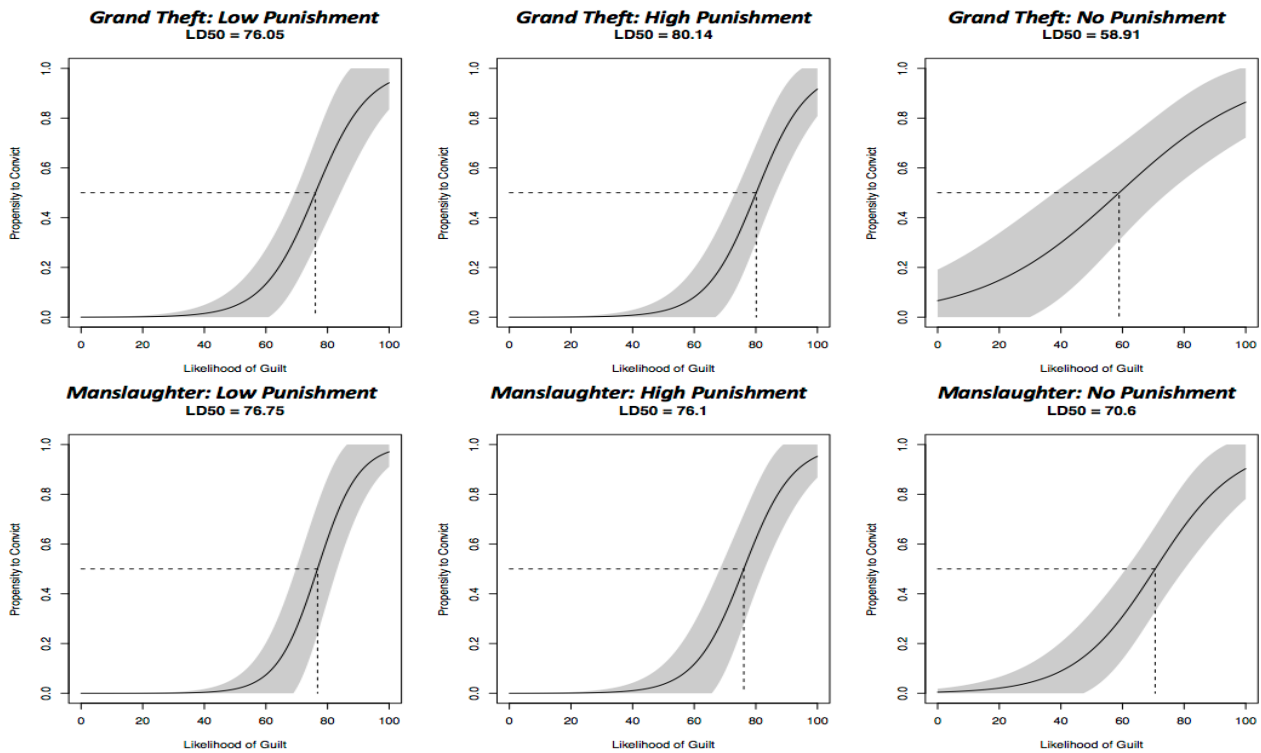

Figure 2. *Predicting the probability of a guilty verdict from the likelihood of guilt based on severity of crime and severity of punishment. Shaded area is the 95% Confidence Interval.*

DISCUSSION

A lively debate exists regarding whether jurors' assumptions do and should affect their propensity to convict criminal defendants. The existing empirical studies have provided mixed results with regard to the effect of potential punishment on verdicts. Some of the inconsistencies in the results across studies could be due to methodological differences (e.g., a small sample of undergraduates versus Canadian adults. See Freedman, 1994). Moreover, the previous studies examined conviction rates, not jurors' threshold for beyond a reasonable doubt per se. Conviction rates are not optimal because they can be noisy (i.e., include unsystematic error) and they confound the likelihood of guilt ratings with the decision threshold. We believe the experiments reported here provide a methodological improvement over the previous studies since they used a relatively large sample of jury-eligible adults and a means to assess jurors' implicit threshold for beyond a reasonable doubt. Notably, this approach did not require participants to introspect about their threshold; rather, the threshold values were derived mathematically based on participants' behavior. Consistent with other commentators (Kaye et al., 2007), we believe this is a superior method to assess how jurors effectuate beyond a reasonable doubt.

