Distributions of Interest for Quantifying Reasonable Doubt and Their Applications *

Katie Evans[†]
Truman State University

David Osthus[‡]
Luther College

Ryan G. Spurrier[§]
Clemson University

Abstract

The concept of reasonable doubt is a standard of our legal system; however, it is a standard that is not well defined. Differences in the way reasonable doubt is applied in different courts and states, as well as ambiguities in its different definitions, suggest that the standard puts pressure on due process and equal protection concerns. This paper explores probability distributions that will aid in the understanding of the American legal system as it is today, what reasonable doubt means under this system, and how reasonable doubt should be defined.

*This research was performed as part of Valparaiso Experience for Research by Undergraduates in Mathematics, 2006, with funding obtained from the National Science Foundation (#DMS-0353870). The original idea for this research came from a conversation roughly twenty years ago between Dr. James Caristi, Professor of Mathematics, and Bruce Berner, Professor of Law, in the Valparaiso University recreation center following their lunch break raquetball games. A special thanks goes out to these professors for their advice and assistance throughout the research.

†k80@truman.edu
‡osthda01@luther.edu
§rspurri@clemson.edu

1 Introduction

Reasonable doubt is a standard of the American legal system that has become a part of the common culture of the United States. In most criminal trials, to be convicted, the defendant must be found guilty beyond a reasonable doubt. However, what reasonable doubt truly means is not general knowledge. Experiments by Simon and Mahan[1] show that there is a large variance in the way jurors interpret the phrase *beyond a reasonable doubt.* This shows that there is little consensus among the American people. It has been suggested that since it is the people who define the standard of reasonable doubt through jury proceedings, it should be the people who establish the standard. The lack of consensus, however, shows that a clear standard does not exist among the people.

In addition to different jurors interpreting reasonable doubt differently, different jurisdictions also define reasonable doubt differently. Even though all jurisdictions inform jurors to convict only if they are convinced of guilt beyond a reasonable doubt, the instructions to the jurors as to how reasonable doubt is defined differ from one jurisdiction to the next. In Victor v. Nebraska, the U.S. Supreme Court upheld two different standards of reasonable doubt[1] that, "read as a whole, satisfy the Constitution's due process requirement."[2] However, the court admitted that some of the language used in these definitions is ambiguous and unclear to the average juror. Much of the debate surrounded the phrase *moral certainty*, a phrase that has fallen out of common usage since it was referred to by Chief Justice Shaw of the Massachusetts Supreme Judicial Court in the 1850 decision, Commonwealth v. Webster[3]. Due to the ambiguities, jurors respond to these definitions in different ways. Thus, the outcome of a trial is dependent upon both the evidence presented in the trial and how the given jury interprets the standard. This stresses both due process and equal protection. Although this stress is not enough to demand a change in the system, a more precise standard would be beneficial. Therefore, a quantified standard is suggested. This standard will be a percentage such that if a juror feels the likelihood the defendant committed the crime is greater than the percentage, the juror should vote to convict; otherwise, the juror should vote to acquit. Critics of the numerical system argue that it would be difficult for jurors to quantify their perceptions of guilt; however, techniques, such as the probability wheel, have been developed for this purpose[2].

[1]The entirety of these definitions can be found in Appendix I.

[2]The probability wheel is a technique to quantify perceptions of likelihood. The technique presents the subject with two hypothetical lotteries from which the subject chooses the preferred lottery. The subject will win the first lottery if his or her belief is correct. In the case of likelihood of guilt, the subject will win the lottery if the defendant committed the crime. The second lottery is based strictly on probabilities, i.e. there is a 50% likelihood that the subject will win the lottery. This percentage is sometimes represented by a fair spinner, i.e. a spinner half blue and half red in which the subject will win if the spinner lands on blue, thus the term probability wheel. In our case, if the first lottery is chosen, then the subject believes the defendant is more likely to have committed the crime than the percentage given in the second lottery. The percentage in the second lottery is then increased until the subject is indifferent as to which lottery to accept. If the second lotterly is selected, the subject believes the defendant is less likely to have committed the crime than the percentage given in the

Many defense attornies feel that the standard, in practice, is too low, making it too easy to convict the innocent. A quantified standard could help protect against this. One way to analyze false conviction errors, and get a feel for the status of the justice system, is by analyzing the results of the innocence projects in which DNA evidence has been used to reevaluate the convictions in trials involving the death penalty. From these studies, the false conviction rate of death penalty cases could be determined. However, it would be difficult to extrapolate the results from death penalty cases to all cases. It is conceivable that error rates are lower for these cases because death penalty cases are high profile. Thus, both the defense and prosecuting attorneys are generally very skilled and extensive resources are exhausted to ensure that as complete a collection of evidence is made as possible. It is also conceivable that jurors may subconsciously lower their standards of reasonable doubt in these cases so that someone is held accountable, even if not the right person, for the more severe crimes against society. Due to these possibilities, extrapolation should not be made from the innocence projects to all cases.

