Psychology, Public Policy and Law

September, 2000

6 Psych. Pub. Pol. and L. 769

ARTICLE: The Conflict between Precision and Flexibility in Explaining "Beyond a Reasonable Doubt"

NAME: Elisabeth Stoffelmayr, American Bar Foundation and Shari Seidman Diamond, American Bar Foundation and Northwestern University

BIO: Elisabeth Stoffelmayr, American Bar Foundation, Chicago, IL; Shari Seidman Diamond, American Bar Foundation, Chicago, IL, and School of Law, Northwestern University.

Correspondence concerning this article should be addressed to Shari Seidman Diamond, American Bar Foundation, 750 N. Lakeshore Drive, Chicago, Illinois 60611. Electronic mail may be sent to sdiamond@abfn.org.

SUMMARY:
... Few legal expressions are as familiar to nonlawyers as the expression *beyond a reasonable doubt*. ... What are the qualities of an optimal jury instruction on the legal standard of proof beyond reasonable doubt? In this article, we describe four criteria that have been delineated explicitly or implicitly by legal authorities and social scientists: (a) Absolute certainty should not be required, (b) a high threshold for conviction should be specified, (c) the beyond a reasonable doubt standard should be distinguishable from lower standards of proof, and (d) the instruction should encourage consistent application by jurors sitting on the same case. ... During the guilt phase of the trial, the jury was instructed to use the proof beyond a reasonable doubt standard, and the instruction defined reasonable doubt as, among other things, "such doubt as would give rise to a grave uncertainty" and "an actual substantial doubt" (p. 40). ... First, if a juror has been exposed to a case with a lower standard of proof, the instruction alerts the juror to the difference. ... In contrast, if jurors shared accurate information on the severity of the potential penalty, it could offer a context-sensitive meaning to the reasonable doubt standard by alerting jurors to the actual costs likely to be incurred by an erroneous conviction. ...

HIGHLIGHT: Though *beyond a reasonable doubt* is a common phrase inside and outside the courtroom, courts have struggled to define the phrase or to set clear standards for an adequate definition. The struggle has important implications because empirical evidence reveals that the choice of definition can influence the jury's verdict. The authors analyze the variety of proposed definitions, including the quantified definitions that some have advocated, setting out 5 criteria for an adequate definition consistent with both legal requirements and empirical evidence. They also suggest that the push for precision, such as that promised by a quantified definition in the form of a single percentage, may produce a loss of desirable flexibility. A structured flexibility permitting the trier of fact to adjust the stringency of the standard to reflect the severity of the offense and the costs associated with error may be a virtue worth fostering.

TEXT:

[*769] Few legal expressions are as familiar to nonlawyers as the expression *beyond a reasonable doubt*. The phrase is used frequently and in a wide variety of contexts. For example, in a recent *Sports Illustrated* article on the Masters Golf Tournament, John Walters (1998) wrote, "In each of the last two years, most of us thought Saturday's would be, *beyond a reasonable doubt* [italics added], the de facto final round of golf's most celebrated tournament" (p. 26). In a letter to the editor published in *Windows Magazine*, Mark Yannone (1996) described a computer user's decision to adopt a newer version of a piece of software: "We don't let go [of the old version] until a newer version proves *beyond a reasonable doubt* [italics added] that it will reliably do what the previous version did" (p. 28). n1 The difficulty for the law is that wide use of and familiarity with a phrase do not ensure accurate legal understanding and appropriate application of the standard. Indeed, a long train of cases and a large set of scholarly writings, empirical and nonempirical, reveal a continuing struggle with the meaning of the phrase and the appropriate way to convey that meaning to the jury.

What are the qualities of an optimal jury instruction on the legal standard of [*770] proof beyond reasonable doubt? In this article, we describe four criteria that have been delineated explicitly or implicitly by legal authorities and social scientists: (a) Absolute certainty should not be required, (b) a high threshold for conviction should be specified, (c) the beyond a reasonable doubt standard should be distinguishable from lower standards of proof, and (d) the instruction should encourage consistent application by jurors sitting on the same case. To these four we add a fifth, perhaps unexpected, criterion and argue that the attractions of flexibility undermine arguments that some scholars have made for the quantified approach to explaining reasonable doubt. We suggest that in the enthusiasm for offering the jury sufficient guidance and the desire to reduce unwarranted inconsistency, some appropriate elasticity in the reasonable doubt standard may be lost. Thus, according to our proposed fifth criterion, an optimal instruction on reasonable doubt should leave some limited room for the fact-finder to adjust the standard in response to consequences of error that may differ across cases. Rigid definitions, such as those that are quantified, may actually produce undesirable consistency across cases that warrant differential interpretation of the standard. We argue that the proposal by the Arizona Supreme Court Committee on the More Effective Use of Juries (1994) to provide jurors with information about the potential penalties associated with conviction reflects a recognition that the standard of reasonable doubt should retain some flexibility so that it can be responsive to varying costs of error. What is reasonable depends on the consequences of the decision, and attempts to provide clear instructions should not define away the flexibility in the beyond a reasonable doubt standard in a blind drive for precision. We begin by discussing the four well-supported attributes of an optimal reasonable doubt standard.

1. Absolute Certainty Should Not Be Required

Courts have long recognized that proof beyond a reasonable doubt means a high degree of proof, but not absolute certainty. The genesis of this distinction was in 17th century England, when John Wilkins and other philosophers rejected the notion that absolute certainty and moral certainty were one and the same (Morano, 1975). Wilkins and other philosophers argued that no one could be absolutely sure of anything and that moral certainty, in contrast to absolute certainty, does not require such a high degree of proof (Morano, 1975). This philosophy influenced later legal thought. Gilbert (1756) reasoned that a juror must be morally certain of a defendant's guilt before handing down a conviction. He viewed moral certainty as based on "proper senses" (p. 2) and concluded that

jurors need not be absolutely certain of a defendant's guilt, that is, certain beyond any doubt. The reasonable doubt test thus replaced the any doubt test that had been previously used in legal decision making (Morano, 1975).

In one of the earliest cases embracing the reasonable doubt rule in the United States, *Giles v. State* (1849), the Supreme Court of Georgia held that the proper threshold degree of belief is not a mere preponderance of the evidence but also not "absolute mathematical or metaphysical certainty" (p. 285) and that a doubt sufficient to preclude conviction is not "mere conjecture, a fancy, a trivial supposition, a bare possibility of innocence" (p. 285). Soon after, Chief Justice Shaw of the Supreme Judicial Court of Massachusetts wrote a definition in [*771] *Commonwealth v. Webster* (1850) that was adopted by many courts in the 1800s. It explicitly recognized that certainty was not required: "What is reasonable doubt? . . . It is not mere possible doubt; because everything relating to human affairs, and depending on moral evidence, is open to some possible or imaginary doubt" (*Commonwealth v. Webster*, 1850, p. 320). Thus, consistent with these philosophical and legal perspectives, an optimal instruction should neither explicitly or implicitly suggest to jurors that absolute certainty is required.

