“The Difference Between Blackstone-Like Error Ratios and Probabilities Standards of Proof”

95-02-01

Michael DeKay
The Difference Between Blackstone-Like Error Ratios and Probabilistic Standards of Proof

Michael L. DeKay*


I wish to thank Ken Hammond, Reid Hastie, Gary McClelland, David Asch, and Peter Ubel for helpful discussions on these issues and for their insightful comments on an earlier version of this manuscript.

Correspondence concerning this article should be addressed to Michael L. DeKay, Division of General Internal Medicine (111GM), Veterans Affairs Medical Center, University and Woodland Avenues, Philadelphia, Pennsylvania, 19104. Electronic mail may be sent via Internet to dekay@opim.wharton.upenn.edu.
Indeede I would rather wish twentie euill [evil] doers to escape death through pittie, then one man to bee vniustly [unjustly] condemnped. 1

In some cases, presumptive evidences go far to prove a person guilty, tho there be no express proof of the fact to be committed by him, but then it must be very waryly pressed, for it is better five guilty persons should escape unpunished, than one innocent person should die. 2

All presumptive evidence of felony should be admitted cautiously: for the law holds, that it is better that ten guilty persons escape, than that one innocent suffer. 3

To many, statements regarding the ratio of erroneous acquittals to erroneous convictions constitute clear declarations of the presumption of innocence and the “reasonable doubt” standard required to convict a defendant in a criminal trial. Such ratios have figured prominently in the history of the reasonable doubt doctrine 4 and in the Supreme Court’s discussion of the presumption of innocence and reasonable doubt standards. 5

More recently, the remarks of Blackstone and his predecessors regarding the relative frequencies of erroneous acquittals and erroneous convictions have been paraphrased in terms of the utilities of these outcomes. For example, Nagel maintained that “William Blackstone once said that it is ten times as bad to convict one innocent defendant as it is to acquit a guilty one.” 6 The appeal of this subtle change lies in the fact that the most popular theory of rational decision making, subjective expected utility theory (SEU), 7 can use information about utilities in a way that it cannot use information about frequencies. Specifically, outcome utilities may be used to

---

1 JOHN PORTESCUE, DE LAUDIBUS LEGUM ANGLIE 62 (London, Companie of Stationers 1616).


5 Coffin v. United States, 156 U.S. 432, 455–56 (1895).

6 Stuart Nagel, Bringing the Values of Jurors in Line with the Law, 63 JUDICATURE 189, 191 (1979).

derive a mathematical expression for the optimal standard of proof required for conviction. These standards, which are expressed as threshold probabilities, clearly influence the relative frequencies of the two types of judicial errors: More stringent standards of proof lead to more erroneous acquittals and fewer erroneous convictions. Unfortunately, there is a popular misconception that the standard of proof completely determines the ratio of errors:

If, as suggested ... in discussing the standard of “clear, unequivocal and convincing evidence,” the probability is about 80%, it means that we would rather have four cases decided in error against the Government than more than one against the defendant. If, in the case of proof “beyond a reasonable doubt,” the figure of 95% or 99% is used, it means that we would rather have, respectively, twenty or one hundred persons go free than more than one innocent person be convicted. Blackstone would have put the probability standard for proof “beyond a reasonable doubt” at somewhat more than 90%, for he declared: It is better that ten guilty persons escape than one innocent suffer. 

Such statements, however compelling they may be, are simply incorrect. The primary purpose of this article is to dispel the notion that probabilistic standards of proof may be equated with ratios of error frequencies in this manner. As a matter of course, a number of related misunderstandings are also resolved.

In the first section of this article, jury decisions are described in terms of signal detection theory (SDT), which provides an excellent framework for understanding the link between standards of proof and ratios of errors. In the second section, the statements of Fortescue, Hale, and Blackstone are interpreted as ratios of error frequencies and expressed in terms consistent with the SDT approach. The conditions under which these statements may be interpreted as


ratios of error utilities are investigated and found to be implausible. In the third section, the “optimal” standard of proof implied by SEU is discussed, along with a dubious simplification of the formula. It is shown that Blackstone’s ratio may be equated with a specific probability threshold (usually 0.91) only if Blackstone is interpreted in terms of error utilities and the simplified formula for the optimal standard of proof is used. In the fourth section, the relationship between probabilistic standards of proof and ratios of error frequencies is considered in greater detail and found to be more complicated than that suggested by the court above. In general, the ratio of judicial errors is only partially determined by the standard of proof. Finally, the possibility of adopting a specific Blackstone-like error ratio as a policy goal is rejected as unworkable and irrational.

**Signal Detection Theory as a Model of the Jury**

Signal detection theory provides an elegant description of systems that discriminate between two classes of objects or events. The theory is applicable to the trial process because, in the simplest case, the goal of juries is to discriminate between truly guilty and truly innocent defendants. Indeed, many authors have conceptualized the jury trial as an exercise in signal detection.\(^{11}\)

The primary benefit of SDT is that it provides a way of conceptually separating two very different components of the jury’s decision: the jury’s accuracy and the jury’s decision criterion. In SDT terms, accuracy refers to the jury’s ability to tell the difference between truly guilty

defendants and truly innocent defendants. The decision criterion refers to the standard of proof used by the jury when deciding whether the evidence against the defendant is strong enough to warrant a conviction.

These ideas are conveyed by the diagram in Figure 1. The horizontal axis may be thought of as "the strength of the jury's belief that the defendant is guilty, after deliberation," or more succinctly, as the defendant's "apparent guilt." There are two distributions in the figure: one for truly innocent defendants and one for truly guilty defendants. The distributions represent the likely values of apparent guilt for the two types of defendants. If the legal process works at all, truly guilty defendants will, on average, appear more guilty than truly innocent defendants. Hence, the distribution for guilty defendants is to the right of that for innocent defendants.

Insert Figure 1 about here.

However, because there is significant variability within each group, some truly innocent defendants appear more guilty than some truly guilty defendants. This overlap implies that errors are inevitable. Because appearance of guilt is imperfectly related to true guilt, there is no way that decisions based on appearances or degrees of belief can correctly classify all defendants as innocent or guilty. The extent of this problem can be quantified using a measure of accuracy. Although there are many measures of accuracy, a popular measure adequate for the present purpose is $d'$, the standardized distance between the means of the two distributions. In Figure 1, $d' = 2$. If $d'$ were larger, there would be less overlap and fewer incorrect verdicts, all else being equal. Anything that makes truly guilty defendants appear more guilty or truly innocent defendants appear less guilty moves the curves apart and increases accuracy. Certainly, the introduction of technology (photography, video and sound recording, fingerprinting, blood typing, DNA testing, etc.) serves this purpose. Thus, one might expect that juries today are more

---

12 For reviews, see John A. Swets, Indices of Discrimination or Diagnostic Accuracy: Their ROCs and Implied Models, 99 PSYCHOL. BULL. 100 (1986) and John A. Swets, Measuring the Accuracy of Diagnostic Systems, 240 SCI. 1285 (1988).
accurate than those in Blackstone's time—not because the juries themselves are any better (they may or may not be), but because they have better information.

