Quantifying Reasonable Doubt: a Proposed Solution to an Equal Protection Problem

Harry D. Saunders

Abstract

In this article we present the case that the Reasonable Doubt standard is in urgent need of repair. Our research reveals that a previously-recognized phenomenon arising from vagueness of the standard is more consequential than thus far realized and creates a serious equal protection problem. We show that the only legally feasible solution to this problem is to quantify the definition of the standard. While others have examined quantified standards, we make a direct case for it and overcome previous objections to it by offering a way to make it practical and workable.

The solution we envision will require new legislation – we show that the problem is unlikely to be corrected within the judicial branch. However, we also show that legal flexibility exists at the U.S. Supreme Court level to permit such a change, and furthermore that flexibility probably exists to prevent undue chaos in the system by allowing past verdicts to be "grandfathered."

Table of Contents

Quantifying Reasonable Doubt: a Proposed Solution to an Equal Protection Problem

Harry D. Saunders

"To be a meaningful safeguard, the reasonable-doubt standard must have a tangible meaning that is capable of being understood by those who are required to apply it. It must be stated accurately and with the precision owed to those whose liberty or life is at risk."

– Justice Harry A. Blackmun[1]

In this article we present the case that the Reasonable Doubt standard is in urgent need of repair. Our research reveals that a previously-recognized phenomenon arising from vagueness of the standard is more consequential than thus far realized and creates a serious equal protection problem. We show that the only legally feasible solution to this problem is to quantify the definition of the standard. While others have examined quantified standards, we make a direct case for it and overcome previous objections to it by offering a way to make it practical and workable.

The solution we envision will require new legislation – we show that the problem is unlikely to be corrected within the judicial branch. However, we also show that legal flexibility exists at the U.S. Supreme Court level to permit such a change, and furthermore that flexibility probably exists to prevent undue chaos in the system by allowing past verdicts to be "grandfathered."

We first summarize the problem and the research that exposes it. We follow this with a review of prior research and show how our research and arguments are different. We then review legal precedent that is relevant to this problem, focusing on Supreme Court precedent as it affects possible changes to the standard. Then we present our research findings in further detail. After that, we describe why case law is unlikely to provide a solution, illustrating the problem with a highly instructive case of our contriving. Then, we go to the proposed solution, arguing that any actual solution will have to look something like it. Finally, we consider objections to the proposed solution.

The Problem

Some colleagues and I undertook an experiment. In this experiment, we asked individuals to consider what, to them, "reasonable doubt" means. The experiment was structured in very precise terms and what it delivered was disturbing. Individuals, in this case college-educated professionals specifically trained in assessing probabilities for decision making, revealed radically different interpretations of reasonable doubt – so different that it calls into critical question what jurors are asked to do when they deliberate and deliver verdicts. Quite clearly, the reasonable doubt standard, as currently described to jurors, is so profoundly vague as to be meaningless. A graphical depiction is given below.

---

[1] Victor v. Nebraska, 511 U.S. 1, 29 (1994) (Blackmun, J., concurring in part and dissenting in part).

Distribution of Responses
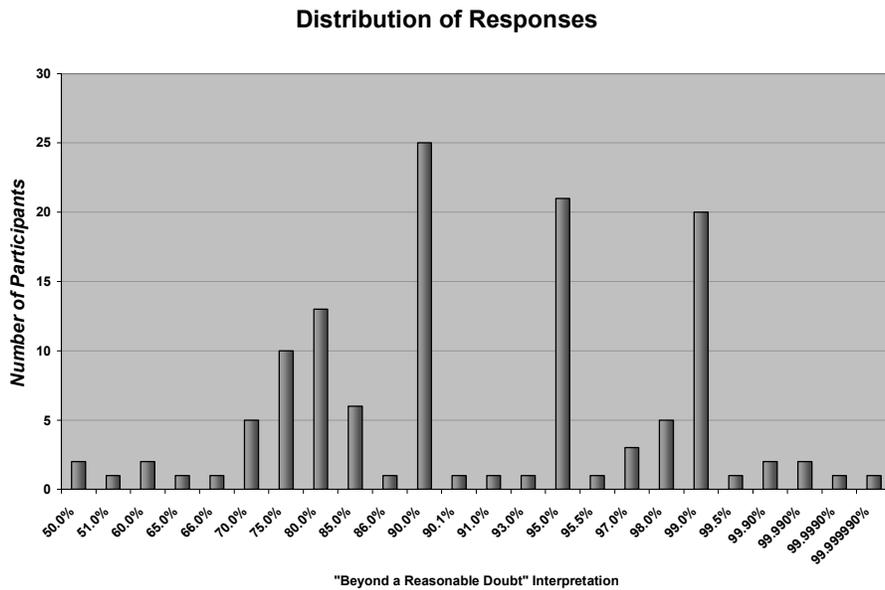
Figure 1. Interpretations of "Beyond a Reasonable Doubt"

The exact nature of our experiment is the subject of a later section, but the results appear to open wide the constitutional interpretation of equal protection. The evidence is strong that juries can, and likely do, deliver verdicts that deviate substantially from the presumed ideal of upholding a standard of proof that is uniform (equal protection) and objective (due process). Specifically, the results show that two different juries, *possessed of identical beliefs about the likelihood of guilt of a defendant*, can easily deliver two different verdicts based solely on different interpretations of reasonable doubt (equal protection). And any particular defendant's fate is in all likelihood far more dependent on individual jurors' interpretation of the standard of proof than on jurors' beliefs about likelihood of guilt (due process). We argue that this situation raises a profound and consequential problem with the reasonable doubt standard (as it is currently applied).

We later show that challenges to the existing standard based on due process are probably legally foreclosed due to precedent, but that equal protection challenges are not. However, we further show that both are probably foreclosed by practical considerations. Nonetheless, we also show that introduction of a quantitative definition of the standard is *not* foreclosed on either basis.

Equal protection and due process considerations aside, we show in a later section there is a deep logical problem with, plus a major legal impediment to, allowing jurors to set their own standard. And short of abandoning altogether the idea of a standard of proof, we demonstrate that a quantitative standard is the only way to assure that all these problems do not arise.

The Experiment

In our experiment, we first trained individuals in some basic probability concepts. The individuals – 130 of them – were participants in corporate training programs designed to improve the quality of decision making in their organizations. Typically they were

college graduates, and typically with degrees in engineering or the sciences. The training was based largely on the discipline of decision analysis, now in widespread use in the business world and there employed by managers and executives to make decisions with large resources at stake that are subject to significant uncertainty.

Our experimental design was influenced by our suspicion that many, maybe most, criminal juries conflate two tasks: determining the likelihood of guilt; and determining if this likelihood passes the reasonable doubt standard. The first task is one that requires a functioning human mind, and human judgment, and everything a human being can bring to the undertaking by way of past experience, discernment, and knowledge of human nature. The second task, in principle, is much easier. It requires only that individuals gauge their judgment against an objective standard. In our experiment, this objective standard was described to study participants using the California jury instruction.[2]

The training helped participants with the equivalent of the first task. In order to make good decisions, business decision makers must make assessments of the likelihood of uncertain events. Modern-day decision makers of major corporations do this quantitatively. They, or their designates, routinely prepare quantitative probability assessments of key uncertainties known or expected to influence the success or failure of the alternatives under consideration. These uncertainties usually involve very complex, multi-determinate events and contain many so-called "intangibles": will the Korean peninsula reunify within the next ten years? Will the economy grow faster than 3% next year? Will the new drug I am researching be successful? Will my competitor decide to build a new plant? Will I strike economic quantities of oil if I drill this well? …and so forth.

