

REASONABLE DOUBT RATCHETING: HOW JURORS ADJUST THE STANDARD OF PROOF TO REACH A DESIRED RESULT

MICHAEL CONKLIN*

ABSTRACT

Research has already been conducted into how jurors alter the reasonable doubt standard based on different legal definitions of the standard provided¹ and perceived harshness of punishment.² This article discusses the findings of research that is the first to analyze how jurors also alter the reasonable doubt standard based on their sympathy toward the defendant—namely, how jurors “ratchet up” reasonable doubt to acquit sympathetic defendants and “ratchet down” reasonable doubt to convict unsympathetic defendants. Political affiliation and gender are also analyzed for potential causal relationships with this effect. The results of this novel research provide valuable information for trial attorneys, judges, and advocates of criminal justice reform.

*†Powell Endowed Professor of Business Law, Angelo State University. I would like to give special thanks to my mentor, Chad Marzen.

1. Mandeep K. Dhimi, Samantha Lundrigan & Katrin Mueller-Johnson, *Instructions on Reasonable Doubt: Defining the Standard of Proof and the Juror’s Task*, 21 PSYCHOL. PUB. POL’Y, & L. 169, 170-72 (2015).

2. See generally Norbert L. Kerr, *Severity of Prescribed Penalty and Mock Jurors’ Verdicts*, 36 J. PERSONALITY & SOC. PSYCHOL. 1431 (1978).

I.	INTRODUCTION	283
II.	METHODOLOGY	284
III.	RESULTS	285
	A. OVERALL	285
	B. GENDER	286
	C. POLITICAL AFFILIATION	287
	D. AMORPHOUS NATURE OF REASONABLE DOUBT	287
IV.	POTENTIAL CRITICISM	288
V.	FUTURE RESEARCH	289
VI.	CONCLUSION	290

I. INTRODUCTION

“Beyond a reasonable doubt” is a legal standard even many nonlawyers are familiar with. However, courts struggle to define exactly what it means. The U.S. Supreme Court has never required that jurors be provided a definition of reasonable doubt, but trial courts are allowed to do so as long as “taken as a whole, the instructions correctly convey the concept of reasonable doubt to the jury.”³

The U.S. Supreme Court has upheld trial court convictions where reasonable doubt was defined as “actual and substantial doubt.”⁴ It has allowed trial courts to warn jurors not to go too far in applying the standard because everything “is open to some possible or imaginary doubt.”⁵ The U.S. Supreme Court has struck down attempts to define the reasonable doubt standard too low, for example, “such doubt as would give rise to a grave uncertainty.”⁶ Furthermore, due to the “inherently qualitative” nature of the standard, efforts to affix a quantified percentage to reasonable doubt have been struck down.⁷ Some courts have maintained that the practice of trying to define reasonable doubt is per se counterproductive because the standard is “self-defining, that there is no equivalent phrase more easily understood . . . that the better practice is not to attempt the definition, and that any effort at further elucidation tends to misleading refinements.”⁸ Reading jury instructions that attempt to explain reasonable doubt provides strong evidence to support this claim. The following are examples:

- “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.”⁹
- “If two conclusions can reasonably be drawn from the evidence, one of innocence and one of guilt, you must adopt the one of innocence.”¹⁰

3. *Holland v. United States*, 348 U.S. 121, 140 (1954).

4. *Victor v. Nebraska*, 511 U.S. 1, 20-21 (1994).

5. *Id.* at 7.

6. *Cage v. Louisiana*, 498 U.S. 39, 40 (1990).

7. *See generally* *McCullough v. State*, 657 P.2d 1157 (Nev. 1983) (holding it was reversible error for a judge to refer to a zero to ten scale where reasonable doubt was described as “seven and a half, if you had to put it on a scale”).