The vertical line in Figure 1 represents the jury's decision criterion. According to SDT, the jury chooses a particular level of apparent guilt as the criterion, convicting only those defendants whose apparent guilt exceeds this criterion and acquitting the remainder. This criterion divides each of the distributions into two regions. These regions correspond to the four possible outcomes of a verdict: acquitting a truly innocent defendant (AI), acquitting a truly guilty defendant (AG), convicting a truly innocent defendant (CI), and convicting a truly guilty defendant (CG).\(^{13}\) The areas of these regions are closely related to the probabilities and frequencies of these four outcomes. If the level of apparent guilt is denoted \(x\), the distributions in the figure are the conditional distributions \(P(x|I)\) and \(P(x|G)\). This means that the areas under the curves represent the conditional probabilities \(P(AI)\), \(P(AG)\), \(P(CI)\), and \(P(CG)\). To get the unconditional probabilities of the four outcomes, \(P(AI)\), \(P(AG)\), \(P(CI)\), and \(P(CG)\), the conditional probabilities must be adjusted to reflect the mix of truly guilty and truly innocent defendants appearing before the court. Specifically, \(P(AI|G)\) and \(P(CI|G)\) must be multiplied by the "prior probability of guilt" \(P(G)\), which is simply the proportion of defendants that are truly guilty, to get \(P(AG)\) and \(P(CG)\). Similarly, \(P(AI|I)\) and \(P(CI|I)\) must be multiplied by the prior probability of innocence \(P(I) = 1 - P(G)\) to yield \(P(AI)\) and \(P(CI)\). Clearly, changing the standard of proof affects the likelihoods of the four possible outcomes. For example, moving the criterion to the right leads to more acquittals [the areas \(P(AI)\) and \(P(AI|G)\) get larger] and fewer convictions [the areas \(P(CI)\) and \(P(CG)\) get smaller].

---

\(^{13}\) Labels for these outcomes differ across disciplines and methodologies. For example, convictions of truly innocent defendants (CI) are equivalent to false positives in medicine, false alarms in SDT, and Type I errors in statistics. Likewise, acquittals of truly guilty defendants (AG) = false negatives = misses = Type II errors, convictions of truly guilty defendants (CG) = true positives = hits, and acquittals of truly innocent defendants (AI) = true negatives = correct rejections.
Blackstone’s Ratio: A Ratio of Error Utilities or a Ratio of Error Frequencies?

Although Fortescue, Hale, and Blackstone conveyed their unwillingness to convict innocent defendants using decidedly different error ratios, these authors’ specificity has usually been overlooked in favor of the underlying theme.\textsuperscript{14} Tempting though it may be, these ratios cannot be taken as recommendations for specific social policies. Indeed, the authors’ use of inequalities precludes this possibility outright. More substantially, there is nothing in these works that suggests a rationale for selecting one ratio over another or a method for converting a particular ratio into a concrete standard of proof. Nevertheless, statements such as Blackstone’s have been interpreted both in terms of error frequencies and error utilities. In this section, the implications of these two interpretations are explored more thoroughly.

The remarks of Fortescue, Hale, and Blackstone are worded almost identically and may be most faithfully translated into mathematics as follows.

\[ U_{AG}(N_{AG}) > U_{CI}(1) \]  \hspace{1cm} (1)

In this expression, \( N_{AG} \) is the number of acquittals of truly guilty defendants (\( N_{AG} = 20, 5, \) and \( 10 \) for the three authors, respectively), \( 1 \) is the number of convictions of truly innocent defendants (\( N_{CI} = 1 \)), and \( U_{AG}(N_{AG}) \) and \( U_{CI}(N_{CI}) \) are the authors’ utility functions for the two types of judicial errors. The first simplification, adopted by all of the authors cited below, is to replace the inequality with equality.

\[ U_{AG}(N_{AG}) = U_{CI}(1) \]  \hspace{1cm} (2)

Although this change clearly misrepresents the original statements, it is necessary for the specification of a precise policy and does not affect the interpretation of error ratios generally.

A Ratio of Frequencies

Note that Equation 2 specifies the relationship between two points and not two utility functions. However, because Blackstone and his predecessors were writing about the entire English legal system, we may safely assume that they intended their ratios to be “scalable”: If Blackstone would be satisfied with a 10:1 ratio, he should also be satisfied with a 20:2 ratio or a

\textsuperscript{14} See, e.g., May, supra note 4, at 653–654.
1,000:100 ratio. In other words, $N_{AG}/N_{CI} = 10$ and the ratio in question is a ratio of error frequencies. Although Connolly\textsuperscript{15} and Hammond et al.\textsuperscript{16} have interpreted Blackstone in this manner, they did not pursue the implications of this interpretation for the utility functions $U_{AG}(N_{AG})$ and $U_{CI}(N_{CI})$. The ratio-of-frequencies interpretation implies that Equation 2 may be generalized as follows.

$$U_{AG}(R \times N_{CI}) = U_{CI}(N_{CI})$$

Thus, $U_{CI}(N_{CI})$ is simply the compound function $U_{AG}[R(N_{CI})]$ where the interior function $R(N_{CI})$ is multiplication by a constant equal to the expressed error ratio. Differentiating both sides of Equation 3 with respect to $N_{CI}$ yields the following.

$$R \times [U_{AG}'(R \times N_{CI})] = U_{CI}'(N_{CI})$$

$$R \times [U_{AG}'(N_{AG})] = U_{CI}'(N_{CI})$$

Thus, specifying the ratio of error frequencies implies a relationship between the slopes of the two utility curves. In Figure 2, Blackstone’s ratio implies that the slope of $U_{CI}(1)$ at point A must be ten times the slope of $U_{AG}(10)$ at point B. Because this constraint must hold for any horizontal line intersecting the two functions $U_{AG}(N_{AG})$ and $U_{CI}(N_{CI})$, the two functions must have similar forms. However, any number of utility function pairs satisfy the slope constraint of Equation 5. Linear, logarithmic, and square-root functions are shown in the figure.

Insert Figure 2 about here.

Notwithstanding the initial assumption of equality, the ratio-of-frequencies interpretation is true to the statements of the original authors. So long as the slopes of the two utility functions maintain the specified relationship, the functions may take on essentially any form. Other constraints (e.g., that both functions be monotonically decreasing) may also be reasonable, but


\textsuperscript{16} Kenneth R. Hammond et al., Making Better Use of Scientific Knowledge: Separating Truth from Justice, 3 PSYCHOL. SCI. 80, 84 (1992).
they are not required by the ratio-of-frequencies interpretation.

