Cowles Foundation Discussion Papers   Cowles Foundation

7-1-1982

Information Processing and Jury Decision making

Alvin K. Klevorick

Michael Rothschild

Christopher Winship

Recommended Citation

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 635

INFORMATION PROCESSING AND JURY DECISIONMAKING

by

Alvin K. Klevorick,

Michael Rothschild

and

Christopher Winship

July 1982

INFORMATION PROCESSING AND JURY DECISIONMAKING

Alvin K. Klevorick
Yale University

Michael Rothschild
University of Wisconsin
National Bureau of Economic Research
Mathematica, Inc.

Christopher Winship
Northwestern University

I. Introduction

The kinds of situations in which a group of people is called upon
to make a decision are numerous. One common setting involves a group of
individuals who all have access to the same data base, or to the same
informational input for the decision, and who are faced with a dichotomous
choice. An example is a corporate investment committee to which a
presentation is made concerning a particular capital project and whose task
is to decide whether or not to proceed with the project. Another is a
personnel committee (perhaps that of an academic department) which inter-
views a candidate for a position (or hears the candidate give a seminar)
and must decide whether or not to hire the individual. There are many
other similar decision settings. The prototype, however, is the criminal
jury, and we will analyze the kind of group decision just described in
terms of the criminal jury's decision process.

At the end of a trial each juror has his/her own perspective on
what has transpired. Each one could, on that basis, render a verdict in
the case. But under the jury system, as we know it, the jury must
deliberate before it--and not the members of the jury acting individually--
renders a verdict. An important claim made in favor of the jury system
(and similar decisionmaking procedures) is that during the deliberation
process, jurors exchange points of view and assemble the evidence into a
coherent picture that is more likely to be correct than is the view of any
one juror. That is, one of the virtues claimed for a jury decision is that
it is based on more complete and better processing of the information
available than the verdict of any one juror deciding alone would be.

1

Given the central and valuable role attributed to the information processing that jury deliberation is supposed to achieve, it is striking that existing models of the jury decision process--including both abstract mathematical formulations and simulation models--are inattentive to this aspect of the jury's work. Although these models depict how a jury might move to a verdict from the initial views of its members, they do not provide any description or specification of how the jurors' views are combined or how their various observations and insights are assimilated. Indeed, some of the more sophisticated mathematical models imply that if a jury deliberates, it will be more likely to err--to convict an innocent defendant _and_ to acquit a guilty defendant--than if it simply decides the case by a simple majority vote _before_ any deliberation occurs!

The purpose of this paper is to explore, in the context of a formal model, the information processing function that jury deliberation and its analogues perform. In particular, we investigate when a jury that deliberates to a unanimous verdict can reach better decisions--ones with lower probabilities of error--than a jury that bases its decision on the view of the majority of its members immediately after the trial is concluded. What is the gap between the quality of decisions reached using a first-ballot, majority-rule procedure and the quality of those that would be generated by a jury making optimal use of the information provided at the trial?

It should be emphasized that we are not saying that juries that deliberate actually process what they have seen and heard in an optimal way. Indeed, in the complicated setting of an actual jury trial, it is not even clear how one would characterize optimal processing of information.

2

Our strategy, instead, is to consider a simple model of juror observations, though one that is richer than characterizations in the literature, for which we can define precisely what an optimal jury decision rule would be. Of course, jurors do not read the books on statistical decision theory that discuss this optimal procedure nor do actual juries aggregate information according to this procedure. Hence, our inquiry is not aimed at measuring the performance of actual juries against that of first-ballot, majority-rule juries. Rather our concern is to gauge, within the context of a particular model, the improvement that deliberation can yield and to consider the circumstances that affect that possible gain. We will see that deliberation has the potential for generating substantial improvement in the quality of decisions, and we will see how that potential arises, especially the central role that heterogeneity among jurors—in terms of what they see and hear, what they believe about the costs of erroneous decisions, and what differences there are in their information processing capacities—plays in determining how much improvement is possible.

We begin, in the next section, with a discussion of the implications that existing models have for the value of jury deliberation. Then in Section III, we present the model of juror observation that is basic to our analysis. Section IV contains our comparison of the jury that makes optimal use of trial information and the one that decides by first ballot majority vote. Central to the analysis in Section IV is the assumption that jurors may differ in what they see and hear at the trial. But that discussion assumes that the jurors share the same view of the relative cost of erroneous convictions and erroneous acquittals and that they have the same individual abilities to process information. Sections V

3

and VI, respectively, consider models in which these assumptions are, each in turn, relaxed. The last section contains some concluding remarks.

The analysis in Sections III-VI is technical and detailed. We conclude this introduction with an informal discussion to convey a sense of the nature of the models we examine and the results we obtain. Formal justifications for the various rules mentioned here are, of course, deferred to the detailed development.

To fix ideas, it is best to have a concrete (if unrealistic) example in mind. Suppose that the jury in a particular trial knows that the defendant is guilty if he is more than six feet tall. The jury also knows that the defendant's height is either six feet or six feet one inch. Thus, the inferential problem each juror faces is one of deciding whether the accused is six feet tall and innocent or six feet one inch tall and guilty. Imagine that the trial provides an opportunity for each juror to view the defendant and thus to try to guess his height. When the juror views the defendant, he/she forms an estimate of his height, which we denote as X_i. This estimate is a normal random variable with variance σ^2, and its mean is six feet if the defendant is innocent and six feet one inch if the defendant is guilty.

First consider the problem that a jury composed of a single individual will have in deciding the case on the basis of his/her information. The juror's decision rule is a simple one: set a threshold Q and convict if the estimate X_i is greater than Q. Thus the juror's problem is to set Q. It is straightforward to calculate that Q is a function of three things: the relative cost of the two kinds of errors (erroneous convictions, erroneous acquittals) the juror can make, the variance σ^2, and

4

the average of the two alternative values of the mean. In fact,

(1) $Q = 6' 1/2'' + \sigma^2 k,$

where k is the log of the ratio of the cost of convicting the innocent to
the cost of acquitting the guilty. Notice that if $k = 0$, so that the cost
of the two kinds of errors is the same, it is not necessary to know σ^2
while if $k \neq 0$, then it is necessary to know σ^2. Consequently, the case
$k = 0$ is much simpler than the case $k \neq 0$.

Now suppose that the jury is composed of several individuals whose
observations are independent. Each juror will make an estimate of the
defendant's height and the jury's problem is to make a decision on the
basis of the observations of all its members $X = (X_1, \ldots, X_n)$. If the
jury deliberates and uses this information to make an optimal decision, it
will base its verdict on the sample mean $\bar{X} = \sum_{i=1}^{N} X_i$, a sufficient statistic
for X. Once again the decision will be to convict if $\bar{X}$ is greater than
or equal to some threshold value Q_n and to acquit if $\bar{X}$ is less than Q_n.
Since $\bar{X}$ is a normal random variable with variance $\sigma^2(\bar{X}, n) = \sigma^2/n$, the
optimal choice will be to set

(2) $Q_n = 6' 1/2'' + \sigma^2(\bar{X}, n)k.$

Again if $k = 0$ the decision is simpler because the threshold value is
independent of σ^2, the variance of the individual observations.

Now consider a jury that decides the same case by voting without
deliberating, a first-ballot jury. Each juror decides on the basis of
his/her observation whether to vote to acquit or to convict. The simplest
voting rule is majority voting and if jurors all apply the same standard,

5

Q, then the value of the median juror's estimate will determine the defendant's fate. The problem then is what this standard should be. If the cost of the two kinds of errors is the same (k = 0), then considerations of symmetry make it clear that each juror should vote to convict if and only if his/her estimate is greater than 6' 1/2", just as if he/she were the only juror. The decision of the first-ballot jury will be less efficient than that of the deliberating jury of the same size because the median contains less information than the mean—the median is not a sufficient statisic for X. But, when errors are equally costly, the decision will be reasonable in the sense that the _relative_ sizes of the two error probabilities will be what the members of the jury would want them to be—namely, equal.

If k ≠ 0, so that false convictions and false acquittals are weighed differently, it is very difficult to see how the first-ballot jury can achieve even this limited goal of desired _relative_ error sizes. For if k ≠ 0, then the variability of the estimate must be taken into account when a threshold is set. If this is not done, the relative costs of the two kinds of errors cannot be properly balanced against one another. Whatever threshold or standard used, it will clearly depend on σ^2. Hence, to apply a reasonable rule when k ≠ 0, jurors must communicate what σ^2 is. How they are to do this without deliberating is unclear.

To this point, we have assumed, unrealistically, that jurors make independent observations. This seems thoroughly incompatible with the basic fact that they have all observed the same trial. It also has an obvious and implausible implication—a jury of infinite size will never err. An

6

independent observations model essentially assumes that the trial generates enough information to settle the question of whether the accused is guilty or innocent. All that is necessary is that there be enough jurors to extract all the information. As we show in Sections III-VI, the most obvious changes in the direction of realism make it even harder for the first-ballot jury, the one that does not deliberate, to reach a reasonable decision. When the amount of information that the trial generates is limited, the tasks of assessing the variability of the sample median and taking it into account when balancing the two kinds of errors become much more difficult.

Suppose, as we do in our analysis, that the information the trial generates is, at best, limited: an infinite number of jurors would only be able to discern the limited information the trial reveals. In the concrete example we have been discussing, imagine that all jurors view the defendant through a foggy one-way mirror. Each juror's observation consists of two parts: the commonly observed events in the trial, which are imperfectly observed by all jurors, and an idiosyncratic error of observation, which is due to independent variations in eyesight, quality of lighting, and the like. The individual errors, which have mean zero, are independent of each other and of the common information in the trial. Again, all random variables are normally distributed.