The first attempt to quantify reasonable doubt was made by Simon[6] in 1970. In the attempt, she presented a trial to groups of students. Half of the students decided the guilt or innocence of the defendant[3]. The other half recorded their perceived likelihood, given as a percentage, that the defendant committed the crime. She then matched the highest likelihoods of guilt with the guilty verdicts and the lowest likelihoods of guilt with the innocent verdicts. From this, she gauged that the cutoff for reasonable doubt fell somewhere between the highest likelihood of guilt matched to an innocent verdict and the lowest likelihood of guilt matched to a guilty verdict. From these samples, Simon concluded that the standard was between 0.70 and 0.74. However, this experiment is based on the assumption that the two halves of the experiment were identical in their values and judgments concerning reasonable doubt. Also, it measures the way people define the standard instead of how the law defines the standard.

Many studies have used a utility approach in an attempt to quantify what the standard of reasonable doubt should be. One such study was performed by Tribe[4] [4] in 1971. In the utility approach, negative values are assigned to the utilities of convicting the innocent and acquitting the guilty. Values of zero or positive values are assigned to the utilities of convicting the guilty and acquitting the innocent. The standard is then chosen to maximize utility. Though this approach appears sound, it has its problems. For one, this process requires four subjective judgments to define the values of the four utilities–a large degree of

second lottery. The percentage will then be lowered until the subject is indifferent between the two lotteries. When indifference occurs, the percentage given by the second lottery is the perceived likelihood of guilt. For more information about probability wheels see Saunders[4] or Winkler[5].

[3] In this paper, guilt refers to factual guilt; innocence refers to factual innocence. Process guilt and innocence are referred to as conviction and acquittal, respectively. Likelihood of guilt refers to the probability of factual guilt.

[4] The full text of this study can be found in "Trial by Mathematics: Precision and Ritual in the Legal Process" by Laurence H. Tribe.[7]

subjectivity. Also, these utilities are based on judgments of the severity of the crime and length of punishment which, if made by the jurors, violate the legal system. Finally, the utility approach suggests different standards for different crimes. The law, however, demands one uniform standard.

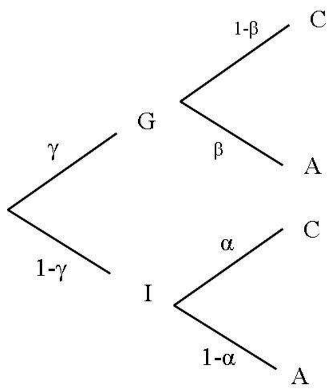
In light of the deficiencies in previous attempts to understand reasonable doubt, we believe a new method is needed. This new approach takes up the issue of reasonable doubt not from the quantified standard itself but from the false conviction and false acquittal rates on which the standard depends. Five probability distributions can be used to understand the relationship between these rates and the standard. They are as follows: the distribution of guilt and innocence, the distribution of likelihood of guilt after trial, the distribution of perceived standard of reasonable doubt[5], the distribution of perceived guilt given perceived standard of reasonable doubt, and the distribution of perceived guilt given likelihood of guilt after trial. The remainder of this paper will focus on these distributions: how they could be found or approximated and what they will disclose concerning the standard of reasonable doubt.

2 Distributions of Interest

2.1 Distribution of Guilt and Innocence

The distribution of guilt and innocence is a binomial distribution in which 0 represents innocence and 1 represents guilt. This distribution represents the proportion of defendants brought to trial who actually committed the crime for which they are accused (shown as γ in Figure 1 below).

Figure 1: Conditional Probability Tree



[5]A distinction needs to be drawn between perceived standard of reasonable doubt and perceived guilt. The perceived standard of reasonable doubt is how sure of guilt a juror must feel to convict beyond a reasonable dobut. Perceived guilt is how sure of guilt the juror actually feels. If a juror feels that the definition of reasonable doubt implies that 80% certainty is required to convict, then 80% is the juror's perceived standard of reasonable doubt. In a given case, if the juror feels that it is 90% likely that the defendant committed the crime, then the juror's level of perceived guilt is 90%.