Fortunately, the ratio of error frequencies may be expressed in a way that is consistent with the SDT framework of Figure 1. For a given decision threshold, the frequency of erroneous acquittals is the conditional probability that a truly guilty defendant is acquitted times the prior probability that the defendant is truly guilty times the total number of defendants appearing before the court. If the frequency of erroneous convictions is computed analogously, the ratio of the two judicial errors is given by the following expression.

$$\frac{N_{AG}}{N_{CI}} = \frac{P(AG) \times P(G) \times N}{P(CI) \times P(I) \times N}$$  \hspace{1cm} (6)

Because the total number of cases does not affect this ratio, the expression may be rewritten as a ratio of error probabilities without loss of generality.

$$\frac{P(AG)}{P(CI)} = \frac{P(AG) \times P(G)}{P(CI) \times P(I)}$$  \hspace{1cm} (7)

According to the ratio-of-frequencies interpretation of Blackstone, this ratio should be 10:1.\(^{17}\)

A Ratio of Utilities

A number of authors have implicitly or explicitly chosen to re-express Equation 2 as a ratio of utilities.\(^{18}\) For example, Nagel’s paraphrase of Blackstone may be generalized as follows.

$$R \times [U_{AG}(1)] = U_{CI}(1)$$  \hspace{1cm} (8)

Figure 2 illustrates the difference between Equations 2 and 8 for the Blackstone condition \(R = 10\). Equation 2 expresses the relationship between points A and B, whereas Equation 8 expresses the relationship between points A and C. It is important to note that none of the authors interpreting Blackstone in this manner rejected the ratio-of-frequencies interpretation. Thus, they did not substitute one constraint for another; they added an additional constraint without realizing

---

\(^{17}\) Connolly, supra note 15, at 104; Hammond et al., supra note 16, at 84.

it. Even so, the ratio-of-utilities interpretation does not preclude the ratio-of-frequencies interpretation because one pair of utility functions (the linear pair) satisfies both constraints for all values of \( N_{CI} \) and \( N_{AG} = R \times N_{CI} \). Of course, there are other pairs of utility functions that satisfy the ratio-of-frequencies constraint but not the ratio-of-utilities constraint and still other pairs (not shown) that satisfy the ratio-of-utilities constraint but not the ratio-of-frequencies constraint. In fact, it is easy to show that only linear utility functions satisfy both constraints. Generalizing Equation 8 to any number of errors (\( N_{AG} = N_{CI} \)) and differentiating yields the following.

\[
R \times [U_{AG}(N_{AG})] = U_{CI}(N_{CI})
\]

(9)

This equation specifies the relationship between the slopes of the two utility functions at points on the same vertical line in Figure 2, whereas Equation 5 specifies the relationship between slopes for points on the same horizontal line. Thus, if one knows the slope \( U_{CI}(N_{CI}) \) at one point (say point A), one also knows the slope \( U_{AG}(N_{AG}) \) at two points (B and C). From this knowledge, one may infer the slope \( U_{CI}(N_{CI}) \) at two new points, and so on. The key is that once one slope is known, all of the others follow. Since the slopes at these new points are always equal to the previously known slopes, both utility functions must be linear.

This is an important point: Blackstone-like statements about ratios of frequencies may not be interpreted as statements about ratios of utilities unless linear utility functions are assumed. Of course, it may not be reasonable to expect linearity. The difference between zero and one acquitted criminals seems (to me) to be much greater than the difference between 100 and 101 acquitted criminals. Marginally decreasing (dis)utility for both types of errors may be a more reasonable assumption.

A more fundamental objection to the ratio-of-utilities interpretation is that it assumes utility is measured on a ratio scale. However, utility scales, as implied by von Neumann and Morgenstern's axioms, have only interval properties.\(^{19}\) In short, Nagel's paraphrase of

---

Blackstone is invalid because it implies that the utility scale has a nonarbitrary zero point. Blackstone’s original statement, on the other hand, is acceptable because it involves only an ordinal comparison. The basic difference between the two statements is that Blackstone’s would remain unchanged across reasonable transformations of the utility scale (such as adding a constant), whereas Nagel’s would not.

In summary, the ratio-of-utilities interpretation deviates substantially from the original authors’ statements. It places greater and less reasonable restrictions on the utility functions for the two types of errors and, indeed, on the nature of the utility scale itself. In light of these difficulties, the ratio-of-frequencies interpretation is strongly preferred.

Subjective Expected Utility and the Optimal Standard of Proof

The conditional probabilities in Equation 7 depend on the placement of the decision criterion in Figure 1. If one knows enough about the distributions in the figure, one may determine the decision criterion that leads to the desired error ratio. Alternatively, one may select the decision criterion in some other manner and calculate the corresponding error ratio. More specifically, it is possible to select the criterion in a way that reflects society’s utilities for the possible outcomes of jury decisions. Subjective expected utility theory\textsuperscript{20} is the most widely accepted normative model of decision making. Briefly, the theory states that one should choose the decision alternative which leads to the highest expected utility.\textsuperscript{21} In the simplified jury trial, the defendant should be convicted if and only if the expected utility of conviction exceeds the expected utility of acquittal.

\[
\text{SEU}(C|x) > \text{SEU}(A|x)
\]  

(10)

These expected utilities depend the defendant’s apparent guilt, \( x \), because apparent guilt is a reliable (but imperfect) indicator of the defendant’s true status. More precisely, the expected

\textsuperscript{20} Savage, supra note 7.

\textsuperscript{21} Although this article is concerned with SEU, Savage, supra note 7, and not its predecessor, expected utility theory, von Neumann & Morgenstern, supra note 19, the term “expected utility” is sometimes used for convenience. The distinguishing characteristic of SEU is that probabilities are estimated by the decision maker.
utilities of conviction and acquittal are given by the following expressions, where \( U_{CG}, U_{CI} \),
\( U_{AG} \), and \( U_{AI} \) are the utilities associated with the four possible outcomes of the verdict.

\[
\begin{align*}
\text{SEU(C|x)} &= P(\text{Glx}) \times U_{CG} + P(\text{Ilx}) \times U_{CI} \\
\text{SEU(A|x)} &= P(\text{Glx}) \times U_{AG} + P(\text{Ilx}) \times U_{AI}
\end{align*}
\]

Substitution of Equations 11 and 12 into the decision rule of Inequality 10 implies a threshold
for the conditional probability of guilt \( P(\text{Glx}) \), denoted \( P_T \). \(^{22}\)

\[
P_T = \frac{1}{U_{CG} - U_{AG} + U_{AI} - U_{CI}} \cdot \frac{U_{AI} - U_{CI}}{U_{AI} - U_{CI} + 1}
\]

The jury should return a guilty verdict if and only if their estimated probability of guilt
\( P(\text{Glx}) \) exceeds this threshold. Of course, this standard of proof may also be expressed on a scale
of 0% (not at all certain) to 100% (completely certain). Such numerical standards communicate
certainty levels much more effectively than traditional verbal instructions. \(^{23}\) \( P_T \) may also be
expressed as a threshold for posterior odds. \(^{24}\)

\[
\left( \frac{P(\text{Glx})}{P(\text{Ilx})} \right)_T = \frac{P_T}{1 - P_T} = \frac{U_{AI} - U_{CI}}{U_{CG} - U_{AG}}
\]

Finally, multiplying by the prior odds of guilt (actually the inverse of prior odds) yields a
threshold likelihood ratio via Bayes Theorem. \(^{25}\)


\(^{23}\) Dorothy K. Kagehiro, Defining the Standard of Proof in Jury Instructions, 1 PSYCHOL. SCI. 194 (1990); Dorothy K. Kagehiro & W. Clark Stanton, Legal vs. Quantified Definitions of Standards of Proof, 9 LAW & HUM. BEHAV. 159 (1985); Nagel, supra note 6, at 194–195; Nagel et al., supra note 18, at 377–78.