Consider the problem of the deliberating jury. We show, once again, that the optimal decision will be based on the sample mean of the jurors' observations and that the problem is where to set the threshold. If $k = 0$, the problem is easy; in our example, the threshold is 6' 1/2". If $k \neq 0$, since $\bar{X}$ is again a normal random variable with a variance equal to

7

$\sigma^2(\bar{X}, n)$ and mean conditional on the innocence or guilt of the defendant, we can again use formula (2) to set the threshold Q_n. In this case, however, $\sigma^2(\bar{X}, n)$ is a more complicated function than in the independent errors case discussed above. In particular, to compute $\sigma^2(\bar{X}, n)$ in this case, each juror must partition the error in what he/she observes into two parts--an idiosyncratic error and a common error. This shows the value of deliberation. Unless jurors get together and talk, it is hard to see how they can correctly make this partition. Deliberation allows jurors to compare observations so that they can correctly aggregate the information they have separately observed. In our example, discussion would bring out the fact that all jurors' vision was distorted by the same foggy mirror. Thus, jurors would know that their errors of observation were not independent.

As for the first-ballot jury, if $k = 0$, then the correct standard is the symmetric one: convict, if the median observation is greater than 6' 1/2". But if $k \neq 0$, the first-ballot jury's problem is very difficult because the median juror must somehow estimate the variance of the median. This is very difficult to do for a correct estimate depends on partitioning the error variance into common and independent components. We do not see how this can be done if jurors do not talk to one another. We note in Section III that if jurors knew the model generating the observations they make, they could devise rules which, if followed, would lead to decisions which were both asymptotically reasonable and as good as those which an optimally deliberating jury could reach. Those decision rules are complicated and they depend on the parameters of the model which will vary

8

from case to case. We do not believe that juries or other groups in
similar decision situations are likely to use them.

II. The Implications of Existing Models

One of the most sophisticated efforts at modeling jury verdicts is that of Gelfand and Solomon. The characteristic that determines the value of the deliberative process in their model is a common feature of models in the literature. Hence, the implications that existing models have for the value of jury deliberation are best introduced by considering Gelfand and Solomon's most recent contributions [Gelfand and Solomon (1977a and b)].

To consider Gelfand and Solomon's results, a brief review of their approach is needed. They begin by partitioning a jury's decision process into two stages. First, they model the determination of the first-ballot distribution of jurors' votes. For this stage, they use a mixed binomial formation that they previously developed as an extension of Poisson's models for jury verdicts [Gelfand and Solomon (1973), (1974)]. Then, in their second stage, they employ some social decision scheme models incorporating varying degrees of majority persuasion to explain how the jury moves from its first-ballot position to its final decision. Following the work or Davis and his colleagues [see Davis (1973) and Davis, Bray, and Holt (1977)], the social decision schemes Gelfand and Solomon use are comparative static, first-ballot/final-verdict models, which do not depict any of the dynamic features of the jury's deliberation process. A social decision scheme is simply a transition matrix that represents the jury's deliberation as a stochastic process moving the jury from a first ballot vote to a final position. ·

To estimate the parameters of their model determining the first-ballot distribution of jurors' votes, Gelfand and Solomon used a unique data set gathered by Kalven and Zeisel as part of the Chicago Jury

project [Kalven and Zeisel (1966)]. These were data on 225 criminal cases in Brooklyn and Chicago for which Kalven and Zeisel were able to use post-trial interviews to reconstruct the juries' first-ballot votes. These were, in fact, the 225 cases that led Kalven and Zeisel to their "majority persuasion" hypothesis that the final verdict is largely determined by the position that the majority of jurors take <u>before</u> any deliberation occurs.[1] Appendix A indicates how the Kalven and Zeisel data on these 225 cases can be used to estimate the parameters in the first stage of the Gelfand and Solomon model.

The second stage of the model specifies how the jury moves from its first-ballot position to a final verdict. Gelfand and Solomon consider two alternatives involving different degrees of majority persuasion. First, they assume that the first ballot majority always prevails and that if the jury is initially evenly split, then with probability 1/2 the final verdict will be innocent and with probability 1/2 it will be guilty. They apply this transition model to both twelve-member and six-member juries.

But Gelfand and Solomon regard this "first-ballot majority decides the outcome" assumption as a crude approximation, and they go on to consider a more refined social decision scheme. Their refined scheme for the twelve-person jury is based on the one Davis (1973) suggested as a result of his experience with a large number of mock jury trials.[2] When they turn to the six-person jury, Gelfand and Solomon provide no elaborate justification for the social decision scheme they present. ·Rather, they simply assert that it is plausible to scale down the Davis scheme for twelve-person juries and apply it to a six-person jury. That is, they assume that the probability of a transition to a particular six-person jury

11

TABLE 1

ERROR PROBABILITIES OF JURY DECISIONS*

| | Conviction of an Innocent Defendant | | Acquittal of a Guilty Defendant | |
	(1) First-Ballot Majority	(2) Post-Deliberation Unanimous Verdict	(3) First-Ballot Majority	(4) Post-Deliberation Unanimous Verdict
6-Person Jury	.0144	.0688	.0138	.0639
12-Person Jury	.0007	.0469	.0007	.0266

*Calculated from Gelfand and Solomon (1977a), Tables 4 and 7.

verdict if k out of 6 people initially vote for conviction is the same as the transition probability to that verdict that Davis specifies for a twelve-person jury when 2k out of 12 people initially vote for conviction.

Gelfand and Solomon use their two-stage model of jury decision-making to analyze the relative magnitudes of the errors six-person and twelve-person juries would make. They conclude that a twelve-member jury is to be preferred to a six-person jury. The basis for their conclusion is shown in Table 1, the entries of which are calculated from the relevant parts of Tables 4 and 7 in the Gelfand and Solomon paper. Columns (1) and (3) are based on the model in which the transition from first-ballot vote to final verdict is determined entirely by the first-ballot majority while columns (2) and (4) use as the transition probabilities from first ballot to final verdict the elements of Gelfand and Solomon's modification of Davis' social decision scheme. The major observations Gelfand and Solomon make are that a six-person jury is much more likely to convict an innocent person than is a twelve-person jury and that a six-person jury is much more likely to acquit a guilty person than is its twelve-person counterpart.[3/]

There is, however, another very important point to observe about the numbers in Table 1, an observation that indicates the need for the kind of further theoretical research on the jury decision process that we present here. Instead of comparing the two rows in the table, as Gelfand and Solomon do, consider each row separately and compare the entry in column (1) in the row with that in column (2) and compare the entry in column (3) in the row with that in column (4). The clear conclusion is that allowing a jury of either size to deliberate until it reaches a unanimous verdict results in a higher probability of convicting an innocent

12

defendant _and_ a higher probability of acquitting a guilty defendant than one would have if the verdict were based instead on the outcome of a predeliberation, first-ballot majority vote. In fact, according to Gelfand and Solomon's results, a six-person jury deciding the case by a first-ballot majority vote would make much better decisions--lower error probabilities for both kinds of errors--than would a twelve-person jury deciding under a unanimity rule.[4/]

This interpretation of the Gelfand and Solomon results summarized in Table 1 suggests that there is a problem, and there is one. The fact that deliberation increases the probability of error is a consequence of an untested (and inexplicit) assumption of the Gelfand and Solomon model. Specifically, they assume that the result of deliberation depends only on the first-ballot vote and a stochastic process, and not on the guilt or innocence of the accused. This means that all of the information about the defendant's guilt or innocence is contained in the first ballot vote. Let $q(k,N;G)$ denote the probability that a jury of N persons that begins with a predeliberation vote of k for conviction eventually convicts a guilty defendant, and let $q(k,N;I)$ be the probability that a jury of N persons with k initial-ballot votes for conviction eventually convicts an innocent defendant. Then Gelfand and Solomon essentially assume that:

$$q(k,N;G) = q(k,N;I) = q(k,N).$$

But this implies that the deliberation process adds noise to the first-ballot vote, which by assumption in the Gelfand and Solomon model is an estimate of the guilt or innocence of the accused. Thus, the deliberative

14

process can only produce decisions that are worse than those resulting from a predeliberation majority vote.

A rather simple formal argument will clarify this point. To facilitate the analysis, assume N is odd. (The same line of reasoning can be followed if N is even, but one must keep careful account of tied votes.) Then, under the first-ballot majority-rule standard, an initial vote of more than N/2 for conviction will result in a conviction while an initial vote of less than N/2 for conviction will result in an acquittal. Let

$f(k,N;G)$ = probability that on the first ballot k of N jurors vote to convict a guilty defendant;

$f(k,N;I)$ = probability that on the first ballot k of N jurors vote to convict an innocent defendant.