Because the conviction rate, and consequently the acquittal rate, is known, if γ is also known, the levels of α (probability of conviction given the defendant is innocent) and β (probability of acquittal given the defendant is guilty) can be found using the following sysem of equations. This will show the error rates of the current legal system.

$$P(C) = (1-\gamma)(\alpha) + (\gamma)(1-\beta) \tag{1}$$

$$P(A) = (1-\gamma)(1-\alpha) + (\gamma)(\beta) \tag{2}$$

P(C) is the conviction rate; P(A) is the acquittal rate. If the error rates, α and β, are in ranges acceptable under the law, there may not be a need for a numerical standard. These error rates could also be used to determine the standard of reasonable doubt in practice as explained in Section 2.2. By knowing γ, we would also know how the conviction rate compares with the percentage of defendants who are guilty (the guilty rate). It may be that α and β errors occur in such a way that they average out so that the conviction rate and guilty rate are similar. Juries, however, could be making systematic errors. These would be noted by a large difference between the conviction rate and guilty rate. Systematic errors in the current system would argue for the addition of a quantified standard to clarify the definition of reasonable doubt.

To calculate the distribution of guilt and innocence exactly, one would have to know the guilt or innocence of each defendant. It is not likely that this could be done by sampling the defendants themselves, as the guilty defendants have an incentive to claim innocence. However, the defendants are often the only persons who truly know whether they committed the crime. Family members or close friends of the defendants might know; however, they would have the same incentive to state innocence in every case. Defense attorneys would not have an incentive to report innocence as opposed to guilt if they did not identify the cases in question. Defense attorneys would also have the best insight into the actions of the defendants due to their relationships with the defendants. However, many defense attorneys do not want to know the true guilt or innocence of the defendant. If the defense attorney knows the defendant is guilty and the defendant testifies, the defense attorney cannot question the defendant because the lawyer knows the defendant will commit perjury. The defendant's testimony is then given as a narrative, and the defense attorney cannot argue in the closing statement concerning the testimony of the defendant. Despite these concerns, the defense attorney will still have the best understanding, after the trial, of the defendant's innocence or guilt. Therefore, it would be possible to sample defense attorneys to get an approximation of the distribution of guilt and innocence. It would be necessary to sample these lawyers anonymously, without reference to specific cases, to protect the defendants and avoid the ethical issues concerning the violation of legal privilege. One possible survey would be as follows: for each of the last ten cases defended, each defense attorney would check a box concerning whether they felt the defendant was factually guilty or innocent. The decisions could be recorded in any order. These responses would not bear the names of the lawyers, so no connection could be made between the lawyer's response and the cases in question.

Even if the lawyers were matched to their cards, their responses could not be matched to individual cases unless all ten of a lawyer's judgments were of the same outcome. The separation between cases and judgments could be improved by having lawyers give their judgment on ten randomly selected cases. However, lawyers would be less likely to respond under these lengthened procedures to ensure random selection. Another way to completely separate the cases and judgments would be for the lawyers to report the number of cases they have tried and the percentage of the cases they judged to be guilty. However, it is doubtful that a lawyer with an extensive career would evaluate each case and calculate the true percentage. Approximations would introduce error into the distribution.

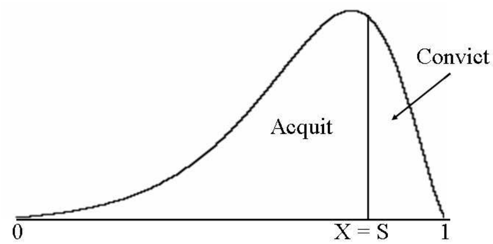
The cooperation of a large public defenders office, the National Association of Criminal Defense Lawyers (NACDL), or a similar institution would help circumvent the privilege concerns while avoiding vague percentages. By teaming up with one or more of these institutions, the integrity of the sample would be aided by the experience of those sampled. This experience would aid in the interpretation of a defendant's guilt or innocence. In addition, the defenders office, or the NACDL, could report the overall number of innocent and guilty defendants on behalf of individual lawyers instead of the lawyers submitting their results independently. This would result in an additional layer of separation between the data and the individual cases. It is likely that such institutions would agree to aid in this survey as many defense attorneys believe that too many innocent defendants are convicted. Working with a nationwide association of defense attornies would also facilitate a national survey. Thus, instead of determining the error rates for merely a county or state, the error rates for the entire judicial system could be found. Because it is suspected that the percentage of guilty defendants is different for public and private defenders, the percentage of the survey given to public and private defenders should equal the percentage of cases argued by each type.

The argument could be put forth that defense attorneys will bias their judgment of the cases based on their incentive to win the case. Therefore, a neutral party, such as the judge or jury should be asked to give their judgment of the guilt or innocence of a client. These results would be different from the conviction rate because a juror could feel that the defendant was guilty but that it was not shown beyond a reasonable doubt. However, it is our opinion that the error introduced by the bias of the defense attorney would be overshadowed by the attorney's additional knowledge. The accuracy of the distribution of guilt and innocence is critical to any inference made from it. Thus, any approximation must be a close approximation in order to draw valid conclusions.