\(^{24}\) Cullison, supra note 8, at 565–66; Grofman, supra note 11, at 310–14.

\(^{25}\) The formula for the threshold likelihood ratio presented in GREEN & SWETS, supra note 10, at 23, COOMBS ET AL., supra note 10, at 170, and Grofman, supra note 11, at 313 has pluses rather than minuses in front of the utilities for misses (AG here) and false alarms (CI here) because the authors assume that error utilities are negative.
\[
L_T = \left( \frac{P(x|G)}{P(x|I)} \right)_T = \frac{U_{AI} - U_{CI}}{U_{CG} - U_{AG}} \frac{P(I)}{P(G)}
\]

(15)

It is easy to see how one would use Equation 15 to determine the placement of the optimal criterion in Figure 1. Because the two distributions in the figure correspond to the conditional probabilities on the left side of the equation, the ratio of the heights of the two distributions corresponds to a likelihood ratio. Because the ratio of heights increases monotonically from left to right in the figure, one may select the value on the horizontal axis that corresponds to the desired likelihood ratio. For the criterion shown in the figure, the likelihood ratio is 10:1. Note that if utilities remain constant (so that \(P_T\) is also constant), increasing the prior odds of guilt \(P(G)/P(I)\) decreases the threshold likelihood ratio and moves the criterion in Figure 1 to the left. Grofman’s assertion that \(P(G)/P(I)\) and \(L_T\) are positively related\(^{26}\) is simply incorrect.

Figure 3 is a graphical representation of the threshold probability \(P_T\). Although similar figures have proved useful for analogous decision problems in other fields (e.g., testing and treatment decisions in medicine\(^{27}\)), such figures have not appeared in the legal literature.\(^{28}\) The left side of the figure represents absolute certainty that the defendant is innocent. Under these circumstances, the only possible outcomes are \(AI\) and \(CI\). The utility difference \(U_{AI} - U_{CI}\) is the cost (primarily to the defendant) associated with the conviction of a truly innocent defendant. Similarly, the right side represents absolute certainty that the defendant is guilty, and the only

This assumption is both confusing and limiting. Moreover, the assumption of an absolute zero point is inconsistent with the interval nature of utility scales. VON NEUMANN & MORGENSTERN, supra note 19, at 23–29, 617–28. Equations 13–15, on the other hand, correctly express thresholds as ratios of utility differences. In these equations, any of the four outcomes may be evaluated positively or negatively. The expression in MACMILLAN & CREEELMAN, supra note 10, at 49, is identical to Equation 15 in this respect.

\(^{26}\) Grofman, supra note 11, at 312.

\(^{27}\) E.g., John C. Hershey et al., Clinical Guidelines for Using Two Dichotomous Tests, 6 MED. DECISION MAKING 68, 71–72 (1986).

\(^{28}\) A description of such a figure does appear in Nagel et al., supra note 18, at 357 n.2.
possible outcomes are CG and AG. The utility difference \( U_{CG} - U_{AG} \) is the benefit (primarily to society) associated with the conviction of a truly guilty defendant.\(^{29}\) The lines \( U_{CI} U_{CG} \) and \( U_{AI} U_{AG} \) are the expected utilities of conviction and acquittal, respectively. Substitution of \( 1 - P(G|x) \) for \( P(I|x) \) in Equations 11 and 12 makes it clear that these expected utilities are linear functions of the defendant's probability of guilt.\(^{30}\)

\[
SEU(C|x) = U_{CI} + (U_{CG} - U_{CI}) \times P(G|x) \\
SEU(A|x) = U_{AI} + (U_{AG} - U_{AI}) \times P(G|x)
\]

If one's goal is to maximize expected utility, one should choose the decision that corresponds to the higher line. For example, if the relative utilities depicted in the graph are correct, and one believes that there is a 70% chance that the defendant committed the crime, one should vote to acquit the defendant because the expected utility of acquittal is higher than the expected utility of conviction. By this logic, it is clear that the intersection of the two lines defines the threshold probability, \( P_T \). One should always acquit when the subjective probability of guilt is less than \( P_T \) and convict when the subjective probability of guilt is greater than \( P_T \). In the figure, the cost associated with convicting an innocent defendant is arbitrarily assumed to be three times as great as the benefit associated with convicting a guilty defendant, so \( P_T = 3/(3 + 1) = 0.75 \).

Insert Figure 3 about here.

Difficulties with the Threshold Probability Formula

A number of difficulties arise if outcome utilities are estimated separately for each defendant. First, because defendants are not identical, the utilities in the expression for \( P_T \) may depend on


\(^{30}\) See Hershey et al., *supra* note 27, at 71.
defendants' reputations, previous criminal histories, and so on.\textsuperscript{31} For example, if jurors believe that defendant is particularly dangerous, the utility difference $U_{AI} - U_{CI}$ will be smaller and the optimal standard of proof will be lower than for other defendants.\textsuperscript{32} Such differences are problematic because using different standards of proof for different defendants seems grossly unfair to many people. Second, the utilities of the outcomes may not be independent of the defendant's probability of guilt. Tribe argues that one may not be able to calculate a single value for $P_T$ in such instances.\textsuperscript{33} Finally, the four utilities are, strictly speaking, marginal utilities. For example, if $N_{CI}$ innocent defendants have already been convicted, the utility $U_{CI}$ is equal to $U_{CI}(N_{CI} + 1) - U_{CI}(N_{CI})$. Even if defendants are assumed to be identical, there is no a priori reason to believe that these utility functions are linear.

In light of these problems, it seems reasonable to use utilities associated with the typical truly guilty defendant and the typical truly innocent defendant. Although this method represents a clear departure from a strict utility-maximization approach (because the utilities used will be poor approximations for atypical defendants), using typical-defendant utilities solves all of the problems mentioned above. For a given type of case, a single threshold probability may be applied in all instances. Of course, this approach still allows different standards of proof.

\textsuperscript{31} See, e.g., Kaplan, \textit{supra} note 8, at 1073–76. However, rules of evidence and instructions to the jury may moderate such effects to a substantial degree. Kaplan, \textit{supra} note 8, at 1074–77; Lempert, \textit{supra} note 18, at 1035–40 & nn.41 & 53.

