

An Experimental Exploration of Reasonable Doubt

Jason A. Aimone Stanton Hudja Wilson Law
Charles M. North Jason Ralston Lucas Rentschler*

April 3, 2025

Abstract

The “beyond a reasonable doubt” standard is a constitutional requirement in U.S. criminal cases, but jury instructions on “reasonable doubt” vary across jurisdictions. We use a controlled experiment to analyze the relationship between the definition of reasonable doubt and juror decisions. In our novel (pre-registered) experiment, we vary the definition of reasonable doubt between subjects and elicit the level of evidence required for subjects to convict a defendant. We analyze juror decisions under two state definitions that are markedly different (Wisconsin and West Virginia) and analyze juror decisions when reasonable doubt is not explicitly defined. We find similar behavior in each treatment. We ran three additional treatments to determine why behavior does not seem to vary across definitions. Our data is consistent with subjects having pre-conceived notions of reasonable doubt that are not affected by jury instructions.

JEL Codes: C91, D91, K40

Keywords: Reasonable Doubt, Criminal Justice, Experiments

*We thank the Charles Koch Foundation for generously funding this study. We appreciate helpful feedback from seminar participants at the 2021 Global Economic Science Association Conference. Aimone: Department of Economics, Baylor University, Waco, TX, 76798, USA; Hudja: Department of Economics, University of Toronto, Toronto, ON, M5S 2E9, CA; Law: Department of Economics, Baylor University, Waco, TX, 76798, USA; North: Department of Economics, Baylor University, Waco, TX, 76798, USA; Ralston: Department of Economics, Whitman College, Walla Walla, WA, 99362, USA; Rentschler: Department of Economics and Finance and the Center for Growth and Opportunity, Utah State University, Logan, UT, 84322, USA.

“It is critical that the moral force of the criminal law not be diluted by a standard of proof that leaves people in doubt whether innocent men are being condemned”¹

1 Introduction

Defendants in the American criminal justice system cannot be convicted unless the prosecution proves every element of the charged crime *beyond a reasonable doubt*.² The ancient common law notion that criminal cases require heightened proof crystallized in the United States as the reasonable doubt standard by the late 1700s (Diamond, 1990). “Proof beyond a reasonable doubt” became a constitutional requirement in 1970, when the Supreme Court held in *In re Winship* that “the Due Process Clause protects the accused against conviction except upon proof beyond a reasonable doubt of every fact necessary to constitute the crime with which he is charged.” While the reasonable doubt standard is ubiquitous in the United States, there is no consensus on whether (and how) it should be defined for a jury.

Instead, U.S. states and federal circuits differ considerably in how they explain the reasonable doubt standard to juries. Indeed, U.S. jurisdictions differ in whether they define reasonable doubt at all. Some jurisdictions require courts to provide a definition of reasonable doubt for jurors, while others allow but do not require a definition. Some jurisdictions even prohibit (or strongly discourage) defining reasonable doubt for jurors. Further, among jurisdictions that do provide reasonable doubt definitions, there is considerable variation.

In this paper, we use an experiment to explore whether reasonable doubt definitions influence juror decisions. We utilize an experiment as it bypasses many of the difficulties of trying to infer reasonable doubt thresholds from court outcome data. We are partially inspired by *Victor v. Nebraska*, where the supreme court acknowledged (but did not find) that differing reasonable doubt definitions may matter. We address two main questions. First, do different reasonable doubt definitions significantly influence juror behavior? Second, which characteristics of the definition are correlated with jurors’ interpretation of reasonable doubt?

We utilize an experiment as it grants us the unique control of varying the reasonable doubt definitions given to subjects. There are two types of sessions in our experiment: (i) theft game sessions and (ii) juror decision sessions. The theft game sessions are run in order to provide real human defendants for the juror decision sessions as this makes jurors’ decisions salient and incentivized. In the theft game sessions, subjects are grouped into pairs. One subject can take money from the other subject in the pair. The subject who can take money may be charged with theft regardless of whether or not she took money. In the juror decision sessions, subjects make decisions about whether

¹*In re Winship*, 397 U.S. 358, 364 (1970)

²Juror instructions include this restriction. For plea bargains, the defendant must admit guilt, which in practice is taken as proof beyond a reasonable doubt. Nonetheless, concerns about the propensity of innocent defendants to accept plea bargains abound. See e.g., Ralston et al. (2021).

to convict a charged defendant. They are asked to report their guilt threshold, which is the lowest probability of guilt that they believe is proof beyond a reasonable doubt. A defendant is found guilty if the defendant's perceived probability of guilt is above a randomly chosen juror's guilt threshold.³ Note that we ran the theft game sessions before the juror decision sessions. After the juror decision sessions, jurors' thresholds are used to determine subjects' payments in the theft game sessions.

In our pre-registered experiment, we vary the definition of reasonable doubt provided to jurors between subjects. Our first round of sessions included three treatments. The first treatment is the No Definition treatment, where subjects are not given any definition of reasonable doubt. The second treatment is the Wisconsin treatment, where subjects are given the Wisconsin definition of reasonable doubt. The third treatment is the West Virginia treatment, where subjects are given the West Virginia definition of reasonable doubt. We conjecture that the Wisconsin treatment will result in more convictions (and lower guilt thresholds) than the No Definition treatment as we believe that the Wisconsin definition focuses heavily on reducing type 2 error (acquitting a guilty individual). We also conjecture that the No Definition treatment will result in more convictions (and lower guilt thresholds) than the West Virginia treatment as we believe that the West Virginia definition focuses heavily on reducing type 1 error (convicting an innocent individual).

We surprisingly find similar guilt thresholds in the No Definition, Wisconsin, and West Virginia treatments and that there are no statistically significant differences between these treatments. However, we do find that certain groups differ in their general interpretation of reasonable doubt. For example, we find lower guilt thresholds among black subjects than among white subjects.

The similarity of observed guilt thresholds in our three treatments inspired us to test two possible explanations in a second pre-registered experiment. The first possible explanation is that states' reasonable doubt definitions use hedging language. It is possible that these definitions both truncate guilt thresholds from above and from below as they both include language that appears to prevent type 1 and type 2 error. We ran two new treatments to address this possibility: (i) the Wisconsin Truncated treatment and (ii) the West Virginia Truncated treatment. In the Wisconsin (West Virginia) Truncated treatment, we truncate the Wisconsin (West Virginia) reasonable doubt definition to remove some of the hedging language. The second possible explanation is that subjects ignore the reasonable doubt standard. We design a new treatment, the No Standard treatment, to address this possibility. In this new treatment, we remove the reasonable doubt language from the No Definition treatment.

The results of the three additional treatments suggest that subjects have pre-conceived notions of reasonable doubt that determine their guilt thresholds. We find that the guilt thresholds in the Wisconsin Truncated treatment are similar to the previously obtained thresholds from the Wisconsin treatment. Additionally, we find that the guilt thresholds in the West Virginia Truncated treatment are similar to the previously obtained thresholds from the West Virginia treatment. Thus, hedging language is not driving the similar

³We explain this in more detail in Section 4.3.

guilt thresholds observed in the original treatments. We do find that the guilt thresholds in the No Standard treatment are different than the previously obtained thresholds from the No Definition treatment. However, this difference is not statistically significant once we control for demographic variables. These additional treatments suggest that subjects have pre-conceived notions of reasonable doubt as their behavior does not change across treatments.

This paper contributes to a growing literature that analyzes how jurors interpret reasonable doubt. This literature has shown conflicting results in the past. There are two groups of papers in this literature. The first group of papers analyze the role of juror instructions on juror decisions and finds inconsistent results. [Kagehiro \(1990\)](#) presents mock jurors a past civil court case and finds that reasonable doubt definitions do not matter and that there is no difference between preponderance of evidence and reasonable doubt. In a series of papers ([Cicchini and White, 2016, 2017](#); [White and Cicchini, 2019](#)), Cicchini and White present hypothetical case descriptions to mock jurors on Amazon Mechanical Turk, who individually decided whether the accused defendant was guilty or not guilty. They find inconsistent evidence as they find that the language used to define reasonable doubt influences juror behavior, but also find that jurors do not differentiate between the preponderance of evidence and beyond a reasonable doubt standards. [Glöckner and Engel \(2013\)](#), using hypothetical cases and university students, find that jurors do differentiate between the preponderance of evidence and beyond a reasonable doubt standards. The second group of papers tries to uncover the minimum threshold for evidence to be beyond a reasonable doubt. These papers ([Dhimi et al., 2015](#); [Tillers, 1993](#); [Kaye, 1986](#)) suggest that the minimum level of guilt required is around 85%. However, [Dhimi \(2008\)](#) shows that estimates of this threshold are dependent on the elicitation method.

Our study differs from this literature in two key ways. First, our methodology improves upon the previous methodology by having jurors make consequential decisions. The non-consequential nature of the previous literature may be leading to inconsistent results.⁴ In our study, jurors in our experiments make decisions that influence the punishment of subjects in the theft game sessions. In fact, the magnitude of the punishments that were implemented in the event of a conviction were extremely large for an online experiment. The consequential nature of juror decisions in our experiment is important as juror behavior may differ between consequential and non-consequential environments.⁵ Second, we ask different questions in this paper than the previous literature. In our original three treatments, we ask (i) whether states' definitions of reasonable doubt are influencing juror decisions and (ii) whether jurors with differing characteristics have different

⁴In experimental economics, hypothetical decisions are often seen as having lower internal validity than incentivized decisions ([Voslinsky and Azar, 2021](#)).

⁵Previous research ([Hudja et al., 2021](#)) suggests that juror behavior may differ in consequential and non-consequential environments. In the experimental economics literature, studies have shown differences between incentivized and hypothetical decisions ([Burke et al., 1996](#); [Harrison, 1994](#); [Clot et al., 2018](#)). While jurors do not get paid based on the outcome of the trial, their decisions in our experiment have real consequences and thus are not hypothetical.

interpretations of reasonable doubt. In response to the results from our first experiment, we ask (i) whether hedging language in state definitions leads to similar juror behavior across definitions and (ii) whether reasonable doubt definitions matter in criminal trials.

Our paper also contributes to the experimental literature on law and economics. This literature has heavily focused on juror behavior. Previous research has generally focused on how judges and jurors make decisions. [Sonnemans and van Dijk \(2012\)](#) suggest that judges tend to make inaccurate verdicts, with judges making errors that are biased towards unfounded conviction. Similarly, a few papers ([Rachlinski et al., 2007](#); [Guthrie et al., 2007](#)) suggest that judges have certain biases that influence their decisions. Other papers ([Hastie et al., 1999](#); [Inbar et al., 2012](#)) suggest that jurors have biases that influence their decision making. [Boudreau and McCubbins \(2008, 2009\)](#) show that competition between experts does not improve subjects' decisions (as jurors) due to their lack of sophistication. Our paper contributes to this literature by focusing on the characteristics that influence jurors' interpretation of reasonable doubt. We find differences between black and white jurors; we additionally find differences between jurors with and without a driver's license.

2 Background on Reasonable Doubt

The “proof beyond a reasonable doubt” standard has been in effect in criminal trials in the United States since at least 1798. This standard had been used by most U.S. jurisdictions for almost two centuries, until it was finally recognized as a constitutional requirement in 1970 ([Diamond, 1990](#)). In *In re Winship*, the Supreme Court held that the due process clauses of the Fifth and Fourteenth Amendments protect an individual accused of a crime from conviction “except upon proof beyond a reasonable doubt of every fact necessary to constitute the crime with which he is charged.”

While *Winship* held that the proof beyond a reasonable doubt standard is constitutionally required in criminal cases, the Supreme Court has never established a constitutionally-required meaning of “reasonable doubt.” Nor has the Court settled on any particular definition of reasonable doubt. To the contrary, in *Victor v. Nebraska*,⁶ the Court reviewed two appeals claiming that state court instructions had improperly defined “reasonable doubt” in ways allowing conviction on proof less than the *Winship* standard. The Court concluded that both contested instructions correctly conveyed the concept of reasonable doubt when taken as a whole, despite the notable flaws in each state court definition.⁷

Winship's requirement of a reasonable doubt standard, reinforced by *Victor*'s approval of varying types of instructions, has led to differences in how states approach defining reasonable doubt to juries. Some states (such as Texas, Illinois, Oklahoma, and Kentucky) either discourage or even expressly prohibit trial judges from defining reasonable doubt for

⁶511 U.S. 1 (1994)

⁷*Victor*, 511 U.S. at 22.

jurors. Some states (such as Indiana, Maryland, and New Jersey) require that reasonable doubt definitions be provided to jurors, while other states (such as Arizona, Connecticut, and Nebraska) allow but do not require that reasonable doubt definitions be provided to jurors.