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

9. PA. SUGGESTED STANDARD CRIM. JURY INSTRUCTIONS NO. 7.01 (2016). This standard is up for vastly different interpretations since many people hesitate before *all* matters of importance in their lives. Furthermore, it is a misleading analogy because most people do not use the reasonable doubt standard when making important decisions in their lives. Michael D. Cicchini, *Instructing Jurors on Reasonable Doubt: It’s All Relative*, 8 CAL. L. REV. ONLINE 72, 75 (2017).

10. *State v. Griffin*, 749 A.2d 1192, 1197 n. 12 (Conn. 2000).

- “Proof beyond a reasonable doubt is proof that leaves you with an abiding conviction that the charge is true.”¹¹

This is perhaps why states like Michigan define reasonable doubt in a rather circular nature, “A reasonable doubt is just that—a doubt that is reasonable”¹² Existing research has demonstrated that variations in juror instructions can result in jurors applying different meanings as to what exactly beyond a reasonable doubt requires.¹³ Research also demonstrates that jurors will adjust the reasonable doubt standard based on perceptions of the punishment’s fairness.¹⁴

II. METHODOLOGY

The survey conducted for this article was administered to 98 undergraduate and graduate students (hereinafter “participants”) in the fall of 2019. Three different versions of the survey were utilized. Each version contained a brief summary of a hypothetical criminal case. Therefore, each participant was presented with only one of the three case summaries. In all three versions the participant was then asked the following three questions:

1. Using only the above information, would you convict the defendant based on the criminal standard of being “beyond a reasonable doubt” as to his guilt?¹⁵
2. How certain are you as to the defendant’s guilt? [0–10 Likert scale is provided]
3. If you had to state the least amount of certainty required to be “beyond a reasonable doubt” as to someone’s guilt, what would you say? Put a number between 0%–100%.

The only difference between the three versions of the survey was the case summary. In order to best measure how the participant’s sympathies for the defendant affect whether they increased or decreased the reasonable doubt standard, case studies involving sympathetic and non-sympathetic defendants were utilized. The first case summary involved the alleged assault of an elderly woman based on a dare. The second and third case summaries involved a gas station clerk who allegedly sold alcohol to someone under the age of twenty-one. This defendant was a college student who, on short notice, was required to work an overnight shift during finals week. The person who

11. CALCRIM No. 220 (2017).

12. MICH. CRIM. JI 1.9(3).

13. See Dhami, *supra* note 1, at 169.

14. See Kerr, *supra* note 2, at 1432.

15. Intentionally, no attempt was made to define the reasonable doubt standard.

allegedly purchased the alcohol was twenty years old, and it was just one can of beer. The second and third versions of the summary both used this case summary. The only difference was that in one version the defendant's name was Lisa and all pronouns were changed to feminine and in the other version the defendant's name was Tom and all pronouns were changed to masculine. This was done to allow for the additional assessment of how male and female participants may alter reasonable doubt differently due to the gender of the defendant.

It was hypothesized that participants would adjust the reasonable doubt requirement to reach their desired verdict. This would mean lowering the standard to convict the unsympathetic defendant accused of assaulting an elderly woman and increasing the standard to acquit the sympathetic clerk. It was also hypothesized that—due to increased feelings of camaraderie from a shared gender—male participants would raise the reasonable doubt standard for the male clerk and females would do likewise for the female clerk. Finally, it was hypothesized that conservative participants would implement a lower reasonable doubt standard than liberal participants.

III. RESULTS

This section will cover the highlights of the survey results. This includes the overall results of how the reasonable doubt standard was either ratcheted up or down based on the prompt provided, and how participant demographics such as gender and political affiliation affected results. Finally, the troubling nature of how disparate the results were is discussed.

A. OVERALL

Survey participants answered consistently with the first hypothesis. They increased the burden of proof required for reasonable doubt in the sympathetic clerk cases and decreased the standard for the unsympathetic elderly woman assailant. The average percentage quantification of reasonable doubt required to reach a guilty verdict in the two clerk cases was 80.2%. The average in the elderly woman assault case was only 61.9%. As predicted, when participants were confronted with the disconcerting facts of a man who allegedly assaulted an elderly woman for no other reason than because he was dared to, they likely felt little sympathy. It is also unlikely that many participants could relate to this defendant. Furthermore, participants probably wanted to err on the side of caution by convicting such a person, thereby reducing the risk of similar, future occurrences.¹⁶

16. Whether through temporarily incapacitating the defendant with a prison sentence, sending a message to the public in general that this behavior is not tolerated, or some combination thereof.