\textsuperscript{32} In an extreme case, the jury might conclude that $U_{CI} > U_{AI}$, in which case the defendant should be convicted regardless of his guilt. Kaplan, \textit{supra} note 8, at 1076. In terms of Figure 3, SEU(convict) would exceed SEU(acquit) for all probabilities, the lines would not cross, and the notion of a threshold probability would be meaningless. Similarly, acquittal may dominate conviction if the penalty is too severe even for guilty defendants (i.e., $U_{AG} > U_{CG}$). Kaplan, \textit{supra} note 8, at 1076.

(preponderance of the evidence, clear and convincing evidence, and proof beyond a reasonable doubt) to be used in different types of cases. Most importantly, the additional utility associated with the consistency and efficiency of a single standard of proof for each type of case is likely to be substantial. Thus, using typical-defendant utilities instead of individual-defendant utilities to derive numerical standards of proof retains the primary benefits of the rational utility-maximization approach without sacrificing important societal values.

**Simplification of the Threshold Probability Formula**

Interestingly, many researchers have chosen to simplify Equation 13 to reduce the number of utilities that must be estimated. Specifically, the utilities of correct decisions are eliminated, leaving only the error utilities $U_{AG}$ and $U_{CI}$. The most straightforward way to accomplish this goal is to ignore the utilities of correct decisions outright. That is, set $U_{CG} = U_{AI} = 0$. Such an assumption leads to a simplified expression for the threshold probability.$^{34}$

$$P_T = \frac{U_{CI}}{U_{AG} + U_{CI}} = \frac{1}{\frac{U_{AG}}{U_{CI}} + 1}$$  \hspace{1cm} (18)

Lempert noted that actual jurors would often violate this assumption, but argued that “the ideal juror should feel no regret at reaching correct decisions.”$^{35}$ However, correctness is certainly not the only admissible component of utility. If it were, one would also have to value the two errors equally, and $P_T$ would always equal 0.5.

A second way to simplify the expression for $P_T$ is to assume that the utilities of incorrect decisions are the opposite of the utilities for the related correct decisions. That is, $U_{AG} = -U_{CG}$ and $U_{CI} = -U_{AI}$. This “equal-absolute-values” assumption also implies the above

---


$^{35}$ Lempert, *supra* note 18, at 1036 n.41 (emphasis added).
simplification. However, note that if $U_{AG} > U_{CI}$, as most authors assume, then $U_{CG} < U_{AI}$, in direct conflict with the assumption that $U_{CG} = U_{AI} = 0$.

The simplified expression for the threshold probability takes on special importance when paired with the ratio-of-utilities interpretation of Blackstone. Specifically, if $U_{CI} = 10 \times U_{AG}$, invoking the above simplification leads to a $P_T$ value of about 0.91. This value is viewed by many as the “correct” numerical equivalent of reasonable doubt. Citing Simon & Mahan’s survey results, Lempert suggested that “a substantial proportion of jurors may, in fact, subscribe to this norm.” This is a provocative suggestion: Some jurors (and perhaps judges and legal scholars) believe that Blackstone’s ratio implies a standard of proof around 90%, and that such a standard is, in Lempert’s words, “arguably normative.”

However, the above assumptions are problematic for two reasons. First, there is disagreement among authors regarding the proper ordering of $U_{CG}$ and $U_{AI}$. Tribe, for example, suggested that $U_{CG} > U_{AI}$, whereas Milanich argued that “acquitting the innocent is the expressed highest value of our legal system.” Connolly noted that the choice was a difficult one and ultimately settled on equality. Although the ordering of $U_{CG}$ and $U_{AI}$ may be an interesting empirical

---

36 Nagel, supra note 6, at 192 n.5; Nagel et al., supra note 18, at 355–57; Grofman, supra note 11, at 314.


38 Rita J. Simon & Linda Mahan, Quantifying Burdens of Proof: A View from the Bench, the Jury, and the Classroom, 5 LAW & SOC’Y REV. 319 (1971).

39 Lempert, supra note 18, at 1038–39.

40 Lempert, supra note 18, at 1039.

41 Tribe, supra note 33, at 1379.


43 Connolly, supra note 15, at 109.
question, the dispute need not be resolved to determine the optimal standard of proof. Figure 3 makes it clear that \( P_T \) is unaffected by the ordering so long as the cost/benefit ratio \((U_{AI} - U_{CI})/(U_{CG} - U_{AG})\) remains constant. For example, increasing \( U_{CG} \) and \( U_{AG} \) by the same amount so that \( U_{CG} > U_{AI} \) has no effect on \( P_T \). Thus, these simplifying assumptions lead to disagreement that is both irrelevant and unnecessary.

A second objection is that utility scales, which are interval in nature, do not have absolute zero points.44 This is why there are no scale values on the vertical axis of Figure 3. For the purposes of determining the optimal threshold, it simply does not matter where zero falls. Although it is often useful to anchor the utility scale by establishing a zero point, doing so may lead researchers to make invalid statements about outcome utilities and derived thresholds. Given interval scales, the only permissible comparisons involve simple utility orderings (as in Blackstone’s original statement) or ratios of utility differences (as in Equations 13–15). These comparisons are permissible because they are unaffected by linear transformations of the scale. The invariance of the thresholds in Equations 13–15 is easily verified by replacing the \( U_i \) with \( \alpha U_i + \beta \). In each case, the \( \alpha \) and \( \beta \) both cancel, leaving the thresholds unchanged. Thus, thresholds calculated from these formulae may be compared freely across respondents, conditions, and studies, regardless of differences in scale construction and usage. On the other hand, individual utilities, utility ratios such as \( U_{CI}/U_{AG} \), and the simplified expression for the threshold probability \( U_{CI}/(U_{AG} + U_{CI}) \) are inappropriately sensitive to such transformations. Direct comparisons involving these measures are risky at best and meaningless at worst. In summary, there appears to be no good reason for abandoning the optimal standard of proof in favor of a simpler expression.

**From Decision Criteria to Error Frequencies**

It is now possible to specify the relationship between a decision criterion and the resulting ratio of errors. For the present purpose, it most useful to express the decision criterion in terms of threshold posterior odds, as in Equation 14. The ratio-of-utilities interpretation of Blackstone and

44 VON NEUMANN & MORGENSTERN, supra note 19, at 23–29, 617–28.
the belief that there is an obvious, fundamental connection between standards of proof and error ratios appears widespread, though not universal. Unfortunately, this appealing view is unfounded and overly simplistic.

A More Rigorous Appraisal

In general, there is no simple relationship between the threshold posterior odds \([P(Glx)/P(I|x)]_T\) and the ratio of erroneous acquittals to erroneous convictions \(P(AG)/P(CI)\). In Figure 4, the relationship between \([P(Glx)/P(I|x)]_T\) and \(P(AG)/P(CI)\) is shown for five different values of the prior odds of guilt \(P(G)/P(I)\), assuming the same level of accuracy \((d' = 2)\) depicted in Figure 1.