In addition to differences in whether states define reasonable doubt, there is also variation in the reasonable doubt definitions that states use. In developing our experiment, we compiled the pattern jury instructions and/or case law on reasonable doubt for 48 states, the District of Columbia, and nine federal circuits. There were several common themes across all of these jurisdictions, but there was no single approach that most courts use to explain the meaning of reasonable doubt.

In designing our study, we reviewed our extensive list of pattern jury instructions on reasonable doubt to identify which ones were likely to be favorable to the prosecution or favorable to the defendant. We concluded that Wisconsin's reasonable doubt definition seemed on its face to push jurors toward convicting the defendant.⁸ The Wisconsin pattern jury instruction that we use reads as follows:

The term "reasonable doubt" means a doubt based upon reason and common sense. It is a doubt for which a reason can be given, arising from a fair and rational consideration of the evidence or lack of evidence.

It means such a doubt as would cause a person of ordinary prudence to pause or hesitate when called upon to act in the most important affairs of life.

A reasonable doubt is not a doubt which is based on mere guesswork or speculation. A doubt which arises merely from sympathy or from fear to return a verdict of guilt is not a reasonable doubt. A reasonable doubt is not a doubt such as may be used to escape the responsibility of a decision.

While it is your duty to give the defendant the benefit of every reasonable doubt, you are not to search for doubt. You are to search for the truth.

The first two paragraphs of this definition explain the meaning of reasonable doubt in fairly common ways - a tautological definition of reasonable doubt as a doubt based on reason or for which a reason can be given, followed by a comparison to a person of ordinary prudence. The third paragraph primes jurors away from thinking their doubts are reasonable and instead toward convicting the defendant. The fourth paragraph tells jurors to "search for truth" instead of searching for doubt.

In contrast to the Wisconsin instruction, we view West Virginia's reasonable doubt definition as pushing jurors toward acquittal.⁹ This perception that arises from the language of the instruction is unsurprising, because attorneys in West Virginia's Public Defender Services authored the pattern jury instructions. West Virginia's pattern jury instruction reads as follows:

⁸This definition can be found at <https://wilawlibrary.gov/jury/files/criminal/0140.pdf>. Note that we break up the first paragraph of the Wisconsin definition into two paragraphs.

⁹We obtained this definition from the seventh edition of the West Virginia Criminal Law Instructions Manual.

The State is not required to prove guilt beyond all possible doubt. The test is one of reasonable doubt. A reasonable doubt is one based upon reason and common sense—the kind of doubt that would make a reasonable person hesitate to act. Proof beyond a reasonable doubt, therefore, must be proof of such a convincing character that a reasonable person would not hesitate to rely and act upon it.

The jury will remember that a defendant is never to be convicted on mere suspicion or conjecture.

The burden is always upon the State to prove guilt beyond a reasonable doubt. This burden never shifts to a defendant, for the law never imposes upon a defendant in a criminal case the burden or duty of calling any witnesses or producing any evidence.

So, if you, after careful and impartial consideration of all the evidence in the case, have a reasonable doubt that the defendant is guilty of the charge, you must acquit. If you view the evidence in the case as reasonably permitting either of two conclusions—one of innocence, the other of guilt—you should adopt the conclusion of innocence.

Like Wisconsin, the West Virginia definition begins by explaining reasonable doubt as a doubt based on reason, and it invokes the “reasonable person.” Following this, however, West Virginia’s instruction reminds jurors not to convict on mere suspicion or conjecture, it emphasizes that the burden of proof is always on the prosecution and never the defendant, and it explicitly tells jurors they must acquit the defendant if they have a reasonable doubt or if they think the evidence could reasonably allow two different interpretations. This language nudges jurors toward a finding of not guilty.

The Wisconsin and West Virginia definitions of reasonable doubt differ from each other because they try to avoid different types of error.¹⁰ Each instruction provides a similar initial definition of reasonable doubt in the opening paragraph. But then the Wisconsin definition shifts, explaining that “a doubt which arises merely from sympathy or from fear to return a verdict of guilt is not a reasonable doubt,” and that a “reasonable doubt is not a doubt such as may be used to escape the responsibility of a decision.” This language is focused on avoiding type II error (acquitting a guilty person). The West Virginia definition shifts in the other direction: “The jury will remember that a defendant is never to be convicted on mere suspicion or conjecture. . . . If you view the evidence in the case as reasonably permitting either of two conclusions—one of innocence, the other of guilt—you should adopt the conclusion of innocence.” This language is focused on avoiding type I error (convicting an innocent person).

¹⁰We are unaware of conviction rate data that can be used to compare trial outcomes between these two states. In the case that we could obtain conviction rate data, it would be difficult to measure a direct impact of reasonable doubt definitions given the many possible differences between the two states that may affect conviction rates.

3 Model

Formally, we define the reasonable doubt threshold as in [Feddersen and Pesendorfer \(1998\)](#). We assume jurors care about the outcomes of the trials. They prefer the guilty being convicted and the innocent being freed. No juror prefers the innocent being convicted and the guilty being freed. Since the evidence of a crime is not a perfect signal of guilt, the juror faces a decision under uncertainty. To simplify our analysis, we normalize the latent utility of convicting a guilty defendant and freeing an innocent defendant to zero as in [Feddersen and Pesendorfer \(1998\)](#). We let the cost of convicting an innocent defendant to be $q \in (0, 1)$ and acquitting a guilty defendant to be $1 - q$. We investigate to what extent the perceived costs can be influenced by the jury instructions. We let the jury instruction be $t \in T$ where T is a set of possible instructions. The jurors' decisions become a cost minimization problem where they convict if and only if the cost of conviction is lower than the acquittal in expectation:

$$q(t)(1 - \mathbb{P}(\text{Guilty}|\text{Charged})) < (1 - q(t))\mathbb{P}(\text{Guilty}|\text{Charged})$$

which implies that we only convict if and only if $\mathbb{P}(\text{Guilty}|\text{Charged}) > q(t)$. The term $q(t)$ can now be interpreted as the reasonable doubt threshold for the juror. If we let

$$q_i = \alpha + \sum_{\tau \in T} 1(t_i = \tau) \beta_\tau + g(z_i) + \epsilon_i.$$

where z_i is a vector of characteristics, we can recover the influence of instructions from our experiment (β_τ).

4 Experimental Design

The experiment is designed with two goals in mind. The first goal is to uncover whether reasonable doubt definitions influence juror decisions. The second goal is to uncover the characteristics that are correlated with subjects' interpretation of reasonable doubt.

4.1 Sessions and Treatments

The experiment consists of two types of sessions. The first type are the theft game sessions, which provide defendants for jurors to make decisions about. We did not vary any aspect of these sessions. These sessions were designed to ensure that juror decisions would actually be used to determine the payoffs of subjects in the role of defendants, so that juror decisions were salient. The second type are the juror decision sessions, where subjects make decisions about whether to convict a defendant based on the strength of the evidence against the defendant. There are several different treatment variations of juror sessions. Since juror decisions are actually used to determine whether "defendants" from the theft game sessions are convicted (and subsequently punished), juror decisions

0	0	0	0	0
0	0	0	0	1
0	0	1	0	0
0	1	0	0	0
1	1	1	1	0

Figure 1: An example of the matrix task that subjects completed.

are not hypothetical. This is a key asset to our experimental design, as most previous work in this area involves the evaluation of hypothetical vignettes.¹¹

We run three initial treatments of juror decision sessions. The first treatment is the No Definition treatment, where we do not give subjects a definition of reasonable doubt (comparable to states like Texas, Illinois, and Oklahoma that discourage or prohibit definitions of reasonable doubt by trial judges). The second treatment is the Wisconsin treatment, where we give subjects the Wisconsin definition of reasonable doubt. The third treatment is the West Virginia treatment, where we give subjects the West Virginia definition of reasonable doubt.

4.2 Theft Game Sessions

Theft game sessions consist of two parts. In the first part, subjects earn \$50.00 by completing three matrix tasks. Figure 1 contains a screenshot of this task. In each matrix task, a subject must count the number of “1s” in a 5×5 grid that is randomly populated with 1s and 0s. Subjects are given an unlimited number of tries in each matrix task. We allow subjects to have unlimited tries because we want them to all earn \$50.00, while putting actual effort into the task. Since subjects must complete the tasks in order to earn the \$50.00, perceived property rights are likely to be greater than if we simply allocated money to them at the beginning of the session.

After completing the matrix tasks, subjects move to the second part of the session, in which subjects are randomly grouped into pairs.¹² One subject in the pair could take \$30.00 from the other subject in the pair.¹³ The other subject in the pair could do nothing. We refer to the subject who could take money as the “chosen subject”. While the chosen subject had the ability to take money, they could also be charged with theft (the not chosen subject could not be charged). The chosen subject could be charged with theft regardless of whether or not she actually took money from the other subject.¹⁴ We

¹¹An exception is [Aimone et al. \(2019\)](#), which experimentally examines juror decision making when jurors are able to donate their juror pay to charity.

¹²The pairing was done ex-post as we ran these experiments on Prolific ([prolific.co](#)).

¹³We use larger stakes than may be expected because we want the theft to be salient to jurors in our online environment.

¹⁴As we discuss in footnote 15, the charge probabilities were determined after the theft game sessions were run. The realized charge probabilities were 0.75 if the chosen subject took money and around 0.20

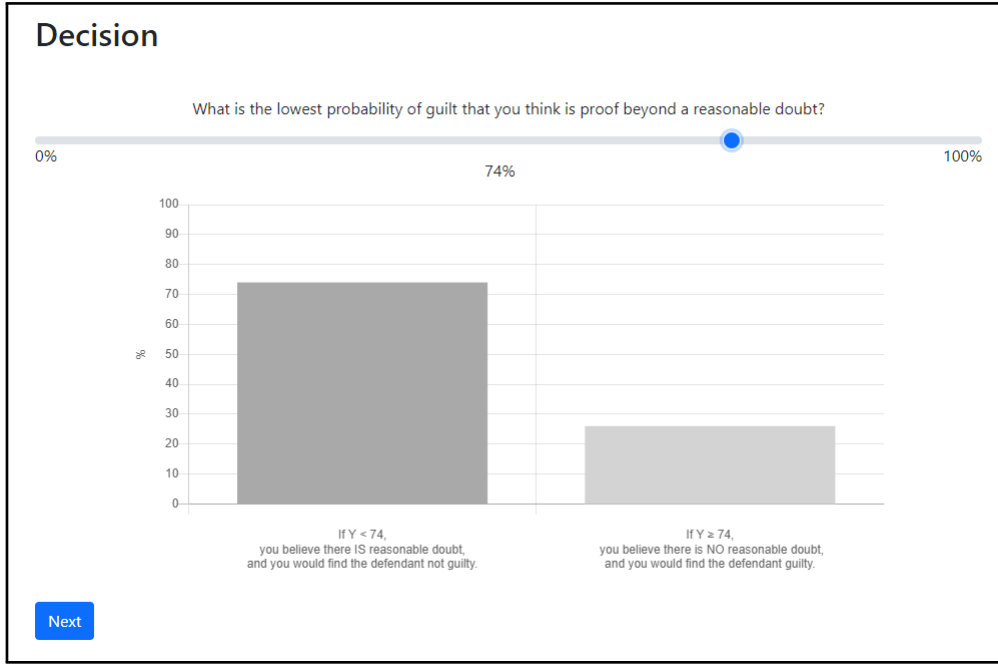


Figure 2: A screenshot showing how subjects revealed their guilt threshold in the juror decision sessions.

used the strategy method in which each subject indicated whether or not they would like to take money if they were designated as the chosen subject.

After the experiment ended, a chosen subject who was charged could be punished if she was found guilty of theft. If this subject was punished, she lost \$40.00. This punishment decision was determined in the juror decision sessions. Note that this means that subject payments for the theft game sessions was delayed until after the juror decision sessions.

4.3 Juror Decision Sessions

In the juror decision sessions, each subject independently makes a single decision that is used to determine whether or not to convict and punish a defendant from the theft game sessions. In the instructions, subjects are told how the theft game sessions are conducted. Subjects are then told that there is a defendant that has a probability Y of being guilty. They are also told that this probability Y is randomly chosen from $\{0\%, 1\%, \dots, 99\%, 100\%\}$. Importantly, the random choice of Y does in fact correspond with the strength of the evidence against the defendant.¹⁵

Subjects were asked to state their guilt threshold, which is denoted as GT and is the lowest value of Y that the subject believes is proof beyond a reasonable doubt. Subjects reported their guilt threshold using a slider. The initial placement of this slider

if the chosen subject did not take money.

