

Reasonable Doubt and Relativity

Michael D. Cicchini*

ABSTRACT

In theory, the Constitution protects us against criminal conviction unless the state can prove guilt beyond a reasonable doubt. But in reality, this lofty standard is only as strong as the words used to explain it to the jury.

Unfortunately, attempts to explain reasonable doubt often create confusion or even diminish the burden of proof. Many courts therefore believe that the better practice is not to attempt a definition. However, empirical studies demonstrate that reasonable doubt is not self-defining, i.e., when it is not explained to the jury, it offers defendants no greater protection against conviction than the two lower, civil burdens of proof.

To solve this dilemma, courts should explain reasonable doubt on a relative basis, within the context of the civil burdens of proof. A relative, context-based instruction will allow jurors to compare and contrast the different standards, thus giving them the necessary reference points to appreciate how high the state's burden actually is.

This approach is rooted in a psychological principle called "contrast effects," and is now supported by empirical evidence as well. In this Article I present the results of my controlled experiment where mock jurors read the identical case summary of a criminal trial and were then randomly assigned to two groups, each of which received a different reasonable doubt instruction. The group that received the relative, context-based instruction acquitted at a rate 30 percent higher than the group that received a simple, undefined instruction. This result was significant at $p < .05$. Further, participants that received this relative, context-based instruction required a higher subjective confidence level in the defendant's guilt before they were willing to convict.

Drawing on this and other behavioral research, this Article presents a comprehensive jury instruction on the presumption of innocence and burden of proof that is designed to fulfill the Constitution's promise: to ensure that defendants remain free of conviction "except upon proof beyond a reasonable doubt."

* J.D., *summa cum laude*, Marquette University Law School (1999); C.P.A., University of Illinois Board of Examiners (1997); M.B.A., Marquette University Graduate School (1994); B.S., University of Wisconsin—Parkside (1990). I am a criminal defense attorney and author; I have published four books and twenty law review articles on criminal law and procedure. Many thanks to Lawrence T. White, Ph.D., for his assistance with data collection and his valuable comments on earlier drafts of this Article. Thanks to Peter White for his assistance with MTurk programming.

TABLE OF CONTENTS

INTRODUCTION 2

I. DUE PROCESS AND REASONABLE DOUBT 4

 A. *Misleading Refinements* 5

 1. An alternative hypothesis 5

 2. Searching for truth 5

 3. More than a feeling 7

 4. The important affairs of life 7

 5. Unreasonable doubts 8

 B. *Reasonable Doubt is Not Self-Defining* 9

 1. What’s in a name? 9

 2. The 60/65 rule 12

II. A LEGAL THEORY OF RELATIVITY 13

III. THE STUDY 16

 A. *Hypotheses* 16

 B. *Research Platform* 16

 C. *Participants* 17

 D. *Methodology* 18

 E. *Findings* 19

IV. EXPLAINING REASONABLE DOUBT: IT’S ALL RELATIVE 22

V. STUDY LIMITATIONS AND POTENTIAL CRITICISMS 26

 A. *Case Summary Method* 27

 B. *Single Fact Pattern* 28

 C. *Lack of Juror Deliberations* 29

 D. *Participant Attention Level* 31

 E. *Participant Bias* 32

CONCLUSION 33

APPENDIX: THE CASE SUMMARY 34

INTRODUCTION

The Constitution protects a criminal defendant from conviction unless the state can prove his or her guilt beyond a reasonable doubt.¹ The problem, however, is that this burden of proof is only as formidable as the words used to describe it to the jury.

When instructing juries on reasonable doubt, many courts go to great lengths to explain the concept. But unfortunately, their definitions often do more harm than good. Some definitions create confusion; others actually diminish the state’s burden of proof below the constitutionally-mandated standard; and yet others are so flawed they actually shift the burden to the defendant.²

Given these risks, other courts have decided that reasonable doubt should not be explained at all, as there is no better way to describe the concept than the two words themselves. However, a wealth of empirical research demonstrates that reasonable doubt

¹ See Part I.

² See Part I.A.

is *not* self-defining. That is, when left unexplained, the reasonable doubt standard offers no greater protection against conviction than the preponderance of evidence standard or the clear and convincing evidence standard.³

How, then, should courts explain the criminal burden of proof to jurors? This Article advocates for a relative, context-based approach to instructing jurors on reasonable doubt. In other words, to provide the jury with necessary points of reference to appreciate how high this burden of proof actually is, the reasonable doubt standard should be explained on a relative basis by comparing and contrasting it with the two lower, civil burdens of proof.⁴

This approach is rooted in the psychological principle called “contrast effects,”⁵ and is now supported by empirical evidence as well.⁶ In this Article I present the results of my controlled experiment where 379 mock jurors read the *same* case summary of a hypothetical criminal trial.⁷ Participants were then randomly assigned to two groups, each of which received a *different* reasonable doubt instruction: Group A (N=181) received an instruction that left reasonable doubt unexplained; Group B (N=198) received an instruction that explained the concept on a relative basis, within the context of the two lower, civil burdens of proof.⁸

Group A, which received the undefined instruction, acquitted the defendant at the rate of 32.6 percent; Group B, which received the relative, context-based instruction, acquitted the defendant at the higher rate of 42.4 percent.⁹ This was a 30 percent increase in the acquittal rate and was statistically significant at $p < .05$, with an exact p -value of 0.0496.¹⁰ Further, participants in Group B also required a higher subjective confidence level in the defendant’s guilt before they were willing to convict.¹¹

Given these findings, there is now strong empirical evidence to support the use of a relative, context-based approach to instructing jurors on reasonable doubt. Drawing on this study and on other empirical research, this Article presents a comprehensive jury instruction on the presumption of innocence and burden of proof for use in criminal trials.¹² The instruction is designed to fulfill the Constitution’s guarantee that defendants will not be convicted of a crime unless the state proves guilt beyond a reasonable doubt.¹³ This Article also addresses some study limitations and potential criticisms, and offers methodological suggestions for future researchers.¹⁴

³ See Part I.B.

⁴ See Part II.

⁵ See *id.*

⁶ See Part III.

⁷ See Part III.C.

⁸ See Part III.D.

⁹ See Part III.E.

¹⁰ See *id.*

¹¹ See *id.*

¹² See Part IV.

¹³ See *id.*

¹⁴ See Part V.

I. DUE PROCESS AND REASONABLE DOUBT

The Supreme Court has held that, in criminal trials, “the Due Process Clause protects the accused against conviction except upon proof beyond a reasonable doubt.”¹⁵ This high burden is designed to protect us “from dubious and unjust convictions, with resulting forfeitures of life, liberty and property.”¹⁶ Further, requiring proof beyond a reasonable doubt “is indispensable to command the respect and confidence of the community in applications of the criminal law.”¹⁷ Despite such grand language, however, this theoretical protection is only as strong as a trial judge’s burden of proof instruction to the jury.

Lower courts have largely been left to their own devices when instructing jurors on reasonable doubt.¹⁸ In the process, two divergent philosophies have emerged. In one camp, many courts believe “that a jury must be given some assistance in understanding the concept.”¹⁹ This is the majority view, as most states attempt to define or explain reasonable doubt in some fashion.²⁰ In the other camp, several courts refuse to define or explain reasonable doubt, claiming that attempts at “elucidation tend[] to misleading refinements.”²¹ This is the view of a substantial minority, including several federal jurisdictions²² and states.²³

Which view is correct? The answer is both. It is true that many attempts to explain reasonable doubt have resulted in “misleading refinements” or, worse yet, have lowered the government’s burden of proof or even shifted it to the defendant. However, it is also clear that juries “must be given some assistance in understanding the concept.” When reasonable doubt is left unexplained, jurors fail to distinguish between it and the two lower, civil burdens of proof. In other words, as the next two Parts demonstrate, both

¹⁵ *In re Winship*, 397 U.S. 358, 364 (1970). For the origins of our criminal burden of proof, see generally Miller W. Shealy, Jr., *A Reasonable Doubt About “Reasonable Doubt”*, 65 OKLA. L. REV. 225 (2013), and Steve Sheppard, *The Metamorphoses of Reasonable Doubt: How Changes in the Burden of Proof Have Weakened the Presumption of Innocence*, 78 NOTRE DAME L. REV. 1165 (2003).

¹⁶ *In re Winship*, 397 U.S. at 362 (quoting *Bringar v. United States*, 338 U.S. 160, 174 (1949)).

¹⁷ *Id.* at 364.

¹⁸ See *Victor v. Nebraska*, 511 U.S. 1 (1994) (giving courts tremendous leeway when instructing jurors on proof beyond a reasonable doubt); Shealy, *supra* note 15, at 229 (“the Supreme Court has inexcusably failed to give definition or substance to this concept, which it regards as fundamental to our system of justice.”).

¹⁹ *State v. Aubert*, 421 A.2d 124, 127 (N.H. 1980).

²⁰ See Bobby Greene, *Reasonable Doubt: Is It Defined by Whatever is at the Top of the Google Page?*, 50 J. MARSHALL L. REV. 933, 941 (2017) (in 39 states, instruction committees or courts make some attempt to explain the concept in their jury instructions); see also Part I.A. for definitions and explanations used in several different states.

²¹ *United States v. Lawson*, 507 F.2d 433, 443 (7th Cir. 1974).

²² See, e.g., *id.*; *United States v. Vavlitis*, 9 F.3d 206, 212 (1st Cir. 1993) (attempting to define the term “is unnecessary, could confuse the jury, and provides fertile grounds for objections”); *United States v. Headspeth*, 852 F.2d 753, 755 (4th Cir. 1988) (“We have frequently admonished district courts not to attempt to define reasonable doubt in their instructions to the jury . . .”).

²³ See, e.g., ILL. CRIM. JURY INSTRUCTIONS NO. 2.03 (2017). This instruction is based on a long line of Illinois cases holding that “neither the trial court nor counsel should define reasonable doubt for the jury.” *People v. Downs*, 69 N.E.3d 784, 788 (Ill. 2015). Several other states, including Texas, also leave the term unexplained. See Greene, *supra* note 20, at 941, n. 44 (naming ten states, in addition to Illinois, that do not define the term).

approaches to instructing jurors on reasonable doubt have failed to fulfill the Constitution's grand promise.

A. *Misleading Refinements*

Many attempts to explain reasonable doubt to the jury have actually lowered the government's burden of proof or even shifted it to the defendant. Five such examples are discussed below.

1. An alternative hypothesis

When explaining reasonable doubt, many courts include an alternative-hypothesis clause in their jury instruction. This clause instructs jurors that “[i]f two conclusions can reasonably be drawn from the evidence, one of innocence and one of guilt, you must adopt the one of innocence.”²⁴

The problem with this language—and with similarly worded versions of the alternative-hypothesis clause—is it “suggests that a preponderance of the evidence standard is relevant, when it is not.”²⁵ That is, from a jury's perspective, “if conviction of a crime fits the facts better than acquittal, it is extremely difficult to overcome the desire to match the facts with the better of the two models, even if the [state's] case is not very strong.”²⁶

Even worse than lowering the state's burden of proof, this clause puts the spotlight “on the defendant's ability to produce alternatives to the government's case, and thereby shift[s] the burden of proof to the defendant.”²⁷ Ironically, this type of jury instruction is especially harmful to an innocent defendant who “knows nothing about the crime,” and is therefore unable to produce an alternative hypothesis for the jury's consideration.²⁸

2. Searching for truth

In some jury instructions on the burden of proof beyond reasonable doubt, courts have strangely described the jury's duty as a “search for truth,”²⁹ or something similar.³⁰

²⁴ State v. Griffin, 749 A.2d 1192, 1197 (Conn. 2000). Connecticut has since changed to an awkward and difficult to understand, but technically superior, instruction: “Proof beyond a reasonable doubt is proof that precludes every reasonable hypothesis except guilt and is inconsistent with any other rational conclusion.” CONN. CRIM. JURY INSTRUCTIONS No. 2.2-3 (2007).

²⁵ United States v. Kahn, 821 F.2d 90, 93 (2d Cir. 1987).

²⁶ Lawrence Solan, *Refocusing the Burden of Proof in Criminal Cases: Some Doubt About Reasonable Doubt*, 78 TEX. L. REV. 105, 108-09 (1999).

²⁷ Id. at 105.

²⁸ Id. at 108.

²⁹ Commonwealth v. Allard, 711 N.E.2d 156, 159 (Mass. 1999); see also People v. Walos, 229 A.D.2d 953, 954 (N.Y. App. Div. 1996) (“search for the truth”); State v. Needs, 508 S.E.2d 857, 866 (S.C. 1998) (“in search of the truth”).

³⁰ See, e.g., United States v. Gray, 958 F.2d 9, 13 (1st Cir. 1992) (“find the truth”); United States v. Pine, 609 F.2d 106, 108 (3d Cir. 1979) (“evolve the truth”); State v. Dunkel, 466 N.W.2d 425, 430 (Minn. Ct. App. 1991) (“[d]etermine what you think the truth of the matter is and act accordingly.”); State v. Marshall, 586 A.2d 85, 154 (N.J. 1991) (“let your verdict declare the truth”).