2.2 Distribution of Likelihood of Guilt after Trial

Although each defendant's probability of guilt is either 0, if innocent, or 1, if guilty, the evidence given in the trial changes the likelihood of guilt that should be perceived. In essentially every trial in which the defendant is guilty, there is some evidence presented in favor of innocence. Therefore, it should not be perceived that the defendant is guilty with a probability of 100%. Similarly, if the defendant is innocent, the guilt after the trial should not be perceived as 0%. Thus, there is a likelihood of guilt after trial for each case that ranges from 0 to 1[6]. The compilation of these likelihoods is the distribution of likelihood of guilt after trial. On this distribution (p(x)), the standard of reasonable doubt (S) could be superimposed. For all trials such that the likelihood of guilt after trial is greater than the standard, jurors should vote to convict. For all trials such that the likelihood of guilt is below or equal to the standard, jurors should vote to acquit.

Figure 2: Distribution of Likelihood of Guilt after Trial (p(x))
Note: Distribution not to Scale



The percentage of trials that end in acquittal (P(A)) and the percentage of trials that end in conviction (P(C)) are:[7]

$$P(A) = \int_0^S p(x)\,dx \tag{3}$$

$$P(C) = \int_S^1 p(x)\,dx \tag{4}$$

The expected value of this distribution is the expected likelihood of guilt for all trials. In the long run, this is the percentage of defendants who are guilty (P(G)).

$$P(G) = \int_0^1 xp(x)\,dx \tag{5}$$

Because innocence is the complement of guilt, the long run percentage of defendants who are innocent (P(I)) is:[8]

$$P(I) = \int_0^1 (1-x)p(x)\,dx \tag{6}$$

[6] If there are 100 cases whose trials suggest the same likelihood of guilt, and of these 100 cases, 90 of the defendants committed the crime, then the likelihood of guilt after trial for each case is 90%.

[7] Hung juries, which occur at a negligible rate, are excluded from p(x)

[8] The integrals for percentage of guilt and innocence can also be thought of as follows: for each value of the random variable X, the likelihood of guilt after trial, the proportion of the defendants at that level x that are guilty, is x (i.e. of all defendants in cases for which the

The joint probabilities are the integrals for guilt or innocence restricted to the region of interest of the curve. They are:

$$P(C \cap G) = \int_S^1 xp(x)\,dx \tag{7}$$

$$P(A \cap G) = \int_0^S xp(x)\,dx \tag{8}$$

$$P(C \cap I) = \int_S^1 (1-x)p(x)\,dx \tag{9}$$

$$P(A \cap I) = \int_0^S (1-x)p(x)\,dx \tag{10}$$

From these integrals, the conditional probabilities α, $(1-\alpha)$, β, and $(1-\beta)$ are:

$$\alpha = P(C|I) = \frac{P(C \cap I)}{P(I)} = \frac{\int_S^1 (1-x)p(x)\,dx}{\int_0^1 (1-x)p(x)\,dx} \tag{11}$$

$$(1-\alpha) = P(A|I) = \frac{P(A \cap I)}{P(I)} = \frac{\int_0^S (1-x)p(x)\,dx}{\int_0^1 (1-x)p(x)\,dx} \tag{12}$$

$$\beta = P(A|G) = \frac{P(A \cap G)}{P(G)} = \frac{\int_0^S xp(x)\,dx}{\int_0^1 xp(x)\,dx} \tag{13}$$

$$(1-\beta) = (C|G) = \frac{P(C \cap G)}{P(G)} = \frac{\int_S^1 xp(x)\,dx}{\int_0^1 xp(x)\,dx} \tag{14}$$

Therefore, if the judicial system decided upon a maximum value for either the α or β error[9], or a ratio of α and β errors, then the above integrals could be solved for S[10]. Also, once the standard is found, the other conditional probabilities could be evaluated. If these values are acceptable to the judicial system,

---

likelihood of guilt after trial is 0.75, 75% actually committed the crime). Thus the integral of $xp(x)$ across the domain of X gives the overall percentage of guilt. At the given value x, the proportion of the defendants that are innocent is $(1-x)$, and the overall percentage of innocence is the integral of $(1-x)p(x)$ over the domain of X.

[9] Ideally, α and β would be 0 if the legal system were perfect; however, juries are required to produce a decision (excluding hung juries which occur at a negligible rate). Juries cannot say, "We don't know."[8] If they are unsure, they must make their best guess, largely contributing to the α and β errors.