¹⁵After randomly determining Y , we choose $\mathbb{P}(\text{Charge}|\text{Guilty})$ and $\mathbb{P}(\text{Charge}|\text{Innocent})$ such that $\mathbb{P}(\text{Guilty}|\text{Charge}) = Y$ according to Bayes rule. Essentially, $\mathbb{P}(\text{Charge}|\text{Guilty})$ and $\mathbb{P}(\text{Charge}|\text{Innocent})$ are chosen such that $Y = \frac{\mathbb{P}(\text{Guilty})\mathbb{P}(\text{Charge}|\text{Guilty})}{\mathbb{P}(\text{Guilty})\mathbb{P}(\text{Charge}|\text{Guilty}) + \mathbb{P}(\text{Innocent})\mathbb{P}(\text{Charge}|\text{Innocent})}$, where Y is the probability an individual is guilty conditional on being charged, $\mathbb{P}(\text{Guilty})$ is the probability a randomly chosen individual from the theft game session took money, and $P(\text{Innocent})$ is the complement of $P(\text{Guilty})$.

was randomly determined in order to avoid systematically biasing decisions. To ensure understanding, the screen also displayed a graph which indicated the implications of the current slider position.¹⁶ Figure 2 displays a screenshot of the interface that a subject uses to reveal her guilt threshold. Each defendant will be either found guilty (and punished) or acquitted based on Y and a randomly selected guilt threshold from the subjects in the juror decision sessions. If $Y < GT$, then that defendant is acquitted and not punished. If $Y \geq GT$, then that defendant is found guilty and is punished.

4.4 Questionnaires and Demographics

At the end of each session, subjects complete five different questionnaires. The first questionnaire regards subjects' ability to serve as a juror and their history as a juror. The second questionnaire elicits a subject's big five characteristics. The third questionnaire is the cognitive reflection task. The fourth questionnaire contains questions regarding subjects' political affiliation. The fifth questionnaire elicits general demographics such as marital status, race, and religion.

4.5 Procedures

Our experimental design, analysis plan, and conjectures about results were pre-registered prior to data collection.¹⁷ The experiments were run on Prolific (prolific.co) during June and July of 2021. We recruited United States citizens who were 18 years of age and older. Our goal was to recruit subjects who could theoretically serve on a jury in the United States. There were 293 subjects who participated in the experiment. There were 99 subjects in the Wisconsin treatment, 96 subjects in the No Definition treatment, and 98 subjects in the West Virginia treatment. The experiment lasted about eleven minutes on average. Each subject earned \$3.50 for completing the experiment. Subjects in the theft game sessions earned more money (\$49.50 on average with completion fee); the amount of money they made depends on their actions and the actions of other subjects.

4.6 Conjectures

In the results section, we test two conjectures based on the reasonable doubt definitions employed in our experimental design. These conjectures were pre-registered.¹⁸

Our first conjecture regards the Wisconsin Definition treatment and the No Definition treatment. We expect differences between these two treatments since the Wisconsin definition seems to nudge subjects towards focusing on reducing type 2 error (i.e., reducing false acquittals). We thus expect more convictions in the Wisconsin Definition treatment. This leads us to Conjecture 1.

¹⁶This graph was empty until the subject first moved the slider.

¹⁷This can be found at <https://osf.io/2vrfp/>.

¹⁸These conjectures can be found at <https://osf.io/2vrfp/>.

	<u>Wisconsin Definition</u>	<u>No Definition</u>	<u>West Virginia Definition</u>
Male:	47.96%	53.68%	48.45%
White:	77.55%	80.00%	76.29%
Black:	10.20%	7.37%	9.28%
Registered to Vote:	90.82%	87.37%	91.75%
Previous Juror Experience:	13.27%	11.58%	13.40%
Lost Juror Eligibility:	1.02%	0.00%	2.06%
Driver’s License:	88.78%	86.32%	91.75%
Number of Subjects	98	95	97

Table 1: Demographics in each treatment. Registered to vote, lost juror eligibility, and driver’s license variables all have “Prefer not to answer” options. Gender variable has “Male”, “Female”, “Other”, and “Prefer not to answer” options. Race variable has “White or Caucasian”, “Native American”, “Multiracial”, “Black/African American”, “Asian”, “Other”, and “Prefer not to answer” options. Previous juror experience variable receives a value of “1” if subjects answered “yes” to a question about previous criminal juror experience or “yes” to a question about previous civil juror experience. Any subject who did not answer all demographic questions is ignored.

Conjecture 1: *There are more convictions (lower guilt thresholds) under the Wisconsin definition of reasonable doubt than when no definition is provided.*

Our second conjecture regards the West Virginia Definition treatment and the No Definition treatment. We expect differences between these two treatments since the West Virginia definition appears to nudge subjects towards focusing on reducing type 1 error (i.e., reducing false convictions). We thus expect fewer convictions in the West Virginia Definition treatment. This leads us to Conjecture 2.

Conjecture 2: *There are more convictions (lower guilt thresholds) when no definition of reasonable doubt is provided than under the West Virginia definition of reasonable doubt.*

5 Results

In this section, we go over the results of the experiment. The first subsection goes over the treatment comparisons. The second subsection goes over the characteristics that are correlated with subjects’ interpretation of reasonable doubt.

We first display the demographics for each treatment to uncover whether there are any treatment differences that may be driven by our randomization of subjects into treatments. Table 1 displays demographic characteristics in each treatment. We display characteristics that are relevant for our regressions in later subsections of the results section. Table 1 shows that the demographics are similar in each treatment.

	Wisconsin Definition	No Definition	West Virginia Definition
Mean Guilt Threshold:	71.19	73.6	71.00
Std. Dev Guilt Threshold:	21.53	21.28	23.15
Observations:	99	96	98
Mann-Whitney P-value:	0.334	—	0.510

Table 2: Displays the basic statistics for the guilt thresholds in each treatment. “Mann-Whitney P-value” refers to the p-value in each treatment when comparing that treatment’s guilt thresholds to the guilt thresholds in the No Definition treatment.

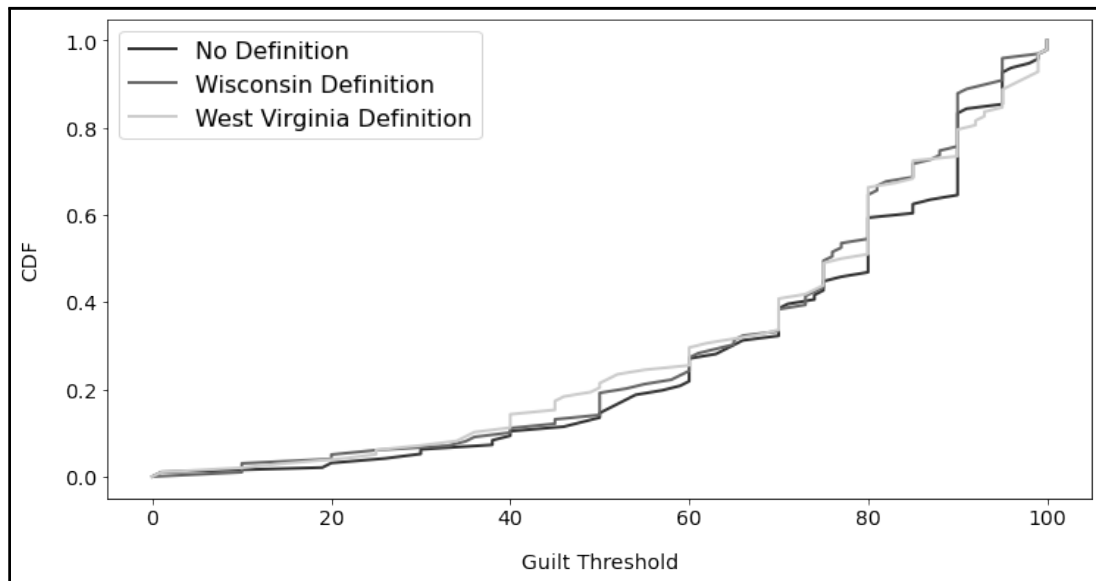


Figure 3: CDFs of the guilt thresholds in each treatment.

5.1 Treatment Comparisons

Table 2 displays the general results from the experiment.¹⁹ Table 2 shows similar behavior between the three treatments. The mean guilt threshold in each treatment is between 71 and 74. We display the distribution of guilt thresholds in the three treatments in Figure 3, which also suggests that behavior is similar between the three treatments.

We use Mann-Whitney tests to test our conjectures, with each subject in the juror sessions taken to be an independent observation.²⁰ We fail to find a significant difference between the Wisconsin treatment and the No Definition treatment at the ten percent level (p-value=0.334). We also fail to find a significant difference between the No Definition treatment and the West Virginia treatment at the ten percent level (p-value=0.510). Additionally, we test whether the West Virginia and Wisconsin treatments differ from each other. We fail to find a significant difference at the ten percent level (p-value=0.858).

¹⁹There are three subjects who reported a guilt threshold, but did not complete all of the demographics questions. They are included everywhere in this section except for the demographics table and the regressions. Our Mann-Whitney tests are robust to dropping these subjects.

²⁰We do not correct for multiple hypothesis testing in this section or the next section as correcting for multiple hypothesis testing would only re-inforce our results.

	(1)	(2)	(3)
	<u>Guilt Thresh.</u>	<u>Guilt Thresh.</u>	<u>Guilt Thresh.</u>
Wisconsin Definition	-2.94 (3.10)	-3.49 (3.19)	-3.14 (3.16)
West Virginia Definition	-1.72 (3.09)	-3.33 (3.20)	-2.29 (3.14)
Male	2.42 (2.59)	— —	2.52 (2.64)
Black	-15.23*** (4.63)	— —	-15.52*** (4.84)
Registered to Vote	— —	1.29 (4.88)	2.80 (5.07)
Previous Juror Experience	— —	-2.97 (3.90)	-1.54 (3.88)
Previous Lost Juror Eligibility	— —	7.99 (12.89)	10.48 (12.55)
Has Driver’s License	— —	8.19* (4.70)	5.96 (5.06)
Constant	74.39*** (4.02)	66.35*** (5.69)	66.87*** (6.83)
<i>N</i>	290	291	290

Table 3: Regressions of different covariates on subject’s characteristics. Registered to vote, lost juror eligibility, and driver’s license questions all have “Prefer not to answer” options. Gender question has “Male”, “Female”, “Other”, and “Prefer not to answer” options. Race question has “White or Caucasian”, “Native American”, “Multiracial”, “Black/African American”, “Asian”, “Other”, and “Prefer not to answer” options. Previous juror experience variable receives a value of “1” if subjects answered “yes” to a question about previous criminal juror experience or a question about previous civil juror experience. Covariates for religion and political affiliation are included in the first and third regressions, but are not reported in this table. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

These results are supported through regression analysis.

5.2 Subjects' Interpretation of Reasonable Doubt

In this subsection, we analyze the relationship between subjects' characteristics and subjects' guilt thresholds.²¹ Specifically, we focus on two types of characteristics. The first type is demographics. The second type is characteristics that are related to juror eligibility. Table 3 reports the regression results.

We report the results of three specifications in Table 3. In the first specification, we regress the guilt threshold on treatment, gender, race, religion, and political affiliation covariates. Importantly, neither Wisconsin nor West Virginia's treatment is statistically significant in influencing the observed guilt thresholds. The estimated coefficients are small relative to the standard deviation of the guilt thresholds reported in Table 2. The only statistically significant result that we find is that black subjects have lower guilt thresholds than white subjects.²² This result is significant at the one percent level. In the second specification, we regress the guilt threshold on the treatment, and several variables that reflect juror eligibility and the ways jurors are recruited for jury duty. These variables denote whether or not a subject is a registered voter, has previous juror experience, has had their right to be a juror previously taken away, and whether they have a driver's license.²³ The only statistically significant result that we find is that subjects who have a driver's license have a higher guilt threshold at the ten percent level. The third specification includes all of the previously mentioned dummy variables. In this specification, the only statistically significant result that we again observe is that black subjects have lower guilt thresholds than white subjects even after controlling for additional variables. This result is significant at the one percent level.

6 Additional Treatments

The previous results showed that subjects behave similarly in our three treatments. In this section, we explore two possible explanations for why we observe similar behavior in our three treatments. We design three new pre-registered treatments that test these

²¹Note the analysis in this section was pre-registered as "exploratory analysis" without specific conjectures about how any individual characteristic might be related to reasonable doubt thresholds.

²²Our data do not allow us to explore why black subjects have lower guilt thresholds than white subjects because subjects are not told the race of the potential defendant. An intriguing possibility is raised by [Depew et al. \(2017\)](#), who studied juvenile court cases in Louisiana. They found that judges were more likely to give harsher treatment to defendants of the same race than to those of other races. One possible reason for the observed negative in-group bias could have been that in-group punishment is more severe when a group member violates the group's social norms, especially when those harmed are more likely also to be members of the in-group. Thus, black jurors might be more willing to convict black defendants, especially because crimes are disproportionately committed against same-race victims. We thank an anonymous referee for referring us to this paper.