2. A High Threshold for Conviction Should Be Conveyed

The absence of absolute certainty as a requirement does not mean that a conviction should occur without strong evidence that the defendant is guilty. In criminal cases, the Due Process Clause of the Fifth and Fourteenth Amendments, as well as the Sixth Amendment, of the U.S. Constitution require that no defendant be convicted of a crime unless the government proves each and every essential element of the crime beyond a reasonable doubt (*In re Winship*, 1970; *Sullivan v. Louisiana*, 1993). This requirement of proof beyond a reasonable doubt, at least when properly applied, is characterized as a way to set the probability of an erroneous conviction at a low level. n2 It is thus a vital safeguard for protecting the life and liberty of the defendant (*Davis v. United States*, 1895) as well as an indispensable instrument for commanding respect and confidence in the criminal justice system (*In re Winship*, 1970). To ensure that the importance of the standard is impressed on the jury, the Supreme Court has held that juries must be told to use the reasonable doubt standard of proof (*Jackson v. Virginia*, 1979); however, the Court has not mandated that the standard be defined, though it has not forbidden trial courts from doing so (*Victor v. Nebraska*, 1994, Justice Ginsburg concurring in part and concurring in judgment). What is required is that "taken as a whole, the instructions correctly convey the concept of reasonable doubt" (*Victor v. Nebraska*, 1994, pp. 8-9). The instructions should communicate to jurors that they must have a high degree of certainty before they convict a defendant.

On several occasions, the Supreme Court has examined whether particular jury instructions adequately conveyed the appropriately high threshold necessary for conviction. In *Cage v. Louisiana* (1990), the defendant was convicted of first-degree murder and sentenced to death. During the guilt phase of the trial, the jury was instructed to use the proof beyond a reasonable doubt standard, and the instruction defined reasonable doubt as, among other things, "such doubt as would give rise to a grave uncertainty" and "an actual substantial doubt" (p. 40). The Louisiana Supreme Court affirmed Cage's conviction, arguing that jurors would have correctly interpreted the definition taken as a whole. However, the U.S. Supreme Court held that the instruction did not properly express the threshold. Both an actual substantial doubt and grave uncertainty suggested a greater degree of doubt than what is actually required to acquit a criminal defendant. That is, these descriptions impermissibly lowered the standard of proof. In a later case, [*772] *Sullivan v.*

Louisiana (1993), the defendant protested use of an instruction that was almost identical to the *Cage* instruction. The Court again held that these instructions impermissibly lowered the standard of proof, violating the defendant's due process rights as well as denying the defendant his Sixth Amendment right to a jury finding of guilt.

On other occasions, the Court has rejected complaints about the level of certainty specified in instructions on reasonable doubt. In *Sandoval v. California* (1994), the Court rejected the defendant's claim that describing a reasonable doubt as "not a mere possible doubt" (p. 8) overstated the degree of doubt necessary for an acquittal. The Court did not find fault with this phrase, concluding that jurors would have interpreted mere possible doubt to be a doubt that is imaginary or fanciful and thus not sufficient to warrant an acquittal, rather than a doubt based on reason. The Court concluded that the jury's confusion would have been eliminated because jurors were also instructed that everything "is open to some possible or imaginary doubt" (p. 17). In a companion case to *Sandoval*, *Victor v. Nebraska* (1994), the defendant claimed that the use of the phrase "actual and substantial doubt" (p. 18) was ambiguous, potentially interpretable as a large degree of doubt, and thus unconstitutional. Citing *Webster's Third New International Dictionary*, the Court acknowledged (p. 19) that *substantial* has two meanings: "not seeming or imaginary" and "that specified to a large degree" (*Victor v. Nebraska*, 1994, p. 19). The latter interpretation would have overstated the degree of doubt necessary to acquit. The Court reasoned that in this case, context made it clear that *substantial* referred to the kind of doubt rather than to the degree of doubt. In the instruction, a substantial doubt was contrasted with "a doubt arising from mere possibility, from bare imagination, or from fanciful conjecture" (p. 20). The Court distinguished this instruction from the one found unconstitutional in *Cage* (1990), in which no contrast was provided. Further, the Court argued that in any event, the instruction in *Victor* provided jurors an alternative definition they could have relied on: that a reasonable doubt is one that would cause one to "hesitate to act" (p. 20). Finally, the defendant also claimed that the reference to "strong probabilities" (p. 22) understated the standard of proof, but the Court held that the phrase was acceptable because the instruction went on to require that probabilities be strong enough to meet the standard of proof.

Writing separately in *Victor* (1994), Justice Ginsburg agreed with the majority that the instructions satisfied the Constitution, although she found fault with the hesitance-to-act language and instructions that define reasonable doubt as "doubt . . . that is reasonable" (p. 25). Justice Ginsburg endorsed a definition proposed by the Federal Judicial Center (1987), which she described as both clear and accurate. That instruction reads in part,

> Proof beyond a reasonable doubt is proof that leaves you firmly convinced of the defendant's guilt. There are very few things in this world that we know with absolute certainty, and in criminal cases the law does not require proof that overcomes every possible doubt. If, based on your consideration of the evidence, you are firmly convinced that the defendant is guilty of the crime charged, you must find him guilty. If on the other hand, you think there is a real possibility that he is not guilty, you must give him the benefit of the doubt and find him not guilty. (Federal Judicial Center, 1987, Instruction 21).

[*773] The majority in *Victor* found that the instructions "taken as a whole" (p. 22) expressed the standard of proof properly, though the Court did not condone parts of the instruction, such as the use of the phrase *moral certainty*. The implication of this set of cases is that language conveying a

standard that potentially requires less than substantial certainty is constitutionally objectionable but that courts may permit instructions that contain somewhat ambiguous language if qualifying language appears to supply sufficient context to overcome that ambiguity. In all of these cases, the Court relied on its own speculations about how jurors were likely to interpret the instructions.