Then, the probability that an N-person jury operating under the first-ballot majority-rule standard will convict an innocent defendant is:

$$\sum_{k>N/2} f(k,N;I).$$

The probability that the same error will be made by a jury that reaches a verdict only when the jurors are unanimous in their position is

$$\sum_{k=0}^{N} f(k,N;I)q(k,N).$$

Hence, the result that the deliberative process increases the probability of an erroneous conviction--that is, the probability of a guilty verdict when the defendant is truly innocent--which appears in the Gelfand and Solomon estimates in Table 1, can be expressed as:

15

$$(1) \quad \sum_{k=0}^{N} f(k,N;I)q(k,N) > \sum_{k>N/2} f(k,N;I).$$

We will now present two assumptions that, taken together, imply that (1) is true _if_ the jury necessarily reaches a verdict--that is, if the stochastic process describing the jury's movement from its first ballot position does _not_ permit hung juries. The first is essentially a symmetry assumption:

$$(2) \quad q(k,N) + q(N-k,N) = 1.$$

This can be interpreted more easily if it is written as:

$$(2') \quad q(k,N) = 1 - q(N-k,N).$$

The left-hand side of (2') is the probability that an N-person unanimity-rule jury eventually convicts a defendant (_whether he/she is innocent or guilty_) when there are initially k votes for conviction. The right-hand side of (2') is the probability that an N-person unanimity-rule jury eventually acquits a defendant when there are initially k votes for acquittal. (Since q(N-k,N) is the probability that a jury eventually convicts given that N-k initially vote to convict and since we assume here that the jury eventually acquits _or_ eventually convicts the defendant, 1-q(N-k,N) is obviously the probability that the jury with N-k initially voting to convict--and hence k initially voting to acquit--eventually acquits.)

The second assumption, which is satisfied by the Gelfand and Solomon model, is that it is more likely that on the first ballot a majority of the jurors will vote correctly and a minority will err than it

16

is that the predeliberation majority will be in error. In terms of our notation, the assumption is that:

(3) $\begin{cases} f(k,N;I) > f(N-k,N;I) \text{ for } k < N/2 \\ \text{and} \\ f(k,N;I) < f(N-k,N;I) \text{ for } k > N/2. \end{cases}$

To see that the conditions in (3) conform to the preceding verbal statement, consider the case of k < N/2. Then (3) says that the probability that on the first ballot k jurors vote to convict an innocent defendant is greater than the probability that N-k jurors vote to convict where N-k > k since k < N/2. The second half of (3), for the case where k > N/2, can be similarly interpreted.

The proposition suggested by the Gelfand-Solomon results--that deliberation produces decisions that are worse than those yielded by a predeliberation majority-rule vote--can be proved as follows.

<u>Proposition</u>: Under conditions (2) and (3) above, it follows that:

(1) $\sum_{k=0}^{N} f(k,N;I)q(k,N) > \sum_{k>N/2} f(k,N;I)$

<u>Proof</u>:

$\sum_{k=0}^{N} f(k,N;I)q(k,N) = \sum_{k>N/2} [f(k,N;I)q(k,N) + f(N-k,N;I)q(N-k,N)]$

$> \sum_{k>N/2} f(k,N;I)[q(k,N) + q(N-k,N)] \text{ by (3).}$

But, by (2), q(k,N) + q(N-k,N) = 1, and we have

17

$$\sum_{k=0}^{N} f(k,N;I)q(k,N) > \sum_{k>N/2} f(k,N;I). \qquad \text{Q. E. D.}$$

This Proposition concerns the probability of an erroneous conviction. But it is easily shown that assumptions (2) and (3) also yield the analogous result for the probability of an erroneous acquittal. Namely, the probability of such an error is higher if the jury deliberates subject to a unanimity rule than if it simply follows the dictates of its first-ballot majority.

The argument leading to the Proposition above applies to a model in which there are no hung juries. Hence it cannot directly explain the results in Table 1 since the Gelfand and Solomon model of the jury deliberating subject to a unanimity rule includes the possibility that such a jury can fail to reach a verdict. At a minimum, introducing the possibility that the jury might fail to reach a verdict requires that the symmetry assumption in (2) and (2') be re-stated because (2) and (2') are based on the assumption that conviction and acquittal are the only outcomes. Let $r(k,N)$ denote the probability that a jury of N persons with k first-ballot votes for acquittal eventually acquits the defendant, and let $\ell(k,N)$ be the probability that a jury of N persons that begins with a predeliberation ballot of k votes for conviction eventually fails to reach a verdict. Then, the corresponding symmetry assumption when juries can hang is:

$$(2'') \quad \begin{cases} q(k,N) = r(N-k,N) \\ \qquad \underline{\text{and}} \\ \ell(k,N) = \ell(N-k,N). \end{cases}$$

This symmetry assumption is satisfied in the Gelfand and Solomon model except for a jury that is evenly split on the first ballot. That is, when the twelve-person jury has six people voting for conviction and six voting for acquittal on the first ballot, Gelfand and Solomon specify that the probability of eventual conviction is 1/2, the probability of eventual acquittal is 1/4, and the probability of a hung jury is 1/4.

When hung juries are possible, there is also some question about how to measure a jury's accuracy, or alternatively how to define a jury error, even given that one begins by conditioning on the true state of nature--the defendant's guilt or innocence. One possibility is to measure accuracy by the probability that a jury convicts a guilty defendant and the probability that it acquits an innocent defendant. In this case, the jury errs (1) when the defendant is innocent and the jury fails to acquit--that is, it votes to convict or fails to reach a verdict, and (2) when the defendant is guilty and the jury fails to convict--that is, it acquits or it hangs. Under this definition of jury error and the symmetry condition (2"), the exact analogue to the Proposition obtains--deliberation produces decisions that are worse than those yielded by a predeliberation majority-rule vote.

A second possible approach to defining jury error when hung juries are possible is to retain the definition used when a decision was assured: the jury errs when it convicts an innocent defendant or acquits a guilty one. Under this second definition, a jury that fails to reach a verdict is not regarded as having made an error. With this definition and with symmetry condition (2") replacing condition (2), we can conclude that the

errors of the deliberating unanimity-rule jury exceed those of the first-ballot majority-rule jury _if but only if_ "hung juries do not occur too often," in a sense that can be made precise. The need for such a proviso should be clear, and it can be made obvious by considering an extreme example: Suppose deliberating juries never reach a verdict—they always hang. Then under this second definition of jury error, a deliberating jury never makes a mistake, while first-ballot majority-rule juries would.

The critical assumption underlying the proposition, its analogues when hung juries are possible, and the manifestation of these results in Table 1, with higher error probabilities for juries that deliberate than for those that do not, is that $q(k,N;I) = q(k,N;G)$. This assumption is quite common in the literature on juries, including in the reverse Ehrenfest model of the jury decision process, which two of us previously analyzed.[5] But the assumption is generally implicit rather than explicit and has obviously not been tested.

III. A Model of Juror Observations

It is certainly no surprise that a first-ballot majority vote results in lower error probabilities than does a unanimous, post-deliberation verdict when deliberation is unrelated to the truth. But how does the outcome of the first-ballot majority-rule procedure compare with the post-deliberation outcome when the jury's deliberation effectively uses the information provided at trial? How do the two decision procedures compare when the jury's discussion provides the opportunity for the jurors to share the various observations they have made during the trial, for the strength of different jurors' views to be considered, and for the jurors to discuss the standards to be applied in reaching a verdict? To answer these questions, we need to develop a characterization of juror observation and jury deliberation that is richer than those in existing models.

We assume that n jurors observe the trial. They may see different things in the same set of in-court proceedings, and they may process what they have seen in different ways. Consequently, each juror enters the jury room not only with a view about whether the defendant is innocent or guilty but also with an idea about how strongly he/she holds that view. The following model of correlated normal observations captures this structure.

Assume that the ith juror observes

$$(1) \qquad X_i = \frac{\hat{Y}_i + a\hat{Y}_0}{[1 + a^2]^{1/2}} , \quad 0 \leq a \leq \infty$$

where $\hat{Y}_0, \hat{Y}_1, \ldots, \hat{Y}_n$ are independent normal observations with unit variance; and for $i = 0, \ldots, n$

$$(2) \qquad \hat{EY_i} = \begin{cases} 0 \text{ if guilty} \\ -B \text{ if innocent.} \end{cases}$$

The observation $\hat{Y_0}$ represents what all the jurors saw, while $\hat{Y_i}$ is what only the ith juror observed. The parameter a indicates the common weight all the jurors place on the common "signal" they extracted from the trial relative to the weight each attaches to his/her individual observation. Finally, $B > 0$ measures the information content of the signal provided by each observation. If B is large, it is relatively easy to tell the innocent from the guilty. If B is small, it is very difficult to do so.

It is important, from both a conceptual and an analytical point of view, to recognize that this model of juror observation is equivalent to one in which each juror receives the same information from the trial but different jurors make independent errors of observation. In the latter model, except for the different errors they make, all the jurors are viewed as having the same information. To see the equivalence between the two models, rewrite X_i as follows:

$$X_i = \frac{(\hat{Y_i} - \hat{EY_i}) + a\left[\hat{Y_0} + \dfrac{\hat{EY_i}}{a}\right]}{(1 + a^2)^{1/2}}.$$

Let $Y_i = \hat{Y_i} - \hat{EY_i}$ for $i = 1, \ldots, n$ and $Z = \hat{Y_0} + (\hat{EY_i}/a)$, and recall that $\hat{EY_i}$ is the same for all i so that Z does not depend on i. Then the ith juror's observation is:

$$(3) \qquad X_i = \frac{Y_i + aZ}{(1 + a^2)^{1/2}}$$

22

where

(4) $Y_i \sim N(0,1)$ for $i = 1, \ldots, n$;

(5) $\begin{cases} Z \sim N(0,1) & \text{if guilty} \\ Z \sim N(-D,1) & \text{if innocent}; \end{cases}$

the Y_i's are independent ($i = 1, 2, \ldots, n$) of each other and of Z; and
$D = \frac{1+a}{a} B$. It is clear that D is the bound on the ability of the _jury_ to
distinguish between the innocent and the guilty.