²³A common way for courts to construct lists of potential jurors is to use lists of registered voters supplemented with other sources of names. Driver's license records are a common way to add potential jurors to the pool when they are not registered to vote. E.g., Cal. Civ. Pro. Code § 197; N.Y. Jud. Law § 506; Tex. Gov't Code § 62.001. See also 28 U.S.C. § 1863.

possible explanations.

The first possible explanation that we explore is that the Wisconsin and West Virginia definitions are written using hedging language. We interpret the first two paragraphs of the Wisconsin definition as trying to prevent type 1 error (convicting an innocent person). We interpret the last two paragraphs as trying to prevent type 2 error (acquitting a guilty person). We conjecture that the combination of these four paragraphs is decreasing variation in guilt thresholds. The first additional treatment, the Wisconsin Truncated treatment, looks to uncover if this is happening. In this treatment, we remove the first two paragraphs of the Wisconsin definition to uncover whether hedging language is leading to the similar behavior in the Wisconsin and No Definition treatments.

We next look at the West Virginia definition. We interpret the first sentence of the West Virginia definition as trying to prevent type 2 error (acquitting a guilty person). We interpret the last three paragraphs of the West Virginia definition as trying to prevent type 1 error (convicting an innocent person). We conjecture that the combination of these four paragraphs is decreasing variation in guilt thresholds. In the second additional treatment, the West Virginia Truncated treatment, we remove the first paragraph of the West Virginia definition to uncover whether hedging language is leading to the similar behavior in the West Virginia and No Definition treatments.

The second possible explanation that we explore is that subjects ignore the reasonable doubt standard. It is possible that subjects are just choosing guilt thresholds that they think are appropriate regardless of the standard presented. We test this explanation by designing an additional treatment, the No Standard treatment, that does not mention reasonable doubt.²⁴ In this treatment, we tell subjects to convict if they think there is enough evidence to warrant conviction. We use this treatment to uncover whether the standard put before subjects influences behavior.

We first start analyzing the additional treatments by looking at the demographics in each treatment.²⁵ Table 4 displays the demographics for the additional treatments. There appear to be some differences in the demographics of these additional treatments. The No Standard treatment has more male subjects and more black subjects than the other two additional treatments (and the original three treatments). We will control for these differences when comparing the No Standard treatment to the other treatments (through regression analysis).

Figure 4 displays comparisons between the original treatments and the additional treatments. As Figure 4 shows, there is little difference between the Wisconsin and Wisconsin Truncated treatments. This is supported by a Mann-Whitney test (p -value=0.722).

²⁴The instructions for this treatment are in Appendix C. The instructions are similar to the No Definition treatment except that we now define the guilt threshold as being “the lowest probability of guilt (the lowest value of Y) that you think is enough to convict” instead of “the lowest probability of guilt (the lowest value of Y) that you think is proof beyond a reasonable doubt”.

²⁵There are three subjects from the additional treatments who reported a guilt threshold, but did not complete all of the demographics questions. These three subjects, as well as the similar three subjects from the first three treatments, are included everywhere in this section except for the demographics table and the regressions. Our Mann-Whitney tests are robust to dropping these subjects.

	No Standard	Wisconsin Truncated	West Virginia Truncated
Male:	65.31%	57.29%	60.23%
White:	67.35%	82.29%	81.82%
Black:	19.39%	10.42%	7.95%
Registered to Vote:	94.90%	91.67%	92.05%
Previous Juror Experience:	19.39%	14.58%	12.50%
Lost Juror Eligibility:	0.00%	1.04%	1.14%
Driver’s License:	95.92%	93.75%	94.32%
Number of Subjects	98	96	88

Table 4: Demographics in each additional treatment. Registered to vote, lost juror eligibility, and driver’s license variables all have “Prefer not to answer” options. Gender variable has “Male”, “Female”, “Other”, and “Prefer not to answer” options. Race variable has “White or Caucasian”, “Native American”, “Multiracial”, “Black/African American”, “Asian”, “Other”, and “Prefer not to answer” options. Previous juror experience variable receives a value of “1” if subjects answered “yes” to a question about previous criminal juror experience or “yes” to a question about previous civil juror experience. Any subject who did not answer all demographic questions is ignored.

Figure 4 also shows little difference between the West Virginia and West Virginia Truncated treatments. This is supported by a Mann-Whitney test (p-value=0.601).

Figure 4 does suggest that there is a difference between the No Definition and No Standard treatments. This is backed up by a Mann-Whitney test (p-value=0.036). To see whether these results are robust to the inclusion of demographics, we run regressions (see Table 5) that are similar to the regressions in Table 3. It is possible that the difference between the No Definition and No Standard treatments is driven by the randomization into treatments. The first specification in Table 5, which includes basic demographics, shows that there is not a significant difference between the No Definition and No Standard treatments after accounting for basic demographics. The second specification in Table 5, which includes characteristics related to juror eligibility, suggests that there is a significant difference between the No Definition and No Standard treatments. Lastly, the third specification in Table 5, which includes all of the characteristics of the first two specifications, shows that there is not a significant difference between the No Definition and No Standard treatments.

The regressions in Table 5 suggest that the difference between the No Definition and No Standard treatments is driven by the randomization into treatments. One possibility for why this arises is the number of black and white subjects in the two treatments. Only 7.37% of subjects in the No Definition treatment are black, while 19.39% of subjects in the No Standard treatment are black. This may be leading to the difference in the two treatments as black subjects have lower guilt thresholds than white subjects. We run a simple regression that compares the No Definition and No Standard treatments to test this possibility. We only additionally include race dummy variables in this regression. The coefficient on the No Standard treatment is insignificant at the ten percent level (p-value=0.223).

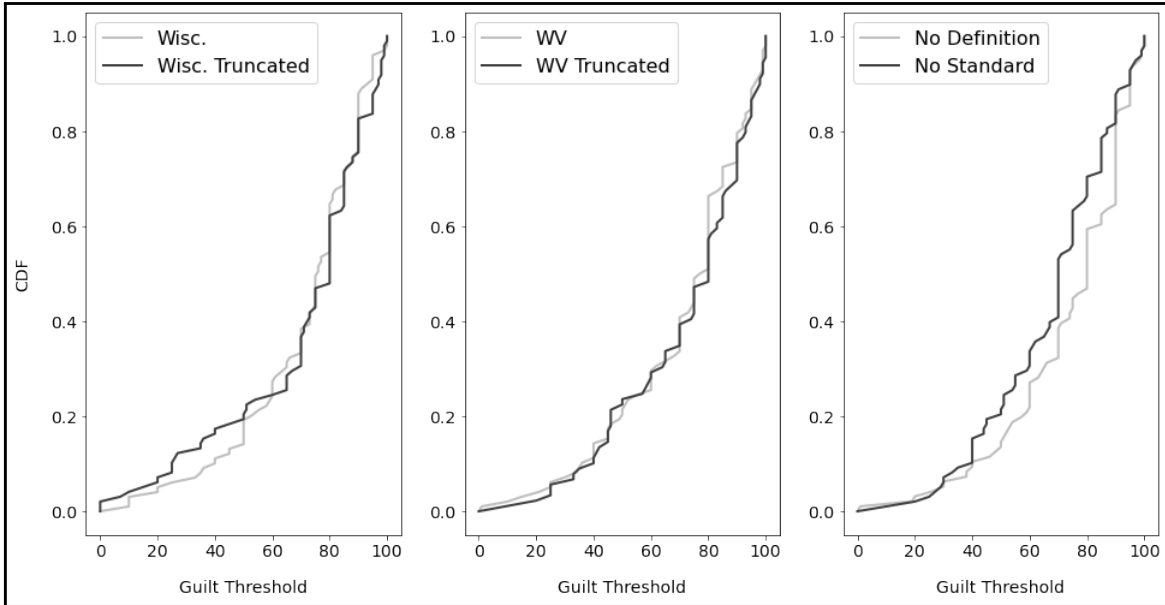


Figure 4: CDFs comparing the original treatments to the additional treatments. A Mann-Whitney test comparing guilt thresholds between the Wisconsin and Wisconsin Truncated treatments results in a p-value of 0.722. A Mann-Whitney test comparing guilt thresholds between the West Virginia and West Virginia Truncated treatments results in a p-value of 0.601. A Mann-Whitney test comparing guilt thresholds between the No Definition and No Standard treatments results in a p-value of 0.036.

We can also use Table 5 to explore the characteristics correlated with subjects’ interpretation of reasonable doubt. These results generally build on the results of Table 3. The coefficient for black subjects, when included, is negative and significant at the one percent level. This bolsters the previous result that black subjects have lower guilt thresholds than white subjects. The coefficient on having a driver’s license, when included, is positive and significant at the one percent level. This suggests that subjects with a driver’s license have higher guilt threshold than subjects that do not. This coefficient was significant in the regression that only considered juror eligibility demographics in Table 3.

7 Discussion

In this study, we analyze whether reasonable doubt definitions influence juror decisions. In our three original treatments, we did not find differences between providing no definition of reasonable doubt, the Wisconsin definition of reasonable doubt, or the West Virginia definition of reasonable doubt. In response to the similarity of behavior across our three original treatments, we ran three additional treatments. These three additional treatments truncated the Wisconsin definition, truncated the West Virginia definition, and removed mention of the reasonable doubt standard from the instructions. Behavior in these additional treatments was similar to the original three treatments. In this section, we discuss the implications of the results from these six treatments.

We interpret our results as suggesting that subjects behave similarly in a criminal trial regardless of the definition of reasonable doubt provided to them in the juror instructions.

	(1)	(2)	(3)
	Guilt Thresh.	Guilt Thresh.	Guilt Thresh.
Wisconsin Full	-2.49 (3.17)	-3.54 (3.27)	-3.05 (3.18)
West Virginia Full	-1.57 (3.17)	-3.52 (3.28)	-2.45 (3.18)
Wisconsin Truncated	-2.56 (3.19)	-4.59 (3.28)	-3.20 (3.19)
West Virginia Truncated	-0.42 (3.26)	-2.71 (3.36)	-0.98 (3.26)
No Standard	-3.02 (3.22)	-6.99** (3.29)	-3.95 (3.23)
Male	2.43 (1.91)	— —	2.15 (1.91)
Black	-16.05*** (3.06)	— —	-15.71*** (3.11)
Registered to Vote	— —	-0.25 (3.70)	1.13 (3.67)
Previous Juror Experience	— —	-0.53 (2.74)	0.77 (2.66)
Previous Lost Juror Eligibility	— —	4.17 (10.26)	4.87 (9.88)
Has Driver's License	— —	9.99*** (3.81)	11.26*** (3.81)
Constant	76.64*** (3.40)	65.88*** (4.71)	66.26*** (5.27)
<i>N</i>	572	576	572

Table 5: Regressions of different covariates on subject behavior. Registered to vote, lost juror eligibility, and driver's license questions all have "Prefer not to answer" options. Gender question has "Male", "Female", "Other", and "Prefer not to answer" options. Race question has "White or Caucasian", "Native American", "Multiracial", "Black/African American", "Asian", "Other", and "Prefer not to answer" options. Previous juror experience variable receives a value of "1" if subjects answered "yes" to a question about previous criminal juror experience or a question about previous civil juror experience. Covariates for religion and political affiliation are included in the first and third regressions, but are not reported in this table. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Our additional treatments suggest that subjects' behavior is also not driven by the hedging language in the Wisconsin and West Virginia definitions. These additional treatments also suggest that subjects behave similarly regardless of any mention of reasonable doubt. There are two potential explanations for these results that we consider. The first possibility is that subjects have a pre-conceived notion of how to behave in criminal trials (i.e., any language used to describe reasonable doubt does not affect their belief of the burden of proof required). The second possibility is that subjects ignore the burden of proof requirement altogether and simply make the decision that they personally prefer.

We consider the possibility that subjects have a pre-conceived notion of reasonable doubt to be more likely than subjects ignoring the burden of proof. This view is shaped by recent research (such as [Glöckner and Engel \(2013\)](#)) that suggests that jurors respond to the standard of preponderance of evidence differently than the standard of proof beyond a reasonable doubt. Thus, we consider it likely that subjects have a pre-conceived notion of how to behave in criminal trials that is based on their own experience with the notion of reasonable doubt. It is quite possible that subjects have pre-conceived notions due to the prevalence of reasonable doubt in popular culture. For example, numerous movies and television shows focus on the standard of proof in criminal trials.

As mentioned earlier, we interpret our results as suggesting that subjects behave similarly in a criminal trial regardless of the reasonable doubt definition put forward. One question is whether subjects' behavior would differ under different definitions if these definitions were shown side-by-side. While this question has no bearing on criminal trials (where jurors see only one definition), it is a stronger test of whether subjects' behavior depends on reasonable doubt definitions.