Although misplaced, this mandate seems innocent enough at first. After all, in a hypothetical world where all relevant facts are known and are presented to the jury, we would want verdicts to reflect the truth of what happened.³¹ But in reality, there are numerous barriers that prevent defendants from obtaining relevant evidence,³² and, even when they do, legislatures and courts have developed numerous trial rules to prevent them from ever presenting such evidence to a jury.³³ Therefore, describing trials as a search for truth is at best naïve, and at worst disingenuous.

More accurately, given the real world in which we must operate, “truth is not the jury’s job.”³⁴ Instead, “[t]he question for any jury is whether the burden of proof has been carried by the party who bears it. In a criminal case . . . [t]he jury cannot discern whether that has occurred without examining the evidence for reasonable doubt.”³⁵ But when courts instruct jurors to search for the truth, this mandate “suggests determining whose version of events is more likely true, the government’s or the defendant’s, and thereby intimates a preponderance of evidence standard.”³⁶

Worse yet, some courts explicitly pit this search for the truth *against* the jury’s duty to examine the state’s case for reasonable doubt. One state, for example, concludes its pattern instruction by telling jurors: “you are *not* to search for doubt. You are to search for the truth.”³⁷ This mandate “impermissibly portray[s] the reasonable doubt standard as a defense tool for *hiding* the truth,”³⁸ and completely eviscerates the state’s burden.³⁹

³¹ Even this tempered truth-related claim is overly simplistic, as it ignores at least two things. First, in some trials the facts are undisputed, and the jury’s sole duty is to determine, for example, whether what happened creates a certain level of risk or “*tends to* cause or provoke a disturbance.” WIS. STAT. § 947.01 (2015-16) (emphasis added). In such cases, there is no truth to find; rather, the jury must render an opinion on a theoretical, even hypothetical, inquiry. Second, the mandate to search for truth also ignores the jury’s power of nullification—an increasingly important concept given today’s over-criminalization of behavior. In Wisconsin, for example, a person’s second or subsequent simple possession of even a small amount of marijuana is a felony. See WIS. STAT. § 941.41(3g)(e) (2015-16). In this type of case, jurors may not want their verdict to reflect the truth of what really happened, and may instead wish to exercise their power of nullification.

³² See Keith A. Findley, *Adversarial Inquisitions: Rethinking the Search for the Truth*, 56 N.Y. L. SCH. L. REV. 912, 914 (2011) (“only one side—the State—has access to all of the crime scene evidence and all of the government’s resources to collect the evidence. Typically, the accused has few resources to permit a serious independent investigation.”).

³³ See, e.g., *State v. Carter*, 782 N.W.2d 695, 709, 716–17, n.6 (Wis. 2010) (Walsh Bradley, J., concurring) (child accuser’s prior sexual conduct not admissible, even though “the prosecutor repeatedly emphasized [the child accuser’s] detailed sexual knowledge as proof of [the defendant’s] guilt”). Cf. *State v. Colburn*, 366 P.3d 258 (Mont. 2016) (accuser’s prior sexual contacts may be relevant to show an alternative source of sexual knowledge).

³⁴ *State v. Berube*, 286 P.3d 402, 411 (Wash. Ct. App. 2012).

³⁵ *Id.*

³⁶ *United States v. Gonzalez-Balderas*, 11 F.3d 1218, 1223 (5th Cir. 1994); see also *United States v. Pine*, 609 F.2d 106, 108 (3d Cir. 1979) (defendant arguing that the truth language “tend[ed] to dilute and thereby impair the constitutional requirement of proof beyond a reasonable doubt”); *United States v. Harper*, 662 F.3d 958, 961 (7th Cir. 2011) (defendant arguing “that the ‘truth’ language might have been misunderstood by the jurors as an invitation to convict by a mere preponderance of the evidence”); *State v. Avila*, 532 N.W.2d 423, 429 (Wis. 1995) (defendant arguing that the truth language “would be reasonably likely to impose a lesser burden than reasonable doubt upon the State”).

³⁷ WIS. CRIM. JURY INSTRUCTIONS No. 140 (2018) (emphasis added).

³⁸ *State v. Berube*, 286 P.3d 402, 411 (Wash. Ct. App. 2012) (emphasis added); see also *Avila*, 532 N.W.2d at 429 (defendant argued the instruction implied “that finding doubt would mean not finding the truth”).

3. More than a feeling

In another example of misleading refinements, some courts explain the concept of reasonable doubt to their juries by distinguishing it from other things. One court, for example, concluded its burden of proof instruction this way: “So, to summarize, you may not find the defendant guilty based on a mere suspicion of guilt.”⁴⁰

The danger of this comparison is demonstrated through an analogy. Describing proof beyond a reasonable doubt as being greater than a mere suspicion of guilt is like describing Roger Federer’s tennis game as being better than mine. The claim is technically true, as Federer *is* better at tennis than I am. However, the comparison does nothing to explain how good he really is, and, worse yet, it is grossly misleading. Why? Because it implies that Federer might be only marginally better than an untrained weekend tennis hack.⁴¹

Similarly, telling a jury that proof beyond a reasonable doubt is greater than a mere suspicion of guilt is technically true; however, it does nothing to explain how strong the state’s evidence must be to win a conviction. Worse yet, the comparison is grossly misleading, as it implies that anything marginally stronger than a mere suspicion might be sufficient to convict. This would include clear and convincing evidence, a preponderance of evidence, and even a *reasonable* suspicion of guilt. Of course, these levels of proof fall well short of what is required.⁴²

4. The important affairs of life

The variety of ways courts have dreamed up to mislead jurors is, seemingly, limitless. Yet another example is the important-affairs-of-life analogy. Some courts—subtly shifting the inquiry from whether the state presented *proof beyond* reasonable doubt to whether the defendant raised *reasonable doubt*—instruct their jurors that “[a] reasonable doubt is a doubt that would cause a reasonably careful and sensible person to hesitate before acting upon a matter of importance in his or her own affairs.”⁴³

Given a strict reading, this language appears pro-defendant. For example, consider the purchase of a home—surely a “matter of importance” for nearly every juror. If a doubt merely caused the juror to “hesitate before acting,” it would be a reasonable doubt even if the juror ultimately moved forward with the purchase. However, empirical studies demonstrate that jurors do not interpret the language in such a literal way.⁴⁴

³⁹ This evisceration of the state’s burden of proof has been empirically demonstrated. See Michael D. Cicchini & Lawrence T. White, *Truth or Doubt? An Empirical Test of Criminal Jury Instructions*, 50 U. RICH. L. REV. 1139 (2106); Michael D. Cicchini & Lawrence T. White, *Testing the Impact of Criminal Jury Instructions on Verdicts: A Conceptual Replication*, 117 COLUM. L. REV. ONLINE 22 (2017).

⁴⁰ Commonwealth v. Lane, No. 2033 EDA 2015, 5 (Penn. Sup. Ct. 2016).

⁴¹ See Michael D. Cicchini, *Roger Federer, Michael Cicchini, and Pennsylvania’s Burden of Proof*, THE LEGAL WATCHDOG (June 17, 2017), at <http://thelegalwatchdog.blogspot.com/2017/06/roger-federer-michael-cicchini-and.html>.

⁴² See *In re Winship*, 397 U.S. 358, 363 (1970) (stating that civil burdens of proof are not constitutionally sufficient to support a criminal conviction).

⁴³ PA. CRIM. JURY INSTRUCTIONS NO. 7.01 (2016).

⁴⁴ See Mandeep K. Dhimi, et al., *Instructions on Reasonable Doubt: Defining the Standard of Proof and the Juror’s Task*, 21 PSYCHOL. PUB. POL’Y & L. 169, 175 (2015) (“by reducing the standard of proof below

Rather, the “decisions we make in the most important affairs of our lives—choosing a spouse, a job, a place to live, and the like—generally involve a heavy element of uncertainty and risk-taking. They are wholly unlike the decisions jurors ought to make in criminal cases.”⁴⁵

As another example, choosing a spouse is probably the most important of all our decisions. Yet even that analogy is defective, as “the decision to marry is often based on a standard far less than reasonable doubt, as reflected in statistics indicating 33 to 60 percent of all marriages end in divorce.”⁴⁶ This demonstrates that “[t]he judgment of a reasonable [person] in the ordinary affairs of life, however important, is influenced and controlled by the preponderance of evidence.”⁴⁷ This, once again, is far less than what the Constitution requires in order to convict a defendant of a crime.

5. Unreasonable doubts

As a final illustration of misleading refinements, when attempting to explain the concept of reasonable doubt to their juries some courts take the inverse approach: they go to great lengths to enumerate the kinds of doubt that are *not* reasonable. This language also shifts the jury’s focus to what the defendant, rather than the state, must establish. Worse yet, such an instruction strongly deters acquittals.

For example, many modern instructions warn jurors that “[a] doubt which arises merely from sympathy or from fear to return a verdict of guilt is *not a reasonable doubt*.”⁴⁸ This is a contemporary adaptation of an older instruction that outright challenged jurors to convict, warning them not to be a “weak-kneed, timid, jellyfish of a juror who is seeking to avoid . . . convict[ing] another human being[.]”⁴⁹ Similarly, other instructions warn jurors that “[a] reasonable doubt is *not* a mere possible doubt, a speculative, imaginary or forced doubt.”⁵⁰ This language is obviously inaccurate: “Doubting, after all, *is* a matter of speculation and imagination. It requires one to imagine alternative models consistent with the evidence.”⁵¹

that intended by the law, the ‘doubt-hesitate’ instruction is more likely to lead to false convictions”). In the real world of criminal trials, however, defense counsel can use the important-affairs-of-life analogy to the defendant’s advantage by stressing that the doubt need only cause a juror to “hesitate before acting” in order to be a reasonable one. In other words, it doesn’t matter that the juror would ultimately act; rather, the issue is whether the juror would merely *hesitate* before doing so.

⁴⁵ FED. JUD. CTR. PATTERN CRIM. JURY INSTRUCTIONS No. 21 (1987); *see also* *United States v. Jaramillo-Suarez*, 950 F.2d 1378, 1386 (9th Cir. 1991) (criticizing the important-affairs-of-life analogy).

⁴⁶ *People v. Nguyen*, 46 Cal. Rptr. 2d 840, 845 (1996).

⁴⁷ *Id.* (quoting *People v. Brannon*, 47 Cal. 96, 97 (1873)).

⁴⁸ WIS. CRIM. JURY INSTRUCTIONS No. 140 (emphasis added). One problem with this warning is that it paints only half of the picture, i.e., it neglects to warn the jury that it should not convict out of sympathy for the *alleged victim* or fear of rendering a *not guilty* verdict. Some prosecutors are quick to take advantage of this oversight by playing on jurors’ sympathies and fears. *See, e.g.*, *Rhodes v. State*, 547 So. 2d 1201, 1206 (Fla. 1989) (attempting to invoke sympathy, the prosecutor urged jury to “show [the defendant] the same mercy shown to the victim on the day of her death.”); *Commonwealth of Northern Mariana Islands v. Mendiola*, 976 F.2d 1475, 486–87 (9th Cir. 1992) (attempting to invoke fear, the prosecutor warned the jury “that gun is still out there. If you say not guilty, [the defendant] walks right out the door, right behind you.”).

⁴⁹ *People v. Feldman*, 71 N.E.2d 433, 443 (N.Y. 1947) (condemning the trial court’s jellyfish analogy).

⁵⁰ FLA. CRIM. JURY INSTRUCTIONS No. 3.7 (1997) (emphasis added).

⁵¹ Solan, *supra* note 26, at 143.

The underlying problem with all of these unreasonable doubt warnings is that “[t]he weight of the instruction conveys a message to the jurors: The judge would not have presented so many ways in which the juror’s doubts can be used improperly if this were not the main problem to avoid.”⁵² This clear message from the bench minimizes the jury’s constitutional obligation to acquit, absent proof beyond a reasonable doubt, and greatly increases the government’s odds of winning a conviction by a lower standard of proof.

B. *Reasonable Doubt is Not Self-Defining*

With so many pitfalls awaiting the trial judge who attempts to define or explain reasonable doubt, many courts have determined “that the better practice is not to attempt the definition.”⁵³ Their justification is this: reasonable doubt is already “self-defining,” and, therefore, jurors require no further explanation to understand it.⁵⁴ However, this assumption has now been thoroughly tested and debunked.

1. What’s in a name?

Several studies have tested whether there is something inherently descriptive in the words “beyond a reasonable doubt” that makes the label self-explanatory. For example, in a state that leaves the concept unexplained, real-life jurors were surveyed after their jury service was concluded.⁵⁵ Quite incorrectly, nearly one-third of respondents “were either very sure or pretty sure” that “once the state has come forward with evidence of a defendant’s guilt, it becomes the defendant’s responsibility to persuade the jury of his innocence.”⁵⁶

Similarly, in the laboratory “a series of empirical studies . . . has shown that in the absence of a definition of ‘beyond a reasonable doubt’ mock jurors find it hard to apply the standard[.]”⁵⁷ These empirical studies, or controlled experiments, essentially follow the same general format. To test the effect of jury instructions on verdicts, researchers recruit test participants to serve as mock jurors in hypothetical criminal cases. Sometimes the case is presented to participants in video form, sometimes through actors who read portions of a transcript from a real-life trial, and other times through case summary materials which the participants read themselves.

After watching, listening to, or reading the identical case, mock jurors are then randomly assigned to different groups, each of which receives a different instruction on

⁵² Id. at 144.