In the formulation given by (3)-(5), Z is the information conveyed
by the trial, Y_i is the observation error the ith juror makes, and X_i is a
measure of the ith juror's perception of the defendant's guilt. The
jurors' errors of observation are independent, identically distributed
normal random variables. Letting

(6) $T = \dfrac{Da}{(1 + a^2)^{1/2}}$,

(7) $\begin{cases} X_i \sim N(0,1) & \text{if guilty} \\ X_i \sim N(-T,1) & \text{if innocent}. \end{cases}$

The correlation between the X_i's, which we denote ρ, is:

(8) $\rho = \dfrac{a^2}{1 + a^2}$

so that $T = D\sqrt{\rho}$. As (7) shows, T is a measure of the individual _juror's_
ability to make a correct decision. Hence, for example, the larger ρ is,

the larger is T, and the easier it is for a juror to distinguish between guilt and innocence in the case of the particular defendant.

Before we examine the ways in which an n-person jury might use the observations it has--the values of the n X_i variables--let us consider the way a decisionmaker would make optimal use of the information conveyed by the trial, Z. Define the random variable W as follows:

$$(9) \qquad W = \frac{aZ}{(1 + a^2)^{1/2}} \quad .$$

Its distribution is

$$(10) \qquad W \sim \begin{cases} N(0,\rho) & \text{if guilty} \\ N(-T,\rho) & \text{if innocent,} \end{cases}$$

and it should be clear that W, as a transform of Z, contains all the information in the trial. Society will surely want the process that uses this trial information, in deciding the defendant's guilt or innocence, to reflect the society's tradeoff between the two possible kinds of errors that can be made--namely, convicting an innocent defendant and acquitting a guilty defendant. Assume the terms of that tradeoff are given by:

$$(11) \qquad K = \frac{(1-\theta)\Gamma(G|I)}{\theta\Gamma(I|G)}$$

where

θ = the prior probability that the defendant is guilty, and

$\Gamma(i|j)$ = the cost associated with declaring state i when state j is the true state.

Suppose that W is observed. Then, given the distribution in (10) and the fact that K specifies the relative magnitudes of the a priori expected costs of errors of misclassification, the optimal rule if society wants to minimize the expected costs of such errors is to declare the defendant guilty if and only if

$$(12) \quad \frac{W}{\rho}T + \frac{1}{2}\frac{T^2}{\rho} \gtreqless k$$

where $k = \log K$. That is, the optimal decision is:

$$(13) \quad \begin{cases} \text{Convict if } W \gtreqless \frac{k\rho}{T} - \frac{1}{2}T \\[2ex] \text{Acquit if } W < \frac{k\rho}{T} - \frac{1}{2}T. \end{cases}$$

Contrast the optimal social decision rule in (13) with what the ith juror would do if he/she adopted the social tradeoff K as his/her own. Given the distribution of X_i in (7), the optimal decision for the ith juror acting alone is to convict if and only if

$$X_i T + \frac{1}{2}T^2 \gtreqless k.$$

That is, the ith juror would

$$(14) \quad \begin{cases} \text{Convict if } X_i \gtreqless \frac{k}{T} - \frac{1}{2}T \\[2ex] \text{Acquit if } X_i < \frac{k}{T} - \frac{1}{2}T. \end{cases}$$

The crucial difference between the optimal decision based on all the information in the trial and the optimal decision based solely on the ith

juror's observation is that the former takes into account the correlation

between the "non-noisy" components of what the jurors see.

IV. <u>The Optimal Use of Trial Information and the First Ballot Majority Vote</u>

If the model of juror observation in Section III applies to the members of an n-person jury, what is the optimal way for that jury to reach a verdict? How can the jury make the best use of the information available to it and, in particular, take account not only of the initial views of each of its members but also of the varying strength of those views? And, how does the performance of such a jury compare with that of a jury following the first-ballot majority rule?

The answer is that if the model presented in Section III applies, the jury faces a problem in discriminant analysis. The jury must "classify" the defendant as guilty or innocent on the basis of the "measurements"--one per juror--it has on the defendant and the knowledge that the jury observes

(1) $X \sim N(0, \Sigma_n)$ if the defendant is guilty and

(2) $X \sim N(-Te, \Sigma_n)$ if the defendant is innocent,

where $X' = (X_1, X_2, \ldots, X_n)$, $e' = (1, 1, \ldots, 1)$ and Σ_n is the variance-covariance matrix of X. From III(7) and III(8), we know that each diagonal element of Σ_n is 1 and each off-diagonal element is ρ.

The solution to the problem of classifying an observation into one of two known multivariate normal populations with equal covariance matrices is given by Anderson (1958). Following Anderson's approach, the jury makes optimal use of the information it has by forming the statistic

(3) $U_n = X' \Sigma_n^{-1} eT + \frac{1}{2} T^2 e' \Sigma_n^{-1} e.$

It can be shown that

$$
(4) \qquad U_n \sim \begin{cases} N(\frac{1}{2}\alpha, \alpha) & \text{if guilty} \\[2em] N(-\frac{1}{2}\alpha, \alpha) & \text{if innocent} \end{cases}
$$

where

$$
(5) \qquad \alpha = T^2 e' \Sigma_n^{-1} e = \frac{T^2 n(1 + a^2)}{na^2 + 1} = \frac{T^2 n}{(n-1)\rho + 1} \; .
$$

Then, if K is, as defined in Section III, the ratio of the a priori expected cost of convicting an innocent defendant to the a priori expected cost of acquitting a guilty one, and k = log K, the optimal rule for the n-person jury we have modelled is:

$$
(6) \qquad \begin{cases} \text{Convict if } U_n \geq k \\[2em] \text{Acquit if } \; U_n < k. \end{cases}
$$

For the variance-covariance matrix Σ_n of the jury's observations as we have modelled them, the test in (6) reduces to:

$$
(7) \qquad \begin{cases} \text{Convict if } \bar{X}_n \geq \dfrac{k}{T}\left[\dfrac{(n-1)\rho + 1}{n}\right] - \dfrac{1}{2}T \\[2em] \text{Acquit if } \; \bar{X}_n < \dfrac{k}{T}\left[\dfrac{(n-1)\rho + 1}{n}\right] - \dfrac{1}{2}T \end{cases}
$$

where $\bar{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ is the mean of the n jurors' observations.

The error probabilities of the optimally deliberating jury, which are calculated using the distribution of U_n in (4) and the decision rule in (6), are:

$$(8) \quad \begin{cases} \Pr(\text{Acquit/Guilty}) = \phi\left[\dfrac{k - \frac{1}{2}\alpha}{\sqrt{\alpha}}\right] \\[3em] \Pr(\text{Convict/Innocent}) = \phi\left[-\dfrac{k + \frac{1}{2}\alpha}{\sqrt{\alpha}}\right], \end{cases}$$

where ϕ is the cumulative distribution function of the standard normal variate S

$$\phi(S) = \int_{-\infty}^{S} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds.$$

In our analysis, we shall be particularly interested in what happens to jury performance as the jury increases in size. Note first that, from (5):

$$(9) \quad \lim_{n \to \infty} = \frac{T^2(1 + a^2)}{a^2} = \frac{T^2}{\rho}.$$

Hence, there is a limit to the quality of the decisions that the optimally deliberating jury--or any jury that uses the information generated by the trial--will make, and this limit is independent of the size of the jury.

Second, and most important, as $n \to \infty$ the decision rule in (7) approaches the optimal decision rule in III(13), the decision rule that optimally uses all the information in the trial. To see this, observe

29

that as $n \rightarrow \infty$, the term $[(n-1)\rho + 1]/n$ in (7) approaches ρ, and the critical value on the right-hand side of (7) approaches $\frac{k\rho}{T} - \frac{1}{2} T$, which is the right-hand side of III(13). Furthermore,

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i + aZ}{(1 + a^2)^{1/2}} = \frac{\bar{Y}_n + aZ}{(1 + a^2)^{1/2}}, \text{ where } \bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i.$$

But by the strong law of large numbers, it follows from the fact that $E(Y_i) = 0$ that $\bar{Y}_n \rightarrow 0$ as $n \rightarrow \infty$. Hence, as $n \rightarrow \infty$, $\bar{X}_n \rightarrow aZ/(1+a^2)^{1/2} = W$, as defined in III(9). But then as the size of the jury increases, $\bar{X}_n \rightarrow W$ so that a decision based on $\bar{X}_n$ is based on all the information contained in the trial, and the left-hand side of (7) approaches the left-hand-side of III(13).