We ask the reader at this point to temporarily suspend your disbelief that individuals can formulate quantified probabilities that are sound; we only mention here that our methodology enables individuals to specify precise probabilities that they believe in strongly enough that they would be willing to bet a large sum of money – or their life if required – that the probability is exactly as they say. Let us also add that we have used our methods to extract from a superior court judge the likelihood he saw that a jury in a case he was familiar with (unknown to us) would deliver a guilty verdict – an event we have heard legal professionals describe as "imponderable." With our help, this judge was able to deliver a very precise quantitative declaration of his subjective probability regarding this future event. Later on, we will explain a tool used by business managers that enables such precision but is simple enough for a child.

The individuals we trained learned that such probability assessments are necessarily subjective – others with different information, knowledge or experience might assess the probabilities differently. But in the business world, as elsewhere, decision makers must rely on their own beliefs – informed, perhaps, by the judgments of those around them – because decision makers bear direct consequences of their own decisions. Decision professionals call such an approach to subjective uncertainty

---

[2] Prior to the recently-adopted new "plain English" instruction to take effect January 1, 2006 (CALCRIM 220, available at http://www.courtinfo.ca.gov/jury/criminaljuryinstructions/calcrim_juryins.pdf). We believe our results would change in no substantive way were we to use this newly-proposed California instruction.

quantification Bayesian, after the 18th century cleric and scholar Thomas Bayes.[3] Jurors, likewise, must make inherently subjective judgments.

Armed with this background, the individuals were asked to place themselves in the following situation: You (the study participant) are serving on a jury in a murder case. You are required first to judge, based on the evidence, the likelihood that the defendant is guilty. You know that you can never be 100% certain of the defendant's guilt or innocence, since the only ones who can possibly know for certain are: the victim, the perpetrator, the defendant (if different from the perpetrator), or an eye witness. None of these can serve on a legally constituted jury. You have seen the evidence and listened to the arguments, we told them, and you have reached some judgment of the likelihood the defendant is guilty. It doesn't matter what this likelihood is, but you have precisely quantified it using methods you are now familiar with.

The question before you is: *what probability of guilt is sufficient for you to vote a verdict of guilty?* To assist you in this, we will read you the California jury instruction:

> *"Reasonable doubt is defined as follows: It is not a mere possible doubt; because everything relating to human affairs is open to some possible or imaginary 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 of the truth of the charge."[4]*

The participants were asked to put on a piece of paper a single number giving their interpretation of "beyond a reasonable doubt" – the standard they would use – and return it anonymously. We recorded the results.

The results are remarkable. Among the 130 participants, the bulk of responses ranged from 50.0% to 99.999990%. Two individuals reported numbers of 30% and 100% (more about these individuals later). Twenty-five individuals recorded 90%; twenty recorded 99%. Note that the difference between 90% and 99% is a *ten-fold* difference in what they would consider "reasonable doubt" – the difference between a 1% level of doubt for one person and a 10% level of doubt for another.

The results are reported in more detail later, but the upshot is clear: two different juries, *possessed of identical beliefs about the likelihood of guilt of the defendant*, could deliver different verdicts simply because jurors interpret the standard of proof differently. We argue that this is violative of equal protection.

It is critical at this point to emphasize that we are not arguing the jury system itself is violative of equal protection. In fact, our distinctions allow a clarification of the jury's role. It is entirely possible and legitimate for two different juries to come to different verdicts on the same facts. This is a simple, unavoidable fact of the jury system. Determining the likelihood of guilt of a defendant is necessarily a subjective undertaking and will always be so. It is the standard itself, against which jurors must measure their subjective beliefs, that we claim assails equal protection.

It is notable that all participants, despite their differences about what the standard should be, rationally could weigh their subjective judgments of guilt against such a standard if – but only if – it were quantified precisely.

---

[3] Thomas Bayes, *Essay towards solving a problem in the doctrine of chances,* 54 PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY OF LONDON, at 296 (1764).
[4] CALJIC 2.90, available at http://www.courtinfo.ca.gov/presscenter/documents/before_after.pdf .

Previous Research

Our research is similar to that of previous researchers, but different in important ways. Seminal work by Simon and Mahan [5] explored the idea of quantifying jurors' judgments of likelihood of guilt, and additionally surveyed judges, prospective jurors, and students as to their quantitative interpretation of the reasonable doubt standard. Like us, they observed a large variance among those surveyed and they additionally reported significant discrepancy between judges' and jurors' interpretation of the standard. Unlike us, they used a 1 to 10 scale instead of percentages and so did not uncover the differences at the high end of the range we later show to be crucially important. Nor did they explicitly draw the conclusions from their results that we would draw. Nonetheless, their research strongly supports our case.

Others have done similar research aimed either at directly assessing jurors' implied reasonable doubt standard or inferring it from mock jury experiments. An excellent reference summarizing these studies is found in Hastie[6] who reported that "…estimates of the beyond-reasonable-doubt standard in criminal cases ranged from .51 to .92"[7] [on a scale of zero to one – equivalent to 51% to 92%]. It is important to note that none of these studies used individuals trained in assessing subjective probabilities. In our case, we used individuals specifically trained in assessing subjective probabilities for use in business and personal decisions, using methods developed by decision analysts. We argue that such individuals, employed by companies that routinely require employees to quantify uncertainties for actual decision making, are much more credible than untrained individuals when it comes to reliably quantifying their beliefs.

An important contribution to the field comes from Kagehiro & Stanton,[8] who offer evidence that quantified standards of proof are superior to verbal descriptions of these standards because they have the intended effect of altering verdicts in the right direction as the standards are made more or less strict, whereas the verbal standards do not. Included in Kagehiro & Stanton's article is an excellent rebuttal[9] to objections against quantified standards of proof raised by Tribe[10] that we recommend to the reader. While Kagehiro & Stanton do not argue for quantification of the standards, they do argue that their results call for re-examining the current verbal definitions.[11] We go a step further and argue that the reasonable doubt standard must be quantified. We augment this argument with a way to do this practically – politically, legally, and workably for the individual juror – and we present the case that to do otherwise violates equal protection and due process.

---

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

[6] Reid Hastie, *Algebraic Models of Decision Processes,* in INSIDE THE JUROR: THE PSYCHOLOGY OF JUROR DECISION MAKING 84, 101-108 (Reid Hastie ed., 1993).

[7] *Id*. at 101.

[8] Dorothy K. Kagehiro & W. Clark Stanton, *Legal vs. Quantified Definitions of Standards of Proof,* 9 LAW & HUM. BEHAV. 159 (1985).

[9] I*d.* at 174-175.

[10] Laurence H. Tribe, *An Ounce of Detention: Preventive Justice in the World of John Mitchell,* 56 VA. L. REV. 371 (1971) and *Trial by Mathematics: Precision and Ritual in the Legal Process,* 84 HARV. L. REV. 1329 (1971).

[11] Kagehior & Stanton, *Supra* note 8*,* at 175.