Several interesting results are evident from the figure. First, for Blackstone’s ratio, \(P(AG)/P(CI) = 10\), to imply the same threshold posterior odds, \([P(Glx)/P(I|x)]_T = 10\), the prior odds of guilt must also be 10:1. This fact is illustrated by point A in the figure. Although the prior odds of guilt in United States courts may be as high as 2:1\(^{52}\) or even 4:1,\(^{53}\) it seems unlikely that they are as high as 10:1.

If the prior odds of guilt are 1:1, \(P(AG)/P(CI) = 10\) implies that \([P(Glx)/P(I|x)]_T = 4.48\) and \(P_T = 0.82\) (point B in Figure 4). This is an interesting result given previous authors’ concerns about the discrepancy between the public’s low standards for reasonable doubt and the “correct” threshold value of 0.91.\(^{54}\) Hastie’s recent review indicates that mean values for the reasonable doubt standard range from 0.51 to 0.91 \((M = 0.82)\) for direct ratings and from 0.50 to 0.90 \((M = 0.60)\) for indirect methods.\(^{55}\) If equal priors are assumed, these “low” thresholds imply Blackstone-like ratios that are surprisingly high—perhaps as high as the 10:1 ratio that generated


\(^{53}\) Robert D. Sorkin, personal communication (Sept. 7, 1994).

\(^{54}\) Nagel, supra note 6; Nagel et al., supra note 18, at 368–69, 376–81.

the $P_T$ value of 0.91 in the first place. Conversely, threshold posterior odds of 10:1 imply a Blackstone-like ratio of 36:1 if the prior odds of guilt are 1:1 (point C in Figure 4). Under these conditions, the “correct” standard of proof is clearly inconsistent with Blackstone’s remark.

Insert Figure 4 about here.

The relationship between the chosen decision threshold and the resulting ratio of decision errors clearly deviates from the expectations of previous authors. This deviation may be more easily appreciated with reference to Figure 1. For the present purpose, assume that the prior odds of guilt are 1:1. Under this assumption, the Blackstone-like ratio given by Equation 7 is simply the ratio of the areas of the two distribution tails $P(A|G)/P(C|I)$. On the other hand, the threshold posterior odds is the ratio of the heights of the two distributions $P(G|X)/P(I|X)$ at the decision criterion. There is no good reason to expect that the ratio of areas be equal to the ratio of heights. In Figure 1, the decision criterion is placed so that the ratio of heights is 10:1, but the resulting ratio of areas is 36:1, as noted above.

The steepness of the curves in Figure 4 may also be explained with reference to Figure 1. Imagine a new threshold placed at the intersection of the two distributions in Figure 1. This new threshold indicates indifference between the two types of errors. The ratio of areas and the ratio of heights are both 1:1. As one moves between the old criterion and the new criterion (or between any two possible criteria), the ratio of areas changes much more rapidly than the ratio of heights. Hence, the curves in Figure 4 have slopes that are much greater than one.

The effects of prior odds. Although the above results are clearly inconsistent with traditional expectations, more severe discrepancies are also apparent. When the prior odds of guilt are high, decision thresholds that appear biased against erroneous convictions, \([P(G|X)/P(I|X)]_T > 1\), may actually imply Blackstone-like ratios in which erroneous convictions outnumber erroneous acquittals, $P(A|G)/P(C|I) < 1$. All points in the lower right quadrant of Figure 4 have this property. Points in the upper left quadrant of the figure have the opposite but equally nonintuitive property that $[P(G|X)/P(I|X)]_T < 1$ and $P(A|G)/P(C|I) > 1$. Although decision theorists are familiar with the
powerful effects of prior probabilities, these results may come as a shock to those who believe that the "clear and convincing evidence" standard guarantees that the more egregious error is also the less common error.\textsuperscript{56} Indeed, no standard short of absolute certainty can ensure such a result.

Figure 4 indicates that Blackstone-like ratios are dependent on priors as well as decision thresholds. Referring to the relationships expressed in Equations 6 and 7, Connolly also argued that changes in prior odds lead to changes in the error ratio:

While the court may well have some control over the two probabilities [P(A|G) and P(C|I)] (for example by the value of [P_r] it adopts), it does not control the numbers of innocent and guilty with whom it deals. Other things being equal, a doubling of the ratio of guilty to innocent defendants brought before the court will double Blackstone's ratio, while leaving unchanged the balance of utilities used in setting [P_r].\textsuperscript{57}

Connolly went on to provide an example in which the prior odds of guilt are 1:1 and the decision criterion has been selected so that the ratio of errors is 10:1, as specified by Blackstone (see point B in Figure 4). He argued that if, "as a result of improved criminal procedure, or other circumstance beyond the court's control," the prior odds of guilt are increased to 1.5:1, the ratio of erroneous acquittals to erroneous convictions will increase to 15:1.\textsuperscript{58} However, in Figure 4, increasing the prior odds of guilt while holding the standard of proof constant results in a downward move and a lower value for the error ratio. Specifically, if the prior odds are increased to 1.5:1, the resulting Blackstone-like ratio decreases from 10:1 to about 8:1. Thus, the analysis presented here conflicts directly with that presented by Connolly.

The problem is that Connolly did not consider the distinction between the different expressions of the optimal decision rule in Equations 13–15. Equations 13 and 14 express the decision rule solely in terms of utilities, whereas the likelihood ratio of Equation 15 also includes prior odds. Connolly was correct in assuming that increasing the prior odds increases the ratio of errors by the same amount if the likelihood ratio is held constant. This is easily seen by referring

\textsuperscript{56} See supra pp. 19–20 and note 50.

\textsuperscript{57} Connolly, supra note 15, at 104–05.

\textsuperscript{58} Connolly, supra note 15, at 105.
to Figure 1. In this figure, maintaining a constant likelihood ratio means keeping the vertical line in the same place. Because the ratio of errors in Equation 7 is simply the ratio of the areas under the tails of the curves multiplied by the prior odds of guilt, it is clear that raising the prior odds also raises the error ratio $P(AG)/P(CI)$. However, it is impossible to maintain a constant threshold likelihood ratio while increasing prior odds without also increasing the threshold posterior odds of guilt and $P_T$ (see Equation 15). Thus, Connolly’s result cannot be obtained “while leaving unchanged the balance of utilities used in setting [P_T],” as he suggests. If the utilities are held constant, the optimal likelihood ratio decreases as the prior odds of guilt increase. Thus, the decision criterion in Figure 1 moves to the left and the ratio of erroneous acquittals to erroneous convictions decreases rather than increases. This result is consistent with the goal of maximizing expected utility, whereas Connolly’s result is not.