Finally, we can use (9) above to calculate the asymptotic error probabilities under optimal use of trial information. Denoted with an asterisk, they are:

$$(10) \quad \begin{cases} \Pr^*(\text{Acquit/Guilty}) = \phi\left[k \frac{\sqrt{\rho}}{T} - \frac{1}{2} \frac{T}{\sqrt{\rho}} \right] \\[2em] \Pr^*(\text{Convict/Innocent}) = \phi\left[-k \frac{\sqrt{\rho}}{T} - \frac{1}{2} \frac{T}{\sqrt{\rho}} \right]. \end{cases}$$

One special case of the jury that makes optimal use of the information at trial is worthy of note. It is the case in which the jury attaches equal weight to the a priori expected costs of the two kinds of errors it can make. For such a jury, $K = 1$, hence $k = \log K = 0$, and the jury optimizes by equating the probabilities of making the two kinds of

errors. Setting k = 0 in (7), (8), and (10), we see that such a jury *follows the decision rule:*

$$(7')\quad \begin{cases} \text{Convict if } \bar{X}_n \geq -\dfrac{T}{2} \\[2ex] \text{Acquit if } \bar{X}_n < -\dfrac{T}{2} \; ; \end{cases}$$

the probability it errs is:

$$(8')\qquad \text{Pr(Acquit/Guilty)} = \text{Pr(Convict/Innocent)} = \phi\left[-\frac{1}{2}\sqrt{\alpha}\right] \; ;$$

and the asymptotic value of this error probability is:

$$(10')\qquad \text{Pr}^*(\text{Acquit/Guilty}) = \text{Pr}^*(\text{Convict/Innocent}) = \phi\left[-\frac{1}{2}\frac{T}{\sqrt{\rho}}\right].$$

We will return to this special case in a moment.

Now consider the performance of a jury that renders a verdict based on a simple majority vote of the jurors before any deliberation has occurred. Each juror observes a normal random variable X_i with the distribution given by III(7). Then, given the tradeoff K between the two kinds of errors, the ith juror follows the decision rule in III(14), which we repeat here:

$$(11)\quad \begin{cases} \text{Convict if } X_i \geq \dfrac{k}{T} - \dfrac{1}{2} T \\[2ex] \text{Acquit if } X_i < \dfrac{k}{T} - \dfrac{1}{2} T. \end{cases}$$

Under the first-ballot majority-rule procedure, the jury will convict the defendant if and only if the number of guilty votes is greater than n/2. Thus, the probability that the jury will render a guilty verdict is the

probability that more than n/2 of the X_i's are greater than or equal to $\frac{k}{T} - \frac{1}{2} T$ so that

(12)
$$
\begin{cases}
\text{Pr(Conviction)} = \text{Pr}\left[X_{med,n} \geq \frac{k}{T} - \frac{1}{2} T \right] \\
\\
\text{Pr(Acquittal)} = \text{Pr}\left[X_{med,n} < \frac{k}{T} - \frac{1}{2} T \right]
\end{cases}
$$

where $X_{med,n}$ is the median observation in a jury of size n.

Hence, as one would expect, the performance of the jury acting under the first-ballot majority-rule procedure depends on the median observation $X_{med,n}$ and how the juror who makes that observation decides on his/her vote. Let $Y_{med,n}$ be the median of the juror errors of observation Y_1, Y_2, ..., Y_n. Then, from III(3), we have:

(13)
$$
X_{med,n} = \frac{Y_{med,n} + aZ}{(1 + a^2)^{1/2}} .
$$

As $n \to \infty$, the variance of $Y_{med,n}$ approaches zero and since $Y_i \sim N(0,1)$ for all i, $Y_{med,n}$ itself approaches zero. But then, as $n \to \infty$,

$$
X_{med,n} \to aZ/(1 + a^2)^{1/2} = W.
$$

Consequently, since, as we have seen, $\lim_{n \to \infty} \bar{X}_n = W$, tests based on the median could be asymptotically as good as tests based on the mean. That is, as the size of the jury increases, the median observation approaches W, which contains all the information in the trial, just as the mean of the observations approaches W.

Thus, a _large_ jury that decides its verdict by a first-ballot majority vote has the information to perform as well as it would if it followed the rule for optimal deliberation. But if the median voter follows the decision rule in (11) above, applying society's error tradeoff K to his/her own observation, the large jury deciding by first-ballot majority rule will _not_--except for one special case--perform as well as it would under the optimal deliberation rule in (7). Following the decision rule in (11), the large jury's median voter will vote to convict if and only if $X_{med,n} \geq \frac{k}{T} - \frac{1}{2} T$ so that the large jury applying a first-ballot majority rule is essentially applying the criterion: convict if and only if $W \geq \frac{k}{T} - \frac{1}{2} T$. The optimal decision criterion, which we have seen the optimally deliberating jury approaches as it increases in size, indicates instead [see III(13)] that a guilty verdict should be rendered if and only if $W \geq \frac{k\rho}{T} - \frac{1}{2} T$. Hence, the large jury deciding cases on the basis of first-ballot majority rule would achieve asymptotic optimality if its median voter applied the rule: convict if and only if $X_{med,n} \geq \frac{k\rho}{T} - \frac{1}{2} T$, which differs from what the rule in (11) dictates.

It is conceptually clear why the first-ballot rule generally fails to produce an asymptotically optimal result. The median voter in the large jury makes an observation in which the noise factor approaches zero. He/she should take this into account and, when formulating his/her decision rule, recognize that his/her observation has a lower variance than an observation chosen at random. If the median juror does not take this into account, his/her vote will lead the jury to a nonoptimal decision. [6/]

This can be made more precise. Suppose that a decisionmaker who is to declare the defendant guilty or innocent observes V where

$$(14) \quad \begin{cases} V \sim N(0,\sigma^2) & \text{if guilty} \\ \\ V \sim N(-T,\sigma^2) & \text{if innocent.} \end{cases}$$

Suppose, too, that the decisionmaker's tradeoff between the a priori expected costs of errors is K, with $k = \log K$. Then the decisionmaker's optimal rule is to convict if and only if

$$(15) \quad V \geq \frac{\sigma^2 k}{T} - \frac{1}{2} T.$$

(This follows, for example, from applying the rule in (6) given the information in (14).) Thus, in the optimal decision criterion, the variance weights the term k/T, which reflects the decisionmaker's tradeoff between the two types of errors. Overestimating the variance has the same effect as using a larger absolute value of k.

Thus, suppose that a large jury decides the case on the basis of a first-ballot simple majority vote and that each juror—including the one with the median observation—follows the decision rule in (11). Then if the jury regards the expected cost of convicting the innocent as higher than that of acquitting the guilty, so that $k > 0$, it will tend to acquit too many truly guilty defendants. That is, if $k > 0$, then the asymptotic error probabilities satisfy the following inequality:
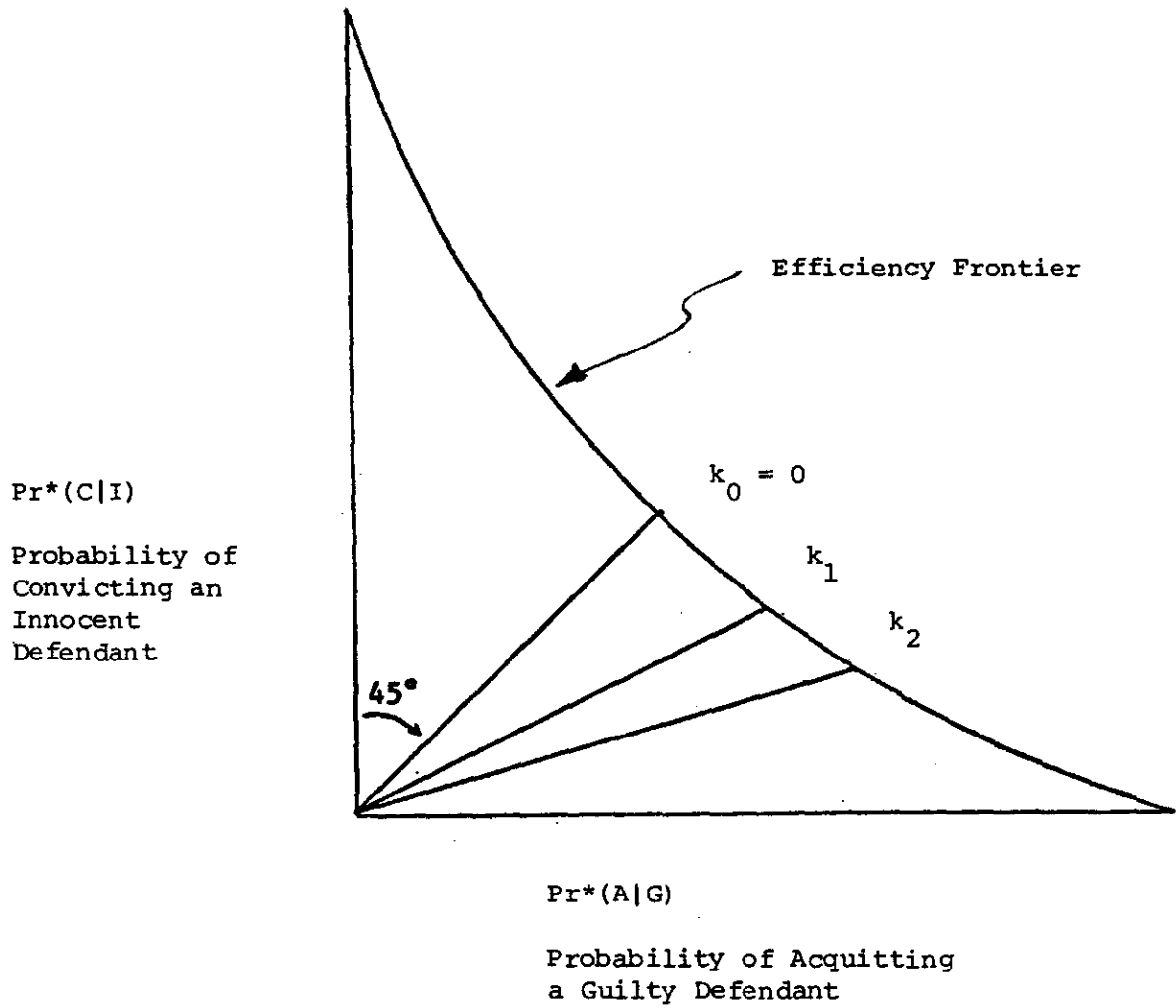
$$(15) \quad Pr_F(A|G) > Pr_D(A|G) > Pr_D(C|I) > Pr_F(C|I) ,$$

where F denotes the first-ballot simple majority rule procedure, D optimal deliberation, A acquittal, and C conviction. If $k < 0$, then the inequalities in (15) are reversed.