Conversely, the participants utilized in this research likely felt sympathy for the fellow college student who was called in to work on short notice during finals week. The case summary described how the situation “resulted in Lisa/Tom being very tired and stressed out during her/his shift. Lisa/Tom tried to study during the shift, but the last-minute nature meant she/he was unable to be productive studying.” This likely caused participants to commiserate with the defendant, recalling times in their lives when they experienced similar difficulties studying.

B. GENDER

The hypothesis that participants would alter the reasonable doubt standard due to increased sympathy for defendants who share their same gender was not supported by the results. Surprisingly, the opposite result occurred. Male participants had, on average, a higher reasonable doubt standard than female participants for the female clerk. And female participants had a higher reasonable doubt standard than male participants for the male clerk. While it is safe to say that the hypothesis of participant/defendant gender similarity resulting in increased sympathy did not occur, any further conclusion from this subset of the data is likely unwarranted as the differences were minor.¹⁷

Although the two versions of the gas station case were originally implemented only to measure differences based on participant gender, this methodology uncovered a peculiar, ancillary finding. Disregarding the gender of the participant, the male clerk received highly favorable treatment when compared to the female clerk. Despite receiving higher predicted levels of guilt, the male clerk was convicted at a lower rate.¹⁸ This is consistent with the significantly higher interpretation of the reasonable doubt standard the male clerk received. Participants who read about the female clerk required only 73.6% certainty to convict while those who read about the male clerk required 86.8% certainty.

The reason for this preferable treatment received by the male defendant is difficult to explain. Common sense would suggest that a female being made to work an overnight shift at a gas station would receive more sympathy than a similarly situated male.¹⁹ One possible explanation is found in the

17. With the male clerk, female participants averaged 91.8%, while male participants averaged 87.8%. With the female clerk, male participants averaged 74.0% while female participants averaged 72.1%.

18. The male clerk was convicted 55.6% of the time while the female clerk was convicted 60% of the time.

19. Although the case summary did not explicitly point out, the most likely interpretation is that the gas station clerk was working the overnight shift alone.

gender differences from perceived complaining.²⁰ In order to increase sympathy for the gas station clerk defendant, the case summary included language regarding the hardship endured by the defendant.²¹ Perhaps this language was interpreted by participants as the voicing of a legitimate concern by the male clerk and perceived as trivial complaining from the female clerk.

C. POLITICAL AFFILIATION

As predicted, liberal participants demonstrated higher standards for reasonable doubt (81.7%) than conservative participants (75.2%). This disparity is likely due to different views of the criminal justice system generally.²² Conservatives, who favor more “law and order” policies, are more likely to interpret the reasonable doubt standard in a way that allows for more convictions. Conversely, liberals are more likely to interpret the standard in a way that allows for more acquittals. This is consistent with the general principle that prosecutors prefer conservative jurors, while defendants prefer liberal jurors.²³

D. AMORPHOUS NATURE OF REASONABLE DOUBT

Although not the primary focus of this research, it is interesting to note that this study confirms the problem of the inherently amorphous nature of the reasonable doubt standard. When asked to affix a minimum percentage to the reasonable doubt standard, 14.5% of participants reported a number of 50% or less. Courts have struck down efforts to affix a percentage to the beyond a reasonable doubt standard.²⁴ Nevertheless, reasonable doubt at least requires the defendant to be more likely guilty than not—in other words, more than 50% likely to be guilty.²⁵

20. See generally Joanna Wolfe & Elizabeth Powell, *Gender and Expressions of Dissatisfaction: A Study of Complaining in Mixed-Gendered Student Work Groups*, 29 *WOMEN & LANGUAGE* 13 (2006).