Recently a school of thought has arisen that argues against quantifying reasonable doubt on the grounds that it precludes needed flexibility in jury decision making. This is an important development and requires explicit attention. Stoffelmayr & Diamond[12] argue that the merits to leaving the reasonable doubt standard vague are that jurors can adjust their verdict to reflect the severity of the offense and the costs associated with error. Similarly, Lillquist[13] argues that juries will require more proof in some cases than in others, that jurors probably apply utilities to wrongful and rightful decisions that reflect those of society, and that the flexibility to consider outcomes specific to the case being decided is preferable to a fixed standard of proof.

But here is the problem with their reasoning: while these researchers adhere to solid principles of decision theory, they do not adhere to principles of the law. The difficulty is that current law prohibits jurors from considering the consequences of their verdict and instructs them only to weigh the evidence. By way of example, the model jury instructions for the Eighth Circuit say, "You may not consider punishment *in any way* [emphasis ed.] in deciding whether the Government has proved its case beyond a reasonable doubt."[14] The Ninth Circuit instructions are identical to this,[15] and the Eleventh Circuit instructions are almost identical.[16] The Fifth Circuit instructions say, "You should not be concerned with punishment in any way. It should not enter your consideration or discussion."[17] The Sixth Circuit states it even more strongly: "It would violate your oaths as jurors to even consider the possible punishment in deciding your verdict."[18]

This practice is commonly carried through to the state level. For instance, the Judicial Council of California's recommended instructions include the statement, "You must reach your verdict without any consideration of punishment."[19] Hawaii includes an instruction that penalty or punishment is not to be discussed.[20]

For these researchers' proposal to work, jurors must be able to consider the utilities of possible outcomes from their verdict (wrongful or rightful acquittal; wrongful or rightful conviction). We show mathematically in Appendix A that this is logically impossible if they are barred from considering punishment. The plain English version of

---

[12] Elisabeth Stoffelmayr & Shari Seidman Diamond, *The Conflict between Precision and Flexibility in Explaining "Beyond a Reasonable Doubt,"* 6 PSYCHOL. PUB. POL'Y & L. 769 (2000).

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

[14] MANUAL OF MODEL JURY INSTRUCTIONS FOR THE DISTRICT COURTS OF THE EIGHTH CIRCUIT (2000 EDITION) 3.12, available at http://www.ca8.uscourts.gov/rules/criminal2000.pdf.

[15] NINTH CIRCUIT MODEL CRIMINAL JURY INSTRUCTIONS (2003 EDITION) 7.4, available at http://www.ce9.uscourts.gov/web/sdocuments.nsf/criminal+jury.

[16] ELEVENTH CIRCUIT PATTERN JURY INSTRUCTIONS (CRIMINAL CASES) (2003 EDITION) 10.1, available at http://www.ca11.uscourts.gov/documents/jury/crimjury.pdf.

[17] FIFTH CIRCUIT DISTRICT JUDGES ASSOCIATION PATTERN JURY INSTRUCTIONS (CRIMINAL CASES) (2001 EDITION) 1.20, available at http://www.lb5.uscourts.gov/juryinstructions/crim2001.pdf.

[18] SIXTH CIRCUIT CRIMINAL PATTERN JURY INSTRUCTIONS (2005 EDITION) 8.05, available at http://www.ca6.uscourts.gov/internet/crim_jury_insts.htm.

[19] CALCRIM 200 available at http://www.courtinfo.ca.gov/jury/criminaljuryinstructions/calcrim_juryins.pdf.

[20] HAWAI'I CRIMINAL JURY INSTRUCTIONS 8.01, available at http://www.courts.state.hi.us (follow this sequence of hyperlinks: "Legal References"/"Internet Resources"/"Jury Instructions"/"Hawai'i Civil and Criminal Jury Instructions"/"Hawai'i Criminal Jury Instructions").

this proof is that, without being able to consider punishment, jurors cannot logically, or legally, trade off the social utilities/disutilities of wrongful vs. rightful verdicts.

Mathematics aside, we argue that weighing the social utility of rightful verdicts vs. the disutility of wrongful ones properly, and legally, belongs in the hands of state legislators, not in the hands of individual jurors. Even where punishment is not explicitly mentioned, most courts emphasize that jurors must confine themselves to the evidence before them and most jury instructions contain directives such as that contained in the Eighth Circuit's instruction: "[Y]our verdict must be based *solely* [emphasis ed.] on the evidence and on the law which I have given to you in my instructions."[21] It is difficult to escape the conclusion that most jury instructions legally preclude jurors from considering anything about the individual or social costs and benefits of their decisions (including for example the cost of releasing a murderer back to society or convicting an innocent person). The clear sense of most or all jury instructions is that jurors should weigh only the evidence, and be pure "finders of fact," not arbiters of social utility.

However, to be fair to Stoffelmayr & Diamond and Lillquist, the method for creating a sensible standard at the state level should involve considerations and methodology precisely as they propose. In setting the standard of proof, legislators will need to consider both the social utility of rightful verdicts and the disutility of wrongful ones. Importantly, part of what motivates these researchers' call for flexibility is that society likely places different utilities on verdicts for different crimes. Ironically, however, the very nature of quantifying the standard (as we suggest) enables exactly this – different quantified standards could be applied to different classes of crimes.

Even if we exclude consideration of utilities by jurors, there exists a logical conundrum. One might argue that jurors should nevertheless determine the standard of proof themselves, based on their interpretation of the jury instruction. Objection # 3 in a later section explains why this is infeasible. Briefly stated, the problem is that jurors, if they are aware of the phenomenon underlying our results, cannot rationally assign meaning to the reasonable doubt standard.

On a final note, it is instructive to look at the reaction of judges to the idea of quantifying the standard as reported by Simon and Mahan. They cite a representative judge's reaction as follows:[22]

> "Percentages or probabilities simply cannot encompass all the factors, tangible and intangible, in determining guilt – evidence cannot be evaluated in such terms."

This statement is patently false, and bespeaks a (probably widespread) unfamiliarity in the legal community with the modern tools of decision analysis. Individuals can be readily trained to quantify their subjective uncertainty about the most subtle and complex of situations replete with so-called "intangibles," and corporations count on this ability daily.

---

[21] MANUAL OF MODEL JURY INSTRUCTIONS FOR THE DISTRICT COURTS OF THE EIGHTH CIRCUIT (2000 EDITION) 3.12, *supra* note 14.

[22] Rita James Simon & Linda Mahan, *Supra* note 5*, at 329.

Legal Foundations

Legal Non-foreclosure of a Quantitative Standard and Foreclosure of Challenges to the Existing Standard

A key question is whether quantifying the reasonable doubt standard falls afoul of – or is indeed precluded by – legal precedent. In this section we address that question and also examine whether there is legal foundation for challenging the *current* application of the standard.

Because our results, we claim, challenge both due process and equal protection, we address both issues. We begin with due process.

The United States Supreme Court found, in the landmark case *Victor v. Nebraska*,[23] that aside from certain specific phraseology such as "moral certainty" some justices found troubling, the most commonly used jury instructions on reasonable doubt did not violate the due process provision of the Fourteenth Amendment. The Court's focus was narrowly on due process; equal protection was not addressed in these cases. Importantly for us, there is nothing in this ruling that of necessity precludes the substitution of another standard for the current standard. Simply because it has not found the current standard to be violative of due process does not mean the Court would find a different standard (and to be specific, a more precise version of the same standard) to be violative of it. Accordingly, if the standard were quantified, the Supreme Court would not be bound by precedent in any challenge to it on due process grounds.