34

The result in (15) is illustrated in Figure 1, which depicts the efficiency frontier for the jury's decision problem. For any given level of one of the error probabilities, the corresponding point on the frontier gives the minimum attainable level of the other error probability--that is, the level that an optimally deliberating jury would achieve. The frontier can be generated by varying k in the expressions for the error probabilities given in (10) above. It can be verified that the frontier is negatively sloped and, if each of the error probabilities is less than one half, convex. In terms of the figure, the inequalities in (15) show that a large jury applying a first-ballot majority rule when, for example, $k = k_1 > 0$ will, in fact, attain a point like k_2 on the frontier where $k_2 > k_1$. The jury _will_ reach a point on the efficiency frontier but one where the probability of false acquittals is higher than desired.

How far from its target on the efficiency frontier does a jury stray when it reaches a verdict on the basis of a first-ballot, simple majority vote? Alternatively, if such a jury were to deliberate in an optimal way, by how much could it reduce the expected loss due to its misclassification of defendants? Of course, the answer depends on the values of the various parameters. But to get some sense of the disparity between the jury's desired combination of error probabilities and the combination it attains, consider a very special case of the social tradeoff between the two types of errors a jury can commit (K). Specifically, suppose that $\Gamma(G|I) = \Gamma(I|G)$, so that the cost of a false conviction is equal to the cost of a false acquittal. In this case, minimizing the expected loss due to errors of misclassification is equivalent to minimizing the (unconditional) probability of an error:

FIGURE 1

Pr*(C|I)

Probability of
Convicting an
Innocent
Defendant

Efficiency Frontier

$k_0 = 0$

k_1

k_2

45°

Pr*(A|G)

Probability of Acquitting
a Guilty Defendant

$$(1-\theta)\Pr(C|I) + \theta\Pr(A|G).$$

The social tradeoff K then equals $\frac{1-\theta}{\theta}$, or the prior probability that the defendant is innocent divided by the prior probability that he/she is guilty, so that K equals the prior odds that the defendant is innocent.

For this specification of K, Table 2 compares the unconditional error probability for infinitely large juries that decide on the basis of a first-ballot, majority vote with the unconditional error probability for similarly sized, but optimally deliberating juries.[1/] The comparison is made for different values of K and ρ for two cases, one in which the asymptotic probability of error of an optimally deliberating jury with K = 1 is .3 and the other in which that probability is .1. We see that in both cases, when K = 1 (so that k = ln K = 0), the asymptotic error probability of the first-ballot, majority rule jury is the same as that of the optimally deliberating jury. As we shall demonstrate in a moment, this result is completely general. If K = 1, a large jury will do equally well applying a first-ballot majority rule or engaging in optimal deliberation. For K = 2 and K = 4, however, the difference in performance that results from optimal deliberation is considerable. As one would expect, the differences are greater for the smaller values of ρ because the smaller ρ is, the larger is the correction that the median juror should make in his/her decision rule. In the worst instance--Case ii with K = 2, ρ = .1--the absolute difference between the asymptotic error probabilities is .22 and the first-ballot majority rule jury is nearly 3 1/2 times more likely to err than is its optimally deliberating counterpart. Thus, if a jury chooses to decide on a verdict on the basis of the outcome of a first

TABLE 2

ASYMPTOTIC PROBABILITY OF ERROR

K		Case 1			Case ii		
		1	2	4	1	2	4
$\rho = .1$	D	0.3	0.2620	0.1645	0.1	0.0890	0.0582
	F	0.3	0.3333	0.2000	0.1	0.3100	0.2000
	A	0.0	0.0713	0.0355	0.0	0.2211	0.1418
$\rho = .5$	D	0.3	0.2620	0.1645	0.1	0.0890	0.0582
	F	0.3	0.2841	0.1972	0.1	0.1022	0.0939
	A	0.0	0.0221	0.0328	0.0	0.0133	0.0357
$\rho = .8$	D	0.3	0.2620	0.1645	0.1	0.0890	0.0582
	F	0.3	0.2652	0.1859	0.1	0.0954	0.0767
	A	0.0	0.0032	0.0214	0.0	0.0064	0.0185

D = asymptotic probability of error for optimally deliberating jury

F = asymptotic probability of error for first-ballot majority rule jury

A = absolute difference between the asymptotic error probabilities of the two kinds of juries

ballot majority vote, with each juror following the decision rule in (11) above, rather than to engage in optimal deliberation, then the jury can have a considerably higher chance of making an error.

The preceding discussion points us to the one special case in which the large jury performs as well under first-ballot simple majority rule as it does under the optimal deliberation rule in (7). It is the special case we considered earlier in which the jury attaches equal weight to the two kinds of errors it can make so that: $K = 1$ and $k = 0$. In this case, the median voter applying the rule in (11) votes to convict if and only if $X_{med,n} \geq -\frac{1}{2} T$. We have observed that as $n \rightarrow \infty$, $X_{med,n} \rightarrow W$. Hence, for large juries, the median voter's criterion and, consequently, the jury's decision rule, is equivalent to: convict if and only if $W \geq -\frac{1}{2} T$. As can be seen from III(13), this is the optimal decision rule for the case in which $k = 0$. Hence, when $k = 0$, the decision criteria of the first-ballot simple majority jury and the optimally deliberating jury converge asymptotically to the same rule. Consequently, for $k = 0$, the asymptotic error probabilities of the two kinds of juries are the same, and they are given in (10'). In terms of Figure 1, this means that when $k = 0$, the large jury that decides the case by a simple majority on the first ballot attains the frontier point, k_o, for which it aims.

For this special case in which the jury attaches equal weight to the two kinds of errors, there is a particularly simple relationship between the size of a large jury that decides by first-ballot majority rule and the size of the optimally deliberating jury that achieves the same degree of accuracy. Namely, if the large jury that deliberates

optimally--that is, follows the rule in (7')--is of size n, then a

first-ballot, majority rule jury will have to be of size $\frac{\pi}{2}n$, or approxi-

mately 57 percent larger, if it is to achieve the same error probabilities,

that is, the ones given in (8'). The reason for this is as follows. For

large n, the median of a set of independent, identically distributed normal

variables with mean (and median) μ and variance σ^2 will be approximately

normally distributed with mean μ and variance $\frac{\pi\sigma^2}{2n}$. $\underline{8/}$ Applying this

result to approximate the distribution of $Y_{med,n}$, we can use

(13) to calculate the variance of $X_{med,n}$, which is $\frac{\pi}{2}\left(\frac{1-\rho}{n}\right) + \rho$. But

then the error probabilities of the large jury that decides the case by

first-ballot majority rule can be computed using the expressions in (12).

One finds that for large n, with k = 0, the probability of error for a jury

that uses the first ballot majority rule procedure is approximately equal

to:

$$(16) \qquad \Phi\left[-T \bigg/ 2\sqrt{\frac{\pi}{2}\left(\frac{1-\rho}{n}\right) + \rho}\,\right].$$

Equating the probability in (8') to that in (16) and solving for the n in

(16) as a function of the jury size in (8'), it is straightforward to show

that if the optimally deliberating jury is of size n, the first ballot jury

must be of size $\frac{\pi}{2}n$ if the two are to be of equal efficiency.

Before leaving the special case in which k = 0, let us consider

some evidence about how quickly the error probabilities of the optimally

deliberating and first-ballot majority-rule juries converge and how large

the discrepancy between the errors is for small n. Table 3 contains a

comparison of the error probabilities of the two kinds of juries as jury

size varies. The illustrative comparisons are made for two values of the

40

TABLE 3

PROBABILITY OF ERROR
COMPARISON OF FIRST BALLOT AND OPTIMALLY DELIBERATING JURIES
WHEN k = 0

n_E		1	5	9	11	15	25	35	∞
				$Pr^* = .3$					
$\rho = .1$	D	0.4341	0.3770	0.3554	0.3487	0.3392	0.3265	0.3200	0.3
	F	0.4341	0.3906	0.3698	0.3628	0.3523	0.3371	0.3287	0.3
	A	0.0000	0.0136	0.0144	0.0141	0.0131	0.0106	0.0087	0.0
$\rho = .5$	D	0.3554	0.3161	0.3094	0.3078	0.3058	0.3036	0.3026	0.3
	F	0.3554	0.3219	0.3136	0.3114	0.3087	0.3054	0.3039	0.3
	A	0.0000	0.0058	0.0042	0.0036	0.0029	0.0018	0.0013	0.0
$\rho = .8$	D	0.3195	0.3044	0.3025	0.3020	0.3015	0.3009	0.3006	0.3
	F	0.3195	0.3062	0.3037	0.3031	0.3023	0.3014	0.3010	0.3
	A	0.0000	0.0018	0.0012	0.0011	0.0008	0.0005	0.0004	0.0
				$Pr^* = .1$					
$\rho = .1$	D	0.3426	0.2219	0.1824	0.1709	0.1555	0.1359	0.1265	0.1
	F	0.3426	0.2488	0.2084	0.1955	0.1771	0.1521	0.1393	0.1
	A	0.0000	0.0269	0.0260	0.0246	0.0216	0.0162	0.0128	0.0
$\rho = .5$	D	0.1824	0.1210	0.1120	0.1099	0.1073	0.1044	0.1032	0.1
	F	0.1824	0.1293	0.1177	0.1147	0.1110	0.1068	0.1049	0.1
	A	0.0000	0.0083	0.0057	0.0048	0.0037	0.0024	0.0017	0.0
$\rho = .8$	D	0.1258	0.1055	0.1031	0.1025	0.1019	0.1011	0.1008	0.1
	F	0.1258	0.1079	0.1046	0.1038	0.1028	0.1017	0.1012	0.1
	A	0.0000	0.0024	0.0015	0.0013	0.0009	0.0008	0.0004	0.0