21. The case study explained, “This resulted in Lisa/Tom being very tired and stressed out during her/his shift. Lisa/Tom tried to study during the shift, but the last-minute nature meant she/he was unable to be productive studying.”

22. Justin McCarthy, *Americans Divided on Priorities for Criminal Justice System*, GALLUP (Oct. 14, 2016), <https://news.gallup.com/poll/196394/americans-divided-priorities-criminal-justice-system.aspx> (finding that 77% of Republicans and 32% of Democrats agree that strengthening law and order is a bigger priority than reducing bias against minorities.); *The Public, the Political System and American Democracy*, PEW RES. CTR. (Apr. 26, 2018), <https://www.people-press.org/2018/04/26/9-the-responsibilities-of-citizenship/> (finding that 79% of Republicans and only 61% of Democrats agree that “good citizens [should] always follow the law”).

23. Barry P. Goode, *Religion, Politics, Race, and Ethnicity: The Range and Limits of Voir Dire*, 92 KY. L.J. 601, 627 (2004).

24. See, e.g., *McCullough v. State*, 657 P.2d 1157, 1159 (Nev. 1983).

25. See *Cunningham v. California*, 549 U.S. 270, 281 (2007) (noting that beyond a reasonable doubt is a more stringent standard than preponderance of the evidence).

To further complicate matters, “Empirical research indicates that jurors may have some difficulty distinguishing among the standards [reasonable doubt, preponderance of evidence, clear and convincing evidence].”²⁶ A group of studies found that participants were no more likely to find for the defendant whether the standard given was beyond a reasonable doubt or clear and convincing evidence.²⁷ Remarkably, one study found that mock jurors returned the same verdict regardless of whether the standard given was beyond a reasonable doubt or preponderance of the evidence.²⁸ These results lead some scholars to conclude that jurors have “difficulty distinguishing among the standards.”²⁹ An alternative explanation is that jurors simply mend whatever standard they are given to arrive at their desired outcome.

Not all scholars are troubled by the wide variations in how jurors interpret the reasonable doubt standard. Some even view this variation as a beneficial aspect of the criminal law system. They argue that flexibility in interpreting the standard allows jurors to customize the standard given the circumstances of the case.³⁰ By using the inherently subjective term “reasonable” in the standard, it does seem designed to contain a certain degree of flexibility. It is not hard to see why a juror may find an 80% likelihood of guilt to be beyond reasonable doubt in a trial on a minor crime but require 90% in a capital murder case.

IV. POTENTIAL CRITICISM

A potential criticism of this study is that the case summary is overly simplistic; real-life cases contain far more variables. While true, there is no reason to believe that adding ancillary information—and therefore unnecessarily complicating the case—would somehow enhance the ability of jurors to apply a more uniform reasonable doubt standard.

Relatedly, one might attempt to criticize this study by pointing out that real cases are resolved by a consensus of deliberating juries, not by averaging individual responses done in private. Again, this criticism points to a legitimate difference, but its relevance to this study is unfounded. Research into individual versus group deliberation “suggests that juries almost always select that verdict favored by a sizable majority of the members at the outset of

26. Elisabeth Stoffelmayr & Shari Seidman Diamond, *The Conflict Between Precision and Flexibility in Explaining ‘Beyond a Reasonable Doubt’*, 6 PSYCH. PUB. POL. & L. 769, 774-75 (2000).

27. *Id.*

28. *Id.*

29. *Id.*

30. Stoffelmayr & Diamond, *supra* note 26, at 778.

deliberation.”³¹ In the landmark study on how jurors adjust reasonable doubt based on severity of punishment, only individual responses were utilized.³² Mathematical models were implemented to measure how the individual findings would have likely varied if group deliberations were used instead.³³ The results showed that group deliberations would have produced even greater effects on reasonable doubt adjustments.³⁴

Others may attempt to minimize the relevance of this study’s findings by claiming that the differences were only minor. This objection demonstrates a lack of understanding regarding the relevance of variations in how reasonable doubt is applied. The difference between a 75.2% and an 81.7% reasonable doubt standard results in only a 6.5% difference in the likelihood of an acquittal—although most trial attorneys would likely view that as enough to be considered “significant.” Because easy cases are more likely to be settled without a trial, this leaves most of the cases that go to trial in a somewhat narrow range of conviction probability.³⁵ This narrow range results in even slight changes in the reasonable doubt standard producing drastic differences in conviction probability.