Importantly for our proposed solution, the Court here explicitly affirmed that the use of *any* reasonable doubt definition is neither compelled upon the state courts, nor prohibited them: "The beyond a reasonable doubt standard is a requirement of due process, but the Constitution neither prohibits trial courts from defining reasonable doubt nor requires them to do so as a matter of course."[24] They only require that if a jury instruction is given, that, " 'taken as a whole, the instructions [must] correctly conve[y] the concept of reasonable doubt to the jury.' "[25]

The Court's ruling in *Victor v. Nebraska* would appear to foreclose a challenge to the current standard on due process grounds, although clearly the Court did not have evidence such as ours in hand, nor would the underlying cases in question necessarily have called for it. In fact, our results *do* challenge the due process viability of the standard in a fundamental way, but, even aside from the Court's implied recognition in the *Victor v. Nebraska* ruling of the standard as currently applied, there is the intimidating challenge of finding any actual means of obtaining clarity on this question through the courts. Our data do show an arguably lethal-to-due-process span of different interpretations of "reasonable doubt." And on this basis a rational argument would be that taken as a whole, the instructions do *not* "correctly convey the concept of reasonable doubt to the jury,"[26] accordingly violating due process. But that is only an argument, not a case.

Only two possibilities exist: One, a case can be found that on the facts alone is so directly and clearly condemnatory of the current use of the reasonable doubt standard that

---

[23] Victor v. Nebraska, 511 U.S. 1 (1994).

[24] Victor v. Nebraska, 511 U.S. 1, 5 (1994) (O'Connor, J., writing for the majority).

[25] *Id.* at 5 (O'Connor, J., quoting Holland v. United States, 348 U.S. 121, 140 (1954)).

[26] *See id.*

it compels change. We argue that such is unlikely, and later show the fanciful nature of attempting to construct such a case. Or two, one could envision interviewing jurors who had reached a verdict, using methods like ours to demonstrate that they carried different interpretations of the reasonable doubt standard. The problem with this idea is that jurors' interpretation of the standard could inadvertently be corrupted by their adjudicating the case. Once an individual has reached a verdict of, say, guilty, he or she might be reluctant to admit they could have convicted against a loose standard that perhaps calls into question their judgment, or even their decision, and the interview process would in all likelihood convey this possibility to the individual. We argue that this is an impractical alternative.

That was due process. While the United States Supreme Court has ruled on other equal protection issues that are jury process-related,[27] it has not ruled on equal protection issues surrounding the reasonable doubt standard. This means, in the first place, that there is no precedent that would require the Supreme Court to prohibit use of a quantitative standard based on equal protection considerations. It means, in the second place, that the door is technically open to a challenge to the current standard on equal protection grounds.

However, as with due process, we argue that equal protection is not practical to preemptively litigate absent an actual change in the standard. As it happens, the practical challenge here is even greater than for due process.

One could attempt to find a case that works on the facts alone, but, as above, we believe this unlikely. The second possibility would require the courts to allow a highly controlled experiment in an actual, unadjudicated case involving outside access both to jurors themselves and to a "sample" jury pool, who also view the proceedings as hypothetical jurors. This would be necessary to develop data on the differences between potential juries; to make a case it would also require real jurors deciding a real case. We argue that this possibility is unlikely, too.

There is precedent for courts attempting to apply a quantitative definition of reasonable doubt. In *McCullough v. Nevada*, the Nevada Supreme Court overturned a district court ruling where the trial judge provided jurors a numerical description of reasonable doubt that on a scale of ten defined reasonable doubt as "seven and a half, if you had to put it on a scale." [28] However, the Nevada Supreme Court did not find the attempt by the trial judge to clarify the reasonable doubt standard to *by itself* be reversible error. Rather, the ruling was based in part on the finding that the instruction "may impermissibly lower the prosecution's burden of proof,"[29] and the finding that the instruction "is reversible error when *coupled with* [emphasis ed.] *any* other attempt to supplement, change, or clarify the statutory reasonable doubt definition."[30] Thus, this finding was by no measure a condemnation of applying a quantitative definition, in and of itself.

In summary, at no point has the U.S. Supreme Court or any other court in the United States made any ruling that closes the door to a quantitative standard. As for

---

[27] For instance, Johnson v. Louisiana 406 U.S. 356 (1972) (where the Court upheld a Louisiana law requiring less-than-unanimous jury verdicts in criminal cases).
[28] McCullough v. State, 657 P.2d 1157, 1157 (1983).
[29] *Id.* at 1159.
[30] *Id.* at 1158.

challenging the current application of the standard, it is theoretically possible to challenge it on equal protection grounds, and possibly even due process grounds, but we do not believe it is practical to do so.

Attempts to Clarify the Standard

It is worthwhile describing attempts by the courts to bring further clarity to the reasonable doubt standard. The method has been to accord different phrases different probabilistic weights. (Within quotations in this paragraph, italicization is ours.) In *Cage v. Louisiana*, the Supreme Court found that "[i]t is plain to us that the words '*substantial*' and '*grave*,' as they are commonly understood, suggest a higher degree of doubt than is required for acquittal under the reasonable doubt standard."[31] They also, as noted previously, expressed deep concern with use of the term "moral certainty." On the other hand, the Court took no issue with "near certitude" being used in instructions to describe beyond a reasonable doubt: "As used in this instruction … [the language] …'impressed upon the factfinder the need to reach a subjective state of *near certitude* in the guilt of the accused."[32] The Court also described this as "the *very high level of probability* required by the Constitution in criminal cases."[33] It further allowed that the use of "strong probabilities" was appropriate so long as " 'it emphasized the fact that those probabilities must be *so strong* as to exclude any reasonable doubt.' "[34] Justice Ginsberg in her separate opinion complained that if these words "save this part of the instruction from understating the burden of proof, see *ante*, at 19, they do so with uninstructive circularity. Jury comprehension is scarcely advanced when a court 'defines' reasonable doubt as 'doubt…that is reasonable.' "[35] About the only thing the "reasonable doubt" standard apparently requires is that the doubt must be reasonably constructed, however improbable the chain of events involved in that construction.

The core problem with all this drawing of verbal distinctions and parsing of words is that defining the standard in words will never suffice. Defining one set of words in terms of a beguiling but equally vague set of other words will not overcome the issue our results uncover, and only buries the problem deeper. The standard needs quantification. Despite huge instinctive aversion to our proposal to quantify the standard, we believe most legal professionals will come to terms with it once they understand that is not so difficult to implement as they might imagine and once they see a change is inevitable. There may be those who oppose losing the ability to subtly sway the definition of reasonable doubt in a jury's mind. But their loss will be a gain for jurors, who will now have greater clarity on what is being asked of them. Our proposal actually removes a burden from the juror. For those who understand how it works, converting one's deeply held beliefs into a quantified probability is an easy and quickly accomplished task. The much harder task for a juror is actually developing that belief in the first place. Removing from the juror any confusion about the standard they must apply would make their most important task – determining the likelihood of guilt – an easier one to focus on.

---

[31] Cage v. Louisiana, 498 U.S. 39, 41 (1990).

[32] Victor v. Nebraska, 511 U.S. 1, 15 (1994) (quoting their opinion on Jackson v. Virginia 443 U.S. 307, 315 (1979)).

[33] Victor v. Nebraska, 511 U.S. 1, 15 (1994) (quoting Dunbar v. United States, 156 U.S. 185, 199 (1895).

[34] *Id.* at 22.