The Court's focus on the terminology in instructions is justified by empirical evidence on jury comprehension of instructions involving reasonable doubt, although some of the Court's conclusions about how juries would interpret the various instructions have been challenged by recent empirical research. Horowitz and Kirkpatrick (1996) gave six-person juries one of five different instructions on reasonable doubt and also varied whether the case favored either a conviction (a strong case) or an acquittal (a weak case). They reasoned that if the reasonable doubt instructions were effective, then jurors given the weak case should have been less likely to convict the defendant than jurors given the strong case. Results of individual predeliberation measures revealed that some instructions, such as those that left reasonable doubt undefined and those that described the requisite belief in guilt as that which is stable and does not waver or vacillate, did not lead jurors to give higher guilt scores when the case was strong as opposed to weak. In contrast, jurors given the Federal Judicial Center instructions (those that Justice Ginsburg endorsed in *Victor*, 1994) and instructions informing jurors that they must have a real doubt had higher guilt scores in the strong case than in the weak case, as one would expect if instructions were performing well. Moreover, after deliberating on both the strong and weak cases, jurors given the Federal Judicial Center's instructions reported higher minimum probabilities for conviction than did jurors given any of the other four instructions. This pattern suggests that when jurors are not provided with adequate criteria, their verdicts are less likely to reflect the appropriate level of certainty or be sensitive to differences in the strength of the evidence. Thus, both Judge Ginsburg in *Victor* and recent empirical evidence indicate that the Federal Judicial Center's recommended instruction fares significantly better than other possible definitions in appropriately signaling to jurors that beyond a reasonable doubt establishes a high threshold for conviction. n3

3. The Reasonable Doubt Standard Should Be Distinguishable From Lower Standards

The third property of an optimal reasonable doubt instruction is that it produces a standard that is distinguishable from standards of proof intended to convey a less stringent degree of certainty. This feature is actually an extension of the requirement that the instruction convey a high threshold for conviction, [*774] reinforcing language conveying a high threshold with a concrete comparison to alternative and lower standards. The value of such comparisons is that they directly confront the potentially misleading expectations and experiences that jurors may bring to their task.

Although jurors in criminal cases are instructed to use the standard proof beyond a reasonable doubt, jurors in most civil cases are given the standard *preponderance of the evidence*. In other special civil cases, jurors are instructed to use a *clear and convincing* standard of proof that is presumably a higher threshold than preponderance of the evidence and a lower threshold than beyond a reasonable doubt. The stringency of the standard is analogous to the stringency of an alpha level in statistical analysis: The standard or *p* level is set to reflect one's interest in ensuring that there are few false positives (Kerr et al., 1976). n4

In criminal cases, society's interest in ensuring that innocent individuals are not convicted is presumed to be greater than society's interest in ensuring that guilty individuals are convicted. Accordingly, jurors are asked to apply the standard of proof beyond a reasonable doubt, and the state bears more of a risk of an erroneous acquittal than if the standard were lower (*Addington v.*

Texas, 1979; *In re Winship*, 1970; *Santosky v. Kramer*, 1982). Proof beyond a reasonable doubt has been estimated by some judges to mean that the trier of fact must find a fact at issue to be at least 90% to 95% likely (McCauliff, 1982; Simon, 1969; *United States v. Fatico*, 1978). In contrast, most civil cases are purely private actions in which society's concern that a defendant may erroneously be held liable is not as severe (*Addington v. Texas*, 1979; *United States v. Fatico*, 1978). Triers apply the preponderance of the evidence standard so that plaintiffs and defendants share the risk of an erroneous decision more equally than they would if the standard were higher (*Santosky v. Kramer*, 1982). n5 This standard has been interpreted to mean that the trier must believe that the plaintiff's claim is more than 50% likely if the trier of fact is to find for the plaintiff (McCauliff, 1982; Simon, 1969; *United States v. Fatico*, 1978). In some civil cases, the standard is higher than the preponderance of the evidence standard but not as high as the beyond a reasonable doubt standard. When the government initiates a civil suit, a "quasi-criminal" case in which individual liberty is at stake and stigma may be attached, societal interest in avoiding an erroneous verdict in favor of the plaintiff is more substantial, and triers are instructed to apply the "clear and convincing evidence" standard or the "clear, unequivocal and convincing evidence" standard (*United States v. Fatico*, 1978, p. 405). In those cases, it is anticipated that the state will bear more of the risk of error than if the standard were preponderance of the evidence, but not as much risk as it would if the standard were proof beyond a reasonable doubt. In probability terms, judicial surveys suggest that triers of fact applying a clear and convincing evidence standard must find the plaintiff's claim to be at least 70% to 80% likely (McCauliff, 1982; *United States v. Fatico*, 1978).

Empirical research indicates that jurors may have some difficulty distinguishing [*775] among the standards. In five experiments, Kagehiro and her colleagues (Kagehiro, 1986; Kagehiro & Rosen, 1986; Kagehiro & Stanton, 1985, Experiments 1 & 2; for a summary, see Kagehiro, 1990) provided student participants with legal definitions for one of the three standards of proof: preponderance of the evidence, clear and convincing, or beyond a reasonable doubt. The particular wording of the verbal instructions on the standards varied across the experiments. Participants read a trial summary describing a civil case and then rendered individual verdicts. Although the results varied across the five experiments, participants in general were no more likely to find for the defendant when instructed that the standard was beyond a reasonable doubt than when they were instructed using a clear and convincing standard; in some instances, conviction rates did not differ between conditions in which instructions provided a preponderance of the evidence standard and those in which instructions provided a beyond a reasonable doubt standard. Some, but not all, of the definitions of beyond a reasonable doubt used in these studies included language that the Supreme Court has found objectionable (e.g., moral certainty).

As a whole, the set of studies suggests that the terminology often used to define the clear and convincing standard frequently is not perceived as distinguishable from terminology often used to define the beyond a reasonable doubt standard. For example, in two experiments, Kagehiro and Stanton (1985) used as the definition for clear and convincing "that measure or degree of proof which will produce in the mind of the juror a firm belief or conviction as to the truth of the allegation sought to be established" (*State v. Addington*, 1979, p. 570). In their first experiment, the reasonable doubt standard used the moral certainty definition that the Supreme Court has expressed concern about, and in their second experiment, the reasonable doubt standard was a somewhat abbreviated version of the 1982 Federal Judicial Center's definition, n6 which used overlapping language on the crucial degree of certainty required (i.e., "proof that leaves you *firmly* [italics added] convinced of the plaintiff's case," p. 27). In neither experiment did respondents differ in their

willingness to find for the plaintiff or in their assessment of how difficult it was for the plaintiff to meet that standard of proof.

In a third experiment, Kagehiro and Stanton (1985) found evidence that some sets of legal definitions can perform better than others, at least on some measures. They compared the performance of five different sets of definitions for the three standards, reasonable doubt, clear and convincing, and preponderance of the evidence. Each respondent evaluated the meaning of one set of definitions for the three standards of proof. Participants who received the set of instructions that included the Federal Judicial Center's (1982) definition of reasonable doubt, n7 the federal pattern definition of preponderance of the evidence (Devitt & Blackmar, [*776] 1977), n8 and the Kansas pattern definition of clear and convincing evidence (Kansas District Judges' Association Committee on Pattern Jury Instructions, 1997, § 2.11) n9 were able to differentiate among the standards of proof better than did those who received any of the four other sets of instructions. Further, the mean rating of the difficulty required for the plaintiff to win under a reasonable doubt standard for those given the Federal/Kansas set was higher than for those given any of the other four sets of instructions.