The following example will help illustrate this principle. Assume that the majority of cases that go to trial have a defendant who is 70% to 95% likely to be guilty.³⁶ Therefore, a 75.2% burden of proof would result in convictions in 85.1% of these cases.³⁷ An 81.7% burden of proof, however, would result in a conviction in only 38% of these cases.³⁸ The difference between an 85.1% chance of conviction and a 38% chance of conviction is highly significant.

V. FUTURE RESEARCH

The results of this novel research invite further investigation and discussion. A qualitative version of this quantitative study could address whether the ratcheting phenomenon is a conscious decision by the juror or whether it is subconscious. If conscious, potential motives could be analyzed. And

31. Kerr, *supra* note 2, at 1440.

32. *Id.*

33. *See id.*

34. *Id.*

35. Cases are frequently dropped by prosecutors against defendants with a low probability of conviction, and defendants with a near certainty of conviction at trial frequently plead out rather than go to trial.

36. The percentages utilized in this illustration are purely speculative since it is impossible to ultimately know which defendants are innocent.

37. This is because 85.1% of the cases between 70% and 95% are greater than the 75.2% standard.

38. This is because only 38% of the cases between 70% and 95% are greater than the 81.7% standard.

proposed jury instructions could be tested for their ability to create a more uniform interpretation of the reasonable doubt standard.

No studies have been conducted using judges as participants to measure the type of reasonable doubt ratcheting analyzed in this research. It is likely that judges engage in the same behavior as the participants in this study, just to a lesser extent. This is based on a related study that measured how judges and jurors adjust the reasonable doubt standard based on the severity of punishment.³⁹ It found that trial court judges adjust upward their reasonable doubt standard in hypothetical severe cases and adjust downward in hypothetical minor cases.⁴⁰ The study found that the variance in adjusting up and down was less pronounced among judges than jury-eligible citizens.⁴¹ Research into the behavior of judges adjusting reasonable doubt as tested in this novel research would prove valuable to defense attorneys assessing whether a bench or jury trial would be ideal.

The burden of proof in civil cases—preponderance of the evidence—is a quantifiable standard. Therefore, it is easier for a juror to understand what the standard is, for there to be uniform agreement on this issue, and for the jurors to be able to determine if the standard is met. Unlike “beyond a reasonable doubt,” jurors are more likely to have experience applying a “preponderance of the evidence” standard in their everyday life.⁴² Future research could be conducted into whether the ratcheting phenomena in criminal cases also applies to civil cases—and if not, what lessons can be learned from the civil court standard that might be applicable to criminal cases, thereby creating more uniformity.

VI. CONCLUSION

The beyond a reasonable doubt standard is an amorphous maxim. The research discussed in this article shows the interplay between how jurors perceive the standard and their sympathies toward the defendant. However, this interplay is not universal among jurors. Background factors such as political affiliation and gender produce diverse reasonable doubt ratcheting effects. This research will allow judges and attorneys to better understand the juror thought process and why jurors ultimately reach the verdicts that they do.

39. See generally Rita James Simon & Linda Mahan, *Quantifying Burdens of Proof: A View from the Bench, the Jury, and the Classroom*, 5 L. & SOC'Y REV. 319 (1971).

40. See *id.* at 328.

41. *Id.*

42. For example, if someone is trying to determine whether he is supposed to meet a friend at one of two locations, he simply determines which location is more likely than not to be the correct one and travels there. This simple, more-likely-than-not determination is the preponderance of evidence standard.