[35] *Id.* at 25.

Overall, it would not be as problematic as some may think to institute a modernized quantitative definition of reasonable doubt.

In fact, there may be little choice. Any thoughtful individual cognizant of the phenomenon underlying our results can reasonably argue that they can never serve on a criminal jury. Such a person, knowing the vastly different interpretations that can ("reasonably"!) be put on the standard, is left to ask "which one should I choose?" "…or does the jury instruction allow me to choose my own standard that I think best reflects the norms of society?" Since this latter, as argued above, is not legally available to them, such a person is stuck. He or she cannot rationally execute the duty asked. As more individuals understand this dilemma, the jury system's very functioning is at risk.

The Experimental Results, in Detail

The degree of the problem is substantial. Both the range of responses and the certainty with which each participant initially held that they understood the reasonable doubt instruction make the findings extraordinary.

The graphical depiction of our results was shown in Figure 1, above. Aside from the worrisome range of results, it is worth observing the differences at the high end of the range. *Each* step up, from 99% to 99.9% to 99.99% and so on, is a *ten-fold* change in the level the person assigns to the meaning of the phrase complementary to "beyond a reasonable doubt," namely "reasonable doubt." The highest number recorded, 99.999990%, corresponds to a level of reasonable doubt comparable to errors reported for DNA analysis. This is surprising because one would normally expect in a real case that the overriding source of uncertainty might be chain of custody or other considerations that call into question the veracity of the DNA evidence as presented – the veracity of the reporting expert, for instance.

One might protest that had we asked participants to report their understanding of "reasonable doubt" instead of "beyond a reasonable doubt" we would have found the results to be different. But if so, this only reinforces the case: without a quantitative criterion, the standard is even more prone to conflicting interpretations than our results suggest. Jurors will be in conflict with themselves if the two choices of wording lead the same person to different interpretations of the standard. They will also likely be in conflict with the reasoning of fellow jurors. There is no way to tell, short of a clear directive from the bench, which of these two ways others may be interpreting the instruction. Depending which phrase they choose as their focus, some could be interpreting it one way and others another.

Either way – whether "beyond a reasonable doubt" or "reasonable doubt" – there is a problem.

A word about some extreme, and revealing, data points. One individual reported a 30% value (not shown on the chart). Our initial response was to ask this person if they really meant to say 70% because they had been measuring "reasonable doubt" instead of "beyond a reasonable doubt." The person responded adamantly that, no, this is exactly what they meant: if someone is charged with murder, they actually *are* more likely to be guilty than not guilty, and therefore the standard for releasing suspected murderers should be very high. This of course flies in the face of the presumption of innocence requirement, which had been explained to participants, but this person argued that in *his* homeland (Nigeria), this is not what he would deem a prudent presumption. Another

individual, an African American, argued the opposite. He had reported 100% (not shown on the chart). It had been explained to him that 100% certainty is an impossibility for a participant in a legally constituted jury. Yet he persisted in his declaration that he would have to be *absolutely* sure the person was guilty to vote for conviction. He had seen too many instances, he said, where evidence had been concocted, and police had lied, to believe any person before him in a courtroom had actually committed the crime. Therefore, he said, the standard should be set at the highest extreme (and at a level we would still argue, by the way, is infeasible.)

These examples, to us, indicate very forcefully that interpretation of society's norms to establish the clarified standard should be done to the extent possible in a way that reflects local cultural norms. Especially where liberty or life are at stake, the threshold likelihood of guilt for letting guilty defendants free or the innocent be convicted should be determined by the deepest beliefs of the community upon whom the crime, or alleged crime, has been inflicted and upon whom the standard will be imposed.

Supporting this contention, governing constitutional authorities for declaring an explanation of reasonable doubt, within certain broad limits, reside at the state level.

To complete the background of our experiment, a description needs to be given of the exact methodology we taught participants to use. Participants had been trained in the use of a "probability wheel," a device used by decision analysts to elicit quantitative values for subjective probabilities. The precise procedure for using this device is detailed in an appendix, but in a few words the method presents subjects (participants learned to use themselves as subjects) with a choice rather than requiring from them a direct numerical evaluation. This is what makes it easy for people to use. The choice is constructed in such a way that the subject is presented with a hypothetical decision involving a very large monetary prize (or in another method, the possibility of losing one's life). The choices the subject makes, as the wheel is adjusted to represent different "lotteries," lead a subject to a point of indifference between two particular alternatives. At that point a numerical probability is automatically extracted from the subject – one that the individual himself has implicitly chosen, and the validity of which he would attest to with his money (or his life, in the other method). Note that participants did not use the probability wheel to generate their interpretation of the standard; it was used only to show them they could precisely quantify a judgment of likelihood of guilt to measure against the standard.

We refer the reader to the appendix for a further description, but it is important to note that training individuals to use this tool is not hard. We have trained people from many different walks of life to use it in a very short time, including an eleven-year-old girl. Once they understand the concept, individuals become very comfortable and facile with it. Use of this, and other similar tools, could be taught in any high school or even middle school.

Why Case Law will not Provide a Solution

It would be appealing to rely on case law to force a solution. But as argued before, practicality dictates that, short of a case that can be decided on a unique set of facts, the solution needs to be found elsewhere. Here we explain why such a case, barring an extraordinary circumstance, will not be brought to the courts.

The difficulty is that the facts of the case would have to be a very precise. Because jurors come to their decision by performing two tasks – determining the likelihood of guilt and seeing if this passes the standard – to test the standard it would have to be a case where the probability of guilt is explicitly and precisely determined. We have tried to construct cases where this prevails, but have come up virtually empty-handed. To describe the difficulties, we present a case of our constructing that, fanciful as it is, would meet the criterion of precise determination of the probability of guilt. But we use it to argue that its very fancifulness is strong indication that a case fitting the criterion is highly unlikely to come before a court.

Here is the case: John Q. is a gambler with a bent for taking dangerous risk. He has come upon a secret club that promises him an unequalled thrill, and unequalled risk. He is told to appear at a particular place and enter a room. There he will be presented with a life-and-death game. On a wall facing him is an opaque window with a hand-size hole. Above it is a screen, giving instructions and showing a roulette-type wheel. He is asked to study the wheel and then, if he agrees to the game, to proceed to put his hand through the hole and fire a gun at someone. The game is this: he is presented with a wheel, partly red and partly blue. The red part covers precisely 30% of the wheel and represents the likelihood that, if he continues the game, the gun he will discharge will be loaded with live ammunition and will kill a live human being. The gun he will discharge will be given him when he reaches his hand through the hole. This gun may contain only blanks, or it may contain live ammunition, but it will contain only one or the other. After he discharges the gun, he will not know which.

The wheel has a pointer, and whether the gun he receives will contain live ammunition is determined by the spin of this pointer, which will spin many times on smooth bearings, and land either within the red or the blue, although John Q. will not be able to see which. If John Q. decides to play the game, whether the gun will be live will be determined by a single spin of this wheel: if the spinner ends in red, the gun will be live; if it points to blue, it will contain only blanks. Given the amount of red showing, he recognizes with great clarity that if he decides to play this game he will have a 30% chance of killing a person.

There is a second gun, unbeknownst to John Q. It is armed with live ammunition whenever John's gun is not. If John decides to take the gun and pull the trigger, the person will be shot dead whether or not John's gun was loaded or disarmed, because of this second gun. John cannot see the person behind the screen and does not know, when in the course of our story John takes the gun and shoots it, that the person has in fact been killed.