The effects of accuracy. As might be expected, the relationship between threshold posterior odds and the resulting Blackstone-like ratio also depends on the separation of the distributions in Figure 1. In other words, the relationship between thresholds and error ratios depends on the accuracy of the jury system. Figure 5 illustrates this dependence assuming the prior odds of guilt are 1:1. Briefly, if the jury system is less accurate ($d’$ is low), the curves in Figure 5 are steeper; if the system is more accurate ($d’$ is high), the curves are flatter. Again, Figure 1 provides some insight into this behavior. As noted above, the ratio of areas is more sensitive to changes in the placement of the decision criterion than is the ratio of heights. This differential sensitivity becomes greater as the distributions move closer together. Consider the extreme case in which $d’ = 0$ and the two distributions are superimposed. In this case, the ratio of heights does not vary with the decision criterion, but the ratio of areas continues to vary (because the areas of interest lie on opposite sides of the criterion). The result is that the $d’ = 0$ curve in Figure 5 is a vertical line. On the other hand, as the distributions of Figure 1 move further apart, this differential sensitivity to the placement of the criterion is reduced. However, in no case are the slopes of the curves in Figure 5 less than one. Changes in the threshold posterior odds of guilt always lead to

---

59 Connolly, supra note 15, at 105.
larger changes in the Blackstone-like ratio. Unfortunately, a simple geometric explanation for this limit is not apparent. Investigation of other distribution types will be necessary to determine the generality of this particular finding.

Figure 6 shows the effect of accuracy on the Blackstone-like ratio for one value of threshold posterior odds (10:1) and five different values of prior odds. The graph has a number of interesting features. First, the Blackstone-like ratio decreases as the prior odds of guilt increase, as discussed above. Second, when accuracy is very high, the Blackstone-like ratio is near 10:1 regardless of the prior odds of guilt. In other words, as the methods used to discriminate among truly innocent and truly guilty defendants improve, the ratio of errors becomes less sensitive to the prior odds of guilt and more similar to the threshold posterior odds. Third, if the prior odds of guilt are equal to the threshold posterior odds of guilt (10:1 in this case), then the resulting Blackstone-like ratio is also equal to the same number. Thus, a P value of 0.91 implies Blackstone's 10:1 ratio if and only if the prior odds of guilt are also 10:1. This interesting relation, discovered earlier for the $d' = 2$ case in Figure 4, is now seen to be independent of accuracy. Note that if the threshold posterior odds of guilt are equal the prior odds of guilt, Equation 15 yields an optimal likelihood ratio of one. This decision criterion would be located at the intersection of the two distributions in Figure 1 and the areas under the tails of the two curves would be equal. Equation 7 then implies a Blackstone-like ratio equal to the prior odds of guilt. This logic holds whenever prior odds and threshold posterior odds are equal, regardless of the specific value.

Figure 7 depicts the effects of accuracy and prior odds on the implied threshold posterior odds when the Blackstone-like ratio is equal to 10:1. This graph is useful for determining the decision rule that will lead to an error ratio of 10:1 under different assumptions about priors and accuracy.
Real juries. Unfortunately, it is difficult to specify the properties of the present-day jury system so that it may be located in Figures 6 and 7. The main problem, of course, is that there is no "gold standard" for differentiating truly innocent and truly guilty defendants. Without such a standard, it is difficult to estimate the accuracy of an imperfect jury system. However, there are some relevant data in the literature. For example, Gelfand and Soloman used a Poisson model to derive accuracy estimates for prototypical jurors using French data from the 1820's and 30's and more recent American data from Kalven and Zeisel. Gelfand and Soloman's parameters may be easily converted to $d'$ values of 0.94–1.54 for the French sample and 2.35–2.68 for the American sample. This performance difference is consistent with the effects of increased technology mentioned earlier. More recently, Sorkin and Dai have presented a model that characterizes the accuracy of groups in terms of the accuracies of their individual members. Although the accuracy of groups should generally be higher, groups may be less efficient than expected because (a) they do not weight members' opinions optimally, and (b) combining individual opinions into a group judgment may introduce additional variability or noise. Nevertheless, Sorkin suggests that present-day juries are likely to have $d'$ values in the 2–3 range. If $d' = 2.5$ and the prior odds of guilt are 4:1, a $P_T$ value of 0.91 yields a Blackstone-like ratio of 14:1, whereas a $P_T$ value of 0.89 is required to yield a Blackstone-like ratio of 10:1. These results indicate that the discrepancy between threshold posterior odds and implied error


62 Sorkin & Dai, supra note 11.

63 Sorkin & Dai, supra note 11, at 11.

64 Robert D. Sorkin, personal communication (Sept. 7, 1994).
ratios is small when accuracy and prior odds are both high. However, these results does not imply that the distinction between decision thresholds and error ratios may simply be ignored in today's courts. First, there are some types of trials in which accuracy will be lower because conclusive evidence is difficult if not impossible to obtain. Certainly, cases of past sexual abuse provide fall into this category, as do other cases that rely predominantly on the testimony of adversaries. Second, jury verdicts, important as they are, represent only a small fraction of the decisions made by the judicial system. The distinction between decision thresholds and error ratios is relevant to many other decisions, including those regarding search warrants, preliminary injunctions, interpretation of forensic evidence, and parole. Again, the distinction may be most important in domains where decisions must be made on the basis of limited evidence of dubious quality.

Summary. The decision criterion or standard of proof employed by the jury does not, by itself, determine the ratio of judicial errors. The ratio of errors also depends on the prior odds of guilt and the accuracy of the jury. For a given standard of proof, increasing the prior probability of guilt decreases $P(AG)/P(CI)$, whereas increasing the accuracy of the jury decreases the difference between $P(AG)/P(CI)$ and threshold posterior odds.

Constant Blackstone-Like Ratios as Policy Goals?

Although May cautioned against interpreting Blackstone-like ratios as expressions of specific social policies, the temptation to do so is apparently very strong. Hopefully, the preceding analysis has convinced the reader that error ratios cannot be so easily equated with standards of proof. However, the fundamental assumption that an appropriate numerical translation of Blackstone's statement may be used to guide social policy has not been addressed. In other words, is it reasonable to choose a decision rule so that the ratio of erroneous acquittals to erroneous convictions is a constant equal to 5, 10, 20, or some other socially acceptable value? What would be the consequences of such a policy? What, if anything, would such a policy imply about society's utilities for the outcomes of jury trials? What happens if such a policy,

---65 See May, supra note 4, at 653–654.
established at one location at one point in time, is adopted as a precedent and applied in different circumstances? To date, these questions have not been addressed in the legal literature.

Figure 7 indicates that if the error ratio, the prior odds, and the accuracy of the jury system can all be quantified, the error ratio implies a particular decision rule. This rule can be easily re-expressed as a threshold probability or standard of proof. However, because utilities are not explicitly considered in the selection of the standard of proof, the standard may be very different from the utility-maximizing value given by Equation 13. The fact that the decision rule implied by a particular Blackstone-like ratio may differ from the optimal rule implies that adopting such a ratio as a matter of policy will necessarily violate the axioms of SEU. In other words, basing a decision rule on outcome utilities and basing a decision rule on a particular error ratio are not two different ways of accomplishing the same goal. They are fundamentally different ways of making decisions.