Experiment 1 revealed that mock jurors do make assumptions about the potential punishments, and that these assumptions systematically differed as a function of the crime. Specifically, participants assumed that a conviction for Manslaughter would occasion a longer prison sentence than a conviction for Grand Theft. However, these apparent assumptions did not affect their implicit threshold for conviction, contrary to our hypothesis. Experiment 2 replicated this effect. Participants had different expectations about the potential punishments associated with each crime, and these expectations could be changed when the judge explicitly informed

20

participants about the range of the potential punishments. Yet again, the expectation about the potential punishment did not affect their implicit conviction threshold. This null result calls into question whether jurors need to be informed of the potential punishment, since their assumptions about the punishment do not appear to impact their conviction threshold. Moreover, their uninformed assumptions do not appear to be outside the bounds of the sentencing guidelines (see Table 2).

The results reported here suggest that jurors apply the beyond a reasonable doubt standard of proof consistently across at least two disparate criminal cases. From a legal perspective, this is arguably an appropriate pattern of behavior. Although Justice Harlan's oft-cited concurrence in *In re Winship* (1970) is grounded in the decision-theoretic logic, the holding does not say that beyond a reasonable doubt applies with lesser force in less serious crimes. It simply states that it applies to all criminal cases. Whether this result generalizes to other cases is an important empirical question that must be addressed by future research. Indeed, future research should examine a wide variety of crimes, including for example so-called "third strike felonies" in which a conviction will result in an automatic twenty-five years to life term of incarceration. Capital cases, in which a jury must determine guilt as a perquisite, would also be ripe for testing whether conviction thresholds vary.

The usual limitations associated with controlled laboratory experiments apply (see Diamond, 1997; Wiener, Krauss, & Liberman, 2011). In brief, participants read a synopsis of a criminal trial and rendered a verdict without deliberation. Though the ecological validity may seem compromised, a recent meta-analysis found no difference between video and written stimuli (Bornstein et al., 2016), and other studies have demonstrated that jury deliberation may change the dynamics of the conversation but typically not the verdicts (see Devine, Clayton,

Dunford, & Seying, 2001). Additionally, participants recruited from TurkPrime may not be representative of any particular population, though these participants are likely to be more representative of the average juror than studies that only use undergraduate participants (Burhmester, Kwang, & Gosling, 2011). Studies that have been conducted using participants recruited from Mechanical Turk have replicated other psychological findings and effects (Crump, McDonnell, & Gureckis, 2013). These limitations should be kept in mind when considering the practical implications of the data reported in this manuscript.

Future research might endeavor to replicate our findings, using different samples of participants and different types of crimes. Future research could look at qualitative differences in punishment (e.g., a fine, house arrest, becoming a registered sex offender, etc.), and/or crimes that participants know less about and would have fewer preconceived notions about the potential punishment. Future studies could also examine participants' punishment motives (e.g., retribution or rehabilitation) or their expectations of the long-term effects of imprisonment that could influence their cost/benefit analysis. Also, future research could apply the implicit threshold methodology widely to other decision making tasks in which a binary choice is made under conditions of uncertainty.

Final Thoughts

The data reported in this paper suggest that jurors do make assumptions about the potential punishments that would follow from a guilty verdict, which is consistent with a bevy of anecdotal evidence from jurors and other commentators. The data also reveal, however, that these assumptions do not impact jurors' operationalization of the beyond a reasonable doubt standard. Indeed, jurors in these studies appeared to apply the same standard of proof across

different cases with very different associated penalties. While the normative issue of whether this is appropriate behavior cannot be settled by empirical findings, we believe the debate can be better informed by understanding that jurors' assumptions about potential punishments, contrary to widespread belief, do not appear to directly impact jurors' propensity to convict.