[10] The first approach we explored was to think of the problem in terms of odds of guilt. There is a long run probability of guilt for all defendants brought to trial. This is given by the proportion of defendants who are brought to trial who actually committed the crime. The American legal system requires a higher likelihood of guilt to convict than simply being brought to trial. Thus in order to convict, a jury must be significantly more sure of the guilt in an individual case than the long run probability of guilt. Unfortunately, procedures for significant difference in proportions or odds do not exist when the final odds or proportion is based on only one case. The same problem is present when trying to extrapolate from a minimum level of doubt to a significant level of doubt which would be defined as a reasonable amount of doubt.

then the standard is acceptable. If the judicial profession set maximum levels for both α and β, each could be used to solve for a standard. Each standard could then be used to find the level of the other error. If only one standard sufficiently protected against both types of error, then it would be the standard. If both standards protected against both types of errors, then the standard that had a lower α value would be selected to better protect the innocent.

It is important to note that under the American legal system, the maximum value of α is expected to be much lower than the maximum value of β, as false convictions are viewed as worse than false acquittals. There are three primary ways for α and β to be decided: by the people directly, by the people through their representation in Congress, or by the judicial system. However, in the interest of protecting the rights of the minority[11], these numbers should not be set by the people directly, but possibly by the people through their representatives in Congress.[9] In order to ensure the integrity of the Judicial Branch, however, these values should be set by that branch. The values are still representative of the people through the interactions of the branches of government through the system of checks and balances.

The conditional probabilities given by the integration of the distribution only hold true if juries convict every defendant for whom the likelihood of guilt is above the standard and acquit every defendant whose likelihood is below the standard. If there are systematic differences in the way juries should and do act, then the conditional probabilities predicted by the standard will not be the conditional probabilities given in practice. If jurors' perceived doubt is symmetric about the value of likelihood of guilt after trial and has a small standard deviation about this actual likelihood of guilt, then the above mentioned process would be sound[12].

Also, for the integrals to correctly reflect the legal system, the distribution of likelihood of guilt after trial would need to be correct, or at least a close approximation. Although it would be extremely difficult to find this distribution exactly, it could be estimated in a few ways. First, previous trials could be presented to large samples of potential jurors. The number of jurors who would need to hear each case would depend on the variance of perceived levels of guilt given a trial. After the sample of jurors hears the case, each juror will record their perceived level of guilt. This level of guilt will be quantified using the probability wheel technique or other techniques for quantifying perceptions. These perceived levels of guilt will be averaged to find the mean perceived level of guilt. Because perceived guilt is essentially an error of actual guilt, we assume this mean to be close to the actual level of guilt following trial. This procedure would be repeated for a large number of trials so that the distribution could

[11] The minority in question is those brought to trial. In particular it is the subset of those brought to trial who are innocent of the crime for which they are charged.

[12] This raises a fundamental question: should the justice system be designed to protect the innocent or should it be altered to ensure a standard of protection taking into account the errors made by those involved in the system? If it is the former, the integrals should stand as they are. If it is the latter, then the integrals should be altered to reflect the trends of the juries.

be formed from the means. A way to cheaply obtain this survey would be to enlist the aid of law schools around the country. Each school would be given a trial to be presented to their students to rate the perceived level of guilt. The law schools would be used because both the institutions and the students would have both a greater interest in the results of the experiment and a higher response rate than the average citizen.

The second approach would be for a legal scholar and mathematician, skilled in conditional probability, Bayesian networks, Bayes Theorem, and legal procedures, to reconstruct the trial, one piece of evidence or testimony at a time, to find the likelihood of guilt that the evidence and trial suggest. A panel of these experts could be used instead of a single expert so that each individual's result could be averaged with the results of the other experts to further reduce error. A distribution would then be fit to these likelihoods.

A third approximation of likelihood of guilt after trial would be the distribution of perceived guilt after trial. Although, admittedly not the best representation of actual guilt, it could be possible to achieve this distribution using more computer simulation and less sampling and experimentation. If the studies of perceived guilt given likelihood of guilt after trial, described below, show that the means of these distributions are the likelihoods of guilt after trial, or are close to these levels, then the perceived guilt distribution would be a good approximation of the distribution of likelihood of guilt after trial. Thus, a simulation to find the perceived guilt distribution could be used to produce this distribution of likelihood of guilt. Finding this distribution through simulation as opposed to sampling and experiments could greatly reduce the cost of finding the distribution.

2.3 Distribution of Perceived Standard of Reasonable Doubt

The distribution of perceived standard of reasonable doubt is the distribution of the percentages at which potential jurors quantify reasonable doubt. One way to define the standard of reasonable doubt is by setting the standard as the mean of this distribution. The case could be made that, since it is jurors who decide whether to convict or acquit, it is the jurors who decide the standard of reasonable doubt. Therefore, it is how jurors define reasonable doubt that is of actual importance. The distribution of perceived standard of reasonable doubt can also be used to combine distributions for inference concerning reasonable doubt.