We run an additional treatment, consisting of 97 subjects, to test whether subject behavior differs when reasonable doubt definitions are shown side-by-side.²⁶ In this treatment, we show each subject both the truncated Wisconsin and truncated West Virginia definitions.²⁷ Each subject sees these definitions side-by-side and then chooses a guilt threshold under each definition. We find an average guilt threshold difference of 1.90 percentage points between the two definitions. Similar to our main treatments, this difference is statistically insignificant (p-value=0.112 under a Wilcoxon signed-rank test; p-value \sim 0.370 under different regressions). This result reinforces our previous finding that subject behavior is similar under different reasonable doubt definitions. We present our full regression analysis, which is similar to our earlier treatments, in Appendix B. These regressions also continue to show a significantly lower guilt threshold for black subjects. Notice that these regression results show no significant driver's license correlation. This suggests that the overall driver's license relationship may not be as strong as suggested in our previous analysis (Table 5).

Overall, we interpret the results of our paper as suggesting that the definitions put forward in criminal trials have little effect on jurors. This suggests that there is little variability in trial outcomes that stem directly from reasonable doubt definitions. This

²⁶We thank an anonymous referee for this suggested treatment.

²⁷Instructions can be found in Appendix C.4.

is encouraging for the criminal justice system as it suggests that there is one less channel with which variability in trial outcomes may occur.

The lack of many significant coefficients on the control variables related to juror eligibility and recruitment provides (encouraging) evidence that, all else equal, such recruitment methods may not be biasing baseline guilt thresholds of recruited jurors. However, the significant correlations identified between having a driver's license and guilt threshold (in our main treatments) is potentially concerning, because driver's license lists are often used to expand jury pools beyond the set of registered voters.

Similarly, the significant correlation between race and guilt thresholds we identified is potentially concerning as well. As ensuring diversity and representation on juries is increasingly made focal in the literature and in practice, it is important to recognize and understand how such changes can impact perceptions of the reasonable doubt standard.

8 Conclusion

In this paper, we use a series of salient and financially consequential pre-registered experiments to explore whether reasonable doubt definitions influence juror decisions. We compare juror decisions when they are provided different definitions of reasonable doubt: no definition of reasonable doubt, the Wisconsin definition of reasonable doubt, and the West Virginia definition of reasonable doubt. We find that these variations of reasonable doubt have little effect on juror decisions. In response, we ran additional treatments that show that the hedging language contained in these definitions is unlikely to be driving our results. These additional treatments also suggest that subjects behave similarly in criminal trials regardless of whether reasonable doubt is mentioned. Our results also suggest that different groups (such as black and white subjects) have different interpretations of reasonable doubt.

Our results have consequences outside the laboratory. First, our results suggest that reasonable doubt definitions may have little effect on juror decisions. Thus, it appears that variation in reasonable doubt definitions are not leading to variation in trial outcomes. Second, we find differences between black and white subjects' jury decisions. This suggests that racial composition of juries matter regardless of evidence interpretation and salient characteristics of the defendant. Similarly, we find that whether jurors are recruited using driver's license records may be unintentionally related to reasonable doubt guilt thresholds of those in juror pools. Lastly, our results suggest that jurors have pre-conceived notions of how to behave in criminal trials.

There are many avenues for future research. First, other studies can explore whether other parts of the criminal justice system influence juror decisions. It is possible that jurors make different decisions depending on the crime, expected sentence, and defendant demographics. Second, other studies can analyze the effect of reasonable doubt definitions on juries, judges, or lawyers. It is possible that there may be an effect when jurors deliberate using these definitions. Third, other studies can explore whether preponderance of evidence definitions influence juror decisions. It is possible that preponderance of

evidence definitions may matter due to civil trials appearing less often in popular culture. Fourth, future research may want to explore the influence of the knowledge of the race of the defendant on reasonable doubt thresholds. Lastly, other studies can explore the effects of other parts of juror instructions on juror decisions. It is possible that other parts of juror instructions influence juror decisions.

References

- Aimone, J., North, C., and Rentschler, L. (2019). Priming the jury by asking for donations: An empirical and experimental study. *Journal of Economic Behavior & Organization*, 160:158–167.
- Boudreau, C. and McCubbins, M. (2008). Nothing but the truth? experiments on adversarial competition, expert testimony, and decision making. *Journal of Empirical Legal Studies*, 5:751–789.
- Boudreau, C. and McCubbins, M. (2009). Competition in the courtroom: When does expert testimony improve jurors’ decisions? *Journal of Empirical Legal Studies*, 6:793–817.
- Burke, M. S., Carter, J., Gominiak, R., and Ohl, D. (1996). An experimental note on the allais paradox and monetary incentives. *Empirical Economics*, 21:617–632.
- Cicchini, M. and White, L. (2016). Truth or doubt? an empirical test of criminal jury instructions. *University of Richmond Law Review*, 50(4):1139–67.
- Cicchini, M. and White, L. (2017). Testing the impact of criminal jury instructions on verdicts: A conceptual replication. *Columbia Law Review Online*, 117:22–35.
- Clot, S., Grolleau, G., and Ibanez, L. (2018). Shall we pay all? an experimental test of random incentivized systems. *Journal of Behavioral and Experimental Economics*, 73:93–98.
- Depew, B., Eren, O., and Mocan, N. (2017). Judges, juveniles, and in-group bias. *Journal of Law and Economics*, 60:209–237.
- Dhmi, M. (2008). On measuring quantitative interpretations of reasonable doubt. *Journal of Experimental Psychology: Applied*, 14:353–363.
- Dhmi, M., Lundrigan, S., and Muller-Johnson, K. (2015). Instructions on reasonable doubt: Defining the standard of proof and the juror’s task. *Psychology, Public Policy, and Law*, 21:169–178.
- Diamond, H. (1990). Note: Reasonable doubt: To define, or not to define. *Columbia Law Review*, 90(6):1716–1736.
- Feddersen, T. and Pesendorfer, W. (1998). Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting. *American Political Science Review*, 92:23–35.
- Glöckner, A. and Engel, C. (2013). Can we trust intuitive jurors? standards of proof and the probative value of evidence in coherence-based reasoning. *Journal of Empirical Legal Studies*, 10:230–252.

- Guthrie, C., Rachlinski, J., and Wistrich, A. (2007). Blinking on the bench: How judges decide cases. *Cornell Law Review*, 93:1–43.
- Harrison, G. W. (1994). Expected utility theory and the experimentalists. *Experimental Economics*, pages 43–73.
- Hastie, R., Schkade, D., and Payne, J. (1999). Juror judgments in civil cases: Effects of plaintiff’s requests and plaintiff’s identity on punitive damage awards. *Law and Human Behavior*, 23:445–470.
- Hudja, S., Ralston, J., Wang, S., Aimone, J., Rentschler, L., and North, C. (2021). The effect of gender on tolerance of type 1 and type 2 error in judicial decisions. *Unpublished Manuscript*.
- Inbar, Y., Pizarro, D., and Cushman, F. (2012). When harmless actions are judged to be morally blameworthy. *Personality and Social Psychology Bulletin*, 38:52–62.
- Kagehiro, D. (1990). Defining the standard of proof in jury instructions. *Psychological Science*, 1:194–200.
- Kaye, D. (1986). Do we need a calculus of weight to understand proof beyond a reasonable doubt? *Boston University Law Review*, 66:657–672.
- Rachlinski, J., Guthrie, C., and Wistrich, A. (2007). Heuristics and biases in bankruptcy judges. *Journal of Institutional and Theoretical Economics*, 163:167–186.
- Ralston, J., Aimone, J., North, C., and Rentschler, L. (2021). False confessions: An experimental laboratory study of the innocence problem. *Working Paper*.
- Sonnemans, J. and van Dijk, F. (2012). Errors in judicial decisions: Experimental results. *Journal of Law, Economics, & Organization*, 28:687–716.
- Tillers, P. (1993). Intellectual history, probability, and the law of evidence. *Michigan Law Review*, 91:1465–1490.
- Voslinsky, A. and Azar, O. (2021). Incentives in experimental economics. *Journal of Behavioral and Experimental Economics*, 93:101706.
- White, L. and Cicchini, M. (2019). Is reasonable doubt self-defining? *Villanova Law Review*, 64(1):1–24.

A Theft Game Appendix

	<u>Number of Subjects</u>	<u>Chose to Take Money</u>	<u>Punished</u>	<u>Average Payment</u>
Overall:	20	11	2	49.50
Chosen Subjects:	10	6	2	63.50

Table 6: Displays the basic statistics for the theft game (used for main treatments). “Overall” refers to all of the subjects who participated in the theft game. “Chosen subjects” refer to subjects who could be charged. “Chose to Take Money” refers to the number of subjects in the subgroup who decided to take money using the strategy method. “Punished” refers to the number of subjects who were punished/convicted. “Average Payment” refers to the average payment including the completion fee.

	<u>Number of Subjects</u>	<u>Chose to Take Money</u>	<u>Punished</u>	<u>Average Payment</u>
Overall:	10	6	0	53.50
Chosen Subjects:	5	3	0	71.50

Table 7: Displays the basic statistics for the theft game (used for side-by-side treatment). “Overall” refers to all of the subjects who participated in the theft game. “Chosen subjects” refer to subjects who could be charged. “Chose to Take Money” refers to the number of subjects in the subgroup who decided to take money using the strategy method. “Punished” refers to the number of subjects who were punished/convicted. “Average Payment” refers to the average payment including the completion fee.

Table 6 provides basic statistics for the theft game (used for the original treatments). It shows how many of the 20 overall subjects chose to take money and how many Chosen Subjects chose to take money. Additionally, it displays how often subjects were punished and the average payment (including show-up fee). Because we used a strategy method to gather participants’ decisions, we have responses from all 20 subjects. Table 6 reports data for all subjects as well as those who ex post were Chosen Subjects. Table 7 provides basic statistics for the theft game used for the side-by-side treatment.

B Side-by-Side Treatment Appendix

This appendix displays the results of the treatment mentioned in the discussion section. In this treatment, subjects observe both the truncated Wisconsin and truncated West Virginia definitions side-by-side and then choose a guilt threshold under each definition. This appendix focuses on random effects regressions with subject-level random effects and subject-level clustering. However, our results are similar without random effects and/or without clustering.

	(1)	(2)	(3)
	<u>Guilt Thresh.</u>	<u>Guilt Thresh.</u>	<u>Guilt Thresh.</u>
West Virginia Truncated	1.90 (2.19)	1.90 (2.14)	1.90 (2.23)
Male	12.47*** (4.27)	—	11.61** (4.59)
Black	-21.84** (9.44)	—	-21.67** (9.54)
Registered to Vote	—	5.36 (10.66)	-2.18 (9.87)
Previous Juror Experience	—	2.68 (5.26)	-0.97 (4.91)
Previous Lost Juror Eligibility	—	11.37*** (2.64)	10.66 (7.25)
Has Driver's License	—	-5.02 (8.24)	0.46 (8.06)
Constant	55.57*** (7.71)	67.34*** (13.34)	57.21*** (14.98)
<i>N</i>	194	194	194

Table 8: Random effects regressions of different covariates on subject behavior. Regressions have subject-level random effects and subject-level clustering. Registered to vote, lost juror eligibility, and driver's license questions all have "Prefer not to answer" options. Gender question has "Male", "Female", "Other", and "Prefer not to answer" options. Race question has "White or Caucasian", "Native American", "Multiracial", "Black/African American", "Asian", "Other", and "Prefer not to answer" options. Previous juror experience variable receives a value of "1" if subjects answered "yes" to a question about previous criminal juror experience or a question about previous civil juror experience. Covariates for religion and political affiliation are included in the first and third regressions, but are not reported in this table.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C Instructions

In this section, we display the instructions for the theft game sessions and the juror decision sessions. In the first subsection, we display the instructions for the theft game sessions. In the second subsection, we display the instructions for the Wisconsin definition treatment.

C.1 Theft Game Session Instructions

Introduction

Welcome and thank you for participating. This experiment is a study of economic decision-making. The amount of money you earn depends partly on the decisions you make and thus you should read these instructions carefully. The show-up fee (\$3.50) will be paid privately through Prolific after the experiment ends. You will receive any bonus payment that you make from the experiment in 2-3 weeks through Prolific.

Please turn off your cell phones now and close any program (other than the experiment) that you may have open on the computer. Please do not talk or in any way try to communicate with other people during the experiment.

Task 1

This experiment consists of two parts. In the first part, you will solve three matrices to earn a credit of \$50 for use in the second part of the experiment. To solve a matrix, you count the number of ones in the matrix. For example, in the matrix below, there are 7 ones.

0	0	0	0	0
0	0	0	0	1
0	0	1	0	0
0	1	0	0	0
1	1	1	1	0

You have unlimited tries to solve these three matrices, but you must solve them correctly to earn the \$50 credit and move on to the next part of the experiment.