The Federal Judicial Center's (1987) instruction on reasonable doubt that Justice Ginsburg cited with approval in *Victor* (1994) pointedly includes the same kind of information that the participants received in the third Kagehiro and Stanton (1985) experiment: It directly draws the juror's attention to the contrast with the standard applicable in ordinary civil cases. Thus, the Federal Judicial Center's (1987) instruction explicitly says,

> The government has the burden of proving the defendant guilty beyond a reasonable doubt. Some of you may have served as jurors in civil cases, where you were told that it is only necessary to prove that a fact is more likely than not true. In criminal cases, the government's proof must be more powerful than that. It must be beyond a reasonable doubt. (Instruction 21)

By providing this explicit contrast with a less stringent standard of proof, the definition encourages jurors to adopt an appropriately high threshold for conviction. It could be strengthened even further by adding an additional contrast with clear and convincing, as follows (additions in italics):

> . . . the government's proof must be more powerful than that. It *must even be more powerful than "clear and convincing evidence," which must be persuasive enough to cause you to believe it. The government's proof in this case* must be beyond a reasonable doubt.

Fact-finders naturally draw on their experiences in making sense of the case they are being asked to decide, whether it is in using their image of the prototypical burglary to determine whether a burglary was committed (Smith, 1991, 1993) or in deciding the meaning of a jury instruction (see Diamond, 1993; Diamond & Levi, 1996). The contrast among standards of proof is unlikely to be familiar to most jurors, but explicitly presenting the contrast has two benefits. [*777] First, if a juror has been exposed to a case with a lower standard of proof, the instruction alerts the juror to the difference. Second, the instruction informs the juror directly that other decisions of some consequence are made on the strength of less evidence, reinforcing the idea that the threshold for the reasonable doubt standard that the jury is to apply in the current case is particularly high.

4. The Reasonable Doubt Instruction Should Not Invite Variability in the Standard Applied by Jurors on the Same Case

The fourth quality of an optimal instruction is that it does not leave jurors hearing the same case with unfettered discretion in interpreting how much proof is required for a conviction. Even if jurors understand that the threshold for reasonable doubt is high, although not absolute, they may still vary in their interpretation of what constitutes a high threshold, producing substantial interjuror variability. Potential within-jury (and between-juries) variability was the basis for some of the objections by the defendants in *Cage* (1990) and *Victor* (1994)/*Sandoval* (1994) regarding the use of the words *moral certainty* and *moral evidence*. In *Sandoval*, the defendant argued, among other things, that the terms drew the jury's attention to the ethicality or morality of the crime, not the evidence in the case, and would have caused the jurors to decide the case on the basis of their own individual passions and prejudices, introducing arbitrary decision making. Although stating that it did not condone use of such archaic language, the Court decided that the meaning of moral certainty today is no different than its original meaning, which is that moral certainty is subjective certitude about some event rather than absolute demonstrable certitude in an event. The Court held that the instructions as a whole correctly conveyed the meaning of reasonable doubt because the phrase was used in conjunction both with explicit instructions admonishing jurors to attend to the case facts (and not, for instance, to be influenced by pity, mere sentiment, conjecture, etc.) and with language that equated proof beyond a reasonable doubt to an "abiding conviction" (p. 14). The Court contrasted this with the instructions in *Cage*, which they argued did not provide a similar essential context.

As in *Sandoval* (1994), the defendant in *Victor* (1994) protested use of the phrase *moral certainty*. Again, the Court found that the rest of the instruction, particularly the "hesitate to act" (p. 21) and "abiding conviction" (p. 21) language, lent sufficient meaning to the phrase. Without that context, such references might have invited jurors to refer to their own moral beliefs when deciding the case, potentially inviting an unacceptable level of interjuror variability in interpretations of the standard of proof.

Empirical evidence indicates that some instructions are in fact better than others in discouraging jurors from relying on their moral beliefs or personal experiences in making legal decisions. Horowitz and Kirkpatrick (1996) found that jurors given the full Federal Judicial Center (1987) instructions on reasonable doubt spent more time discussing the evidence and less time discussing, for example, personal experiences and news reports of other cases than jurors given the other instructions. They hypothesized that jurors given moral certainty instructions would give weight to their personal moral standards and in fact found that jurors given the moral certainty instruction discussed punishment more than [*778] did others when hearing the strong case (but not the weak case). Horowitz (1997) also reported that juries given the moral certainty instructions tended to vilify the murder defendant more than did those given other instructions. Finally, in a telling demonstration of the guidance provided by the Federal Judicial Center's instruction, jurors who were given it showed less variability in their verdict scores (on a 1-to-6-point scale of guilt) than jurors who were given other instructions--the most direct indicator of reduced interjuror variability in response to the same case. Thus, the fourth quality of an optimal instruction is one that does not invite unjustified variation in people's interpretations of the standard of proof on the same case. People of course are likely to differ in their perceptions of the probative value of a particular set of facts. They may also differ in their perceptions of the stringency of a standard, whether expressed in words or as a probability value. The point here is that the definition of beyond a reasonable doubt

should not provide a foundation for idiosyncratic interpretations of the level of certainty required for a conviction. Language like *moral certainty* offers precisely such an invitation.

On both theoretical and empirical grounds, the Federal Judicial Center's (1987) instruction appears to perform better than other instructions on all four of the criteria we have identified to this point as necessary for an optimal instruction on reasonable doubt. It does not require absolute certainty, but it does specify a stringent standard for conviction. It includes a specific contrast with one lower standard of proof, and it avoids language that encourages inconsistent application of the standard by jurors sitting on the same case. Empirically, the Federal Judicial Center's definition has performed well when compared with other verbal definitions of the standard. We turn now to the final, more controversial, criterion.

5. The Reasonable Doubt Instruction Should Leave Room for Flexible Tailoring of the Standard to the Costs of Error

Efforts to guide and control jury decision making are a central feature of jury instructions (Diamond, Casper, & Ostergren, 1989). It was in part the failure of jury instructions to provide adequate guidance that led the United States Supreme Court to find the death penalty unconstitutional in *Furman v. Georgia* (1972). n10 Yet the Court was unwilling to remove discretion from the jury and went on to invalidate efforts by some states to guide capital sentencing by delineating offenses that would carry a mandatory death penalty (*Roberts v. Louisiana*, 1976; *Woodson v. North Carolina*, 1976). States were left with the task of determining how to strike a balance between providing guidance that would prevent unwarranted inconsistency and avoiding rigidity that would preclude jurors from considering legally relevant attributes.