One thing John has been told and what is actually true is that the game has been set up to destroy all implicating evidence. The moment after the guns have been discharged, both guns are dropped into a vat and melted.

Just as John Q. is preparing to leave, a police team arrives. They seize him and all available evidence, including a videotape that has recorded John Q. reaching his hand through the hole and discharging the gun. In the course of their investigation the police determine that the software used to randomly "spin the wheel" was in fact working perfectly and was indeed random, though any evidence of where the spinner "landed" has been extinguished.

John Q.'s attorneys contend that John's gun was not the one used to kill the victim. The prosecution contends that it was, and is preparing charges of first degree murder. John's attorneys argue that to be first degree murder, the People must prove that it was John's gun specifically that killed the victim. Both eventually stipulate to the fact that there was a 30% chance that his gun was used to murder the victim.

John Q. contends that this is not sufficient proof to pass the reasonable doubt standard and convict him of a capital crime.

The question is, what can be done then? Is it sent to a jury, whose only question will be whether a 30% likelihood of guilt is sufficient to pass the "beyond reasonable doubt" standard? But the standard is allegedly set by law, not by juries. Where exactly is it written that the standard must be interpreted as greater than 30%? How are we to proceed?

Even if there is legal evidence that the Constitution requires that the standard be above 30%, the question is not answered. What if the wheel John Q. saw was set to 60%,? …or 90%? …or 99.9999%? Where does the Constitution require this threshold be set?

This is a highly contrived case. It is possible there may be other more realistic cases to be thought of. But we believe the example favors the conclusion that it is unlikely for the courts to force a resolution of our difficulty, or not at least any time soon, and rather that a case meeting such requirements will probably never be brought forward, or at least not one with the clarity sufficient to adequately compel the argument for a revised standard.

Our Solution

Our research, while it brings to light this problem with the reasonable doubt standard, also suggests a solution. We have developed a recommendation:

We suggest that state legislatures each commission a panel to develop a recommendation for a quantified reasonable doubt standard and that the states then enact enabling legislation. Because revising the standard would require respecting societal norms – specifically the willingness to tolerate both wrongful releases and wrongful convictions – we recommend participation of the public, perhaps through open hearings. We further recommend that the panel include the highest quality legal minds available and in other appointments be representative in some way of the breadth of that state's social norms.

We recommend that each such panel deliberate a full spectrum of alternatives for quantifying the standard, including setting different standards for different classes of crimes. This will no doubt take them on a long journey, but their conclusions and recommendations will need to be viewed by the legislature as valid, and right for their state. We also recommend that legislatures "grandfather" convictions prior to the date of implementation.

Part of any solution will no doubt be to institute educational initiatives geared at training the jury pool at large. If such education has been initiated during the time it will take the panel to deliver its recommendations and the legislature to act, the transition will be smoothed. In the longer term, education through the school system may be found to be the lasting solution to providing qualified juries.

Public support is likely to take time. But the case for a just system is compelling. Most compelling to a lay person, in our thinking, is the existing risk to equal protection. Violation of equal protection will probably be seen as constituting the most egregious assault on the sensibilities of society. The notion that people are treated differently simply because of lack of clarity in the law is distasteful. Nonetheless, we think a solution that secures due process, not just equal protection, will be asked for as well.

Objection!

Several objections have been raised to our proposal.

1. You can't put a number on *that*.

It is difficult for some to imagine that even their deepest and most subtle feelings can be translated into a number. But we would tell them that they every day make little decisions based on their hidden assessments of subtle and complex conditions. And that whether or not they are aware of it, they are quantitatively gauging the possible outcomes of their choices continuously. The quantifications they make are already there and only need to be extracted, we would say. We have found no instance of being unable to convince people, given willingness to learn about it, that they can quantify the most "imponderable" uncertainties: whether it will rain on this day a month from now; whether a particular sports team will reach the championship; whether a loved one will survive a dangerous illness; or will marry another; whether a stock price will move to a certain level; or whether, to put the sharpest point on it, a particular jury will deliver a conviction or not.

The use of subjective probabilities in the criminal justice system is not without precedent. We have used this methodology to create software that is currently in use in a California superior court public defender's office. The software involves the use of a probability wheel, used by defense attorneys to assess the likelihood of favorable (or adverse) verdicts. We believe others will follow and that the public and legal professionals will in due course familiarize themselves with the concepts and grow comfortable that it is entirely possible, and indeed easy, to quantify subjective uncertainty.

2. It will throw us into chaos: everyone will demand retrials.

Legislation enacted by the states could contain provisions that "grandfather" past application of the reasonable doubt standard. The question then is whether such legislation would be deemed constitutional.

There is precedent for the United States Supreme Court to allow changes of a similar nature to apply but not retroactively. In *Schriro v. Summerlin*,[36] the Court determined that state courts are constitutionally required to have a jury decide the sentence for a first-degree murder conviction, not a judge, but that their ruling did not apply retroactively to cases already final on direct review at that time. Importantly for our situation however, the ruling was explicitly authorized on the basis that it was not a "watershed rule of criminal procedure." The Court defined a "watershed" rule as one "without which the likelihood of an accurate conviction is *seriously* diminished."[37] (italics original) The Court was split 5 to 4 on whether the ruling was a "watershed"

---

[36] Schriro v. Summerlin, No. 03-526, slip op. (2004).
[37] *Id.* at 4.

ruling. In our case, it is entirely unclear whether a ruling on such a change as we propose would amount to a "watershed" ruling in the eyes of the Supreme Court.

The rule change was also seen to be "procedural," not "substantive." New substantive rules generally apply retroactively; new rules of procedure generally do not unless they are "watershed." A rule is substantive rather than procedural "if it alters the range of conduct or the class of persons that the law punishes,"[38] whereas "[i]n contrast, rules that regulate only the *manner of determining* the defendant's culpability are procedural."[39] (italics original) Thus a change such as we contemplate is arguably procedural and could be applied non-retroactively.

In a prior rule change they also deemed "procedural," the Supreme Court found in *DeStafano v. Woods*,[40] where the Court refused to give retroactive effect to *Duncan v. Louisiana*,[41] that "the values implemented by the right to a jury trial would not measurably be served by requiring retrial of all persons convicted in the past by procedures not consistent with the Sixth Amendment right to a jury trial."[42] Furthermore, Justice Breyer in *Schriro v. Summerlin* noted that "important interests argue against, and indeed generally forbid, retroactive application of new procedural rules,"[43] and includes among these interests "the fact that society does not have endless resources to spend upon retrials."[44] Clearly there are costs and benefits that must be weighed against each other.

But should events come to such a juncture, any costs are easily argued to be bearable when weighed against the right for all future generations to have criminal cases decided "accurately and with the precision owed to those whose liberty or life is at risk."[45]

3. Jurors *should* be the ones deciding the standard of proof.

In raising this objection, people mean one of two things. In the first instance, they mean jurors should be setting the level for the standard that is appropriate. In the second instance, they mean jurors should decide what the reasonable doubt standard means.