⁵³ *United States v. Lawson*, 507 F.2d 433, 443 (7th Cir. 1974); *see also* *United States v. Hall*, 854 F.2d 1036, 1038 (7th Cir. 1988) (“At best, definitions of reasonable doubt are unhelpful to a jury, and, at worst, they have the potential to impair a defendant’s constitutional right to have the government prove each element beyond a reasonable doubt.”).

⁵⁴ *Lawson*, 507 F.2d at 443; *see also* notes 22 and 23, *supra*, for examples of federal jurisdictions and other states that leave the concept largely undefined.

⁵⁵ *See* Solan, *supra* note 26, at 119 (discussing Bradley Saxton, *How Well Do Jurors Understand Jury Instructions? A Field Test Using Real Juries and Real Trials in Wyoming*, 33 LAND & WATER L. REV. 59 (1998)).

⁵⁶ Id. at 119-20.

⁵⁷ Federico Picinali, *The Threshold Lies in the Method: Instructing Jurors About Reasoning Beyond Reasonable Doubt*, 19 INT’L J. EVIDENCE & PROOF 139, 140 (2015).

the burden of proof they should apply. One group receives a preponderance of evidence instruction (the lowest burden), another receives a clear and convincing evidence instruction (an intermediate burden), and another receives a reasonable doubt instruction (the highest burden). With all conditions being held constant across test groups—except, of course, the burden of proof instruction—mock jurors are then asked to render a verdict. Researchers then compare the groups’ verdicts to see if the reasonable doubt instruction provided the defendant with greater protection against conviction, i.e., a higher acquittal rate.

These controlled experiments have the advantage of isolating the effect of a single variable—the burden of proof jury instruction—on verdicts. And the experiments have consistently demonstrated that reasonable doubt is *not* self-defining.⁵⁸ That is, when one group receives a simple reasonable doubt instruction that leaves the concept largely unexplained, there are no significant differences in conviction rates between groups. In some cases, the lower burdens of proof even offer defendants *more* protection from conviction than does the undefined, or minimally defined, reasonable doubt standard.

For example, in a 1973 study, mock jurors heard one of two cases: sexual assault or theft.⁵⁹ They were then instructed on one of the three different burdens of proof.⁶⁰ In the theft case, the different burdens of proof did not produce statistically significant differences in verdicts.⁶¹ In the sexual assault case, once again the differences between groups were not statistically significant.⁶² Further, the verdicts did not even order themselves properly: mock jurors who received the reasonable doubt instruction acquitted at a *lower* rate than those who were instructed to apply the clear and convincing evidence standard.⁶³

Likewise, a 1985 study published the results of two experiments.⁶⁴ In the first experiment, mock jurors read the same fact pattern based on a civil trial and were then assigned to groups, each of which was instructed on a different burden of proof.⁶⁵ Despite the three different burdens, there were no statistically significant differences in

⁵⁸ Id. at 140, n. 6 (citing multiple empirical studies).

⁵⁹ W. R. Cornish & A. P. Sealy, *Juries and the Rules of Evidence*, 1973 CRIM. L. REV. 208, 210-11 (1973).

⁶⁰ Id. at 213-14. The study also tested two slightly different versions of a reasonable doubt instruction, each offering only a minimal explanation of the concept. The first version described reasonable doubt as “a doubt that might affect you in daily business or domestic decisions,” which is superficially similar to the important-affairs-of-life analogy discussed in Part I.A.4. However, the words “might affect you” are vague and, if anything, may even strengthen the instruction. The second version described reasonable doubt as “a doubt for which reasons can be given.” Merely requiring that “reasons . . . be given” for the doubt to be a reasonable one offers little elucidation and, if anything, seems to broaden the category of reasonable doubts, thus strengthening the instruction. These instructions were brief, totaling only 32 and 35 words respectively. By comparison, many of today’s reasonable doubt instructions that attempt to define or explain the concept run well into the hundreds of words. *See* Part IV. In short, these minimalist explanations failed to provide the defendant with any protection beyond the lower, civil burdens of proof.

⁶¹ Id. at 219.

⁶² Id.

⁶³ Id. This study labeled the intermediate burden of proof as “sure and certain” rather than “clear and convincing.” The researchers concluded, with regard to the sexual assault case, that the instruction “to be ‘sure and certain’ will cause [participants] to require a higher degree of proof than is the case with either of the other [instructions] tested.” Id.

⁶⁴ Dorothy K. Kagehiro & W. Clark Stanton, *Legal v. Quantified Definitions of Standards of Proof*, 9 L. & HUM. BEHAV. 159 (1985).

⁶⁵ Id. at 163.

group conviction rates.⁶⁶ The second experiment—a replication of the first—produced the same null effect.⁶⁷ That is, the instruction on reasonable doubt did not offer defendants significantly greater protection against conviction than did the two lower, civil burdens of proof.⁶⁸

Similarly, a 1991 study published the results of two additional experiments.⁶⁹ In the first experiment, mock jurors watched a video reenactment of a murder trial involving an insanity defense.⁷⁰ They were then assigned to groups which received instructions on the different burdens of proof.⁷¹ Despite the three different burdens, there were no statistically significant differences in group conviction rates.⁷² The second experiment—a replication study—again produced the same null effect.⁷³ In other words, the reasonable doubt instruction was not shown to offer the defendant greater protection against conviction.

Finally, in a 2019 study, mock jurors read one of four criminal cases: battery with weak evidence of guilt; battery with strong evidence; trespass with weak evidence; and trespass with strong evidence.⁷⁴ Participants were then assigned to one of three groups, each of which was instructed to apply a different burden of proof.⁷⁵ Once again—and regardless of the case type (battery or trespass) and strength of evidence (weak or strong)—the three different burdens of proof did *not* produce significantly different verdict patterns.⁷⁶ In fact, similar to the findings of the 1973 study, the verdicts did not even order themselves properly: mock jurors who received the undefined reasonable doubt instruction acquitted the defendant at a *lower* rate than those who were instructed to apply the mere preponderance of evidence standard.⁷⁷

⁶⁶ Id. at 164.

⁶⁷ Id. at 169.

⁶⁸ This study tested a reasonable doubt instruction that included very little elaboration and totaled only 79 words. It described being convinced beyond a reasonable doubt as feeling “an abiding conviction, to a moral certainty . . .” Id. at 163. Just as in the previous study, such minimal guidance was insufficient to convey the concept, as it did not provide the defendant with significant protection beyond the lower, civil burdens of proof.

⁶⁹ James R. P. Ogloff, *A Comparison of Insanity Defense Standards on Juror Decision Making*, 15 L. & HUM. BEHAV. 509 (1991).

⁷⁰ Id. at 514.

⁷¹ Id. at 515. This study’s reasonable doubt instruction was described as a “standard jury instruction[] a judge would give jurors in a similar case.”

⁷² Id. at 516.

⁷³ Id. at 519.

⁷⁴ Lawrence T. White & Michael D. Cicchini, *Is Reasonable Doubt Self-Defining?*, 64 VILL. L. REV. 1 (forthcoming, 2019).

⁷⁵ Id. at 11. In addition to testing three different burdens of proof, the study also tested three different instructions on the reasonable doubt standard. One instruction left the concept completely unexplained; a second merely equated being convinced beyond a reasonable doubt with having “a firm belief” in the defendant’s guilt; and a third merely told jurors that, if they had a reasonable doubt, they should acquit “even if you think that the charge is probable true.” The three reasonable doubt instructions, which totaled 50, 63, and 60 words respectively, did not produce significantly different verdict patterns and therefore were combined into a single reasonable doubt group. Id. at 11, n. 74.

⁷⁶ Id. at 12-13.

⁷⁷ Id. at 12.

2. The 60/65 rule

In addition to the above studies showing no significant differences in acquittal rates under the three burdens of proof, researchers have also been testing the impact of reasonable doubt jury instructions in a different way: they seek to determine the subjective confidence level that jurors require before they are willing to convict. “This research has consistently shown that the jurors in criminal cases will often be satisfied with much less certainty than is conventionally assumed.”⁷⁸

When *judges* and others trained in the law are asked to equate the “proof beyond a reasonable doubt” standard with a corresponding confidence level in the defendant’s guilt, most reply that it should be somewhere near 90 percent. In fact, in a large survey of federal judges that produced 171 respondents, 74 percent (or 126 judges) set the corresponding number at “90% or higher.”⁷⁹ Similarly, in a real-life example, an appellate court reversed a conviction because the trial judge, when instructing the jury, had equated proof beyond a reasonable doubt with “anything more than a 75% chance” the defendant was guilty.⁸⁰ The court held that setting such a low threshold, coupled with another mistake, “constituted prejudicial error.”⁸¹

Contrary to the beliefs and desires of most judges, however, *jurors* equate our highest burden of proof with a much lower confidence level in guilt. In a 1996 study, when asked to quantify proof beyond a reasonable doubt, test participants set an average threshold that ranged from a mere 54 percent to 70 percent, depending on the particular version of the reasonable doubt instruction they received.⁸² Similarly, in a 2007 study, test participants set the conviction threshold at a mere 63 percent chance that the defendant was guilty.⁸³ And in 2014, researchers found that laypersons were willing to convict at a 68 percent probability of guilt.⁸⁴

The 2019 study discussed earlier also tested the threshold for conviction and found that 90 percent of mock jurors followed what the researchers called “the 60/65 rule.”⁸⁵ That is, *regardless of the burden of proof instruction they received*, when participants believed that less than 60 percent of the evidence favored the state, they voted to acquit; but when they believed that more than 65 percent of the evidence favored

⁷⁸ Erik Lillquist, *Recasting Reasonable Doubt: Decision Theory and the Virtues of Variability*, 36 U.C. DAVIS L. REV. 85, 88 (2002).

⁷⁹ Solan, *supra* note 26, at 126 (discussing C. M. A. McCauliff, *Burdens of Proof: Degrees of Belief, Quanta of Evidence, or Constitutional Guarantees?*, 35 VAND. L. REV. 1293, 1325 (1982)). See also Jon O. Newman, *Quantifying the Standard of Proof Beyond a Reasonable Doubt: A Comment on Three Comments*, 5 L. PROBABILITY & RISK 267, 267 (2006) (discussing the subjective threshold set by a judge and several law professors as being “no less than 95%,” “more than a 95% chance,” and no less “than 0.8.”).

⁸⁰ *McCullough v. State*, 657 P.2d 1157, 1159 (Nev. 1983).

⁸¹ *Id.*

⁸² Irwin A. Horowitz & Laird C. Kirkpatrick, *A Concept in Search of a Definition: The Effects of Reasonable Doubt Instructions on Certainty of Guilt Standards and Jury Verdicts*, 20 L. & HUM. BEHAV. 655, 666 (1996).

⁸³ Daniel B. Wright & Melanie Hall, *How a “Reasonable Doubt” Instruction Affects Decisions of Guilt*, 29 BASIC & APPLIED SOCIAL PSYCHOL. 91, 96 (2007).

⁸⁴ Svein Magnussen, et al., *The Probability of Guilt in Criminal Cases: Are People Aware of Being ‘Beyond Reasonable Doubt’?*, 28 APPLIED COGNITIVE PSYCHOL. 196, 199 (2014).

⁸⁵ White & Cicchini, *supra* note 74, at 14.

the state, they voted to convict.⁸⁶ Or, as another researcher put it, “[r]ather than having to move jurors from 0% to 90% certainty, all prosecutors need do is move the needle on the scale from 50% to perhaps 65% certainty.”⁸⁷

II. A LEGAL THEORY OF RELATIVITY

As demonstrated earlier, attempting to explain reasonable doubt to a jury is fraught with peril. Many attempts are a “grand conglomeration of garbled verbiage and verbal garbage.”⁸⁸ Worse yet, many such attempts have actually diminished the burden of proof or even shifted it to the defendant.⁸⁹ But on the other hand, the empirical evidence shows that when reasonable doubt is left undefined or is only minimally explained, the standard offers defendants no greater protection than the two lower, civil burdens of proof.⁹⁰

The question, then, is this: How should courts explain the concept of reasonable doubt to accurately convey the heightened standard? At first blush, the answer might seem obvious. “Blackstone would have put the probability standard for proof ‘beyond a reasonable doubt’ at somewhat more than 90%, for he declared: It is better that ten guilty persons escape than one innocent suffer.”⁹¹ Therefore, courts could simply instruct jurors that, in terms of confidence level, proof beyond a reasonable doubt means they must be 90 percent certain the defendant is guilty before they may convict.⁹²

This, however, is not a viable solution. As a practical matter, the numeric-based approach to defining proof beyond a reasonable doubt is dead on arrival. Even though controlled experiments demonstrate that quantified definitions are effective and produce the desired, relative acquittal rates, “The legal profession appears to be adverse to the expression of standards of proof in probability terms.”⁹³ In other words, there is “no jurisdiction in which quantified definitions are given.”⁹⁴

But even if courts were willing to adopt a numeric-based definition, not everyone would agree about what number should be assigned to the reasonable doubt standard. It is well-settled that the preponderance of evidence standard means “any amount of certainty greater than 50%.”⁹⁵ But what level of confidence is needed to be convinced of

⁸⁶ *Id.*

⁸⁷ Irwin A. Horowitz, *Reasonable Doubt Instructions: Commonsense Justice and Standard of Proof*, 3 PSYCHOL., PUB. POL. & L. 285, 294 (1997).