Task 2

Congratulations. You have earned a \$50 credit to use during the second part of the experiment. You will be paired with another subject who has also earned a \$50 credit in

the same manner.

One of you will have access to the other person's \$50 credit and will be able to take \$30 of it. For example, if you are the chosen group member, you will be able to take \$30 from your partner's \$50. If you are not the chosen group member, your partner will be able to take \$30 from your \$50.

For the remainder of the instructions, imagine that you are the chosen group member. As the chosen group member, you may be accused of theft whether or not you have taken \$30. The computer will randomly select a percentage of people who actually took money to be accused of theft. The computer will also randomly select a percentage of people who did not take money to be accused of theft.

In a later experiment, participants in the role of jurors will determine whether each individual who has been accused of theft gets a "guilty" verdict. In this future experiment, jurors will know that some people who did not take money may be accused anyway. Jurors will also know that some people who did take money may not be accused. Any individuals that receive a "guilty" verdict will have \$40 subtracted from their earnings as a fine. Individuals who are not accused or found "not guilty" will not be fined.

If you are the chosen group member, do you want to take \$30 from your partner?

- Yes
- No

Your total earnings from the Matrix Task and the Group Task will make up your "bonus payment" for the experiment.

C.2 Wisconsin Treatment Instructions

Introduction

Welcome and thank you for participating. This experiment is a study of economic decision-making. You will earn \$3.50 for participating in today's experiment. The money that you earn will be paid privately to you through Prolific after the experiment ends.

Please turn off your cell phones now and close any program (other than the experiment) that you may have open on the computer. Please do not talk or in any way try to communicate with other people during the experiment.

Overview and Quiz

You will make a decision that may determine whether or not to convict someone based on the probability that they committed theft in another experiment. In a typical criminal trial, you are instructed to convict only if the defendant is proven guilty beyond a reasonable doubt. In this experiment, we will ask you what probabilities of guilt provide proof beyond a reasonable doubt.

Information about reasonable doubt:

The term “reasonable doubt” means a doubt based upon reason and common sense. It is a doubt for which a reason can be given, arising from a fair and rational consideration of the evidence or lack of evidence.

It means such a doubt as would cause a person of ordinary prudence to pause or hesitate when called upon to act in the most important affairs of life.

A reasonable doubt is not a doubt which is based on mere guesswork or speculation. A doubt which arises merely from sympathy or from fear to return a verdict of guilt is not a reasonable doubt. A reasonable doubt is not a doubt such as may be used to escape the responsibility of a decision.

While it is your duty to give the defendant the benefit of every reasonable doubt, you are not to search for doubt. You are to search for the truth.

In a previous experiment, on Prolific, real-human subjects were placed into two-person groups where each group member had already earned \$50 of “bonus pay”. One group member had the opportunity to take \$30 from their partner. If that subject took \$30, they are said to be guilty of theft. If they did not take \$30, they are said to be innocent. The subject who could take from the other group member was told that they may be charged with theft regardless of whether or not they took money.

The computer will determine the probability that an innocent person is charged and the probability that a guilty person is charged in such a way to ensure that the probability of a charged person being guilty is a certain set percentage that we will call “Y”. For example, a “Y” of 90 means that there is a 90% chance that a person charged is guilty of taking \$30 from their partner and a 10% chance that the person is innocent (that is, they did not take \$30 from their partner). As another example, a “Y” of 5 means that there is a 5% chance that a person charged is guilty of taking \$30 from their partner and a 95% chance that the person is innocent (that is, they did not take \$30 from their partner).

We will ask you for your guilt threshold, which is the lowest probability of guilt (the lowest value of Y) that you think is proof beyond a reasonable doubt.

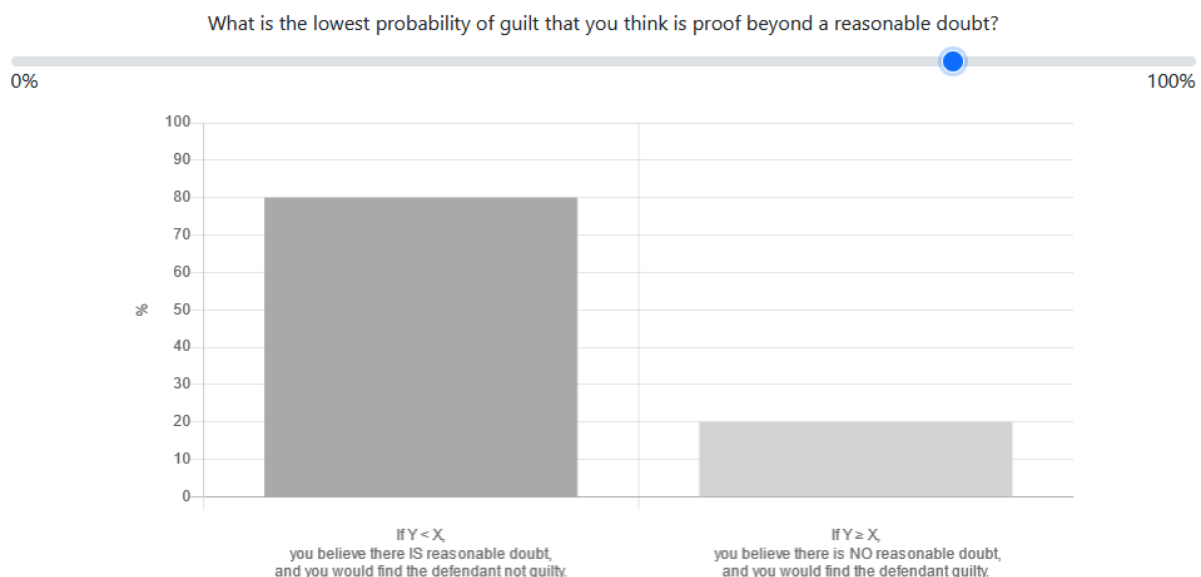
Your decision will be entered into a pool of juror decisions. For each person charged with a crime, one decision will be randomly drawn from the pool. If the decision says that the evidence is proof beyond a reasonable doubt, the person charged will be considered to be found “guilty” and fined \$40 by having \$40 subtracted from their bonus pay. If the decision says that the evidence is not proof beyond a reasonable doubt, the person

charged will not be punished. In this way, you, and the other participants making the same types of decisions as you, are serving as a juror.

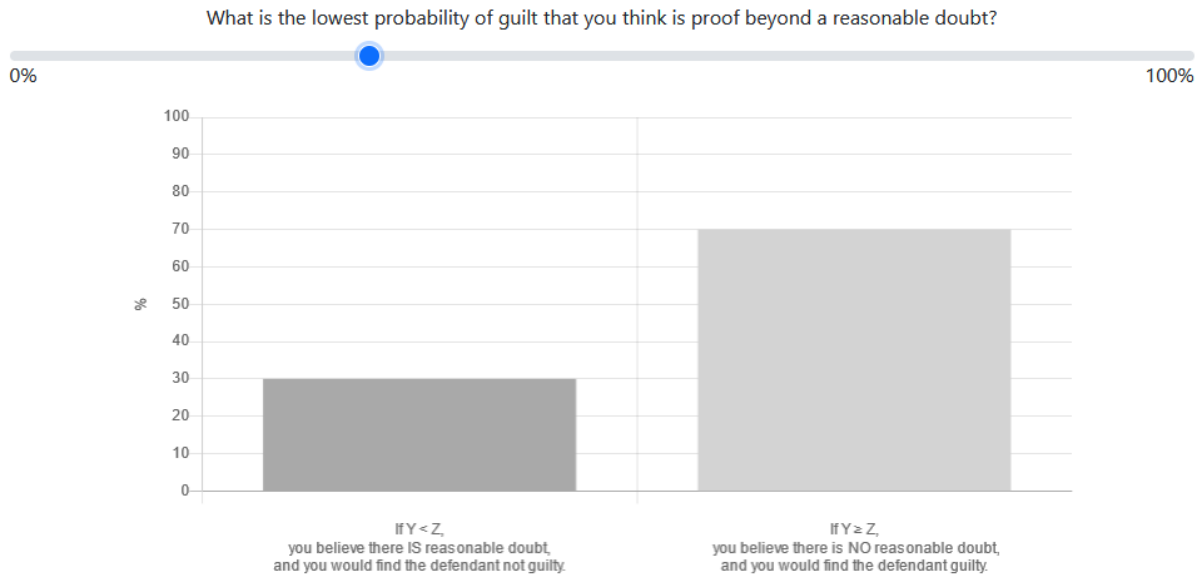
The value of “Y” that will be used in the experiment will be randomly determined from between 0% and 100%. Any probability in 0%, 1%, 2%, ..., 98%, 99%, and 100% is possible. Instead of asking you different questions for each probability, we will ask you to position the slider bar to your personal guilt threshold. That way, we will know whether you think a person is guilty beyond a reasonable doubt for every probability of guilt in the given range.

Overview and Quiz

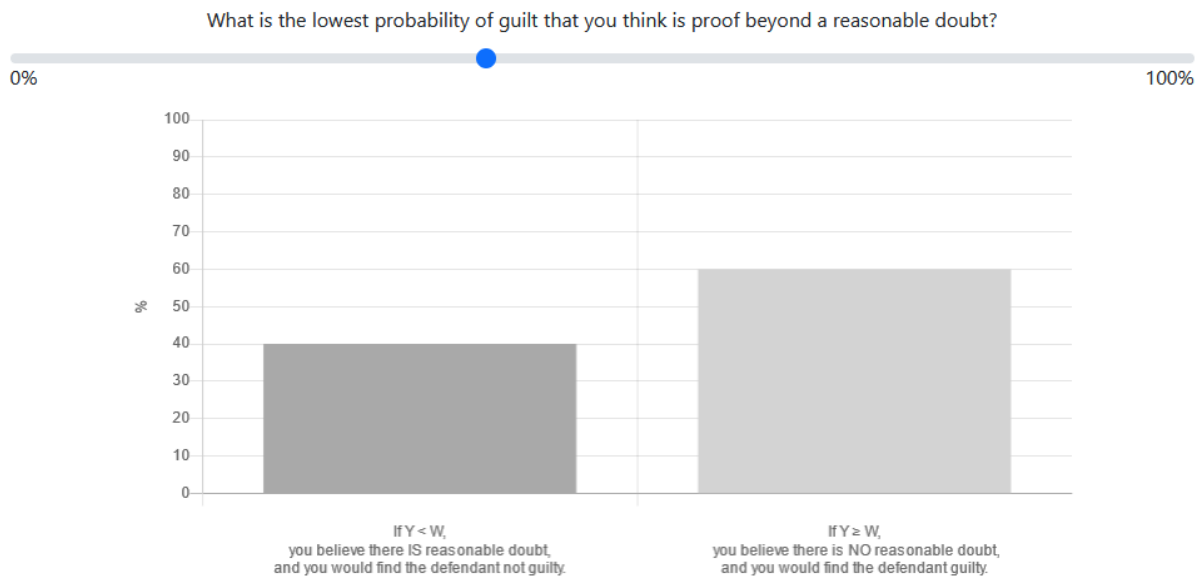
For example, if you place the slider at X%, as in the picture below, it indicates that you will convict the defendant if the probability of guilt is X% or greater. For any probability of guilt less than X%, you do not think they are guilty beyond a reasonable doubt.



For another example, if you place the slider at Z%, as in the picture below, that indicates that you think they are guilty beyond a reasonable doubt if there is a Z% probability or greater that a person charged is guilty. For any probability of guilt less than Z%, you do not think they are guilty beyond a reasonable doubt.



For another example, if you place the slider at $W\%$, as in the picture below, that indicates that you think they are guilty beyond a reasonable doubt if there is a $W\%$ probability or greater that a person charged is guilty. For any probability of guilt less than $W\%$, you do not think they are guilty beyond a reasonable doubt.



Please note the individuals in the previous experiment have been paid their Prolific experiment pay of \$3.50 already, but have not received their bonus pay yet. The decisions of jurors in this Prolific experiment will finalize those bonus pay amounts and only then will those participants be paid their bonus payments.

Suppose that you place the slider at 75%, and the computer chooses a Y of 60%. Would you find the defendant guilty or not guilty?

- Guilty

- Not Guilty

Suppose that you place the slider at 75%, and the computer chooses a Y of 80%. Would you find the defendant guilty or not guilty?