A similar issue arises if the standard for reasonable doubt is viewed as a problem [*779] of both comprehension and fit. Jurors apparently do have difficulty applying standards of proof as the law intends, at least when they are offered some verbal descriptions of those standards. As a result, some scholars have suggested that jurors should be provided with an explicit probability level that describes the standard they are to apply (see, e.g., Kagehiro & Stanton, 1985). Kagehiro and Stanton (1985) found that when they described the three standards of proof with quantitative definitions, these definitions produced differentiated means in the expected direction when respondents were asked to rate and assign percentages to the difficulty a plaintiff would have in proving his or her case. Kagehiro and Stanton concluded that quantified definitions were superior to the legal definitions in their ability to increase defense verdicts as the standard becomes stricter, just as the law intends. n11

Courts, in contrast, have been reluctant to use quantified definitions (Kagehiro & Stanton, 1985; Simon, 1969; Simon & Mahan, 1971). In *McCullough v. State* (1983), the trial judge located the standards on a scale from 0 to 10 to indicate the relative certainty required. The judge placed preponderance of the evidence at just above 5 and put reasonable doubt at 7.5. The Nevada Supreme Court reversed, not only objecting to the particular number the judge chose for reasonable doubt but also holding that the term reasonable doubt is "inherently qualitative" (p. 75) and reasoning that quantification would confuse jurors and lower the standard of proof.

Simon and Mahan (1971; see also Simon, 1969) surveyed judicial reactions to a procedure similar to quantified reasonable doubt instructions. In this hypothetical system, jurors would be asked to estimate the likelihood that a defendant committed the act with which he or she was charged, and then the court would compare the assigned probability to the probability required by

the standard of proof. A guilty verdict would be handed down only if the jury's assigned probability exceeded the court-determined threshold probability. Over 90% of the judges in the survey objected to this procedure. One judge argued, "Percentages or probabilities simply cannot encompass all the factors, tangible or intangible, in determining guilt--evidence cannot be evaluated in such terms" (Simon, 1969, p. 113).

Judges in other contexts have shown some general resistance to quantitative and probabilistic bases for deciding cases. For example, Wells (1992) demonstrated judicial reluctance to find a defendant liable on the basis of a variety of forms of probabilistic evidence. An unwillingness to adopt a single quantitative standard for reasonable doubt may be justifiable, however, if it rests on an alternative view that jurors must apply a flexible standard that is tailored to reflect the characteristics of the case. n12 If it is desirable for jurors to make small adjustments in the threshold for reasonable doubt depending on the consequences of both incorrect convictions and incorrect acquittals, it may be advisable to allow for that adjustment in the standard either directly or indirectly. n13

[*780] Evidence that legal decision makers do in fact modify thresholds depending on potentially relevant case characteristics such as the severity of the crime is provided by some survey evidence. Simon (1969) instructed a sample of state and federal trial court judges to "imagine that a . . . defendant has been charged with each [of a list of various criminal acts]. In arriving at a verdict for each offense, what would the probability that the defendant committed the act have to be before you declared him guilty" (p. 110)? The crimes listed included murder, embezzlement, grand larceny, forcible rape, petty larceny, and fraud. Simon argued that consistent application of the standard would mean judges would assign equal probability thresholds to the crimes regardless of the severity of crime. Instead, Simon found systematic variation rather than consistency: Judges on average assigned higher threshold probabilities to more severe crimes as opposed to less severe crimes. The probability thresholds that students and jury-eligible citizens who participated in lieu of jury duty assigned to the various crimes varied even more than the probability thresholds judges assigned to the various crimes (Simon & Mahan, 1971).

Some experimental data also support the notion that juror decision making is influenced by crime and penalty severity. For instance, Kerr (1978) asked participants to read a trial summary describing a case of first-degree murder, seconddegree murder, or manslaughter. Participants were also told the penalty that the defendant would receive on conviction: either a moderate penalty (1--5 years or 5--20 years) or a severe penalty (25 years to life or capital punishment). As predicted, participants were less likely to convict when told that the defendant would incur the more severe penalty. Further, participants were more likely to convict defendants accused of second-degree murder than defendants accused of first-degree murder. Participants were no more likely to convict defendants accused of manslaughter than defendants accused of second-degree murder, although the means were in the predicted direction. Further, the conviction criterion that participants assigned was higher (that is, participants required more proof) when the penalty was severe than when it was mild.

Freedman, Krismer, MacDonald, and Cunningham (1994; Freedman, 1994) pointed out that some previous studies of the effects of severity of crime on verdicts confounded severity with the amount of evidence required for a conviction. They argued that although some studies have shown that there are fewer convictions when crimes are more severe, those results may be due to the fact that severe crimes typically have more elements that need to be proven beyond reasonable doubt,

and therefore it is not surprising that jurors would be less likely to convict defendants who commit severe crimes than those who commit milder crimes. Moreover, the cost of an erroneous conviction for the defendant may rise with the severity of the offense, but society may also incur greater costs from the erroneous acquittal if the defendant has actually committed a more serious offense.

Freedman et al. (1994) attempted to disentangle severity and the number of elements of the crime and did not find that verdicts varied as a function of crime severity. More surprisingly and inconsistent with others' research, they found no independent effect of penalty severity. As Kaplan (1994a; 1994b) pointed out, however, several variables may moderate the effect of penalty severity on verdicts, producing these apparently inconsistent results. For example, Kaplan and [*781] Krupa (1986) found evidence suggesting that penalty severity influences verdicts when the evidence of guilt is relatively weak (rather than strong), the case is real (rather than simulated), and a higher authority (rather than the fact-finders) will decide the sentence if the defendant is convicted.

If triers of fact adjust the standard of proof that they apply in response to the severity of the offense they are judging or the potential penalty that they expect a convicted offender to receive, that empirical reality does not answer the question of how the legal system should respond. If such adjustments are deemed inappropriate, a single quantitative probability value might offer the prospect of reducing them. The remaining and by no means insignificant challenge would be to determine just what probability level should be applied across all criminal cases. Alternatively, if a response sensitive to offense severity or potential penalty is justifiable, a definition that leaves room for such a context-specific response may be not only permissible but preferable.

The potential influence that sentence severity can have on the standard used for conviction highlights these two sources of variability, one undesirable and the other potentially appropriate. The undesirable effect arises from the varying expectations that jurors are likely to carry about the probable penalty associated with the case they are deciding. Some jurors (and some juries) may assume that a defendant convicted of sexual assault will be released after a brief incarceration, while others may anticipate a long prison term. If these different expectations induce different jurors (and juries) to employ different thresholds for reasonable doubt, then identical defendants tried by juries that happen to have different penalty expectations would have different chances for acquittal. In contrast, if jurors shared accurate information on the severity of the potential penalty, it could offer a context-sensitive meaning to the reasonable doubt standard by alerting jurors to the actual costs likely to be incurred by an erroneous conviction. If the standard then varied so that, for example, jurors deciding cases that they knew could carry a life sentence applied a higher standard than jurors deciding cases that could carry a 2-to-5-year prison sentence, such variability in the standard might be appropriate.