⁸⁸ *State v. Aubert*, 421 A.2d 124, 127 (N.H. 1980) (internal citations omitted).

⁸⁹ *See* Part I.A.

⁹⁰ *See* Parts I.B.1. and I.B.2.

⁹¹ *United States v. Fatico*, 458 F. Supp. 388, 411 (E.D.N.Y. 1978) (internal citation omitted).

⁹² Instead of asking jurors to quantify their confidence level, another way of wording the question is to ask jurors to quantify the strength of the state’s evidence. Wording the question that way invokes the weight-of-evidence or scales-of-justice analogy. For example, one journalist used such an analogy in explaining the preponderance of evidence standard: “If . . . 50.1% of the evidence supports a claim but 49.9% does not, that 50.1% is still enough to tip the scale, to prove the claim.” Alan Abrahamson, *Tragedy at Sea Pits What-Ifs Against Legal Proof*, L.A. TIMES (Aug. 26, 1991), at http://articles.latimes.com/1991-08-26/local/me-820_1_circuitcourt/2.

⁹³ Kagehiro & Stanton, *supra* note 64, at 174.

⁹⁴ *Id.*

⁹⁵ Lillquist, *supra* note 78, at 87.

something beyond a reasonable doubt? “Is 90% certainty required? 95%? 99%? Or could the amount of certainty be much lower, say perhaps 75%?”⁹⁶

Rather than defining the criminal burden of proof on a numeric scale, the better approach is to explain the reasonable doubt standard on a relative basis by comparing it to the two lower, civil burdens of proof.

As a preliminary matter, unlike attempts to assign a number to proof beyond a reasonable doubt, the accuracy of the relative, context-based explanation is not open to debate. It is uncontroversial that the reasonable doubt standard is *higher* than the clear and convincing evidence standard which, in turn, is *higher* than the preponderance of evidence standard.⁹⁷

But the primary benefit of discussing the reasonable doubt standard on a relative basis is this: it gives the jury some necessary points of reference to understand how high the state’s burden of proof actually is. This relative approach is rooted in the psychological principle known as “contrast effects.”

Most judgments in everyday life are evaluative in nature. People may want to know whether a particular grade is good or bad, whether a person is trustworthy, how well someone performed on a test, or what a person’s athletic abilities are like. Rarely can such questions be answered in absolute terms (e.g., running 1 mile in 5 minutes). Rather than absolute, judgments are usually *relative* and result from comparisons. That is, judgments are mostly evaluations of a target with respect to some *comparison standard*. . . . [J]udgments may differ significantly depending on the comparison standard they are contrasted to, a phenomenon that social psychologists refer to as *contrast effects*.⁹⁸

We saw this principle in effect earlier when a jury instruction compared proof beyond a reasonable doubt to a mere suspicion of guilt.⁹⁹ This approach was correct in principle, but absolutely horrible in its execution. Why? Because it utilized the defective comparison standard of mere suspicion, which is something that shouldn’t even be in the same discussion with proof beyond a reasonable doubt.

⁹⁶ *Id.* The difficulty in answering this question is likely due to its theoretical economic underpinnings. That is, the number associated with proof beyond a reasonable doubt is sometimes derived from the perceived disutility of convicting an innocent person compared with the perceived utility of convicting the guilty. These, of course, are matters that are the subject of considerable debate. However, such debate often leads to conclusions that are questionable or even highly suspicious. *See id.* at 91 (arguing that serious crimes such as “terrorism” should have a lower burden of proof due to the high cost of acquitting the guilty, but arguing that petty crimes such as “traffic offenses” should *also* have a lower burden of proof due to the low cost of convicting the innocent).

⁹⁷ *See e.g., Fatico*, 458 F. Supp. at 402-06 (discussing the different burdens of proof, their justifications, and the types of cases in which they are applied). Few things are uncontroversial in the legal arena, but there are countless legal authorities asserting the fundamental principle that proof beyond a reasonable doubt is the highest burden of proof. This principle is to criminal law what offer and acceptance are to contract law.

⁹⁸ *Contrast Effects*, PSYCHOLOGY: PSYCHOLOGY RESEARCH AND REFERENCE, at <https://psychology.iresearchnet.com/social-psychology/social-cognition/contrast-effects/> (emphasis added) (accessed Dec. 18, 2018).

⁹⁹ *See* Part I.A.3.

Rather, in order to accurately convey its meaning and ensure that it provides defendants with more protection than the two civil burdens of proof, jurors must be instructed that proof beyond a reasonable doubt is higher than a preponderance of evidence and higher than clear and convincing evidence. Providing the proper comparison standards may not guarantee that jurors will approach a 90 percent confidence level before they convict; however, providing these points of reference will likely mean the prosecutor will have to “move the needle on the scale” beyond a mere “65% certainty” to win a conviction.¹⁰⁰

Another benefit to this relative, context-based approach is its familiarity. It is not a completely foreign concept, as some courts already instruct their jurors this way. For example, one state’s instruction reads in relevant part: “In civil cases, it is only necessary to prove that a fact is more likely true than not or that its truth is highly probable. In criminal cases such as this, the State’s proof must be more powerful than that. It must be beyond a reasonable doubt.”¹⁰¹ Similarly, another state’s instruction reads in relevant part: “It is not enough to establish a probability, even a strong probability, that the defendant is more likely to be guilty than not guilty. That is not enough.”¹⁰²

Some researchers have advocated for this relative, context-based approach to jury instructions. In a 2000 article, researchers argued that “[b]y providing this explicit contrast with a less stringent standard of proof [preponderance of evidence], the definition encourages jurors to adopt an appropriately high threshold for conviction.”¹⁰³ Better still, the instruction “could be strengthened even further by adding an additional contrast with [the] clear and convicting” standard of proof.¹⁰⁴ Similarly, in the 2019 study previously discussed, the researchers argued: “In criminal cases, defining ‘proof beyond a reasonable doubt’ by comparing it to other, lower standards would provide the necessary context for jurors to appreciate this high standard.”¹⁰⁵

There is even some empirical support for this relative, context-based approach to defining proof beyond a reasonable doubt. The 1985 study of two experiments, discussed earlier, actually included a third experiment. Participants read a set of jury instructions that included all three standards of proof, compared and contrasted the standards, and then rated each one for how difficult it would be for a plaintiff to win a civil case.¹⁰⁶ The researchers concluded that this “third experiment suggested—but did not demonstrate—that the ineffectiveness of the legal definitions may be due to a *lack of a comparative context*; jury instructions may have to communicate the applicable standard’s ordinal position vis-à-vis the other standards of proof for improved understanding and utilization by jurors.”¹⁰⁷

In sum, a fundamental principle of psychology (“contrast effects”), legal precedent from some states, sound logical argument, and even some empirical evidence all suggest that courts should use a relative, context-based approach when explaining

¹⁰⁰ Horowitz, *supra* note 87, at 294.

¹⁰¹ ARIZ. CRIM. JURY INSTRUCTIONS No. 5(b)(1) (2016).

¹⁰² MASS. CRIM. JURY INSTRUCTIONS No. 2.180 (2015).

¹⁰³ Elisabeth Stoffelmayr & Shari S. Diamond, *The Conflict Between Precision and Flexibility in Explaining “Beyond a Reasonable Doubt”*, 6 PSYCHOL., PUB. POL. & L. 769, 776 (2000).

¹⁰⁴ *Id.*

¹⁰⁵ White & Cicchini, *supra* note 74, at 17.

¹⁰⁶ Kagehiro and Stanton, *supra* note 64, at 172.

¹⁰⁷ *Id.* at 174 (emphasis added).

reasonable doubt to their juries. However, this collection of evidence may not be compelling enough to spur legal reform. The following study therefore empirically tests the claim that such an instruction will provide defendants with more protection than an instruction that leaves reasonable doubt undefined.

III. THE STUDY

In order to test this legal theory of relativity, I designed a controlled experiment similar to those described earlier. Test participants read the identical material about a single criminal trial. They were then randomly assigned to groups, each of which received a different reasonable doubt instruction: Group A's instruction left the concept undefined; Group B's explained it using the relative, context-based approach discussed above. Participants rendered a verdict and then indicated their confidence level in the defendant's guilt.

Keeping all other things constant, the random assignment of sufficiently large samples to groups that received different jury instructions will isolate the impact of the instructions on verdicts.

The following sections formally state the specific hypotheses being tested, describe the research platform and the test participants, explain the study design and methodology, and discuss the study's findings.

A. Hypotheses

The first hypothesis is that when reasonable doubt is explained on a relative basis—i.e., by putting it in context with the two lower, civil burdens of proof—this relative, context-based instruction will produce significantly more acquittals than a simple instruction that leaves reasonable doubt undefined. (The null hypothesis is that the relative, context-based instruction has *no impact* on mock jurors' verdict choices.)

The second hypothesis is that, if there is an increase in acquittals, the relative, context-based instruction will also shift the tipping point—i.e., “the point at which the majority of participants, for the first time, vote[] guilty instead of not guilty”¹⁰⁸—thus providing defendants with more protection when jurors perceive the state's case as weak.

B. Research Platform

To test these hypotheses, participants were recruited through Amazon's Mechanical Turk (MTurk) online research platform to serve as mock jurors in a hypothetical criminal case.

MTurk has many advantages, including “easy access to a large, stable, and diverse subject pool, the low cost of doing experiments, and faster iteration between developing theory and executing experiments.”¹⁰⁹ Consequently, “thousands of

¹⁰⁸ White & Cicchini, *supra* note 74, at 13.

¹⁰⁹ Winter Mason & Siddharth Suri, *Conducting Behavioral Research on Amazon's Mechanical Turk*, 44 BEHAV. RES. 1, 1-2 (2012).

researchers across the social sciences have conducted research using MTurk.”¹¹⁰ In 2015 alone, “social science journals with an impact factor greater than 2.5” published “more than 500” papers using data collected from MTurk.¹¹¹

Historically, test participants for behavioral research experiments have been recruited from a readily available pool of college students. MTurk samples are, unsurprisingly, “more demographically diverse than typical undergraduate populations.”¹¹² Further, researchers have found that “MTurk participants provided data that met or exceeded the psychometric standards set by data collected using other means (e.g., undergraduate samples).”¹¹³

C. Participants

MTurk workers were required to have an approval rating of 90 percent or higher to participate in this study.¹¹⁴ Additionally, because workers were recruited to serve as mock jurors, they were required to be jury eligible, i.e., U.S. citizens and 18 years of age or older.¹¹⁵ To further enhance data quality, participants were monitored during the experiment and their work was rejected if, after rendering their verdict, they failed an attention-check question,¹¹⁶ a reading comprehension question,¹¹⁷ or both.

After data collection was completed, I discovered that one participant was not a U.S. citizen and several others failed to affirm that they were, in fact, citizens. Their data were discarded, leaving a sample of 379 mock jurors for the study.¹¹⁸

This sample was large and diverse. Participants hailed from 45 different states. Forty-eight percent were female. Their ages ranged from 19 to 71 years, with a mean (average) age of 38 years and a median age (50th percentile) of 35 years.¹¹⁹ The ethnic composition consisted of 76 percent non-Hispanic whites; 10 percent African Americans;

¹¹⁰ Michael D. Buhrmester, et al., *An Evaluation of Amazon’s Mechanical Turk, Its Rapid Rise, and Its Effective Use*, 13 PERSPECTIVES ON PSYCHOL. SCI. 149, 149 (2018).

¹¹¹ Id. at 150.

¹¹² Id.

¹¹³ Id. at 149 (parenthetical original).

¹¹⁴ See id. at 151 (discussing MTurk worker approval rating and its relationship to data quality).

¹¹⁵ At least one state, however, requires its citizens to be twenty-one years of age before serving on a jury. *Juror Basics*, MISSOURI COURTS: JUDICIAL BRANCH OF GOVERNMENT, at <https://www.courts.mo.gov/page.jsp?id=994> (accessed Dec. 7, 2018).

¹¹⁶ The attention-check question simply asked jurors to select the highest value in a row of numbers. Twenty-one workers selected the wrong number; their data were rejected.

¹¹⁷ In addition to a simple attention-check question, participants answered a more complex reading comprehension question about the study’s case summary. This multiple choice question included four possible answers, and participants were told to select all that were correct. Participants who selected A, the correct choice, were approved. Those who failed to select A were rejected. Those who selected A but also selected B and/or C, two incorrect choices, were also rejected. Initially, those who selected A but also selected D were rejected. However, answer D included two different components or claims, one of which was correct. Several workers who were initially rejected for selecting both A and D wrote to complain of the confusing sentence structure of D. This was a legitimate criticism; therefore, all 84 workers who selected A and D (and had not been rejected for other reasons, e.g., citizenship) also had their work approved and were included in the study.

¹¹⁸ To request the dataset of approved participants who were included in the study, email the author at mdc@cicchinilaw.com.

¹¹⁹ Of the 379 approved participants, six failed to provide their age. However, because all MTurk workers are required to be at least eighteen years old, their data were included in the study.

seven percent Asian-Americans; three percent Hispanics; three percent mixed race; and one percent other. Fifty-three percent of participants reported having at least a college degree, while an additional 35 percent completed some college.