- Guilty
- Not Guilty

Decision

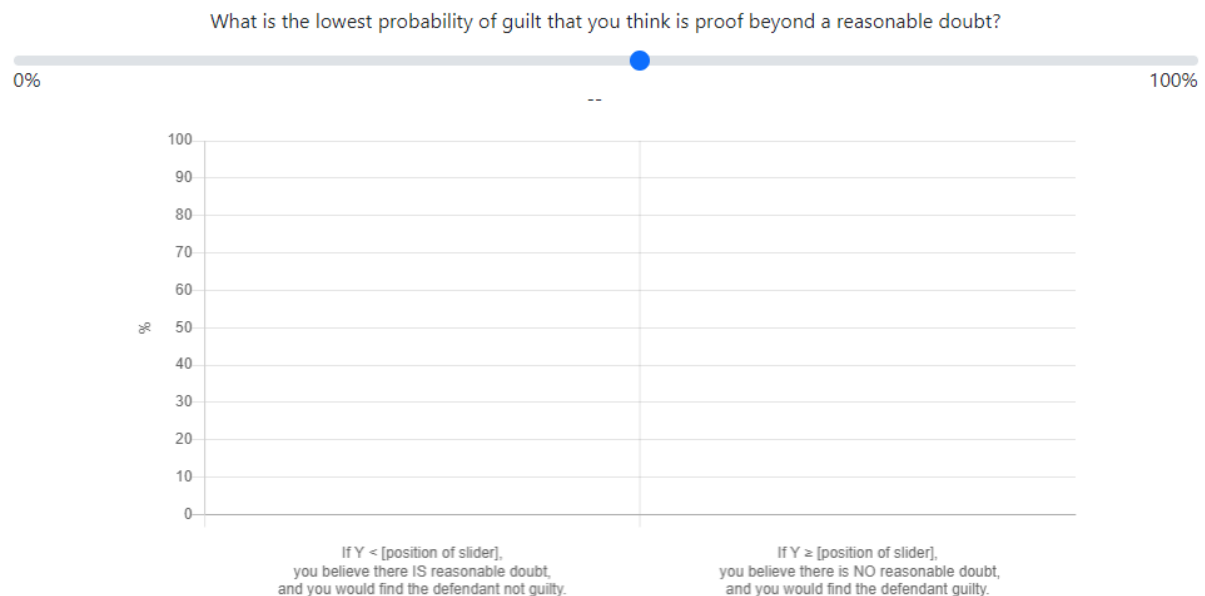
Information about reasonable doubt:

The term "reasonable doubt" means a doubt based upon reason and common sense. It is a doubt for which a reason can be given, arising from a fair and rational consideration of the evidence or lack of evidence.

It means such a doubt as would cause a person of ordinary prudence to pause or hesitate when called upon to act in the most important affairs of life.

A reasonable doubt is not a doubt which is based on mere guesswork or speculation. A doubt which arises merely from sympathy or from fear to return a verdict of guilt is not a reasonable doubt. A reasonable doubt is not a doubt such as may be used to escape the responsibility of a decision.

While it is your duty to give the defendant the benefit of every reasonable doubt, you are not to search for doubt. You are to search for the truth.



C.3 No Standard Treatment Instructions

Introduction

Welcome and thank you for participating. This experiment is a study of economic decision-making. You will earn \$3.50 for participating in today's experiment. The money that you earn will be paid privately to you through Prolific after the experiment ends.

Please turn off your cell phones now and close any program (other than the experiment) that you may have open on the computer. Please do not talk or in any way try to communicate with other people during the experiment.

Overview and Quiz

You will make a decision that may determine whether or not to convict someone based on the probability that they committed theft in another experiment.

In a previous experiment, on Prolific, real-human subjects were placed into two-person groups where each group member had already earned \$50 of "bonus pay". One group member had the opportunity to take \$30 from their partner. If that subject took \$30, they are said to be guilty of theft. If they did not take \$30, they are said to be innocent. The subject who could take from the other group member was told that they may be charged with theft regardless of whether or not they took money.

The computer will determine the probability that an innocent person is charged and the probability that a guilty person is charged in such a way to ensure that the probability of a charged person being guilty is a certain set percentage that we will call "Y". For example, a "Y" of 90 means that there is a 90% chance that a person charged is guilty of taking \$30 from their partner and a 10% chance that the person is innocent (that is, they did not take \$30 from their partner). As another example, a "Y" of 5 means that there is a 5% chance that a person charged is guilty of taking \$30 from their partner and a 95% chance that the person is innocent (that is, they did not take \$30 from their partner).

We will ask you for your guilt threshold, which is the lowest probability of guilt (the lowest value of Y) that you think is enough evidence to convict.

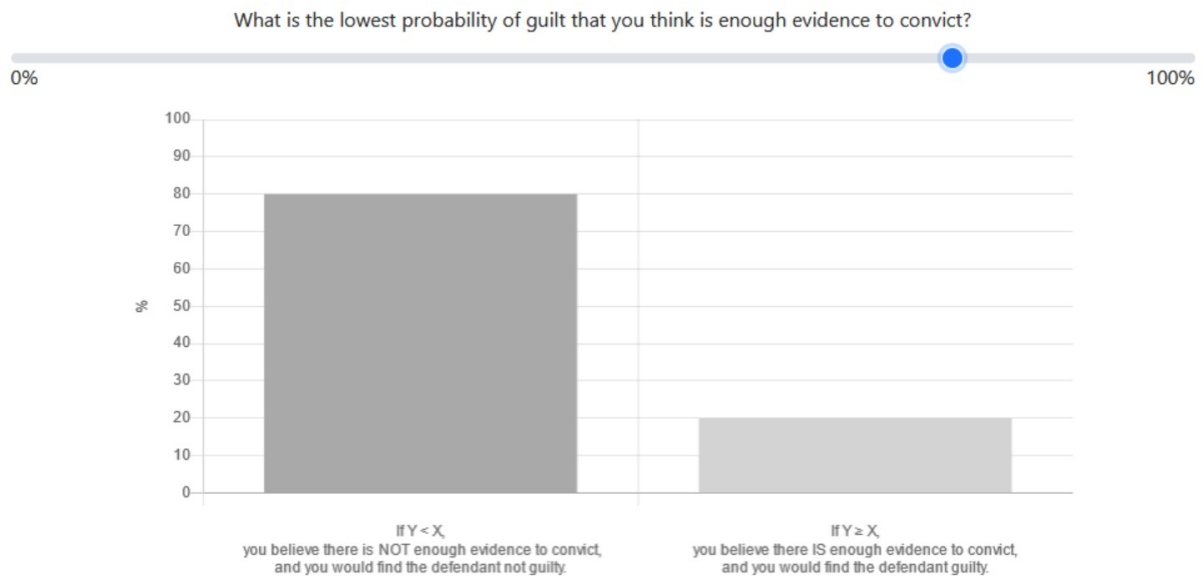
Your decision will be entered into a pool of juror decisions. For each person charged with a crime, one decision will be randomly drawn from the pool. If the decision says that the evidence is enough to convict, the person charged will be considered to be found "guilty" and fined \$40 by having \$40 subtracted from their bonus pay. If the decision says that the evidence is not enough to convict, the person charged will not be punished. In this way, you, and the other participants making the same types of decisions as you, are serving as a juror.

The value of "Y" that will be used in the experiment will be randomly determined from between 0% and 100%. Any probability in 0%, 1%, 2%, ..., 98%, 99%, and 100% is

possible. Instead of asking you different questions for each probability, we will ask you to position the slider bar to your personal guilt threshold. That way, we will know whether you think a person should be found guilty for every probability of guilt in the given range.

Overview and Quiz

For example, if you place the slider at $X\%$, as in the picture below, it indicates that you will convict the defendant if the probability of guilt is $X\%$ or greater. For any probability of guilt less than $X\%$, you do not think there is enough evidence to convict.



For another example, if you place the slider at $Z\%$, as in the picture below, that indicates that you think there is enough evidence to convict if there is a $Z\%$ probability or greater that a person charged is guilty. For any probability of guilt less than $Z\%$, you do not think there is enough evidence to convict.



For another example, if you place the slider at $W\%$, as in the picture below, that indicates that you think there is enough evidence to convict if there is a $W\%$ probability or greater that a person charged is guilty. For any probability of guilt less than $W\%$, you do not think there is enough evidence to convict.



Please note the individuals in the previous experiment have been paid their Prolific experiment pay of \$3.50 already, but have not received their bonus pay yet. The decisions of jurors in this Prolific experiment will finalize those bonus pay amounts and only then will those participants be paid their bonus payments.

Suppose that you place the slider at 75%, and the computer chooses a Y of 60%. Would you find the defendant guilty or not guilty?

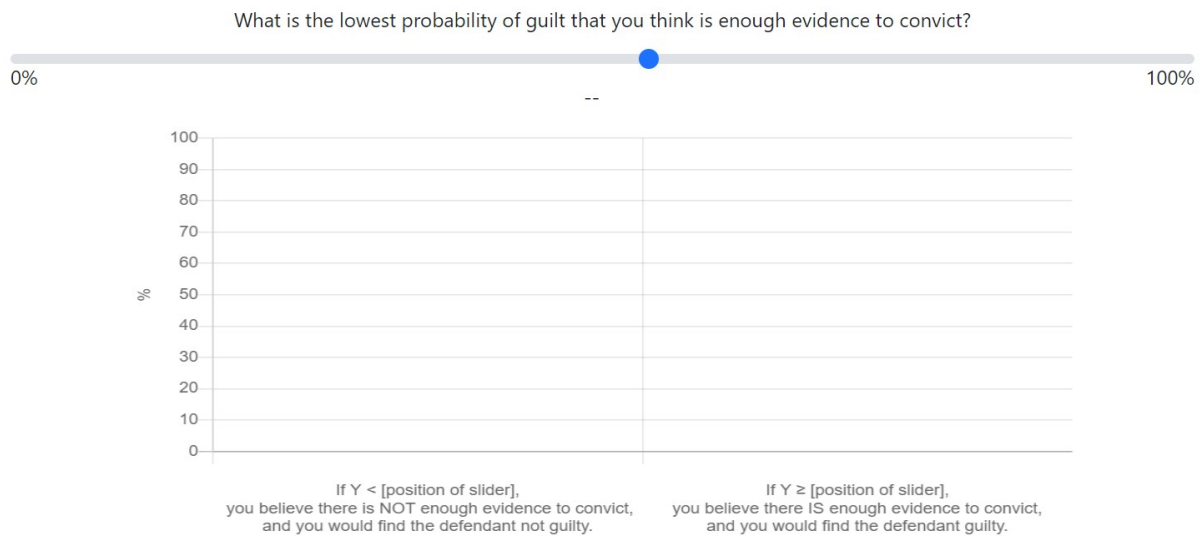
- Guilty

- Not Guilty

Suppose that you place the slider at 75%, and the computer chooses a Y of 80%. Would you find the defendant guilty or not guilty?

- Guilty
- Not Guilty

Decision



C.4 Side-by-Side Treatment Instructions

The instructions for the Two Definition Treatment start below. We randomize the order of the definitions. There is a 50% chance that the Wisconsin Truncated definition is definition #1 (West Virginia Truncated definition is definition #2) and a 50% chance that the Wisconsin Truncated definition is definition #2 (West Virginia Truncated definition is definition #1). After the definitions have been assigned numbers, there is a 50% chance that definition #1 is presented first on the decision page and a 50% chance that definition #2 is presented first on the decision page. We randomize to avoid any possible order effects of showing the definitions. We show one example of a set of what the instructions would look like. Starting points for decision sliders in each treatment are random.

Introduction

Welcome and thank you for participating. This experiment is a study of economic decision-making. You will earn \$3.50 for participating in today's experiment. The money

that you earn will be paid privately to you through Prolific after the experiment ends.

Please turn off your cell phones now and close any program (other than the experiment) that you may have open on the computer. Please do not talk or in any way try to communicate with other people during the experiment.

Overview and Quiz

You will make a decision that may determine whether or not to convict someone based on the probability that they committed theft in another experiment. In a typical criminal trial, you are instructed to convict only if the defendant is proven guilty beyond a reasonable doubt. In this experiment, we will provide you with two different definitions of reasonable doubt and ask you what probabilities of guilt provide proof beyond a reasonable doubt under each definition.

Information about reasonable doubt (definition 1):

A reasonable doubt is not a doubt which is based on mere guesswork or speculation. A doubt which arises merely from sympathy or from fear to return a verdict of guilt is not a reasonable doubt. A reasonable doubt is not a doubt such as may be used to escape the responsibility of a decision.

While it is your duty to give the defendant the benefit of every reasonable doubt, you are not to search for doubt. You are to search for the truth.

Information about reasonable doubt (definition 2):

The jury will remember that a defendant is never to be convicted on mere suspicion or conjecture.

The burden is always upon the State to prove guilt beyond a reasonable doubt. This burden never shifts to a defendant, for the law never imposes upon a defendant in a criminal case the burden or duty of calling any witnesses or producing any evidence.

So, if you, after careful and impartial consideration of all the evidence in the case, have a reasonable doubt that the defendant is guilty of the charge, you must acquit. If you view the evidence in the case as reasonably permitting either of two conclusions—one of innocence, the other of guilt—you should adopt the conclusion of innocence.