In the first instance, the current jury instructions would have to be abandoned. Jurors can determine an appropriate standard only by considering the consequences and utilities of their verdict, but today they are barred from doing this. And were the instructions changed so that they are *not* barred from it, the requirement that the standard be "beyond a reasonable doubt" would be redundant. As shown in Appendix A, jurors need only consider the utilities of verdict outcomes (whether their personal utilities or utilities reflecting their beliefs about societal values would be a matter for debate) to determine the implied threshold of proof. (Not that they would do an explicit calculation, but that these utilities are all they need be thinking about when deciding a standard.) If "reasonable doubt" is nonetheless retained in the instructions, it must be defined, if it is to retain any rational meaning at all, as something like "the degree of certainty that you believe should be required of the government in this case," to conscript language offered

---

[38] See *id.*
[39] *Id.* at 5.
[40] DeStefano v. Woods, 392 U.S. 631 (1968) *(per curiam)*.
[41] Duncan v. Louisiana, 391 U.S. 145 (1968).
[42] DeStefano v. Woods, 392 U.S. 631, 634 (1968).
[43] Schriro v. Summerlin, No. 03-526, slip op. (2004) (Breyer, J., dissenting – page 5 of dissent).
[44] *Id.* at 7.
[45] Victor v. Nebraska, 511 U.S. 1, 29 (1994) (Blackmun, J., concurring in part and dissenting in part).

(and rejected) by Lillquist.[46] It is a solution, but one that in effect discards any standard of proof that is remotely objective or replicable from jury to jury. The consequence, as we have argued, is that such a solution violates due process and equal protection.

In the second instance, jurors would be asked to weigh the evidence, without consideration of the consequences, and decide if the weight of evidence exceeds the stated standard. One could visualize jury instructions that would direct individual jurors to decide what "reasonable doubt" about the evidence "means to *you*." But this creates a logical conundrum. I can only assign meaning to a word if I know what others mean by it. Once I am aware that other thoughtful individuals ascribe radically different interpretations to the phrase "reasonable doubt" I am at a complete loss as to how to ascribe meaning to it for myself. The word "meaning" can be the legitimate subject of semantic and epistemological debate, but one thing it certainly implies is "meaning held in common." Once individuals abandon the idea that words carry a common meaning, all hope for rational societal discourse is abandoned along with it.

The problem in this instance might not be so serious were individuals to interpret the meaning of "reasonable doubt" in roughly similar ways, but our results demonstrate that the differences in perceived meaning span *several orders of magnitude* (powers of ten). Because I have no idea what it means for others, I can have no idea what it logically means "for me." It is clear the phrase "reasonable doubt" has no meaning in any practical sense of the word "meaning." This proposal to have jurors interpret it "for themselves" is a logical cul-de-sac from which the only exit is quantifying the standard. With quantification, what you mean by "reasonable doubt" and what I mean by it is precise and identical. You and I ascribe a common meaning to it, and we can rationally communicate with one another. Most importantly, I know how to assign it meaning and work with it, and with you. The conclusion is, of course, that allowing jurors to decide "for themselves" the meaning of "reasonable doubt" is neither a logical nor a practical possibility.

Once jurors begin realizing this logical conundrum, they will throw up their hands at what is being asked of them. The only jurors left will be those who have abandoned logic and rationality, hardly the ideal prescription for due process and equal protection.

4. Trials will become like computers and jurors like calculators.

The role of the juror will stay effectively unchanged. Jurors will still be required to do the hardest task of all, which is to absorb all the evidence, the demeanor of witnesses, and the arguments of the attorneys, and form out of that a judgment as to the likelihood of guilt of the defendant. The only change will be to take that judgment and compare it to a truly objective standard of proof, to see if their judgment passes the required threshold. It is still a sacred and heavy duty, but it will now be a clearer, more logically accomplished one.

Some worry that this change will cause trials to become non-stop parades of statistical and other experts toting endless tedious charts, and that the role of the jury will either become one of making complex mathematical judgments, or simply be reduced to making calculations. To attorneys that pursue such a strategy we would caution that only to the extent the evidence directly affects the juror's final weighed-in assessment of the likelihood of guilt will such strategies prove effective. To the extent they do, they belong

---

[46] Erik Lillquist, *supra* note 13, at 187.

in the courtroom. But the juror's final weighed-in assessment will undoubtedly contain many elements and will be accomplished by a process more akin to intuition than to mathematical reasoning. It is only at the last step, when the juror's final weighed-in assessment is quantified, that anything resembling mathematics is required – and even here a juror needs no knowledge of mathematics or statistics to deliver a quantification using our methods.

5. If it's such a big problem, why hasn't it surfaced before?

Two reasons: one, no case meeting the strict requirements of a clear test has come before the courts; and two, no one with legislative authority has to date seriously considered the indisputable evidence (such as ours) that jurors have such difficulty interpreting the standard, nor the far-reaching consequences of this.

6. It has worked for hundreds of years.

This is the most emotionally arresting of the objections. At first glance, it is deeply compelling. But the fact of the matter is, to the extent defendants have been sentenced or released according to a flawed and imprecise standard, it is not at all clear that it has "worked" as society would like it to work. And certainly if some citizens have been judged according to a different standard than others, who would not call for a change to be made?

A Note on Other Standards of Proof

Not all standards of proof currently in use suffer from the same ambiguity as reasonable doubt, but some do. The standard in civil cases, "preponderance of the evidence" has the happy characteristic of being defined mathematically on its face. Since the English word "preponderance" means "having greater weight," it is safe to conclude that subjective likelihood of liability in the mind of the juror need only be anything exceeding 50% to deliver a verdict of liable.

"Clear and convincing evidence" is more troublesome. About all that can be said quantitatively about this is that, judging from its use, it must reflect a threshold somewhere between the 50% level of "preponderance of the evidence" and whatever probability "beyond a reasonable doubt" might be deemed to carry (a problem in itself, as we have shown). A case can surely be made that the "clear and convincing evidence" standard needs to be quantified too.

Conclusion

We raise a problem and propose a solution. We believe the case for quantifying the reasonable doubt standard is a logical slam dunk: it is necessary to overcome problems of equal protection and due process and to avoid a logical conundrum for jurors; it is sensible; and it is practical to implement. It is far from clear whether society has the desire, or the will, to make such a foundational change in the jury process. But we believe that as more become educated in the language of probability and in critical quantitative thinking, not only will they desire change but will compel it. Whether in two, five, or fifty years, we submit that it is clear society will, of unavoidable necessity, come to terms with this problem and create quantitative standards of proof.