D. Methodology

To test the hypotheses, I used the case summary method. (The case summary method, including its strengths and weaknesses, is discussed in Part V.A.) Every mock juror read the identical fact pattern describing a hypothetical battery case. The case summary began with an instruction on the charged crime, including its elements, followed by an 882-word synopsis of court testimony from three individuals: Emily, the alleged victim; Officer Hamilton, the responding police officer; and Dr. Wilkins, a clinical psychologist.

The synopsis detailed how the police were called to investigate a disturbance at an apartment building. Officer Hamilton discovered that a husband and wife were the cause of the disturbance. The wife, Emily, made a written statement to the officer alleging that her husband, John, had battered her. At trial, however, Emily recanted the accusation and the prosecutor confronted her with her prior, inconsistent statement to the officer. In addition, Dr. Wilkins testified for the state and explained that recantations by abused persons are common. All witnesses were cross-examined by the defense; the defendant elected not to testify and the defense did not present any evidence.

The case summary also included a 100-word general instruction from the judge. This instruction defined “evidence” to include not only sworn testimony but also statements that were read into the record in court. The instruction also explained that the weight to be given to any piece of evidence is up to the jurors. (The actual case summary used in the study is reproduced in the Appendix.)

Before being asked to render a verdict, test participants were randomly assigned to one of two groups, each of which received a different jury instruction on the burden of proof beyond a reasonable doubt. Group A was given a simple instruction where reasonable doubt was *not* defined or explained. The instruction, in its entirety, was as follows:

This is a criminal case, and the State (through the prosecutor) has the burden of proving the defendant’s guilt beyond a reasonable doubt. If, after carefully considering all of the evidence, you are convinced beyond a reasonable doubt, then you should find the defendant guilty. However, if you have a reasonable doubt, you must find the defendant not guilty.¹²⁰

Group B was given the identical instruction, plus additional language that explained the burden of proof on a relative basis by placing it within the context of the two lower, civil burdens of proof. The instruction, in its entirety, was as follows:

This is a criminal case, and the State (through the prosecutor) has the burden of proving the defendant’s guilt beyond a reasonable doubt. If, after carefully

¹²⁰ This instruction closely mirrors the pattern instruction of the Seventh Circuit Federal Court of Appeals. See SEVENTH CIR. CRIM. JURY INSTRUCTIONS No. 1.03 (2012).

considering all of the evidence, you are convinced beyond a reasonable doubt, then you should find the defendant guilty. However, if you have a reasonable doubt, you must find the defendant not guilty.¹²¹

Some civil cases use the preponderance of evidence standard. In those cases, it is only necessary to prove that something is probably true, or more likely true than not. But this is a criminal case, and the State's proof must be more powerful than that.

Other civil cases use the clear and convincing evidence standard. In those cases, it is necessary to prove that the truth of something is highly probable. But this is a criminal case, and the State's proof must also be more powerful than that.

In criminal cases such as this, you can convict the defendant only if the State's proof satisfies you beyond a reasonable doubt that the defendant is guilty. If it does not, you must find the defendant not guilty even if you think the charge is probably true, and even if you think it is highly probable that the charge is true.¹²²

After receiving one of the two reasonable doubt instructions, participants rendered their verdicts. Next, they were asked to "use a scale of 0% to 100%" to answer the following question: "Regardless of whether you voted Guilty or Not Guilty, how confident are you that the defendant committed the crime?" They then answered the attention-check question and the reading comprehension question. Finally, participants provided information about their age, citizenship, gender, state of residence, ethnicity, and education.

E. Findings¹²³

To test the first hypothesis—that the relative, context-based jury instruction will produce a significantly higher group acquittal rate—we must compare the acquittal rates of Group A and Group B.

Group A received the simple reasonable doubt instruction, and 122 of the 181 mock jurors returned verdicts of guilt for a group conviction rate of 67.4 percent and a group acquittal rate of 32.6 percent. Group B received the relative, context-based instruction, and 114 of 198 mock jurors voted to convict for a group conviction rate of 57.6 percent and a group acquittal rate of 42.4 percent. This is illustrated in the table below.

¹²¹ Id.

¹²² The last three paragraphs of this instruction draw heavily from the pattern instructions used in Arizona and Vermont. See ARIZ. CRIM. JURY INSTRUCTIONS No. 5(b)(1) (2016); VT. CRIM. JURY INSTRUCTIONS No. 04-101 (2005). The instruction tested in this Article was also proposed in White & Cicchini, *supra* note 74, and in Michael D. Cicchini, *Instructing Jurors on Reasonable Doubt: It's All Relative*, 8 CALIF. L. REV. ONLINE 72 (2017).

¹²³ In addition to the primary findings discussed in the Article, there are several ancillary findings. For example, there was no relationship between participants' ethnicity and verdicts. With regard to gender, men were more likely than women to vote guilty. With regard to education, participants with some college or a college degree were more likely to vote guilty than were participants who were less educated (high school diploma, GED, or some high school) or more educated (post-graduate degree). See Part V.E. for a discussion of how these differences are addressed through random assignment of participants to groups.

	G	NG	Total	%G	%NG
Group A (unexplained)	122	59	181	67.4%	32.6%
Group B (relative)	114	84	198	57.6%	42.4%
Total			379		

The acquittal rate did increase, as hypothesized. However, the next step is to determine whether this increase in the acquittal rate—a 30 percent increase under the relative, context-based instruction ($.326 \times 1.30 = .424$)—is statistically significant.

Significance is measured by a statistic called the *p*-value.¹²⁴ The *p*-value is the probability of obtaining the observed effect, or one that is more extreme, assuming that the *null* hypothesis is *true*.¹²⁵ A *p*-value of .05 or smaller is generally considered statistically significant; a small *p*-value means the observed effect is inconsistent with the null hypothesis, allowing us to reject the null.¹²⁶

The result of this study—with sample sizes of 181 and 198 and acquittal rates of 32.6 and 42.4 percent—is significant at $p < .05$, with an exact *p*-value of 0.0496.¹²⁷ In other words, if the *null* hypothesis is *true* (i.e., if the relative, context-based instruction does *not* have an impact on verdicts), there is about a five percent chance (*p*) that we would obtain the observed difference or a greater difference (i.e., a 30 percent or more increase in acquittal rates).¹²⁸ This is strong evidence against the null hypothesis, leading us to reject the null.¹²⁹

Alternatively stated, the *p*-value measures the probability of a Type I error, i.e., the risk of obtaining a false positive when testing the hypothesis; in plain language, we are slightly more than 95 percent certain ($1 - p$) that the difference in acquittal rates between Groups A and B is a real difference and did not occur by chance.¹³⁰

¹²⁴ See Raymond Hubbard, *Alphabet Soup: Blurring the Distinctions Between *p*'s and *α*'s in Psychological Research*, 14 THEORY & PSYCHOL. 295, 298 (2004); Arthur Aron & Elaine N. Aron, STATISTICS FOR PSYCHOLOGY 92 (3d ed. 2003).

¹²⁵ See Hubbard, *supra* note 124, at 298.

¹²⁶ For a general discussion of sample size, effect size, statistical significance, and *p*-values, see Aron & Aron, *supra* note 124.

¹²⁷ The *p*-value can be quickly and easily calculated using an online statistics calculator. See, e.g., Comparison of Proportions Calculator, MEDCALC: STATISTICAL SOFTWARE, at https://www.medcalc.org/calc/comparison_of_proportions.php (difference between the two proportions produced a *p*-value of .0496) (accessed Dec. 8, 2018). In addition, a statistical test for the difference between two proportions produced a z-score of -1.97, and the *p*-value is 0.049.

¹²⁸ See Hubbard, *supra* note 124, at 298.

¹²⁹ See *id.* (the “significance test revolves around the rejection of the null hypothesis at the $p \leq .05$ level. . . . ‘Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis.’”).

¹³⁰ While we can be confident there is “a real difference” between groups, this alternative explanation should *not* be read to mean that the difference will always be 30 percent. Further, while this alternative explanation is intuitively easier to understand and accurately conveys the relevant concept, it is open to objection on highly technical, semantic, and historical grounds. For example, many social scientists, psychology journals, APA publication manuals, and even “well-regarded textbooks on statistical methods” use the *p*-value to measure the probability of a Type I error. *Id.* at 304-07. This, apparently, blends concepts from two competing schools of thought: the Fisherians (*p*-values and evidence against the null) and the Neyman-Pearsonians (the Type I error and α). See *id.* at 297-304. Strangely, even though both of these camps “regard observed significance levels, or P values, as error probabilities, we occasionally hear

This finding provides strong empirical support for the first hypothesis: that a relative, context-based jury instruction will produce a higher group acquittal rate than an instruction that leaves reasonable doubt unexplained.

To test the second hypothesis—that the relative, context-based jury instruction also shifted the tipping point, i.e., “the point at which the majority of participants, for the first time, voted guilty instead of not guilty”¹³¹—we must look at Group B’s conviction numbers segregated by the participants’ confidence level in the defendant’s guilt.

To begin, recall that several studies demonstrated that when reasonable doubt is left unexplained, the tipping point occurs somewhere between 63 and 68 percent confidence in the defendant’s guilt.¹³² The most recent published study, which confirmed those findings, found the tipping point to occur somewhere between 60 and 65 percent.¹³³

In our study, Group B, which received the relative, context-based jury instruction on reasonable doubt, was comprised of 198 participants. The point at which the majority of these participants, for the first time, voted guilty instead of not guilty occurred somewhere between 70 and 75 percent confidence in guilt. In other words, the tipping point was indeed shifted, as participants in Group B required a higher confidence level in the defendant’s guilt before a majority of them voted to convict. This is illustrated in the table below.

Confidence	Group B			
	G	NG	Total	%G
Less than 60%	6	25	31	19.4%
60% & 65%	2	9	11	18.2%
70%	5	6	11	45.5%
75%	11	7	18	61.1%
80% & 85%	18	14	32	56.3%
More than 85%	72	23	95	75.8%
	198			

While there was a shift in the tipping point,¹³⁴ it was not dramatic. Far from requiring the prosecutor to move the needle from zero to 90 percent certainty as many

allegations . . . that P values are actually not error probabilities.” Deborah G. Mayo, *Are P Values Error Probabilities? or, “It’s the Test, Stupid” (Second Install)*, ERROR STATISTICS PHILOSOPHY (Aug. 17, 2014), at <https://errorstatistics.com/2014/08/17/are-p-values-error-probabilities-installment-1/>. Mayo sensibly argues that readers should not “be misled into thinking there’s a deep inconsistency” that prevents “using both N-P and Fisherian tests.” *Id.* Conversely, those who object to calling the *p*-value the measure of error rate bring to mind a saying about the narcissism of petty differences. At the very least, this debate is well beyond the scope of this Article and, apparently, even beyond the interest of most scientists, as “[u]sers of statistical techniques in the social and medical sciences are almost totally unaware of the[se] distinctions[.]” Hubbard, *supra* note 124, at 304-05.

¹³¹ White & Cicchini, *supra* note 74, at 13.

¹³² See Part I.B.2.

¹³³ See *id.*

¹³⁴ The tipping point was shifted not only when compared to previous studies, but also when compared to Group A in this study, which received the instruction that left reasonable doubt undefined and unexplained. Group A’s tipping point fell between the 65 and 70 percent confidence levels—slightly higher than that of similarly instructed jurors in previous studies, but lower than that of jurors in Group B.

judges expect,¹³⁵ the majority of jurors first became willing to convict when their confidence level in guilt reached somewhere between 70 and 75 percent.

Interestingly, when compared to the simple, undefined reasonable doubt instruction received by Group A, the relative, context-based instruction received by Group B provided no additional protection for the defendant when mock jurors’ confidence in guilt was *very* low, i.e., when they perceived the state’s case as *very* weak. Mock jurors who were less than 60 percent certain of guilt voted to convict at the nearly identical (and very low) rate, regardless of the instruction they received: five guilty votes out of 31 participants in Group A; six out of 31 in Group B. In other words, when jurors viewed the state’s case as *very* weak, the great majority of them voted not guilty regardless of their instruction.

Once the mock jurors’ confidence level in the defendant’s guilt reached 60 percent, the relative, context-based instruction offered the defendant more protection throughout all confidence levels. This is illustrated in the table below.

Confidence	Group A				Group B			
	G	NG	Tot	%G	G	NG	Total	%G
Less than 60%	5	26	31	16.1%	6	25	31	19.4%
60% & 65%	6	10	16	37.5%	2	9	11	18.2%
70%	4	3	7	57.1%	5	6	11	45.5%
75%	4	2	6	66.7%	11	7	18	61.1%
80% & 85%	26	10	36	72.2%	18	14	32	56.3%
More than 85%	77	8	85	90.6%	72	23	95	75.8%
	181				198			

In sum, given these data, the second hypothesis is confirmed. The majority of mock jurors in Group B were not willing to convict until their confidence level in the defendant’s guilt reached a higher level: somewhere between 70 and 75 percent. However, instead of dramatically shifting the tipping point, the relative, context-based instruction received by Group B shifted it only slightly, but produced lower conviction rates throughout multiple levels of the mock jurors’ confidence.