Several studies have been taken to approximate this distribution. One survey, by Simon and Mahan[13] [1], polled potential jurors and asked what they believed the standard of reasonable doubt should be. They were asked to fill in the following phrase: "I would have to believe that it was a ___ out of ten chance that the defendant committed the act [in order to convict beyond a reasonable doubt]." However, the potential jurors could only respond in increments of 0.5,

[13]The entirety of this study can be found in the article "Beyond a reasonable doubt: An experiment attempt at quantification" by Rita James Simon.[10]

or 5 percentage points. This led to a large number of selections of 100% (36% of the sample). A belief of 100% would lead to no convictions as, in the course of law, nothing is proven to 100% certainty. It is assumed that the vast majority of the people who decided on 100% believed the standard should be closer to 100% than 95%, so they checked the box for 100% instead of the box for 95%. This sample also showed a large variance indicating little consensus among the people on which to base a standard.

A second sample, taken by Saunders[4], simply asked subjects to record the percentage at which they felt reasonable doubt is defined. Using this technique, the majority of the responses fell on whole percentages with most of these ending in 5 or 0. However, it is unlikely that there would be a great difference in the number of people who believe the standard should be 95% and the number who believe it should be 95.1%. Thus, it appears that a sample needs to be taken in which subjects quantify beyond reasonable doubt after receiving some training in techniques, such as the probability wheel, that will aid them in precisely quantifying their beliefs. One difference between these two samples is that the Simon and Mahan sample did not present the subjects with a written standard of reasonable doubt; the Saunders sample presented the subjects with the California standard of reasonable doubt. We believe in order to compare quantified and written standards of reasonable doubt, the distribution of perceived standard of reasonable doubt should be formed from samples using a written definition that has been upheld by the U.S. Supreme Court.

Although some believe that the standard of reasonable doubt should be set by the people, we believe that the precedent of our legal tradition and the language used in the Constitution inherently dictate the standard of reasonable doubt. Therefore, it should be those trained in law, lawyers and judges, in particular, the Supreme Court Justices, who should define the standard. However, it would be better to have these legal scholars define maximum values for the α and β errors, so that the standard could be set to ensure these error rates are not violated. If the scholars simply defined a standard of reasonable doubt, they would be inherently defining error rates without knowing if these rates would be acceptable. Therefore, the distribution of perception of reasonable doubt should not be used to compute the standard of reasonable doubt. However, the distribution of potential juror's perception of reasonable doubt is crucial to the simulation described in Section 3.

2.4 Distribution Class of Perceived Guilt Given Perceived Standard of Reasonable Doubt

The distribution of perceived guilt given perceived standard of reasonable doubt is in fact a class of distributions. There is a distribution for each possible value of perceived standard of reasonable doubt from 0 to 1. By investigating these distributions, it will be known how perceptions of guilt vary based on a given juror's perceived standard of reasonable doubt. For example, will a juror who believes the standard should be 50% judge a case in the same way as a juror who believes the standard should be 95%? If there is a substantial difference between the two, it might not be favorable for defense attorneys to strike jurors with low standards of reasonable doubt. The people with lower standards might, in fact, weigh the prosecution's evidence more lightly than jurors with a high standard. It is possible it would be harder to convince those with low standards to convict than those with high standards. The opposite could also be the case.

The distributions would also be useful in another facet. The sum of these distributions, weighted by the distribution of perceived standard of reasonable doubt, would give the distribution of perceived guilt. Thus, these distributions would be vital in a simulation to find the distribution of perceived guilt. In this simulation, simulated jurors would be created and assigned a randomly selected value of perceived standard of reasonable doubt from the distribution of the perceived standard of reasonable doubt. Then, each juror would be assigned a randomly selected value of perceived guilt from their distribution of perceived guilt given perceived standard of reasonable doubt. A distribution would be fit to the perceived values of guilt to estimate the perceived guilt distribution. The average of several such simulated distributions would be an estimate of the distribution of likelihood of guilt after trial.

2.5 Distribution Class of Perceived Guilt Given Likelihood of Guilt after Trial

The distribution of perceived guilt given likelihood of guilt after trial is also a class of distributions. There is a distribution for each level of likelihood of guilt after trial. By investigating these distributions we will know how perceptions of guilt depend upon the likelihood of guilt. For example, if a trial has a likelihood of guilt of 50%, will perceptions of guilt vary about 50% in the same as they would vary about 75% if the likelihood of guilt was 75%? It is suspected that, since perceptions of guilt are errors about likelihood of guilt, these distributions will be symmetric about the likelihood of guilt after trial. However, it is not clear how these distributions will change as the likelihood of guilt approaches either 0 or 1. In both cases, the proximity of the likelihood of guilt to a cutoff of this likelihood will distort the shape of the distribution. If the distributions of this class were known, as well as the distribution of likelihood of guilt after trial, they could be weighted by the distribution of likelihood of guilt after trial to generate the distribution of perceived guilt. However, since the distribution of the likelihood of guilt after trial is not known, the main function of these distributions will be in the simulation that explores whether a quantified standard of reasonable doubt will better ensure due process and equal protection. This simulation is described in Section 3. Also, if the distribution of perceived guilt was known, as well as these distributions of perceived guilt given likelihood of guilt after trial, then it could be possible to deduce the distribution of the likelihood of guilt after trial, as it is a linear combination of these distributions.