Appendix A: Why Jurors cannot use a "Flexible" (Decision Theoretic) Approach

The standard decision theoretic structure for determining a logically consistent standard of proof is usually credited to Tribe (1971a, 1971b)[47]. This structure can be illustrated as follows:
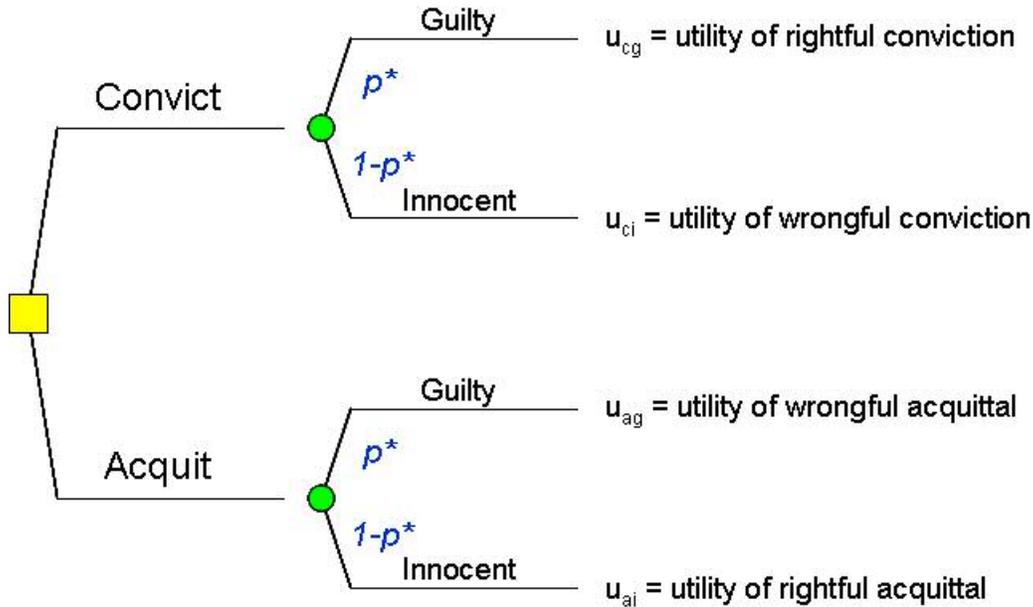
Figure 2. Determining the Standard of Proof

This structure shows a decision (convict, acquit), an uncertainty as to whether the defendant is in fact guilty or innocent, and a set of "utilities" (or "disutilities") on each of the four possible outcomes. Tribe's insight is that jurors (or the state legislature, in our construct) should set the standard of proof ($p*$) at a level such that they are indifferent between convicting and acquitting. In decision theoretic terms, this means setting $p*$ such that the "expected" (probability-weighted) utilities of each decision alternative are equal. Mathematically, the expected utility of convicting is $E(u_c) = p* u_{cg} + (1 - p*)u_{ci}$, and the expected utility of acquitting is $E(u_a) = p* u_{ag} + (1 - p*)u_{ai}$. If the four utilities are known, the equation $E(u_c) = E(u_a)$ can be solved, with a little algebra, to yield the breakeven probability that gives this indifference point between convicting and acquitting:

$$p* = \frac{u_{ai} - u_{ci}}{(u_{cg} - u_{ci}) + (u_{au} - u_{cu})}$$

Given these utilities on the outcome, $p*$ is the logically consistent standard of proof.

The problem with this approach is not theoretical, but legal. As shown previously, jury instructions prohibit the juror from placing any value on two of these

---

[47] Laurence H. Tribe, *supra* note 10.

utilities. Here's why: consider first the (dis)utility of wrongfully convicting an innocent defendant, u_{ci}. This utility will inevitably be a function of the punishment, $u_{ci} = f(punishment)$. That is, jurors, if they reflect social values, will weigh the disutility of a wrongful conviction for murder different from the disutility of a wrongful conviction for shoplifting, which carry different punishments. But jurors are prohibited from considering the punishment "in any way."[48] Similarly, the utility of a rightful conviction will in general be weighed differently for murder than for shoplifting. So these utilities, to the juror, are indeterminate. This creates a fundamental problem: jurors cannot assess the appropriate standard of proof with two variables in the above equation being indeterminate.

This means jurors cannot logically create a "flexible" standard based on the particular case, given the legal constraints. However, there is nothing to prevent state legislatures from creating such standards. While we do not suggest that legislatures should necessarily quantify these utilities to reach their determinations on quantified standards of proof, we do argue that this structure represents the complete set of considerations they must deliberate.

Appendix B: The Probability Wheel

The probability wheel is a tool developed by Dr. Carl Spetzler, co-founder of Strategic Decisions Group headquartered in Palo Alto, California. The tool presents people with a series of hypothetical "bets," from which subjects can extract numerical assessments of confidence. Once the bets have been completed, people agree that the numbers they arrive at reflect all the nuances they want to consider, including knowledge, intuition and even feelings about any given possible event. The event can be, in particular, the event that a defendant committed a crime.

The probability wheel is similar to a pie chart containing two adjustable colors. It can show all red, or it can show mostly red with a slim pie-slice of blue, or any other proportion of red to blue. On the back is a scale that tells what percentage of the wheel is showing red at any given time. Attached to the center of the wheel is a pointer, which can be spun much like the pointer in a board game.

Suppose the subject is considering: how confident am I that this particular defendant is guilty? To help the subject quantify her answer, she plays a game that involves making a large (though fictitious) bet. The player can choose to place her bet in one of two ways: she can put her money on the probability wheel or on a "clairvoyant." The clairvoyant, for purposes of the game, is a hypothetical person who can see into the situation and know with absolute certainty what has transpired. In our case, we could postulate to the subject as our "clairvoyant" the existence of a videotape that shows with undisputed clarity the crime that has occurred for which the defendant is on trial.

The rewards are as follows: if the player chooses to bet on the wheel and the pointer ends up on red, she will win $100 million. If the pointer ends up on blue, she will win nothing. Alternatively, if the player chooses to bet on the clairvoyant, and the clairvoyant (videotape) reveals that the defendant indeed committed the crime, the player wins $100 million. If the videotape reveals that the defendant did not, the player wins

---

[48] See *supra* notes 14 – 18.

nothing. The question for the player, at each stage in the game, is whether to bet on the wheel or the clairvoyant so as to maximize her chance of winning. Note that the reward structure is irrelevant, so long as the reward structure is the same for both alternatives. The reward could be "you get to live" and the cost could be "you are shot dead." The objective is to focus attention on the reward structure as if it were real.

At the beginning of the betting the subject will not be able to state her true confidence level as a percentage, so she first considers some extreme bets, setting the wheel at say 100 percent red and 0 percent blue. Then she asks herself: would I rather bet my money that the spinner will end up on red, or that the videotape will reveal that the defendant committed the crime.

A quick glance at the wheel makes it clear to this player that the odds are overwhelming for the spinner to end on red. If, by contrast, she thinks there is only about a 60 percent chance that the videotape will reveal guilt, she will quickly decide she has a better chance of winning her $100 million by betting on the wheel.

Next, she adjusts the wheel to the opposite extreme: 0 percent red and 100 percent blue. This time the player perceives the assurance that the spinner will end up in blue; she opts for the clairvoyant (videotape), which now seems a safer bet. From this point, a series of increasingly fine adjustments is made, much like the adjustments that take place in an ophthalmologist's offices. The patient being fitted for eyeglasses is asked to read an eye chart through lenses of differing strengths – stronger, then less strong. The subject continues to adjust the amount of red on the wheel, each time asking herself to choose whether she prefers to bet her money on the wheel or the hypothetical videotape. Each time she finds herself choosing the wheel, she decreases the amount of red and asks again. Each time she chooses the videotape, she increases the amount of red and asks again. When the amount of red showing leaves her just as willing to bet on the wheel as on the videotape – when, in other words she is indifferent about her choice of bets – then her degree of confidence has been quantified. It takes about twenty minutes to teach people the probability wheel technique.

Other methods, just as simple, can be used in place of the probability wheel. If the probability is a very small one – such as if the "beyond a reasonable doubt" standard were to be set at a high number like 99.9%, supplementary methods are required to develop the individual's assessment. These methods, too, involve hypothetical bets, but clearly visualized ones.

Other References

H.D. Saunders & J.G. Genser, *Trial and Error*, THE SCIENCES, Sept.-Oct. 1999, 18, published by the New York Academy of Sciences (see also letters to the editor in the subsequent issue).

Acknowledgements