There are two particular ways in which providing jurors with knowledge about the potential sentencing range might affect the jury's threshold for conviction and arguably improve the quality of the decision it reaches. The first arises with increasing frequency under the mandatory minimum sentencing provisions that many states have today. The legislatively sanctioned penalty can be quite severe in light of the facts in a specific case (e.g., a minimal quantity of marijuana sold by a young first offender prosecuted under a statute that provides for a mandatory prison sentence). Jurors who assume that the sentence would be probation might not set their threshold for reasonable doubt at a particularly stringent level. Their expectation about the leniency of the likely punishment may increase their willingness to convict a defendant who is very likely to be guilty in order to teach the likely offender a lesson. The same jurors would be unwilling to take a small chance that the

defendant was innocent if the penalty were as severe as a prison term. While jurors should not lawlessly undercut the decisions of the legislature by nullifying in the face of substantial mandatory penalties, flexibility in the reasonable doubt standard and knowledge of the costs of error leave them [*782] room within the law to manage the inherent uncertainty of the decision in a rational way.

The opposite adjustment in jury decision making may also occur when jurors make inferences about the severity of the potential sentence. Bartels (1981) described a trial in which a juvenile was acquitted by a jury of reckless driving even though the evidence appeared to be very convincing. The judge was left with little doubt as to the juvenile's guilt and was so disturbed by their verdict that he questioned the jurors about their reasoning. The jurors reported that they believed (incorrectly) that the defendant would be sent to the state prison and that such a punishment would be disproportionate to the offense and would increase the likelihood that the juvenile would become a criminal. It is unclear whether the jurors would have acquitted no matter how much proof they faced of the defendant's guilt, in light of their concern over what they believed would be disproportionate punishment. Nonetheless, it is reasonable to assume that to convict, these jurors would have at least required extremely convincing evidence of the defendant's guilt.

Jurors in the United States generally do not participate in criminal sentencing, except in capital cases. Unlike civil disputes in which the jury typically determines both liability and the remedy (i.e., damages), the American criminal jury usually completes its work with a verdict on guilt. Jurors do, however, speculate on sentencing, as they speculate on many consequences of their verdicts (e.g., how much attorneys will be paid, whether there is insurance). The legal system operates as if ignoring those speculations or simply admonishing the jury to refrain from considering them will result in a blindfolded jury uninfluenced by forbidden considerations. As Diamond and Casper (1992) have shown, a more collaborative model of jury treatment can both avoid inaccurate speculation and control unwarranted use of potentially biasing information. Failing to inform jurors about potential penalties thus carries a double cost: It misses the opportunity to correct juror misunderstanding and to channel juror behavior by providing accurate information along with an explanation of how the information is to be used, and it ignores what could be an occasion for tailoring the standard of proof to costs of error in the particular case.

The promise of precision that a specific quantitative standard appears to offer is inconsistent with a more tailored approach to decision making that reflects the different costs of error associated with wrongful convictions for particular offenses. An alternative to the single probability standard for beyond a reasonable doubt might be to provide a range (e.g., .87 to .92) that jurors would be invited to apply according to their assessments of the costs of error associated with a particular offense. Alternatively, a legislature might set different quantitative levels for different offenses. Thus, adopting the median standard (the mean was identical) suggested by the judges in Simon and Mahan's (1971) study, petty larceny would carry a probability standard of .87, whereas murder would require a standard of .92. If, as Simon and Mahan's data suggest, jurors tend to vary more in the standards they apply across offenses (a median of 7.5 out of 10 for petty larceny and 9.5 out of 10 for murder), specifically presenting them with a quantitative standard, albeit one that varies by offense, might reduce some undesirable disparity and still maintain some tailoring of the standard to case severity. The challenge here is to arrive at a carefully calibrated metric that [*783] sufficiently reflects both the severity of the offense and the costs associated with error.

An alternative approach is to dispense with the idea of introducing a quantitative standard at all. Jurors would be given tailored guidance in the form of a carefully crafted nonquantitative definition of reasonable doubt like the one suggested by the Federal Judicial Center, along with a picture of the costs associated with an unwarranted conviction in the form of the potential sentence the defendant could receive (e.g., "The decision you are about to make is a very important one. You should know that if you find the defendant guilty, I will be required by the law to sentence the defendant to a sentence of between ___ and ___").

Refraining from quantification and other precise, rigid definitions allows jurors to bring to their decision-making process a case-specific analysis of the disutilities associated with convicting the innocent and acquitting the guilty. This view does not clash with the defendant's constitutional right to have all elements of the crime with which he or she is charged proven beyond a reasonable doubt. Jurors may dutifully employ a reasonable doubt standard of proof in criminal cases, considering the consequences of error as a basis for adjusting their interpretation of what constitutes a reasonable doubt under the circumstances. Reasonable doubt is sometimes described as a doubt that would cause one to hesitate to act. Although a doubt may be unreasonable and not worthy of hesitation when the sentence is no more than a year in prison, that same doubt may be perfectly reasonable and worthy of hesitation when a conviction produces a sentence to life in prison without the possibility of parole.

The Arizona Supreme Court Committee on the More Effective Use of Juries (1994) recommended that jurors be told prior to deliberation the potential range of sentences a defendant would face on conviction. n14 The committee reasoned that jurors need to know the possible penalties to evaluate more effectively whether the standard of proof has been met, arguing that jurors are likely to be, and that it is normatively appropriate for them to be, more circumspect as the penalties become more severe. The committee also pointed out that jurors often feel shocked, angry, and betrayed after they have convicted a defendant and they learn the penalty that the judge had imposed. Full disclosure of the range of penalties, the committee argued, would produce more fully informed verdicts and greater confidence in both the verdict and the criminal justice system.

The merits of this proposal depend on the pattern and size of the effects that this knowledge is likely to have. n15 Thus, the proposal deserves further empirical investigation. The potential for jury displeasure would not justify providing jurors [*784] with information that would satisfy their curiosity but impermissibly affect their decisions. The examples discussed earlier, however, indicate that part of juror dismay may stem from the fact that jurors recognize that their decisions have been influenced by misinformation about and disagreement with the penalties they thought that the defendant could receive.

A flexible response to information on the severity of the potential penalty does not mean that jurors should treat an erroneous conviction of any defendant as without significant cost--or that jurors should demand an unreasonable level of certainty when the penalty is particularly severe. From the available evidence, however, there is little reason to expect such extreme responses. The proposal offers an innovative way to correct misinformation that currently can influence juror judgments in an arbitrary fashion. At the same time, it reflects the view that reasonable doubt is an inherently flexible standard, one that is not tied to a single probability level. Although consistency across identical cases is an accepted and uncontroversial goal for legal decisions, it does not include consistency across different cases. The goal of consistency is distorted when it is used to justify imposing a rigid single standard for managing uncertainty across different cases that differ

dramatically in the costs of errors that are unavoidable potential consequences of any legal decision. Sharing information with the jury rather than blindfolding it to those consequences offers an intriguing way to encourage the application of an appropriately flexible standard for reasonable doubt. If we recognize that the jury is an active decision maker, rather than a passive recipient of evidence and standards, we can anticipate more accurately what the jury is likely to do and not attempt to channel and control its behavior in undesirable and ineffective ways.