In a previous experiment, on Prolific, real-human subjects were placed into two-person groups where each group member had already earned \$50 of “bonus pay”. One group member had the opportunity to take \$30 from their partner. If that subject took \$30, they are said to be guilty of theft. If they did not take \$30, they are said to be innocent. The subject who could take from the other group member was told that they may be charged with theft regardless of whether or not they took money.

The computer will determine the probability that an innocent person is charged and the probability that a guilty person is charged in such a way to ensure that the probability

of a charged person being guilty is a certain set percentage that we will call “Y”. For example, a “Y” of 90 means that there is a 90% chance that a person charged is guilty of taking \$30 from their partner and a 10% chance that the person is innocent (that is, they did not take \$30 from their partner). As another example, a “Y” of 5 means that there is a 5% chance that a person charged is guilty of taking \$30 from their partner and a 95% chance that the person is innocent (that is, they did not take \$30 from their partner).

We will ask you for your guilt threshold, which is the lowest probability of guilt (the lowest value of Y) that you think is proof beyond a reasonable doubt, under each definition of reasonable doubt. We ask for your guilt threshold under each definition of reasonable doubt because there is a 50% chance that "definition 1" is appropriate for the previous experiment and a 50% chance that "definition 2" is appropriate for the previous experiment.

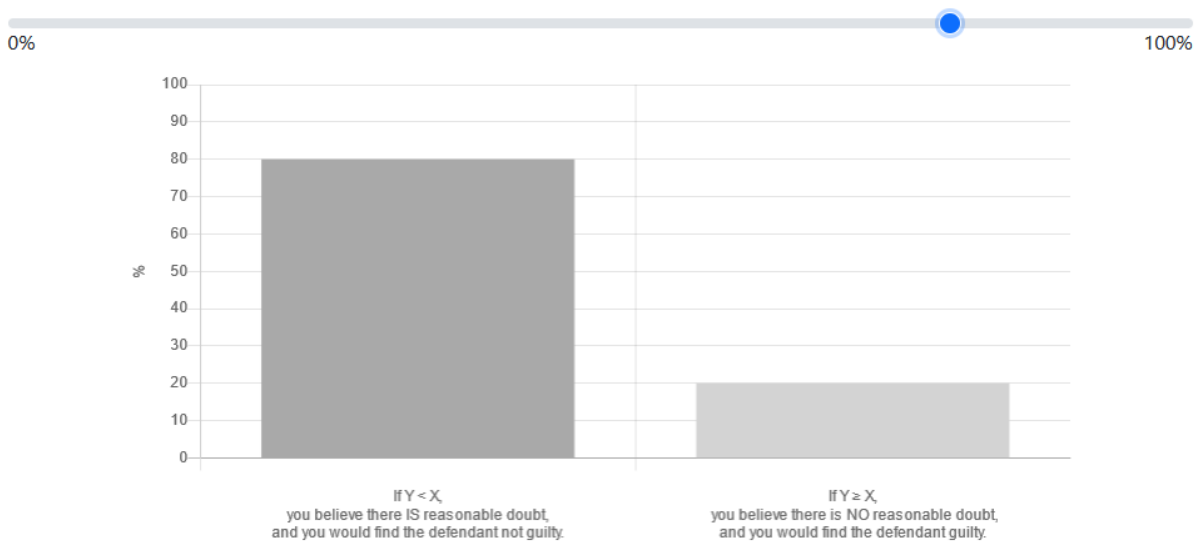
One of your two decisions will be entered into a pool of juror decisions. If "definition 1" is appropriate, your guilt threshold under "definition 1" will be entered. If "definition 2" is appropriate, your guilt threshold under "definition 2" will be entered. For each person charged with a crime, one decision will be randomly drawn from the pool. If the decision says that the evidence is proof beyond a reasonable doubt, the person charged will be considered to be found “guilty” and fined \$40 by having \$40 subtracted from their bonus pay. If the decision says that the evidence is not proof beyond a reasonable doubt, the person charged will not be punished. In this way, you, and the other participants making the same types of decisions as you, are serving as a juror.

The value of “Y” that will be used in the experiment will be randomly determined from between 0% and 100%. Any probability in 0%, 1%, 2%, ..., 98%, 99%, and 100% is possible. Instead of asking you different questions for each probability, we will ask you to position the slider bar to your personal guilt threshold. That way, we will know whether you think a person is guilty beyond a reasonable doubt for every probability of guilt in the given range.

Overview and Quiz

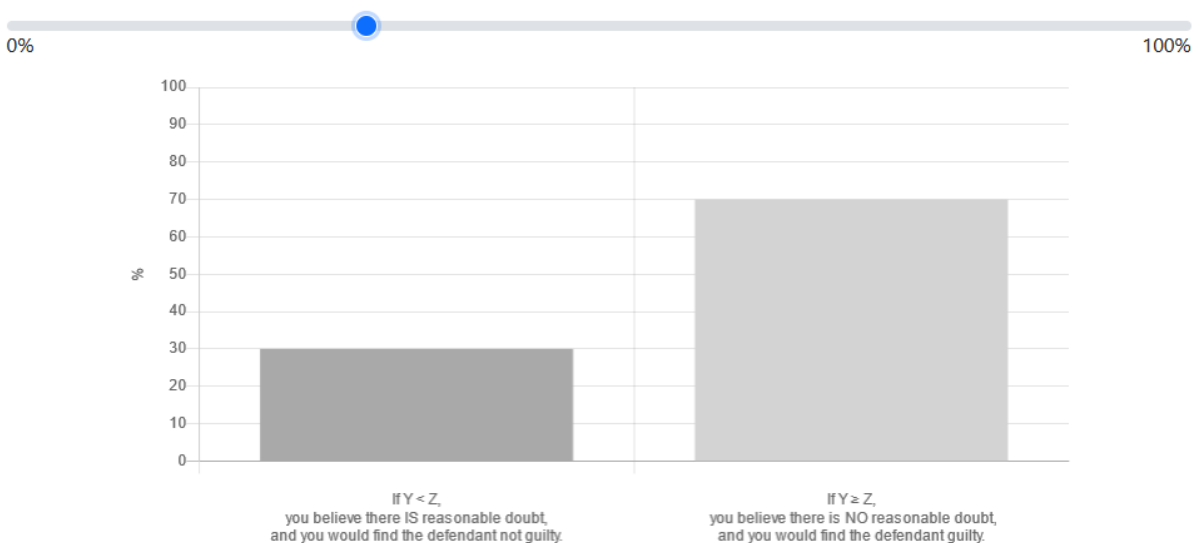
For example, if you place the slider at X%, as in the picture below, it indicates that you will convict the defendant if the probability of guilt is X% or greater. For any probability of guilt less than X%, you do not think they are guilty beyond a reasonable doubt.

What is the lowest probability of guilt that you think is proof beyond a reasonable doubt?



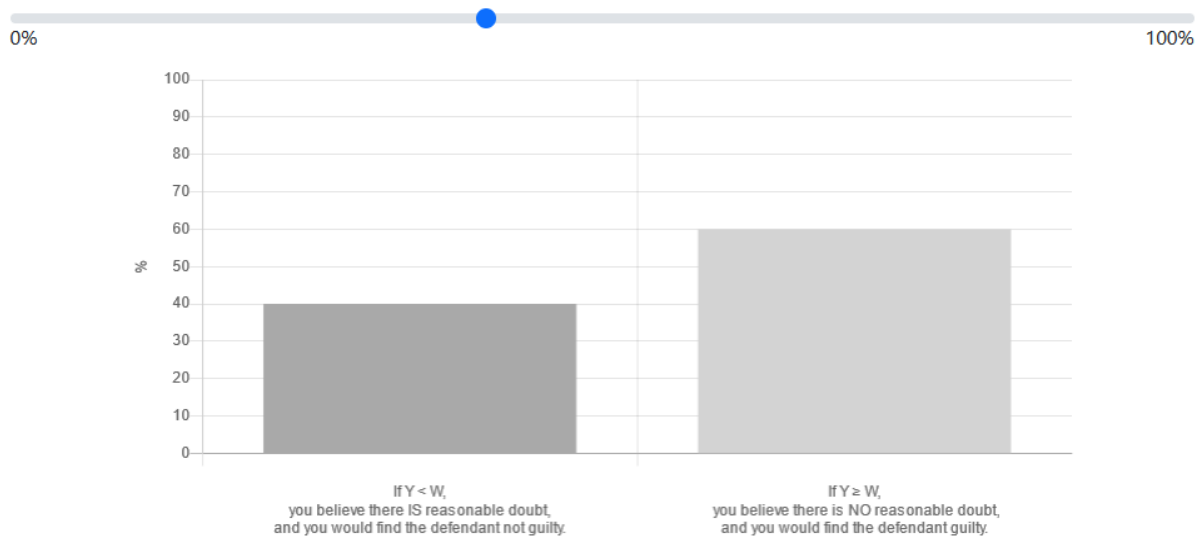
For another example, if you place the slider at Z%, as in the picture below, that indicates that you think they are guilty beyond a reasonable doubt if there is a Z% probability or greater that a person charged is guilty. For any probability of guilt less than Z%, you do not think they are guilty beyond a reasonable doubt.

What is the lowest probability of guilt that you think is proof beyond a reasonable doubt?



For another example, if you place the slider at W%, as in the picture below, that indicates that you think they are guilty beyond a reasonable doubt if there is a W% probability or greater that a person charged is guilty. For any probability of guilt less than W%, you do not think they are guilty beyond a reasonable doubt.

What is the lowest probability of guilt that you think is proof beyond a reasonable doubt?



Please note the individuals in the previous experiment have been paid their Prolific experiment pay of \$3.50 already, but have not received their bonus pay yet. The decisions of jurors in this Prolific experiment will finalize those bonus pay amounts and only then will those participants be paid their bonus payments.

Suppose that you place the slider at 75%, and the computer chooses a Y of 60%. Would you find the defendant guilty or not guilty?

- Guilty
- Not Guilty

Suppose that you place the slider at 75%, and the computer chooses a Y of 80%. Would you find the defendant guilty or not guilty?

- Guilty
- Not Guilty

Decision

Information about reasonable doubt (definition 1):

A reasonable doubt is not a doubt which is based on mere guesswork or speculation. A doubt which arises merely from sympathy or from fear to return a verdict of guilt is not a reasonable doubt. A reasonable doubt is not a doubt such as may be used to escape the responsibility of a decision.

While it is your duty to give the defendant the benefit of every reasonable doubt, you are not to search for doubt. You are to search for the truth.

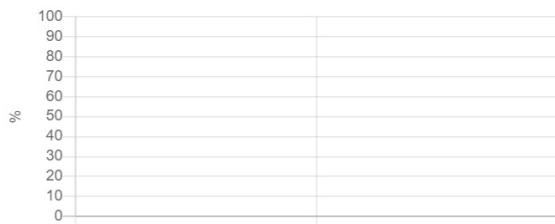
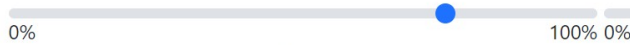
Information about reasonable doubt (definition 2):

The jury will remember that a defendant is never to be convicted on mere suspicion or conjecture.

The burden is always upon the State to prove guilt beyond a reasonable doubt. This burden never shifts to a defendant, for the law never imposes upon a defendant in a criminal case the burden or duty of calling any witnesses or producing any evidence.

So, if you, after careful and impartial consideration of all the evidence in the case, have a reasonable doubt that the defendant is guilty of the charge, you must acquit. If you view the evidence in the case as reasonably permitting either of two conclusions—one of innocence, the other of guilt—you should adopt the conclusion of innocence.

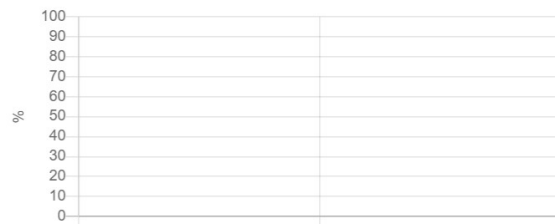
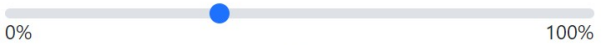
What is the lowest probability of guilt that you think is proof beyond a reasonable doubt under definition 1?



If $Y < [\text{position of slider}]$, you believe there IS reasonable doubt, and you would find the defendant not guilty.

If $Y \geq [\text{position of slider}]$, you believe there is NO reasonable doubt, and you would find the defendant guilty.

What is the lowest probability of guilt that you think is proof beyond a reasonable doubt, under definition 2?



If $Y < [\text{position of slider}]$, you believe there IS reasonable doubt, and you would find the defendant not guilty.

If $Y \geq [\text{position of slider}]$, you believe there is NO reasonable doubt, and you would find the defendant guilty.