IV. EXPLAINING REASONABLE DOUBT: IT’S ALL RELATIVE

The Supreme Court has held that “a person accused of a crime . . . would be at a severe disadvantage, a disadvantage amounting to a lack of fundamental fairness, if he could be adjudged guilty and imprisoned for years on the strength of the same evidence as would suffice in a civil case.”¹³⁶ Given this, and given that jurors find simple reasonable doubt instructions to be indistinguishable from the lower, civil burdens of proof,¹³⁷ courts cannot continue to leave the concept of reasonable doubt undefined and still comply with Due Process.

¹³⁵ See Solan, *supra* note 26, at 126.

¹³⁶ In re Winship, 397 U.S. 358, 363 (1970) (quoting “the dissenters in the New York Court of Appeals.”).

¹³⁷ See Part I.B.1.

Instead, this study’s findings provide strong empirical support for what a small number of states are already doing and what some researchers recommend: to comply with Due Process, the burden of proof beyond a reasonable doubt should be explained to jurors on a relative basis, within the context of the two lower, civil burdens of proof.

This study demonstrated that Group A—which received a simple, undefined reasonable doubt instruction—acquitted the defendant at the rate of 32.6 percent.¹³⁸ By comparison, Group B—which received the relative, context-based instruction—acquitted at the higher rate of 42.4 percent.¹³⁹ This statistically significant 30 percent increase in acquittal rates demonstrates that Group B’s instruction offers more protection against conviction than a simple instruction that leaves the meaning of reasonable doubt to a jury’s imagination.¹⁴⁰

In addition to this finding, recall that previous studies also demonstrated that, when reasonable doubt is not properly explained, mock jurors are willing to convict when their confidence level in the defendant’s guilt reaches a mere 60 to 65 percent.¹⁴¹ This confidence level is far lower than what is expected by those trained in the law, and is lower than what some courts have held to be an appropriate threshold for conviction.¹⁴²

The relative, context-based instruction tested in this study moved the needle on the mock jurors’ confidence gauge—if only a small distance. A majority of mock jurors in Group B did not vote to convict until their subjective confidence level in the defendant’s guilt reached between 70 and 75 percent.¹⁴³ Instead of producing a dramatic shift, Group B’s instruction provided the defendant with greater protection against conviction at multiple confidence levels.¹⁴⁴

The relative, context-based instruction tested in this study is a vast improvement over the seemingly endless variety of defective explanations currently being spouted by the courts,¹⁴⁵ as well the simple instructions that leave reasonable doubt largely or entirely undefined.¹⁴⁶ However, the instruction is not yet complete.

In addition to the concept of reasonable doubt, the instruction must also discuss the closely-related concept of the presumption of innocence. This presumption is far more important in real-life trials, which typically last days and sometimes weeks, than it is in controlled experiments, which are over in a relatively short time. Given the length of real-life trials, it is also important to instruct juries on these concepts at the *beginning*, as well as near the end, of the case.¹⁴⁷ This will hopefully prevent jurors from forming an opinion early in the case, and then taking that preexisting view with them into deliberations.

¹³⁸ See Part III.E.

¹³⁹ See *id.*

¹⁴⁰ See *id.*

¹⁴¹ See Part I.B.2.

¹⁴² See *id.*

¹⁴³ See Part III.E.

¹⁴⁴ See *id.*

¹⁴⁵ See Part I.A.

¹⁴⁶ See Part I.B.1.

¹⁴⁷ See, e.g., WIS. CRIM. JURY INSTRUCTIONS No. 50 (2017) (listing preliminary instructions, which include pattern jury instruction no. 140 on the presumption of innocence and burden of proof which is to be read right before the judge announces that “[t]he lawyers will now make opening statements.”)

Next, while the instruction tested in this study moved the needle to the point where mock jurors were 70 to 75 percent confident in guilt before a majority first voted to convict, this is still not close to the 90 percent certainty that is widely accepted by judges as the requisite confidence level. Therefore, the jury instruction that was tested in this study should be improved in two additional ways.

First, other empirical research demonstrates that, when judges explain the *reason* behind a rule of law, jurors are more likely to follow it.¹⁴⁸ The ideal instruction would therefore inform jurors *why* they must apply such a high burden of proof to the case.

Second, the Supreme Court has held that, in order to convict, the jury must first “reach a subjective state of *near certitude* of the guilt of the accused[.]”¹⁴⁹ Because 70 to 75 percent confidence in guilt is far from near certitude, this important concept should also be conveyed in a criminal jury instruction.

Therefore, a jury instruction on the burden of proof (and the closely-related concept of the presumption of innocence) that passes constitutional muster would read, in its entirety, as follows:

Presumption of Innocence

Defendants are not required to prove their innocence. The law presumes every person charged with a crime to be innocent. This means you must find the defendant not guilty unless, in your deliberations at the end of the case, you find this presumption is overcome by evidence which satisfies you beyond a reasonable doubt that the defendant is guilty.¹⁵⁰ The presumption of innocence is not a mere slogan but an essential part of the law that is binding upon you.¹⁵¹

Burden of Proof

This is a criminal case, and the State has the burden of proving the defendant’s guilt beyond a reasonable doubt. If, after carefully considering all of the evidence, you are convinced beyond a reasonable doubt, then you should find the defendant guilty. However, if you have a reasonable doubt, you must find the defendant not guilty.¹⁵²

Burden of Proof Explained

Some civil cases use the preponderance of evidence standard. In those cases, it is only necessary to prove that something is probably true, or more likely true than

¹⁴⁸ See Nancy Steblay et al., *The Impact on Juror Verdicts of Judicial Instruction to Disregard Inadmissible Evidence: A Meta-Analysis*, 30 L. & HUM. BEHAV. 469 (2006).

¹⁴⁹ Jackson v. Virginia, 443 U.S. 307, 315 (1979) (emphasis added).

¹⁵⁰ This proposed language draws heavily from WIS. CRIM. JURY INSTRUCTIONS No. 140 (2018).

¹⁵¹ HAW. CRIM. JURY INSTRUCTIONS No. 3.02 (2014).

¹⁵² This instruction closely mirrors the pattern instruction of the Seventh Circuit Federal Court of Appeals. See SEVENTH CIR. CRIM. JURY INSTRUCTIONS No. 1.03 (2012). This instruction is identical to the language tested in this study, except it deletes the parenthetical indicating that the prosecutor represents the state, as this would be explained to a real-life jury through other instructions that are not relevant for our purposes. This instruction does not indicate that the state must prove *every element* of the crime beyond a reasonable doubt, as that language should be included in the substantive jury instruction setting forth the elements of the charged crime.

not. But this is a criminal case, and the State’s proof must be more powerful than that.

Other civil cases use the clear and convincing evidence standard. In those cases, it is necessary to prove that the truth of something is highly probable. But this is a criminal case, and the State’s proof must also be more powerful than that.

In criminal cases such as this, you can convict the defendant only if the State’s proof satisfies you beyond a reasonable doubt that the defendant is guilty. If it does not, you must find the defendant not guilty even if you think the charge is probably true, and even if you think it is highly probable that the charge is true.¹⁵³

Reasons for Proof Beyond a Reasonable Doubt

One reason for this high burden of proof is to protect defendants against the risk of wrongful conviction. Another is to promote public confidence in our justice system. Individuals going about their ordinary affairs should be confident that the state cannot convict them of a crime unless it proves their guilt with utmost certainty.¹⁵⁴ Therefore, you must not convict the defendant unless, in your deliberations, you become fully satisfied or entirely convinced of the defendant’s guilt.¹⁵⁵

All of the language in this proposed instruction is accurate. And with the exception of three explanatory sentences that give the *reasons* for the high burden of proof, all of the language in this instruction is already being read to jurors in some states. And the brief explanatory language—a contemporary but accurate paraphrase of the Supreme Court’s reasoning—is included only to increase the probability that jurors will actually apply the reasonable doubt standard.

While the proposed instruction is somewhat lengthy—without subject headings, it tallies 361 words—it also includes the presumption of innocence language on which some courts instruct separately. But this length is not uncommon,¹⁵⁶ and some instructions greatly exceed it.¹⁵⁷ Finally, 361 words is certainly not too long for what “is perhaps the most important aspect of the closing instruction to a jury in a criminal trial.”¹⁵⁸

¹⁵³ The three paragraphs at the heart of this proposed instruction draw heavily from ARIZ. CRIM. JURY INSTRUCTIONS No. 5(b)(1) (2016) and VT. CRIM. JURY INSTRUCTIONS No. 04-101 (2005). These three paragraphs were tested in this study, and were previously proposed by White & Cicchini, *supra* note 74, and by Cicchini, *supra* note 122.

¹⁵⁴ This proposed language is a paraphrase of *In re Winship*, 397 U.S. 358, 363-64 (1970).

¹⁵⁵ This proposed language is a slightly modified version of N.C. CRIM. JURY INSTRUCTIONS No. 101.10 (2008).

¹⁵⁶ Many states’ pattern instructions reach the 300-word range. *See, e.g.*, ALASKA CRIM. JURY INSTRUCTIONS No. 1.06 (2012) (reasonable doubt and presumption of innocence instruction is 329 words).

¹⁵⁷ *See, e.g.*, CJI 2d (NY) MODEL INSTRUCTIONS: FINAL INSTRUCTIONS 6-9 (reasonable doubt and presumption of innocence instructions’ combined total is 632 words), at https://www.nycourts.gov/judges/cji/5-SampleCharges/CJI2d.Final_Instructions.pdf.

¹⁵⁸ *State v. Aubert*, 421 A.2d 124, 127 (N.H. 1980).

V. STUDY LIMITATIONS AND POTENTIAL CRITICISMS

Those who have an interest in preserving the status quo are quick to criticize scientific research if it contradicts their opinions, preferences, or wishes. For example, two recently-published studies demonstrated that when a reasonable doubt instruction concludes by telling participants “not to search for doubt,” but instead “to search for the truth,” they convict at significantly higher rates than those who received the same reasonable doubt instruction *without* the truth-not-doubt mandate.¹⁵⁹ These findings should not have surprised anyone. How could the closing mandate “*not* to search for doubt” do anything *except* lower the burden of proof below the reasonable doubt standard? Nonetheless, the response from prosecutors—a group that relies on that burden-lowering mandate to win convictions—was frantic.

Prosecutorial criticisms of those studies included several poorly reasoned, knee-jerk complaints. By way of example, one prosecutor argued that the studies should be disregarded because they were biased in their design. The prosecutor elaborated:

The first problem is that the entire premise of the [studies] was biased from the start. The authors were not searching for the truth: they were not looking to see what effect various instructions might have in a mock trial situation. *What they were searching for was evidence to back their contention* that an instruction that urges jurors to search for the truth will lead to more convictions than an instruction that urges jurors to search for doubt.¹⁶⁰

Aside from misstating the substance of the studies—neither of them tested an instruction that “urges jurors to search for doubt”—the prosecutor attempted to spin the positing and testing of hypotheses, i.e., “searching for evidence to back their contention,” into a form of bias. However, as the trial judge in that case explained, “The positing of hypotheses is not bias, but is the first step in scientific investigation. The empirical results from sound methods are what inform. If the empirical difference . . . is statistically significant, the null hypothesis is rejected and the posited hypothesis is accepted.”¹⁶¹

Such bizarre criticisms are constrained only by a prosecutor’s imagination and boldness, and are therefore difficult to predict and nearly unlimited in their variety and number. In other words, these types of wild criticisms “are very much like landmines: they are easy to lay, but difficult and time-consuming to cleanup.”¹⁶² Consequently, the balance of this Part will address only potential criticisms that are at least somewhat rooted in scientific thinking.

¹⁵⁹ See Cicchini & White, *supra* note 39.

¹⁶⁰ Michael D. Cicchini, *Spin Doctors: Prosecutor Sophistry and the Burden of Proof*, 87 U. CIN. L. REV. 489, 498-99 (2018) (quoting the prosecutor’s argument in response to the defendant’s motion to modify the reasonable doubt jury instruction) (emphasis added).

¹⁶¹ *Id.* at 500 (citing the judge’s written decision granting the defendant’s motion to modify the reasonable doubt instruction).

¹⁶² *Id.* at 498.

To be sure, all studies, including this one, have inherent limitations and are susceptible to criticism.¹⁶³ Possible limitations and potential criticisms of this study include the use of the case summary method, the use of a single fact pattern to test the hypotheses, the lack of juror deliberations, the possibility of participant inattention, and the possibility of participant bias. Some of these criticisms carry more weight than others, but each is addressed below.

A. Case Summary Method

This study employed the written case summary method. (The actual case summary, in its entirety, is reprinted in the Appendix). Case summaries are commonly used in behavioral research.¹⁶⁴ But this methodology is not the best way to “approximate or mimic features in the real world” of jury trials.¹⁶⁵ Consequently, some researchers have argued for the use of more realistic trial simulations when conducting controlled experiments.¹⁶⁶ The evidence and opinions on this matter are far from settled, however, as other researchers have failed to observe any differences in the decision-making of test participants who receive basic simulations compared with those who receive more elaborate ones.¹⁶⁷

A *written* case summary would not be ideal—or even adequate—for certain types of experiments, such as, for example, testing the impact of the physical attractiveness bias on mock juror decision-making.¹⁶⁸ But the *written* case summary method is certainly appropriate—and possibly even optimal—for testing the effect of *written* jury instructions, as this study did.