REFERENCES

Arkes, H. R., & Mellers, B. A. (2002). Do juries meet our expectations? *Law and Human Behavior, 26*(6), 625-639. doi:10.1023/A:1020929517312

Baguley, C. M., McKimmie, B. M., & Masser, B. M. (in press). Deconstructing the simplification of jury instructions: How simplifying the features of complexity affects jurors' application of instructions. *Law and Human Behavior.* doi: http://dx.doi.org/10.1037/lhb0000234

Bornstein, B. H., Golding, J. M., Neuschatz, J., Kimbrough, C., Reed, K., Magyarics, C., & Luecht, K. (2017). Mock juror sampling issues in jury simulation research: A meta-analysis. *Law and Human Behavior, 41(1),* 13-28. doi: http://dx.doi.org/10.1037/lhb0000223

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data?. *Perspectives on Psychological Science*, *6*(1), 3-5. doi: 10.1177/1745691610393980

Ceci, S.J., & Friedman, R.D. (2000). The suggestibility of children: Scientific research and legal implications. *Cornell Law Review, 86*, 33–108.

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS one*, *8*(3), e57410. doi: http://dx.doi.org/10.1371/journal.pone.0057410

Davis, J. H., Kerr, N. L., Stasser, G., Meek, D., & Holt, R. (1977). Victim consequences, sentence severity, and decision processes in mock juries. *Organizational Behavior and Human Performance*, *18*(2), 346–365. doi: http://dx.doi.org/10.1016/0030-5073(77)90035-6

Devine, D. J., Clayton, L. D., Dunford, B. B., Seying, R., & Pryce, J. (2001). Jury decision

making: 45 years of empirical research on deliberating groups. *Psychology, Public

Policy, and Law*, *7*(3), 622-727. doi: 10.1037//1076-8971.7.3.622

Diamond, S. S. (1997). Illuminations and shadows from jury simulations. *Law and Human

Behavior, 21*(5), 561-571. doi:10.1371/journal.pone.0057410.g001

Freedman, J. L. (1994) Penalties and verdicts: Keeping the record straight. *Law and Human

Behavior, 18(6)*, 699-702. doi:10.1007/BF01499333

Freedman, J. L., Krismer, K., Macdonald, J. E., Cunningham, J. A., & Al, F. E. T. (1994).

Severity of penalty, seriousness of the charge, and mock jurors' verdicts. *Law and Human

Behavior*, *18*(2), 189–202. doi:10.1007/BF01499015

In re Winship, (1970).

Grinberg, E. (2016, Decmber 19). Judge in Brock Turner case cleared of misconduct allegations.

CNN. Retreived from:

http://www.cnn.com/2016/12/19/us/judge-aaron-persky-misconduct-investigation/

Kadvany, E. (2016, June 13). Brock Turner juror to judge: 'Shame on you'. *The Palo Alto

Weekly*. Retrieved from: http://www.paloaltoonline.com/news/2016/06/13/brock-turner-

juror-to-judge-shame-on-you

Kagehiro, D. K., & Stanton, W. C. (1985). Legal vs. quantified definitions of standards of proof.

Law and Human Behavior, 9, 159-178. doi:10.1007/BF01067049

Kaplan, J. (1967). Decision Theory and the fact finding process. *Stanford Law Review, 20*, 1065-

1077.

Kaplan, K. J., & Simon, R. I. (1972). Latitude and severity of sentencing options, race of the

victim and decisions of simulated jurors: Some issues arising from the Algiers Motel

    trial. *Law & Society Review*, *7*, 87.

Kaplan, M. F., & Krupa, S. (1986). Severe penalties under the control of others can reduce guilt

    verdicts. *Law and Psychology Review, 10*, 1-18.

Keeney, R. L., & Raiffa, H. (1967) *Decisions with multiple objectives: Preferences and value*

    tradeoffs. Cambridge: Cambridge University Press.

Kerr, N. L. (1978). Severity of prescribed penalty and mock jurors' verdicts. *Journal of*

    Personality and Social Psychology, *36*(12), 1431–1442.

    doi: http://dx.doi.org/10.1037/0022-3514.36.12.1431

Kozinski, A. (2015). Criminal law 2.0. *Georgetown Law Journal Annual Review Criminal*

    Procedure, 44, iii – xliv.

Krauss, D., & Scurich, N. (2014). The impact of case factors on jurors' decisions in a sexual

    violent predator hearing. *Psychology, Public Policy, and Law*, *20*(2), 135. doi:

    10.1037/law0000007

Laudan, L., & Saunders, H. D. (2009). Re-thinking the criminal standard of proof: Seeking

    consensus about the utilities of trial outcomes. *International Commentary on Evidence,*

    7(2), 1-34. doi: https://doi.org/10.2202/1554-4567.1099

Lempert, R. O. (1977). Modeling relevance. *Michigan Law Review, 75*, 1021-1057.