D = probability of error for optimally deliberating jury

F = probability of error for first-ballot majority rule jury

A = absolute difference between the error probabilities of the two kinds of juries

41

asymptotic error probabilities--.3 and .1--and for a range of values of ρ, the correlation between the jurors' observations. (In the table, we do not distinguish between the error probabilities for false convictions and those for false acquittals since the jury acts to equate those probabilities when $k = 0$.)

Several features of the comparison are noteworthy. First, the differences between the error probabilities of the two kinds of juries are quite small, even for small n. The largest difference in the table is .0269, and it occurs for $\rho = .1$, Pr* = .1, n = 5. Thus, when $k = 0$ and the jury's asymptotic error probability is .1 or .3, the first-ballot majority-rule jury does not seem to be at a great disadvantage relative to one that engages in optimal deliberation even when n is small. Second, the rate of convergence between the two juries' error probabilities is moderate. Third, convergence is not monotonic. For example, for $\rho = .1$, Pr* = .3, the largest absolute difference between the two error probabilities is for n = 9. One's intuition might suggest that the smaller the jury, the greater the disadvantage of the first-ballot majority-rule procedure. The nonmonotonicity of convergence shows that intuition is not correct, and another moment's thought suggests the flaw in that intuition. We know that the error probabilities under the two decision methods must be the same for a single-person jury (n=1) since one person does not engage in deliberation. And we have also shown that the error probabilities are equal for $n \to \infty$. Consequently, the convergence between the two error probabilities cannot be monotonic.

Finally, if a large jury attaches different expected costs to the two kinds of errors --so that $K \neq 1$ and $k \neq 0$--but it still wants to

42

decide on the basis of a predeliberation vote, there are two ways it can make an optimal decision. First, individual jurors can take account of the correlation between their observations and alter the cutoff in the individual decision rule in (11) to $\frac{k\rho}{T} - \frac{1}{2}$ T. Then a decision based on a __simple__ majority vote would converge to the optimal decision as $n \to \infty$. Alternatively, the individual jurors can be instructed to vote as if the two kinds of errors do have equal weight--that is, as if $k = 0$--but the jury as a whole would agree to convict the defendant only if a fraction

$$(17) \qquad \beta = \phi \left[\frac{a^2}{(1+a^2)^{1/2}} \; \frac{k}{T} \right]$$

voted to convict. It is straightforward to show that as k varies, a first-ballot jury following this rule asymptotically traces out the same efficient tradeoffs between the two kinds of error probabilities as the optimally deliberating jury does.

V. The Effects of Differences in Jurors' Standards

To this point, we have implicitly assumed that the social tradeoff between the two types of errors that a jury can commit, K, is universally shared or at least that each juror puts aside personal views and applies the social tradeoff in casting his/her vote. Suppose, on the contrary, that different jurors have different views about the appropriate tradeoff between the expected costs of erroneous convictions and erroneous acquittals, and suppose that they do not simply adopt the social tradeoff when they become jurors. In particular, assume that individual standards $k_i = \log K_i$ are normally distributed among the population with mean k^* and variance v^2. We will also assume that the distribution of standards is independent of the distribution of observations made by jurors. What implications does the existence of this distribution have for the comparative performance of first-ballot, majority-rule juries and optimally deliberating juries?

To focus attention on the new issues introduced by the dispersion of standards in the population, let us assume that a jury adopting the first-ballot majority-rule procedure does so in a "sophisticated" way. Specifically, assume that each individual juror takes account of the correlation between his/her observation and the others' and alters the cutoff in his/her decision rule to reflect that correlation. Hence, if juror i applies the standard k_i, that juror's decision rule is:

(1)
$$
\begin{cases}
\text{Convict if } X_i \geq \dfrac{k_i \rho}{T} - \dfrac{1}{2} T \\[2ex]
\text{Acquit if } X_i < \dfrac{k_i \rho}{T} - \dfrac{1}{2} T .
\end{cases}
$$

44

As we observed at the end of Section IV, a decision based on a simple majority vote of jurors who apply (1) to determine their individual positions will converge to the optimal decision as $n \to \infty$ _if_ each juror uses the social standard k for k_i.

Consider the decision process of a first-ballot, majority-rule jury in which juror i votes according to the decision rule in (1) with k_i representing that juror's _personal_ view of the tradeoff between false convictions and false acquittals. Since the jury is deciding its verdict on the basis of a first-ballot, simple-majority vote, the outcome for the defendant depends on which inequality in (1) holds for more than $\frac{n}{2}$ jurors, that is, whether the median value of $X_i - \frac{k_i \rho}{T}$ is greater than, equal to, or less than $-\frac{1}{2} T$.

Define

$$R_i = \frac{Y_i}{(1+a^2)^{1/2}} - \frac{k_i \rho}{T}$$

so that, from the definitions of X_i and W in Section III, $X_i - \frac{k_i \rho}{T} = R_i + W$. Then the jury's decision rule can be expressed as:

$$(2) \quad \begin{cases} \text{Convict if } W + R_{med,n} \geqq -\frac{1}{2} T \\ \text{Acquit if } W + R_{med,n} < -\frac{1}{2} T \, , \end{cases}$$

where $R_{med,n}$ is the median value of $\dfrac{Y_i}{(1+a^2)^{1/2}} - \dfrac{k_i \rho}{T}$ in a jury of size n. From the definition of R_i, we can determine its distribution:

$$(3) \quad R_i \sim N\left(-\frac{k^*}{T} \, , \, \frac{1}{1+a^2} + \frac{v^2 2}{T^2} \right).$$

45

It follows from the distribution of R_i in (3) that as $n \to \infty$, $R_{med,n}$ approaches $- \frac{k^* \rho}{T}$ and its variance goes to zero. Thus, asymptotically the defendant is convicted by a jury following the first-ballot, majority-rule procedure if and only if $W - \frac{k^* \rho}{T} \geq - \frac{1}{2} T$. This large jury follows the decision rule:

(4)   Convict if and only if $W \geq \frac{k^* \rho}{T} - \frac{1}{2} T$.

Comparing (4) with the optimal decision rule in III(13), we see that the large first-ballot, majority-rule jury in which jurors correct for the correlation of their observations but apply their personal views of the error tradeoff will reach a point on the efficiency frontier. Whether it reaches the socially optimal point depends on the relationship between the social standard k and the mean k^* of the distribution of the k_i's. If $k = k^*$, then as $n \to \infty$, the sophisticated first-ballot, majority-rule jury will approach the socially optimal decision rule. Otherwise, it reaches a point on the efficiency frontier, but not the socially optimal point.

It should be emphasized that in this section we have compared the verdicts of a "sophisticated" first-ballot, majority-rule jury with the results of following a decision rule that makes use of all the information in the trial and applies the society's tradeoff between the two kinds of erroneous verdicts. We have not identified the latter with the results of a particular way of processing the contributions of n jurors, an identification we were able to make in the previous section. There the optimal way to proceed was to apply discriminant analysis to the jurors' observations. The problem with making any such identification in the

current setting is that there is no way to specify how the individual jurors' k_i's _ought_ to be aggregated to arrive at the social standard k.

It is also not possible to compare the performance of the first-ballot, majority-rule jury with that of a jury that first combines its members' standards and then applies discriminant analysis using that composite standard, unless one specifies how that aggregation occurs. Consider a specific aggregation method: namely, suppose the jury simply averages the n jurors' k_i-values, and denote that average $\bar{k}$. If the jury then optimally processes its trial observations, using discriminant analysis, we know from the analysis in Section IV--specifically from IV(7)--that the jury will

(5)   Convict if and only if $\bar{X}_n \geq \dfrac{\bar{k}}{T} \left[\dfrac{(n-1)\rho + 1}{n} \right] - \dfrac{1}{2} T$.

As such a jury grows large, $\bar{k}$ will approach k*, and it follows from the argument in the previous section that the criterion in (5) will converge to that in (4) above. In short, given the specification in this section, a first-ballot, majority-rule jury that is "sophisticated" in correcting for correlated observations will attain the same error probabilities as a large jury that first averages its members' standards and then applies discriminant analysis to their observations with that averaged standard.