The primary benefits of the written case summary method are its brevity and simplicity. While lengthy video-taped reenactments of trials, for example, are more detailed and realistic, they also “provide a myriad of additional legally relevant and

¹⁶³ For example, behavioral research is graded on four different measures of validity, and “when researchers take steps to strengthen one kind of validity, another kind of validity may be weakened.” Michael D. Cicchini & Lawrence T. White, *Educating Judges and Lawyers in Behavioral Research: A Case Study*, 53 GONZAGA L. REV. 159, 183 (2017-18).

¹⁶⁴ See, e.g., Wayne Weiten & Shari S. Diamond, *A Critical Review of the Jury Simulation Paradigm*, 3 L. & HUM. BEHAV. 71, 77 (1979).

¹⁶⁵ Cicchini & White, *supra* note 163, at 183 (discussing the concept of “external validity”).

¹⁶⁶ See Joel D. Lieberman & Bruce D. Sales, *What Social Science Teaches Us About the Jury Instruction Process*, 3 PSYCHOL., PUB. POL. & L. 589, 592 (1997) (“A better methodology is to provide a videotaped trial to participants. The videotaped format provides a highly engaging simulation, and is much more representative of an actual trial.”).

¹⁶⁷ See Geoffrey P. Kramer & Norbert L. Kerr, *Laboratory Simulation and Bias in the Study of Juror Behavior: A Methodological Note*, 13 L. & HUM. BEHAV. 89, 89 (1989) (“Results provided no support for the contention that treatment effects act differently as a function of the length of the stimulus trial in which they are embedded. Rather, it is suggested that treatments used in simplified jury simulations may often show similar effects when examined in more realistic, complex settings if the treatments are comparable.”).

¹⁶⁸ See Marc W. Patry, *Attractive But Guilty: Deliberation and the Physical Attractiveness Bias*, 102 PSYCHOL. REP. 727 (2008). The case summary method is not ideal for testing the impact of other types of evidence as well. See Shari S. Diamond, *Illuminations and Shadows from Jury Simulations*, 21 L. & HUM. BEHAV. 561, 564 (1997) (While certain studies, such as those that test “the credibility of various types of expert testimony,” demand “a fairly elaborate simulation,” other studies can be accomplished using “a less extensive trial stimulus.”)

irrelevant bases on which to make a decision[.]”¹⁶⁹ In the process, a single jury instruction could easily get lost in the clutter of the additional decision factors. By comparison, the written case summary method allows researchers to reduce this risk by increasing the salience of the jury instruction within the test materials.¹⁷⁰

Further, as one judge explained in response to a prosecutor’s criticism of the case summary method:

[The criticism] is a red herring because in no way does not using live witnesses undermine the validity of [the study]. One could have presented live witnesses, but that would have been a different study. As long as the variable of the story told in the study was consistent among groups, *how the story is told makes no difference . . .*¹⁷¹

Nonetheless, future researchers may wish to test the impact of the relative, context-based reasonable doubt instruction on juror decision-making by using different methodologies, including audio-recorded readings of transcripts and video-recorded reenactments of trials.

B. *Single Fact Pattern*

When constructing this study I valued simplicity of design and clarity of findings over unnecessary nuance and complexity.¹⁷² The study therefore tested the impact of a single version of the relative, context-based reasonable doubt instruction, and it tested the instruction’s impact on verdicts using a single fact pattern. In other words, it did not test multiple variations of the relative, context-based instruction; nor did it test the instruction’s impact in cases with varying strengths of evidence, i.e., weak- and strong-evidence cases; nor did it investigate the instruction’s impact on verdicts in cases charging different crimes, e.g., burglary or sexual assault.

The study’s findings demonstrated that the relative, context-based instruction provided the defendant with more protection than the instruction where reasonable doubt

¹⁶⁹ Diamond, *supra* note 168, at 564.

¹⁷⁰ See Kagehiro & Stanton, *supra* note 64, at 162 (“Since the study was concerned with the effects of jury instructions, rather than with evidentiary issues, a trial summary was utilized to maximize the salience of the instructions concerning the standard of proof.”).

¹⁷¹ Cicchini, *supra* note 160, at 509 (quoting the trial judge’s written decision granting the defendant’s motion to modify the burden of proof jury instruction) (emphasis added).

¹⁷² My preference for clarity and simplicity was based, in part, on some of the criticisms that were levied against my previously published work. These criticisms, some of which are discussed briefly in this Article, demonstrated how little some prosecutors and judges understand about behavioral research and, even worse, how eager they were to form strong opinions and even publicly express them. Further, some criticisms were so baseless that they likely crossed the line separating mere scientific illiteracy from outright bad faith. For example, some prosecutors and judges launched multiple *ad hominem* attacks against me personally, and also inaccurately branded the Richmond and Columbia journals in which the research was published as unreliable “defense attorney journals.” See Michael D. Cicchini, *The Battle Over the Burden of Proof: A Report from the Trenches*, 79 U. PITT. L. REV. 61, 77-80 (2017) (refuting these criticisms). Consequently, the scientific illiteracy, and probably even bad faith, exhibited by some members of the legal community influenced my choices in study design and statistical analysis for this project.

was left unexplained; the null hypothesis (that the instruction has *no* effect on verdicts) was rejected. However, we cannot say the extent to which, or the frequency with which, this will be observed in other studies or in real life cases.

This limitation is best explained by analogy. Researchers who survey a random sample of likely voters may be able to forecast a characteristic in the sample, such as support for a certain political candidate, to the larger population to determine the percentage of votes the candidate will receive (and whether he or she will win the election). Such projection is not possible with the controlled experiment presented in this Article. The most obvious reason is that, unlike the population of real-life *voters* who go to the polls and are choosing among the same candidates as the survey participants, real-life *jurors* will be faced with dramatically different fact patterns than those encountered by the test participants in this controlled experiment.

To use two examples, many real-life trials will have very strong objective evidence of guilt (e.g., strong forensic evidence, several reliable eyewitnesses, or a non-coerced confession), and many will have very weak objective evidence of guilt (e.g., no physical evidence, no disinterested witnesses, and a highly impeachable complaining witness). In these strong or weak cases—which will comprise a large percentage, if not the majority, of real-life jury trials—the reasonable doubt instruction won’t have any effect at all on verdicts. That is, the jurors will either convict (due to very strong evidence) or acquit (due to very weak evidence) regardless of the burden of proof instruction and, *hypothetically*, even *regardless of the burden of proof itself*.

However, it is important to keep in mind that, even though the burden of proof instruction won’t have any impact at all in many real-life trials, “that is certainly not a justification for improperly [or sub-optimally] instructing jurors on reasonable doubt, only to hope they will view the evidence as falling near one of the two extremes on the strength-of-evidence spectrum,” thus rendering the instruction irrelevant.¹⁷³ Rather, a trial judge’s duty is to properly, and optimally, instruct the jury from the outset.¹⁷⁴

From a scientific standpoint, however, given the inherent limitations of the study, researchers may wish to assess the generalizability of its findings by testing the impact of such instructions using different fact patterns that involve different charged crimes and varying levels of evidence strength. This is the scientific method at work. “The goal of social science is to arrive at conclusions that are supported by multiple converging lines of evidence, with each contributing study being necessarily flawed, but flawed in a different way.”¹⁷⁵

C. Lack of Juror Deliberations

This study tested the impact of a jury instruction on mock jurors’ *pre*-deliberation verdicts—or, to be very precise, their guilty or not-guilty votes. When criticizing similarly constructed studies, many in the legal community have seized upon this design feature to argue that, in order to be valid, studies must include juror deliberations. This common criticism fails for several reasons.

¹⁷³ See Cicchini, *supra* note 160, at 507.

¹⁷⁴ See *id.* Jurors are often instructed on reasonable doubt not only after the presentation of evidence but also before the opening statements. See *supra* note 147.

¹⁷⁵ Cicchini & White, *Truth or Doubt?*, *supra* note 39, at 1160.

First, numerous published studies examine the impact of jury instructions on decision-making *without* the use of deliberations. In fact, “in mock jury studies, the jurors usually answer without deliberating with other jurors.”¹⁷⁶ This study design feature is common and well accepted in the behavioral research community; the lack of deliberations in no way invalidates a study.¹⁷⁷

Second, in the broader context, beyond the narrow topic of burden of proof jury instructions, the research on whether and how juror deliberations impact verdicts is mixed.¹⁷⁸ And when jurors do deliberate, there is evidence they spend a very small amount of their time discussing the burden of proof.¹⁷⁹ Further, and perhaps unsurprisingly, “The prevailing view . . . is that deliberations play a minor role in determining jury verdicts because the pre-deliberation majority generally prevails in the end.”¹⁸⁰

Third, the purpose of a jury instruction is to “accurately inform jurors about . . . the law that they are to apply in an understandable, conversational, and unbiased manner.”¹⁸¹ Similar to the point made in the previous Part, a judge’s goal is *not* to instruct jurors in a way that is incorrect, or even suboptimal, only to hope that juror misconceptions about the law somehow get cleared up during the course of deliberations. Consequently, individual mock jurors’ *pre*-deliberation verdicts and their thought processes are highly relevant and should be of great interest to judges and other lawmakers.

Fourth, even if jurors were to spend a significant amount of their deliberation time discussing the burden of proof, there is no reason to believe that the deliberative process would iron-out the numerous problems with reasonable doubt jury instructions. As discussed earlier, numerous studies now demonstrate that, unlike judges who are trained in the law and associate proof beyond a reasonable doubt with approximately 90 percent certainty, jurors are willing to convict when their confidence level in the defendant’s guilt reaches a mere 60 to 65 percent.¹⁸²

¹⁷⁶ Ron C. Michaelis, et al., *A LITIGATOR’S GUIDE TO DNA: FROM THE LABORATORY TO THE COURTROOM* 243 (2008).

¹⁷⁷ See, e.g., Ogloff, *supra* note 69, at 515 (“Participants were not given an opportunity to deliberate.”); Kagehiro & Stanton, *supra* note 64, at 163 (same); White & Cicchini, *supra* note 74, at 18 (“mock jurors in our study did not deliberate”); Bette L. Bottoms, et al., *Gender Differences in Jurors’ Perceptions of Infanticide Involving Disabled and Non-Disabled Infant Victims*, 35 CHILD ABUSE & NEGLECT 127, 128, 132 (2011) (same); Joanna D. Pozzulo et al., *The Effects of Victim Gender, Defendant Gender, and Defendant Age on Juror Decision Making*, 37 CRIM. JUST. & BEHAV. 47, 47, 54 (2010) (same).

¹⁷⁸ See, e.g., Richard R. Izzett & Walter Leginski, *Group Discussion and the Influence of Defendant Characteristics in a Simulated Jury Setting*, 93 J. SOC. PSYCHOL. 271, 271 (1974) (deliberations *mitigated* the effect of physical attractiveness bias); Robert J. MacCoun, *The Emergence of Extralegal Bias, During Jury Deliberation*, 17 CRIM. JUST. & BEHAV. 303, 311 (1990) (deliberations *exacerbated* the effect of physical attractiveness bias); and Patry, *supra* note 168, at 731 (deliberations *reversed* the effect of physical attractiveness bias, i.e., mock jurors were biased *against* the attractive defendant).

¹⁷⁹ See James R. P. Ogloff, *Judicial Instructions and the Jury: A Comparison of Alternative Strategies*, FINAL REPORT (BRITISH COLUMBIA LAW FOUNDATION, 1988).

¹⁸⁰ Diamond, *supra* note 168, at 564.

¹⁸¹ Model Crim. Jury Instructions, MICHIGAN COURTS: ONE COURT OF JUSTICE, at <https://courts.michigan.gov/Courts/MichiganSupremeCourt/criminal-jury-instructions/Pages/default.aspx> (accessed Dec. 15, 2018).

¹⁸² See Part I.B.2.

There are numerous real-life examples that support the empirical research on this point. In one state where courts are not permitted to define or explain reasonable doubt, a deliberating jury submitted the following question to the judge: “What is your definition of reasonable doubt? 80%, 70%, 60%?”¹⁸³ The judge could only answer: “We cannot give you a definition[,] it is your duty to define.”¹⁸⁴ The jury convicted the defendant, perhaps with only 60 percent confidence in his guilt. This demonstrates how jury deliberations, at least as they pertain to the burden of proof, are very much like the blind leading the blind.¹⁸⁵

D. Participant Attention Level

A common and legitimate concern about the MTurk online research platform is that, unlike studies using college students as test subjects, it is difficult to ensure the participants are paying sufficient attention to the test materials. However, research has demonstrated that “MTurk participants’ attention is equal to or better than undergraduate participants’ attention.”¹⁸⁶ In addition, as discussed earlier, several measures were taken to ensure data quality for this study.

First, MTurk workers are rated,¹⁸⁷ and participants for this study were required to have an approval rating of at least 90 percent in order to participate.¹⁸⁸ Second, based on the test materials, I determined that participants could complete the experiment in approximately eight to nine minutes. And the approved participants that were included in the study completed it in an average time of 10.2 minutes and a median time of 8.7 minutes.¹⁸⁹ Third, participants were given an attention-check question, and those who did not answer it correctly were rejected.¹⁹⁰ And fourth, participants were also given a more complex reading comprehension question, and those who failed to answer correctly were also rejected.¹⁹¹

In sum, much more was done to ensure participants’ attention to the test materials than judges typically do to ensure jurors’ attention in real-life courtrooms—a place where the stakes are high, yet inattentive and even sleeping jurors are widely tolerated over defendants’ objections.¹⁹²

¹⁸³ *People v. Downs*, 69 N.E.3d 784, 786 (Ill. 2015).