3 Proposed Simulation to Test the Necessity of a Quantified Standard

The purpose of this proposed simulation is to explore the claim that the use of a quantified standard of reasonable doubt will improve the legal system of the United States. As noted in the introduction, there is concern that the current system stresses due process and equal protection. However, to implement a quantified standard, it must be shown that a new standard will better protect the rights of the defendants. One study of the effects of a quantified standard was performed by Kagehiro and Stanton[11]. In their study, Kagehiro and Stanton found that quantified standards are better at describing the differences among preponderance of evidence, clear and convincing evidence, and beyond a reasonable doubt than the written standards established by both the California Committee on Standard Jury Instructions, Criminal, 1979 and the Texas Supreme Court in State v. Addington (1979). However, they also showed that not all standards are equally clear in their meanings and that the California and Addington standards were the least clear of those tested. Thus, more testing is required to fully understand the relationship between a quantified standard and a written standard.

A simulation to test the necessity of a quantified standard could be performed by creating simulated jurors and inserting them into simulated juries. The juries would be composed of 12 members because, although six member juries are legal in some parts of the country, the overwhelming majority of criminal cases involve a 12 member jury. Each juror would be given two attributes: a perceived standard of reasonable doubt and a perceived level of guilt. The simulation would be run under two methods. In the first, the perceived standard of reasonable doubt would be chosen randomly from the distribution of perceived standard of reasonable doubt. In the second method, the standard of reasonable doubt would be a set, quantified standard. In both methods, the perceived level of guilt would be randomly selected from the distribution of perceived guilt given likelihood of guilt corresponding to the simulated case.

The jurors would then be sorted into two groups: those who would convict (perceived guilt $>$ perceived standard of reasonable doubt) and those who would acquit (perceived guilt $\leq$ perceived standard of reasonable doubt). The simulated jury would then work to a final decision from this preliminary vote. We suggest that psychologists aid in the determination of persuasiveness, resistance to change in belief, "strength in numbers," and other similar issues. A simple approximation of persuasiveness would be the difference between a juror's perceived guilt and standard of reasonable doubt. The farther the perception is from the standard, the more strongly the juror holds the belief. The side with the greater total difference would carry the decision. If the sides had equal total differences, the jury would by hung.[14]

In order to show that a quantified standard would improve the current sys-

[14] It must be taken into account that jurors with standards close to 0 or 1 would be limited in their ability to persuade if a strict difference was taken.

tem, it must be shown that cases that should be convicted are convicted more often with the set standard than under the current system. Likewise, it must be shown that cases that should be acquitted are acquitted more often with the set standard. To show this, it is necessary to show that in the simulation, when the likelihood of guilt is higher than the standard being tested, the conviction rate is higher for juries using the quantified standard than for juries using the distribution of the perceived standard of reasonable doubt. Similarly, it is necessary to show that the conviction rate is lower for juries using the set standard for cases in which the likelihood of guilt is less than the standard being tested. Both instances could be tested using two sample Z tests for difference of proportions.

Also, if the distribution of the likelihood of guilt after trial was known, then the number of replications of the simulation at different likelihoods of guilt could be weighted based on this distribution to show the difference between the overall conviction rate with and without a quantified standard. These rates could be combined with the guilty rate, γ, from the conditional probability tree to compare α and β error rates with and without a quantified standard.

For comments on an attempted simulation, see Appendix II.

4 Conclusion

Due to the ambiguous nature of the current defintions of reasonable doubt, we believe that a quantified standard is necessary to clarify the law and give it "the precision owed to those whose liberty or life is at risk."[12] We do not believe that a quantified standard should be used to replace the current legal system; however, we think that it could, and should, be implemented to strengthen and clarify it. To do so, the distribution of guilt and innocence, the distribution of likelihood of guilt after trial, α, and β must be more extensively studied. After these studies have been completed, we will be in a much better position to quantify reasonable doubt. To give the standard the accuracy that the accused deserve, care must also be taken when approximating the distributions described. It is these distributions that will give the vital insight into how reasonable doubt should be quantified.

Appendix I: Written Standards of Reasonable Doubt

Emphasis added to highlight ambiguous language

California Standard as cited in Victor v. Nebraska, p. 1244[13]

Reasonable doubt is defined as follows: It is not a mere *possible doubt*; because everything relating to human affairs, and depending on *moral evidence*, is open to possible doubt. It is that state of the case which, after the entire comparison and consideration of all the evidence, leaves the minds of the jurors in that condition that they cannot say they feel an *abiding conviction*, to a *moral certainty*, of the truth of the charge.