The Federal Judicial Center's proposed definition for the beyond a reasonable doubt standard represents a substantial improvement over earlier and present definitions that suffer from ambiguous wording or suggest a weak standard for conviction. Both legal and empirical work suggest that the Federal Judicial Center's definition substantially meets all of the first four criteria we have described as necessary for an adequate definition of the beyond a reasonable doubt standard. The Federal Judicial Center's definition does not go as far as a single quantitative value might in producing a single uniform standard. As we have suggested, however, this apparent imperfection may yet be an advantage for meeting our fifth criterion. A single uniform standard across cases is not an optimal resolution when the decisions to which the standard is being applied carry different costs. A uniform standard brings with it undesirable rigidity, whereas the flexibility inherent in the Federal Judicial Center's definition preserves the ability of jurors to tailor the meaning of the standard to the costs of error. A jury provided with this definition and informed about the costs of error (see the Appendix for a suggested model instruction) may be optimally situated to make a reasoned choice that is sensitive to context and consistent with the law.

References

Addington v. Texas, 441 U.S. 418 (1979).

Arizona Supreme Court Committee on the More Effective Use of Juries. (1994). *The Power of 12*.

[*785] Bartels, R. D. (1981). Punishment and the burden of proof in criminal cases: A modest proposal. *Iowa Law Review, 66*, 899--930.

Bell, R. S. (1987). Decision theory and due process: A critique of the Supreme Court's lawmaking for burdens of proof. *Journal of Criminal Law and Criminology, 78*, 557--585.

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

Commonwealth v. Webster, 59 Mass. 295 (1850).

Dann, B. M., & Logan, G. (1996). Jury reform: The Arizona experience. *Judicature, 79*, 280--286.

Davis v. United States, 160 U.S. 469 (1895).

DeKay, M. L. (1996). The difference between Blackstone-like error ratios and probabilistic standards of proof. *Law and Social Inquiry, 21*, 95--132.

Devitt, E. J., & Blackmar, C. B. (Eds.). (1977). *Federal jury practice and instructions* (Vol. 2). St. Paul, MN: West.

Diamond, S. S. (1993). Instructing on death: Psychologists, juries, and judges. *American Psychologist, 48*, 423--434.

Diamond, S. S., & Casper, J. D. (1992). Blindfolding the jury to verdict consequences: Damages, experts, and the civil jury. *Law and Society Review, 26*, 513--563.

Diamond, S. S., Casper, J. D. & Ostergren, L. (1989). Blindfolding the jury. *Law and Contemporary Problems, 52*, 247--267.

Diamond, S. S., & Levi, J. (1996). Improving decisions on death by revising and testing jury instructions. *Judicature, 79*, 224--232.

English, P. W., & Sales, B. D. (1997). A ceiling or consistency effect for the comprehension of jury instructions. *Psychology, Public Policy, and Law, 3*, 381--401.

Federal Judicial Center. (1982). *Pattern criminal jury instructions*. Washington, DC: Author.

Federal Judicial Center. (1987). *Pattern criminal jury instructions*. Washington, DC: Author.

Federal Judicial Center. (1988). *Pattern criminal jury instructions*. Washington, DC: Author.

Freedman, J. L. (1994). Penalties and verdicts: Keeping the record straight. *Law and Human Behavior, 18*, 699--702.

Freedman, J. L., Krismer, K., MacDonald, J. E., & Cunningham, J. A. (1994). Severity of penalty, seriousness of the charge, and mock jurors' verdicts. *Law and Human Behavior, 18*, 189--202.

Furman v. Georgia, 408 U.S. 238 (1972).

Gilbert, G. (1756). *The law of evidence*. London: H. Lintot for W. Owen.

Giles v. State, 6 Ga. 276 (1849).

Horowitz, I. A. (1997). Reasonable doubt instructions: Commonsense justice and standard of proof. *Psychology, Public Policy, and Law, 3*, 285--302.

Horowitz, I. A., & Kirkpatrick, L. C. (1996). A concept in search of definition: The effect of reasonable doubt instructions on certainty of guilt standards and jury verdicts. *Law and Human Behavior, 20*, 655--670.

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

Jackson v. Virginia, 443 U.S. 307 (1979).

Kagehiro, D. K. (1986, May). Defining the standard of proof in jury instructions. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, IL.

Kagehiro, D. K. (1990). Defining the standard of proof in jury instructions. *Psychological Science, 1*, 194--200.

Kagehiro, D. K., & Rosen, C. J. (1986, October). Defining the standard of proof in jury instructions: II. Paper presented at the meeting of the American Society of Criminology, Atlanta, GA.

[*786] Kagehiro, D. K., & Stanton, W. C. (1985). Legal vs. quantified definitions of standards of proof. *Law and Human Behavior, 9*, 159--178.

Kansas District Judges' Association Committee on Pattern Jury Instructions. (1977). *Pattern instructions for Kansas--Civil* (2nd ed.). San Francisco: Bancroft-Whitney.

Kaplan, M. F. (1994a). Keeping the record complete. *Law and Human Behavior, 18*, 702--703.

Kaplan, M. F. (1994b). Setting the record straight (again) on severity of penalty: A comment on Freedman et al. *Law and Human Behavior, 18*, 697--699.

Kaplan, M. F., & Krupa, S. (1986). Severe penalties under the control of others can reduce guilt verdicts. *Law and Psychology Review, 10*, 1--18.

Kerr, N. L. (1978). Severity of prescribed penalty and mock juror verdicts. *Journal of Personality and Social Psychology, 36*, 1431--1442.

Kerr, N. L., Atkin, R. S., Stasser, G., Meek, D., Holt, R. W., & Davis, J. H. (1976). Guilt beyond a reasonable doubt: Effects of concept definition and assigned decision rule on the judgments of mock jurors. *Journal of Personality and Social Psychology, 34*, 282--294.

McCauliff, C. M. A. (1982). Burdens of proof: Degrees of belief, quanta of evidence, or constitutional guarantees? *Vanderbilt Law Review, 35*, 1293--1335.

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

Morano, A. A. (1975). A reexamination of the development of the reasonable doubt rule. *Boston University Law Review, 55*, 507--528.

Roberts v. Louisiana, 428 U.S. 325 (1976).

Sandoval v. California, 511 U.S. 1 (1994).

Santosky v. Kramer, 455 U.S. 745 (1982).

Simon, R. J. (1969). Judges' translations of burdens of proof into statements of probability. *Trial Lawyer's Guide, 13*, 103--114.

Simon, R. J., & Mahan, L. (1971). Quantifying burdens of proof: A view from the bench, the jury, and the classroom. *Law and Society Review, 5*, 319--330.

Smith, V. L. (1991). Prototypes in the courtroom: Law representations of legal concepts. *Journal of Personality and Social Psychology, 61*, 857--872.