¹⁸⁴ *Id.*

¹⁸⁵ *See Greene, supra* note 20, at 950-51 (discussing examples of juror misconduct in deliberations where jurors searched the internet or called a personal attorney seeking an explanation of the term “reasonable doubt”).

¹⁸⁶ Buhrmester, et al., *supra* note 110 at 151.

¹⁸⁷ *See id.* at 151 (discussing MTurk worker approval rating and its relationship to data quality).

¹⁸⁸ *See* Part III.C.

¹⁸⁹ To request the dataset of approved participants who were included in the study, email the author at mdc@icchinilaw.com.

¹⁹⁰ *See* Part III.C.

¹⁹¹ *See id.*

¹⁹² *See, e.g., State v. Chestnut*, 643 S.W.2d 343, 346 (Tenn. Ct. App. 1982) (upholding the denial of defendant’s motion for new trial despite undisputed evidence of two jurors sleeping through evidentiary portions of trial). Jurors are also paid miserably for their important work. For example, in Milwaukee County, a large metropolitan area, jurors receive \$8.00 per half-day of service, or about \$2.00 per hour or even less, depending on how late the judge keeps them into the evening. *Juror Pay*, MILWAUKEE CTY. COURTS: JURY DUTY, at <https://county.milwaukee.gov/EN/Courts/Jury-Duty#basics> (accessed Dec. 15, 2018).

E. Participant Bias

Finally, one of the more common prosecutorial criticisms of studies like the one presented in this Article is as follows: the test participants were not randomly selected and were not screened for bias; therefore, because the participants could have been biased, the study is not valid and should be disregarded.¹⁹³

This criticism fails for three reasons. First, test participants cannot be randomly selected *and* screened for bias, as these two things are obviously mutually exclusive.¹⁹⁴ Second, this criticism confuses random *selection* (something that is used, for example, in *surveys*, where researchers use a sample to forecast the frequency of a characteristic in the larger population) with the random *assignment* of participants to groups (something that is used in *experiments* where researchers manipulate variables and test for differences between groups).¹⁹⁵ As explained earlier,¹⁹⁶ controlled experiments like this one do not attempt to forecast actual conviction rates in real-life jury trials; rather, they study the differences in conviction rates between differently instructed groups.

Third, with regard to participant bias, one judge articulated his complaint this way when criticizing a similarly-constructed study: “You know, for example, do [the participants] have an interest in the case? You know, how is their intelligence? Did they essentially look like they had a bias? Well, anything of that sort, and none of that is referred to in this case.”¹⁹⁷ The judge then argued that the participants should have been screened for such bias before they could be included in the study. Similarly, with regard to that same study, a prosecutor complained that because the researchers failed to use a *voir dire* process to “weed-out those with preconceived ideas,” the research was invalid.¹⁹⁸

The call for this type of screening of participants in a controlled experiment is, at best, grossly misinformed. As a different judge explained when responding to this precise complaint: “If *voir dire* would have occurred, the sample would have been biased based on the subjective bias of the person(s) doing the *voir dire* (and striking possible study participants) resulting in the study’s validity being compromised by the subjectivity of those doing the *voir dire*.”¹⁹⁹

Further, this complaint about participant bias is much ado about nothing, as properly constructed experiments already address this issue through the process of random assignment—a topic mentioned above. That is, when sufficiently large samples of participants are *randomly assigned* to test groups, we can expect that each group will

¹⁹³ See Cicchini, *supra* note 160, at 500-02 (discussing this common prosecutorial criticism in the context of similarly constructed studies).

¹⁹⁴ See *id.*

¹⁹⁵ See Beth Morling, RESEARCH METHODS IN PSYCHOLOGY: EVALUATING A WORLD OF INFORMATION 173 (2012) (discussing how sample selection is far more important for a survey, or “frequency claim,” than it is for controlled experiments that seek to detect “associations and causes”).

¹⁹⁶ See Part V.B.

¹⁹⁷ Cicchini & White, *supra* note 163, at 172 (quoting a transcript of a trial judge criticizing two published studies and denying the defendant’s motion to modify a reasonable doubt instruction).

¹⁹⁸ Cicchini, *supra* note 160, at 500 (quoting a prosecutor’s argument from a trial judge’s written decision granting the defendant’s motion to modify a reasonable doubt instruction).

¹⁹⁹ *Id.* at 501.

have roughly the same proportion of men and women, the same proportion of old and young persons, the same proportion of educated and uneducated persons, and the same proportion of “biased” and unbiased persons—however that word might be defined—and so forth.²⁰⁰

The benefits of random assignment are not just theoretical; rather, they are easily verified. With regard to this study and the identifiable personal characteristics discussed above—gender, age, and educational level—women comprised 49 percent of Group A and 47 percent of Group B; the average age of participants was 37.3 years in Group A and 38.5 years in Group B; and participants with some college or a college degree comprised 79 percent of Group A and an identical 79 percent of Group B.²⁰¹

The point is this: when statistically equivalent groups such as these receive an identical case summary followed by different jury instructions, and then convict at significantly different rates, we can be confident that the difference was produced by the manipulated variable, i.e., the jury instruction, and not by the personal characteristics of the test participants.²⁰²

CONCLUSION

The Constitution requires the jury to acquit a criminal defendant unless the state proves guilt beyond a reasonable doubt.²⁰³ However, this constitutional protection is only as strong as the trial court’s burden of proof instruction to the jury.

Instructing jurors on reasonable doubt is risky business. Many attempts at a definition have created confusion or, worse yet, have lowered or even shifted the burden of proof.²⁰⁴ However, the solution is *not* to leave the concept of reasonable doubt unexplained. Empirical research demonstrates that, when left unexplained, the reasonable doubt standard offers defendants no greater protection than the two lower, civil burdens of proof.²⁰⁵ Further, jurors are willing to convict with only a 60 to 65 percent confidence level in the defendant’s guilt—a threshold far lower than what is expected under the proof beyond a reasonable doubt standard.²⁰⁶

To adequately convey the state’s high burden of proof, courts should instruct jurors on a relative basis. By explaining proof beyond a reasonable doubt as being higher than the preponderance of evidence standard, and higher even than clear and convincing evidence, jurors will have the necessary reference points to appreciate how high the standard actually is.²⁰⁷ This relative, context-based approach is supported by a fundamental principle of psychology (contrast effects), the existing jury instructions of

²⁰⁰ See Morling, *supra* note 195, at 251-52 (random assignment “creates a situation in which the experimental groups will become virtually equal . . .”).

²⁰¹ To request the dataset of approved participants who were included in the study, email the author at mdc@cicchinilaw.com.

²⁰² See Morling, *supra* note 195, at 251-52.

²⁰³ See Part I.

²⁰⁴ See Part I.A.

²⁰⁵ See Part I.B.1.

²⁰⁶ See Part I.B.2.

²⁰⁷ See Part II.

some states, sound logical argument, the recommendations of other researchers, and now direct empirical evidence.²⁰⁸

This Article presented the results of my controlled experiment where mock jurors read the *same* case summary and were then randomly assigned to two groups, each of which received a *different* instruction on reasonable doubt.²⁰⁹ Group A received a simple, undefined instruction and acquitted at the rate of 32.6 percent; Group B received a relative, context-based instruction and acquitted at the higher rate of 42.4 percent.²¹⁰

This 30 percent increase was statistically significant at $p < .05$, with an exact p -value of 0.0496.²¹¹ Further, mock jurors that received the relative, context-based instruction required a higher subjective confidence level in the defendant's guilt before the majority of them were first willing to convict, and the instruction also provided the defendant with more protection throughout multiple participant confidence levels.²¹²

Based on these findings, and the findings from other empirical research, this Article presented a comprehensive jury instruction on the presumption of innocence and the burden of proof.²¹³ The instruction is designed to fulfill the Constitution's promise that every accused person remains free of conviction unless there is proof of guilt beyond a reasonable doubt.

APPENDIX

Below is the case summary that was read by all test participants in both groups. Each group, however, received a different reasonable doubt instruction.

Case Summary

I. Charged crime

This is a case where the State charged John, the defendant, with one count of battery. The crime of battery, for purposes of this case, occurs when all of the following elements are met:

- A. The defendant caused bodily harm to the victim. ("Bodily harm" means physical pain or injury or both.)
- B. The defendant intended to cause bodily harm to the victim. ("Intended to cause bodily harm" means that the defendant had the mental purpose to cause harm or was aware that his conduct was practically certain to cause harm.)
- C. The defendant caused the bodily harm without the victim's consent.

II. Evidence presented at trial

The State, through the prosecutor, presented three witnesses at trial. The defendant elected not to testify.

²⁰⁸ See Part III.

²⁰⁹ See Part III.D.

²¹⁰ See Part III.E.

²¹¹ See *id.*

²¹² See *id.*

²¹³ See Part IV.

A. Testimony of Emily

Emily, the alleged victim, identified the defendant, John, as her husband. They have been married for nearly six years. Emily does not work; she stays at home with the couple's two children. On November 1, 2018, after John got home from work, they were arguing about child-related issues, and John got angry. Emily tried to get John to keep his voice down, but he would not. She believes this is why the neighbor in the next-door apartment called 911 on them.

Emily testified, however, that her dispute with John was only verbal. She denied that John ever harmed her physically. When Emily denied physical contact, the prosecutor showed her a written statement and asked whether she had signed that statement on the day of the incident. Emily agreed that she had. The statement was then read in court, as follows:

“Today, November 1, 2018, I got into an argument with my husband John. Things got very loud, and he was very angry. I have never seen him get that angry, and I was afraid. John grabbed my upper arms and shook me. I have red marks and bruising on my upper arms from John doing this. I told John that he was hurting me, and he should stop it and let me go, but he continued to squeeze my arms for about thirty seconds while he yelled in my face. I did not give John consent to cause me pain or injury. I am still afraid and I want the temporary no-contact order enforced.”

John's defense lawyer then questioned Emily. Emily again testified that she had signed the written statement. However, she did not write the statement; rather, it was handwritten for her by one of the police officers. Emily testified that the officer gave her the chance to review the statement before signing it, but she declined. Emily testified that she was very angry at John when the police arrived—in part because of things he said during their argument—and she might have exaggerated some things. Emily also testified that the police officer may have gotten some things wrong, either because he was rushed or possibly to exaggerate the situation for his police report. Emily testified that John did not touch her on that day, and the two only argued.

The prosecutor then asked Emily some more questions. Emily admitted that she had small, mild bruises on her upper arms, but does not know how they got there. She then said that she had a similar bruise on one of her calves. That mark was not caused by John either, and she does not know how she got that mark. She testified that all of the marks on her body were minor.

B. Testimony of Police Officer Hamilton

Police officer James Hamilton testified. He has been a police officer for six years. He responded to the 911 call and went to the apartment of John and Emily on November 1. He handwrote the statement for Emily. His practice is to always take down the alleged victim's accusations as accurately as possible. He had never seen John or Emily before that day and has no reason to want to punish John.

Officer Hamilton testified that he gave Emily the written statement to review and sign, but has no way of knowing how carefully she reviewed it. He testified that Emily appeared frightened and also told him that she was frightened. Emily wanted to make sure John was not allowed back into their apartment until he cooled off, and she signed the temporary no-contact request.

After refreshing his memory with his police report, Officer Hamilton testified that Emily's right upper arm had visible red marks and bruising, and her left upper arm had red marks. Emily declined medical treatment and refused to be photographed, so pictures of the marks are not

available. He testified that Emily was very clear that her upper-arm injuries were caused by John grabbing and shaking her, and that she did not consent to any of that.

C. Testimony of Dr. Samuel Wilkins

Dr. Samuel Wilkins is the director of a local shelter for abused persons. He has a master's degree in social work and a Ph.D. in clinical psychology. He testified as an expert witness.

Dr. Wilkins testified that it is common for an abused spouse to recant, or take back, a truthful accusation of domestic violence in order to protect the abuser from criminal prosecution. The reasons for recanting an allegation include the abused spouse's desire to protect the family income and/or preserve the family unit. Often, the abused spouse will recant the allegation with the genuine belief that the abuser has changed and will not repeat the abuse. However, this often leads to what is known as "the cycle of violence."

John's lawyer cross-examined Dr. Wilkins, who admitted that he had never met either John or Emily and had no insight into either of them as individuals. He also admitted that while there are reasons a person might recant, or take back, a true allegation, there are also reasons a person might make a false allegation in the first place.

III. Instructions from the Judge

A. General instructions

You should base your verdict on the evidence. Evidence includes the testimony of witnesses, both on direct and cross-examination, as well as any prior statements of a witness that were read into the record or testified to in court. The weight to be given to any particular piece of evidence is solely up to you. There is no magic way for you to evaluate testimony or a prior statement; instead, you should use your common sense and experience. In everyday life, you determine for yourselves the reliability of things people say to you. You should do the same thing here.

B. Burden of proof instruction

[See Part III.D. for the two different jury instructions that were tested.]