Nebraska Standard as cited in Victor v. Nebraska, p. 1249[13]

"Reasonable doubt" is such a doubt as would cause a reasonable and prudent person, in one of the *graver and more important transactions of life*, to pause and hesitate before taking the represented facts as true and relying and acting thereon. It is such a doubt as will not permit you, after full, fair, and impartial consideration of all the evidence to have an *abiding conviction*, to a *moral certainty*, of the guilt of the accused. At the same time, absolute or mathematical certainty is not required. You may be convinced of the truth of a fact beyond a reasonable doubt and yet be fully aware that possibly you may be mistaken. You may find an accused guilty upon the strong probabilities of the case, provided such probabilities are strong enough to exclude any doubt of his guilt that is reasonable. A reasonable doubt is an actual and *substantial doubt* reasonably arising from evidence, from the facts or circumstances shown by the evidence, or from the lack of evidence on the part of the state, as distinguished from doubt arising from *mere possibility*, from *bare imagination*, or from *fanciful conjecture*.

Appendix II: Attempted Simulation to Test the Necessity of a Quantified Standard

The simulation described in Section 3 was attempted using approximated distributions. The distribution of perceived standard of reasonable doubt was a continuous approximation of the sample of jurors' perceptions of reasonable doubt taken by Simon and Mahan[1]. By converting this distribution from the discrete distribution to a continuous cumulative distribution, it was believed that the errors caused by only allowing jurors to choose standards that were multiples of 5%, as discussed in the Section 2.3, would be minimized by spreading the probability over the entire range between consecutive multiples of 5%. The distribution of perceived guilt was approximated using a normal distribution centered at the likelihood of guilt of the trial being tested. The distribution had a standard deviation of 5%. The normal assumption was made because perceived guilt is, in effect, error about the actual likelihood of guilt.

The results of the simulation showed that, when the likelihood of guilt was below the standard of reasonable doubt being tested, the juries using the quantified standard convicted a significantly smaller proportion of the cases. However, the results were inconclusive when the likelihood of guilt was above the standard. The results suggested that they were highly dependent on the specific shapes and centers of the distributions used to compute the simulation. Therefore, we do not believe that our approximations are sufficiently justified to draw conclusions from our simulation. However, we do believe that a simulation will show the effects of a quantified standard once the distributions are known.

References

[1] Rita James Simon and Linda Mahan. Quantifying burdens of proof: A view from the bench, the jury, and the classroom. *Law and Society Review.*, 5(3):319–330, 1971.

[2] Ginsburg Decision. *Victor v. Nebraska,.* 114 S.Ct. 1239 (1994).

[3] *Commonwealth v. Webster,.* 59 Mass. 295, 320 (1850).

[4] Harry D. Saunders. Quantifying reasonable doubt: A proposed solution to an equal protection problem. *The Berkeley Electronic Press.*, 2005.

[5] Robert L. Winkler. *Introduction to Bayesian Inference and Decision.* Series in Quantitative Methods for Decision Making. Holt, Rinehart and Winston, INC., New York, 1972.

[6] Francis C. Dane. In search of reasonable doubt: A systematic examination of selected quantification approaches. *Law and Human Behavior.*, 9(2):141–158, 1985.

[7] Laurence H. Tribe. Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review*, 84(6):1329–1393, April 1971.

[8] *Interview with Bruce G. Berner, Professor of Law, Valparaiso University.* (July 2006).

[9] James Madison. The federalist no. 10: The utility of the union as a safeguard against domestic faction and insurrection (continued). *Daily Advertiser*, November 22, 1787.

[10] Rita James Simon. Beyond a reasonable doubt: An experimental attempt at quantification. *Journal of Applied Behavioral Science*, 6:203–209, 1970.

[11] Dorothy K. Kagehiro and W. Clark Stanton. Legal vs. quantified definitions of sandards of proof. *Law and Human Behavior.*, 9(2):159–178, 1985.

[12] Blackmun Decision. *Victor v. Nebraska,.* 114 S.Ct. 1239 (1994).

[13] O'Connor Decision. *Victor v. Nebraska,.* 114 S.Ct. 1239 (1994).

Acknowledgements

Special thanks goes out to Dr. Jim Caristi and Dr. Ashraf El-Houbi of the Valparaiso Department of Mathematics and Computer Science; the faculty of the Valparaiso University School of Law, in particular Professors Bruce G. Berner, Ivan E. Bodensteiner, and David E. Vandercoy; and Debbie Blennerhassett of the law library, for their advice and assistance throughout this project.