Smith, V. L. (1993). When prior knowledge and law collide: Helping jurors use the law. *Law and Human Behavior, 17*, 507--536.

Solan, L. M. (1999). Refocusing the burden of proof in criminal cases: Some doubt about reasonable doubt. *Texas Law Review, 78*, 105--147.

State v. Addington, 588 S.W 569 (1979).

Sullivan v. Louisiana, 508 U.S. 275 (1993).

United States v. Fatico, 458 F. Supp. 388 (E.D.N.Y. 1978), *aff'd*, 603 F.2d 1053 (2d Cir. 1979).

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

Walters, J. (1998, April 13). Golf: The masters. *Sports Illustrated*, 26.

Wells, G. (1992). Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology, 62*, 739--752.

Wiener, R. L., Pritchard, C. C., & Weston, M. (1995). Comprehensibility of approved jury instructions in capital murder cases. *Journal of Applied Psychology, 80*, 455--467.

Woodson v. North Carolina, 428 U.S. 280 (1976).

Yannone, M. (1996, December 1). Keeping it simple. *Windows Magazine*, 28.

[*787] Appendix

Federal Judicial Center Instruction With Modifications

The Federal Judicial Center (1987) recommended instruction of reasonable doubt (Instruction 21), with suggested additions in italics, follows:

Proof Beyond a Reasonable Doubt

As I have said many times, the government has the burden of proving the defendant guilty beyond a reasonable doubt. Some of you may have served as jurors in civil cases, where you were told that it is only necessary to prove that a fact is more likely true than not true. In criminal cases, the government's proof must be more powerful than that. It *must even be more powerful than "clear and convincing evidence," which must be persuasive enough to cause you to believe it. The government's proof in this case* must be beyond a reasonable doubt.

Proof beyond a reasonable doubt is proof that leaves you firmly convinced of the defendant's guilt. There are very few things in this world that we know with absolute certainty, and in criminal cases the law does not require proof that overcomes every possible doubt. If, based on your consideration of the evidence, you are firmly convinced that the defendant is guilty of the crime charged, you must find him guilty. If on the other hand, you think there is a real possibility that he is not guilty, you must give him the benefit of the doubt and find him not guilty.

[Additional instruction to encourage flexible tailoring]

The decision you are about to make is a very important one. You should know that if you find the defendant guilty, I will be required by the law to sentence the defendant to a sentence of between ___ and ___.

Received August 10, 1998

Revision received July 2, 2000

Accepted October 3, 2000

FOOTNOTES:

n1 A search of NEXIS for the 5 years between January 1, 1995, and December 31, 1999, revealed 341 uses of the phrase in major magazines and 4,848 uses in major newspapers. Although the majority of these instances involved legal issues, a substantial number, like those in the text, applied the phrase to a nonlegal context. For example, in the year 1999, 16% of the 57 mentions of *beyond a reasonable doubt* in major magazines used the phrase in a nonlegal context.

n2 As DeKay (1996) pointed out, the ratio of errors of conviction to errors of acquittal is not solely determined by the standard of proof. The ratio also depends on juror accuracy and the proportion of defendants who are in fact guilty, that is, the base rate for guilt. (See also Bell, 1987.)

n3 See also Lawrence Solan's (1999) linguistic analysis of the Federal Judicial Center's recommended instruction. Solan pointed out that the instruction properly focuses attention where it belongs: on the strength of the government's case.

n4 The choice of the standard, of course, also affects the likelihood of a false negative: A more stringent standard for conviction increases the likelihood that a guilty defendant will be acquitted.

n5 Whether they are likely to share the risk of error equally depends on the base rate for actual liability.

n6 The full version of the Federal Judicial Center's 1982 definition of reasonable doubt is presented in the Appendix. The identical definition appears in the Federal Judicial Center's 1987 and 1988 *Pattern Criminal Jury Instructions*.

n7 The specific language used by Kagehiro and Stanton (1985) from the Federal Judicial Center's (1982) definition of reasonable doubt in this study, modified to fit a civil fact pattern, was: "Proof beyond a reasonable doubt is proof that leaves you firmly convinced of the [plaintiff's] case. There are very few things in this world that we know with absolute certainty and the law does not require proof that overcomes every possible doubt." Thus, it did not

include two additional sentences that appear in the Federal Judicial Center instruction that explicitly tell jurors that if they think there is a real possibility that the defendant is not guilty, they must give him or her the benefit of the doubt.

n8 The specific language for the preponderance of the evidence standard was: "proof that something is more likely so than not so. In other words, it means such evidence as, when considered and compared with that opposed to it, has more convincing force, and produces in your minds belief that what is sought to be proved is more likely true than not true" (Devitt & Blackmar, 1977, § 71.14).

n9 The specific language for the clear and convincing standard eliminated the use of "firm belief or conviction" that weakened the contrast with reasonable doubt in the first two experiments. The Kansas pattern instruction used in this set of instructions (along with the Federal Judicial Center's reasonable doubt instruction) was: "proof which is 'clear' in the sense that it is certain, plain to the understanding, unambiguous, and 'convincing' in the sense that it is so reasonable and persuasive as to cause you to believe it" (Kansas District Judges' Association Committee on Pattern Jury Instructions, 1977, § 2.11).

n10 Since *Furman* (1972), a majority of states have created sentencing schemes that the Court has found acceptable, although there is substantial evidence that instructions in death penalty cases still fail to provide jurors with comprehensible instructions that guide their judgments and avoid unconstitutionally impermissible inconsistency (see, e.g., Diamond, 1993; Diamond & Levi, 1996; Wiener, Pritchard, & Weston, 1995). It is unclear whether this represents a consistency ceiling or is merely a failure to write instructions that are sufficiently clear (Diamond, 1993; English & Sales, 1997).

n11 Of course, implementation of a quantitative standard requires that a legislature or court identify what the appropriate number should be.

n12 Note that judges often describe the preponderance of the evidence standard in quantitative terms. That standard is applied when it is assumed that the risks of error are balanced for the opposing parties (but see DeKay, 1996).

n13 Of course, if there were evidence that jurors adjusted their threshold based on illegitimate features of the case, such as the defendant's race, flexibility to engage in that adjustment would have serious negative consequences.

n14 Robert Bartels (1981), who is a member of the committee, proposed that jurors be informed of the potential penalties the defendant could receive. His "modest proposal" found a partner in the progressive members of the committee but was withdrawn after it drew opposition when the tentative recommendations were circulated for comment. The committee decided that the opposition might endanger the rest of the reform proposals (Dann & Logan, 1996).

n15 In addition, as the Arizona Supreme Court Committee on the More Effective Use of Juries (1994) acknowledged, implementation of the procedure will not be easy in view of, for example, enhanced sentences due to prior convictions and possible concurrent and consecutive sentences. Bartels (1981) has provided a thoughtful analysis of some of these issues.