Although it is tempting to draw conclusions regarding utilities from the decision threshold implied by a Blackstone-like ratio, the validity of Equations 13–15 depends upon the maximization of expected utility. Because the adoption of a constant Blackstone-like ratio is also an implicit rejection of SEU, utilities cannot be inferred from decision thresholds in this manner. Thus, Connolly’s cautionary comments regarding the tenuous relationship between Blackstone’s ratio and outcome utilities⁶⁶ may not go far enough. From an expected utility perspective, outcome utilities do imply an optimal Blackstone-like ratio. However, if adopted as a matter of policy, a Blackstone-like ratio does not imply outcome utilities because the relationship between decision thresholds and outcome utilities is not defined in the absence of SEU.

However, it may be that the distinction between Blackstone-like ratios and expected utility has been overstated in the preceding paragraph. After all, expected utility theory was “derived” centuries after Fortescue, Hale, and Blackstone made their now-famous comments. It is at least possible that these authors were attempting to convey something akin to an intuitive utility analysis. Why else would they attempt to specify the appropriate tradeoff between the two types

---

⁶⁶ Connolly, supra note 15, at 104–05.
of errors? If their statements represent initial attempts to balance the costs of the two types of errors, then their *goal* in selecting an error ratio was essentially identical to that of SEU. If this is the case (and I am not saying that it is), it may be inappropriate to disallow the connection between decision rules and outcome utilities simply because fixed error ratios do a relatively poor job of maximizing expected utility. Certainly, these authors cannot be faulted for not having the mathematical expertise to convey their opinions more accurately.

Unfortunately, fixed-error-ratio decision rules do not fare much better when one is allowed to infer outcome utilities (or, more precisely, differences between outcome utilities) from implied decision rules. The first and most obvious problem is that there is great difficulty in determining the decision rule itself. As noted above, it is quite difficult to assess the prior odds of guilt, the accuracy of the jury system, the conditional probabilities $P(A|G)$ and $P(CI)$, or their ratio—which, after all, is to be held constant. If the court does not have the privilege of observing these quantities, there is no way it can establish an appropriate decision rule.

Acquiring such knowledge, however, would not solve the court’s problems. Because the court has little control over the prior odds of guilt or the accuracy of the jury system, its primary means for exerting control over the error ratio is the selection of the standard of proof. If the connection to SEU is allowed, the relationship among the utilities of the four outcomes is given by Equation 13. As has been shown above, there is good reason to believe that the implied utilities will be different from the true utilities (which may be assessed independently). Moreover, if the conditions under which the court operates change, Figures 4 and 7 imply that it must adopt a different decision rule in order to maintain a constant error ratio. For example, if improved technology leads to higher priors, the court must compensate by adopting a more stringent standard of proof (i.e., a higher value of $P_T$).

Thus, the logic behind a constant-error-ratio policy is seen to be exactly backwards. The policy is selected at the outset, and it implies a standard of proof that is unlikely to correspond society’s values. Just as importantly, maintenance of the policy across different conditions

---

(different priors and accuracies) implies that different values are held in these conditions. Thus, Blackstone’s ratio, as a policy, violates the fundamental premise of SEU and all other “consequentialist” theories of decision making. Even if the specifics of SEU are rejected, it seems reasonable to argue that decisions and decision criteria should be based on the relative desirabilities of potential outcomes.\textsuperscript{68} From this perspective, maintaining a constant ratio of errors is decidedly nonnormative because it implies that the relative desirabilities of outcomes follow from the policy itself.

**Conclusion**

There can be no doubt that, in the last 25 years, the analysis of error ratios has far exceeded anything that Blackstone or his predecessors could have intended or even imagined. Certainly, the present investigation continues this trend. However, two differences distinguish this work from previous endeavors. First, the concerted effort to translate the remarks of Fortescue, Hale, and Blackstone faithfully into mathematics has led to a clearer understanding of the properties of error ratios. Second, these venerable ratios have not been distorted to fit the mold of SEU. Rather, important differences between the constant-error-ratio and utility-maximization approaches have been illuminated. Significantly, what has been viewed as a fundamental consistency between these approaches is now seen to be a fundamental inconsistency. The primary results of these analyses may be summarized as follows.

(1) The error ratios of Fortescue, Hale, and Blackstone cannot be fairly interpreted as calls for specific social policies. However, if one desires to express such statements mathematically, the ratio-of-frequencies interpretation should be preferred to the ratio-of-utilities interpretation. The ratio-of-frequencies interpretation is more consistent with the authors’ original wording, places less severe restrictions on the utility functions for the two types of errors, and does not imply that the utility scale has ratio properties.

(2) All else being equal, more stringent standards of proof do imply higher Blackstone-like

Figure 1. Signal detection theory representation of the jury system. The distance between the two distributions is indicative of assumed accuracy ($d' = 2$). The vertical line represents the decision criterion or standard of proof. The conditional probabilities refer to the areas under the appropriate curves on either side of the decision criterion. A = acquit, C = convict, I = innocent, and G = guilty. Although normal distributions are depicted, normality is not an essential component of SDT. Other unimodal distributions (e.g., logistic, Poisson) yield similar results in many situations.
Figure 2. Three sets of utility functions consistent with Blackstone's ratio. Linear (thick), logarithmic (medium), or square root (thin) functions are shown for convicting innocent defendants (left) and acquitting guilty defendants (right). Points A, B, and C are discussed in the text.
Figure 3. Subjective expected utility derivation of the optimal threshold probability, $P_T$ (see Equation 13). Because the utility differences used in this example are arbitrary, the implied threshold is not definitive.
Figure 4. The relationship between threshold posterior odds \([P(G|X)/P(I|X)]_T\) and the Blackstone-like ratio \(P(AG)/P(CL)\). Curves represent prior odds of guilt \(P(G)/P(I)\) as labeled. Accuracy is the same as in Figure 1 \((d' = 2)\) for all cases. Points A, B, and C are discussed in the text.
Figure 5. The relationship between threshold posterior odds $[P(G|x)/P(I|x)]_T$ and the Blackstone-like ratio $P(AG)/P(CI)$. Curves represent different levels of accuracy ($d''$) as labeled. Prior odds of guilt $P(G)/P(I)$ equal 1:1 for all cases.
Figure 6. The relationship between accuracy ($d'$) and the Blackstone-like error ratio $P(AG)/P(\neg AG)$ when the threshold posterior odds of guilt $[P(\neg G|x)/P(G|x)]_{T}$ equal 10:1. Curves represent prior odds of guilt $P(G)/P(\neg G)$ as labeled.
Figure 7. The relationship between accuracy ($d'$) and the threshold posterior odds of guilt $[P(\text{Gl|x})/P(\text{I|x})]$ when the Blackstone-like error ratio $P(\text{AG})/P(\text{CI})$ equals 10:1. Curves represent prior odds of guilt $P(G)/P(\text{I})$ as labeled.