Lillquist, E. (2002). Recasting reasonable doubt: Decision theory and the virtues of

    variability. *University of California, Davis Law Review*, *36*, 85-1285.

Lillquist, E. (2004). The puzzling return of jury sentencing: Misgivings about apprendi. *North*

    Carolina Law Review, 82(621), 635-636.

Litman, L., Robinson, J., & Abberbock, T. (2016). TurkPrime.com: A versatile crowdsourcing

data acquisition platform for the behavioral sciences. *Behavior Research Methods*, 1-10.
doi:10.3758/s13428-016-0727-z

Magnussen, S., Eilertsen, D. E., Teigen, K. H., & Wessel, E. (2014). The probability of guilt in
criminal cases: Are people aware of being 'beyond reasonable doubt'?. *Applied Cognitive
Psychology*, *28*(2), 196-203. doi: 10.1002/acp.2978

Mason, W., & Suri, S. (2012). Conducting behavior research on Amazon's Mechanical Turk.
Behavior Research Methods, 44, 1-33. doi:10.3758/s13428-011-0124-6

Myers, M. A. (1979). Rule departures and making law: Juries and their verdicts. *Law and Society
Review*, 781-797.

Niedermeier, K. E., Horowitz, I. A., & Kerr, N. L. (1999). Informing jurors of their nullification
power: A route to a just verdict or judicial chaos? *Law and Human Behavior*, *23*(3), 331–
351. doi:10.1023/A:1022360632283

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks:
Detecting satisficing to increase statistical power. *Journal of Experimental Social
Psychology, 45*(4), 867-872. doi:10.1016/j.jesp.2009.03.009

Risinger, D. M. (1997). John Henry Wigmore, Johnny Lynn Old Chief, and Legitimate Moral
Force: Keeping the Courtroom Safe for Heartstrings and Gore. *Hastings Law Journal*, *49*,
403-462.

Rocha, V., & Winton, R. (2016, June 08). Light sentence for Stanford swimmer in sexual assault
'extraordinary,' legal experts say. *The Los Angeles Times*. Retreived from:
http://www.latimes.com/local/lanow/la-me-ln-stanford-sexual-assault-sentence-
20160607-snap-story.html

Scurich, N. (2015). Criminal justice policy preferences: Blackstone ratios and the veil of

ignorance. *Stanford Law & Policy Review*, *26*, 23-35.

Simon, D., & Scurich, N. (2011). Lay judgments of judicial decision making. *Journal of Empirical Legal Studies*, *8*(4), 709-727. doi: 10.1111/j.1740-1461.2011.01238.x

Simon, D., Snow, C. J., & Read, S. J. (2004). The redux of cognitive consistency theories: Evidence judgments by constraint satisfaction. *Journal of personality and social psychology*, *86*(6), 814-837. doi: 10.1037/0022-3514.86.6.814

Stoffelmayr, E., & Diamond, S. S. (2000). The conflict between precision and flexibility in explaining "beyond a reasonable doubt". *Psychology, Public Policy, and Law*, *6*(3), 769–787. doi: 10.1037//1076-8971.6.3.769

Vidmar, N. (1972). Effects of decision alternatives on the verdicts and social perceptions of simulated jurors. *Journal of Personality and Social Psychology*, *22*(2), 211-218. doi: 10.1037/h0032605

Weinstein, J.B., & Dewsbury I. (2006). Comment on the meaning of 'proof beyond a reasonable doubt'. *Law, Probability and Risk, 5*, 167-173. doi: https://doi.org/10.1093/lpr/mgl016

Wiener, R. L., Krauss, D. A., & Lieberman, J. D. (2011). Mock jury research: Where do we go from here?. *Behavioral Sciences & the Law, 29*(3), 467-479. doi: 10.1002/bsl.989

Wright, D. B., Strubler, K. A., & Vallano, J. P. (2011). Statistical techniques for juror and jury research. *Legal and Criminological Psychology, 16,* 90–125. doi: 10.1348/135532510X487655