One last observation can be made about the case in which $k = k*$, that is, in which society's standard is equal to the mean of the distribution of individual standards. A possible interpretation of this situation is that all members of the population actually share the same standard k, but when a particular individual, say the ith, applies that

47

standard, some "noise" enters and he/she actually uses the individual

standard k_i. An interesting question is whether a jury that decides by

"sophisticated" first-ballot, majority rule should agree on a common

standard, call it k^0, before casting individual votes or whether each juror

should vote based only on his/her individual standard k_i. For the case in

which $k = k^*$, intuition suggests that if the aggregation procedure is

unbiased, so that $E(k^0) = k$, then the jury should agree on a single

standard before voting if the variance of the agreement procedure is less

than the variance of the individual k_i's, namely, less than v^2. It can

be demonstrated that this is, in fact, a sufficient condition for

preballoting aggregation of standards to be desirable, though we omit the

details here. An example of an aggregation procedure that satisfies this

sufficient condition is averaging the jurors' k_i-values. One that fails

to meet the condition is adopting the view of the most extreme juror.

VI. The Effects of Differences in Jurors' Abilities to Decide

In the previous analysis, we have assumed that all jurors were equally able to follow the testimony and arguments they saw and heard and the instructions they received at the trial. Of course, though, jurors are likely to differ in their abilities to process the information produced by a trial. In this section, we examine the effect of one kind of difference in jurors' abilities: differences in the variances of the jurors' measures of the defendant's guilt or innocence. To focus attention on these differences among jurors in the context of a tractable model, we simplify our earlier formulation by assuming that the jurors' observations are uncorrelated.

Specifically, we suppose that juror i observes X_i where

$$(1) \begin{cases} \text{(a)} \quad X_i \sim N(0,\sigma_i^2) \text{ if the defendant is guilty} \\ \\ \text{(b)} \quad X_i \sim N(-1,\sigma_i^2) \text{ if the defendant is innocent ,} \end{cases}$$

and the X_i's are uncorrelated. The crucial difference between this specification and our previous one is that σ^2, the variance of the juror's measure of the defendant's guilt, can now vary across jurors. We will assume that the precision of a juror's observation, $(\sigma_i^2)^{-1}$, is the same no matter whether the defendant is innocent or guilty, and we will assume that this precision is distributed according to a gamma distribution with parameters $m/2$ and $(2\gamma)^{-1}$. The mean and variance of the precision of the juror's observation are:

$$(2) \qquad E\left[\frac{1}{\sigma_i^2}\right] = \gamma m$$

49

and

(3) $\text{Var}\left[\dfrac{1}{\sigma_i^2}\right] = 2\gamma^2 m.$

The analysis that follows is facilitated by the observation that if H satisfies

(4) $\dfrac{1}{\sigma_i^2} = \gamma H,$

then H is distributed as chi-square with m degrees of freedom.

Suppose the ith juror considers the expected costs of the two kinds of errors he/she can make to be equal so that $K = 1$ and $k = 0$ for this juror. Consequently, he/she adopts a decision rule that equates the probabilities of making the two kinds of errors. Then, given the specification in (1), this juror will vote to convict if his/her observation exceeds $-\dfrac{1}{2}$ and to acquit if that observation is less than $-\dfrac{1}{2}$. The probability that this juror votes incorrectly when the defendant is guilty is

(5) $p = \text{Pr}\left[X_i < -\dfrac{1}{2}\;\middle|\; X_i \sim N(0,\sigma_i^2)\right].$

Since $k = 0$ for this juror, the probability in (5) is also the probability that he/she errs when the defendant is innocent.

This error probability is easily calculated. To evaluate it, we must remember that σ_i^2 is a random variable with density $g(\sigma_i^2)$, calculate the probability of an incorrect decision conditional on a given value of σ_i^2, and then integrate over σ_i^2. Thus, the probability that the ith juror reaches an incorrect decision is:

$$\int \left[\text{Pr} \left[X_i < -\frac{1}{2} \right] \mid \sigma_i^2 \right] g(\sigma_i^2) d\sigma_i^2 = \int \left[\text{Pr} \left[\frac{X_i}{\sigma_i} < -\frac{1}{2\sigma_i} \right] \mid \sigma_i^2 \right] g(\sigma_i^2) d\sigma_i^2$$

$$= \text{Pr} \left[\frac{X_i}{\sigma_i} < -\frac{1}{2\sigma_i} \right],$$

where $X_i \sim N(0, \sigma_i^2)$. Letting S be a standard normal variate and using (4), we see that this probability is:

$$(6) \qquad p = \text{Pr} \left[S < -\frac{1}{2} \sqrt{\gamma H} \right] = \text{Pr} \left[\frac{S}{\sqrt{H}/\sqrt{m}} < -\frac{1}{2} \sqrt{\gamma m} \right] = \text{Pr} \left[t_m < -\frac{1}{2} \sqrt{\gamma m} \right]$$

where t_m is distributed as a t distribution with m degrees of freedom.

From (6) it is clear that the probability that an incorrect decision is made by a juror who attaches equal weight to the expected costs of the two kinds of errors is a function of the two parameters γ and m. In our analysis of the effects of differences in jurors' abilities, we will want this error probability to remain constant. This is achieved by defining γ as a function of m, which we will denote u(m). The properties of u(m) are important because they establish m as a measure of the dispersion of jurors' abilities. Specifically, with the error probability p in (6) held constant, the parameter m measures, inversely, the variability of individual jurors' abilities to process trial information.[9] If m = 1, jurors' abilities are quite variable;[10] while as m $\to \infty$, the variability in ability of individual jurors vanishes, and the model converges to the model of Section IV where jurors did not vary in ability. (To see this rigorously, note that $t_m \to S$ as m $\to \infty$.)

The fact that jurors vary in ability has no effect on the probability of error when the jury decides on the basis of first ballot votes. The errors made by the jury are simply determined by the probability that the individual juror errs, which is given by (6) above as:

$$p(\gamma,m) = Pr\left[t_m < -\frac{1}{2}\sqrt{\gamma m} \right].$$

The probability that an n member jury errs in its decision is then:

$$(7) \qquad \sum_{r>n/2} \binom{n}{r} p(\gamma,m)^r (1 - p(\gamma,m))^{n-r}.$$

Moving on to the deliberating jury, consider an n member jury that makes optimal use of its observations of the n random variables $X = (X_1,\ldots,X_n)$, which are independently distributed with $X_i \sim N(0,\sigma_i^2)$ if the defendant is guilty and $X_i \sim N(-1,\sigma_i^2)$ if he/she is innocent. The jury will form the statistic U_n defined in IV(3) and vote for conviction if and only if $U_n \geq k$. Here we will consider only the case where the jury treats the expected costs of the two kinds of errors as equal so that k=0 and its decision rule is to convict if and only if $U_n \geq 0$. Given the structure of the jurors' observations in the present model, this decision rule is equivalent to finding the defendant guilty if and only if

$$(8) \qquad \bar{X}_w \geq -\frac{1}{2}$$

where $\bar{X}_w$ is the weighted mean of the X_i's:

$$(9) \qquad \bar{X}_w = \sum_i \frac{X_i}{\sigma_i^2} \Big/ \sum_i \frac{1}{\sigma_i^2}.$$

Now $\bar{X}_W$ is distributed as:

(10)
$$\begin{cases} \bar{X}_W \sim N(0, 1/\alpha) & \text{if guilty} \\ \\ \bar{X}_W \sim N(-1, 1/\alpha) & \text{if innocent} \end{cases}$$

where $\alpha = \underset{i}{\Sigma}(1/\sigma_i^2)$. For fixed $(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$, the probability that the jury errs is:

(11) $\quad \Phi\left[-\frac{1}{2}\sqrt{\alpha}\right].$

Since $\alpha = \underset{i}{\Sigma}(1/\sigma_i^2)$, the assumptions about the distribution of σ_i^2 imply $\alpha \sim \Gamma[nm/2, (2\gamma)^{-1}]$. If we let $\alpha = \gamma L$, then $L \sim \chi_{nm}^2$. We can then calculate the optimally deliberating jury's error probability by integrating over α. Letting $y(\alpha)$ denote the density of α and using S as the standard normal variate, we have:

$$\text{Error probability} = \int \left\{ \Phi\left[-\frac{1}{2}\sqrt{\alpha}\right] \Big| \alpha \right\} y(\alpha)d\alpha = \Pr\left[S < -\frac{1}{2}\sqrt{\gamma L}\right]$$

(12) $$= \Pr\left[S/(\sqrt{L}/\sqrt{\gamma nm}) < -\frac{1}{2}\sqrt{\gamma nm}\right] = \Pr\left[t_{nm} < -\frac{1}{2}\sqrt{\gamma nm}\right]$$

where t_{nm} is distributed as a t distribution with nm degrees of freedom.

How should we expect the error probability of the optimally deliberating jury to change with m? Intuitively, one would think that the more variable the abilities of the jurors, the lower the probability of error. The optimal deliberation process pays the most attention to those individuals with high abilities (small σ_i^2), and it discounts the views of

those with low abilities (large σ_1^2). The optimal jury will always do at least as well as the most able juror and usually considerably better. Thus, as the variance in ability increases, the jury would be more likely to get a very high ability juror and that would ensure a low probability of error.

Although we do not have an analytic proof that this intuition is correct, we do have some numerical results that support the following conjecture for the jury whose members' observations were modelled in this section: If the probability of a single juror being in error is held constant, then the greater the variance in ability (the smaller is m), the lower is the probability that the optimally deliberating jury will err. The numerical results appear in Table 4. That table presents the probability of error for optimally deliberating juries with different variances in ability and the probability of error for the first ballot jury. As we noted before, differences in ability do not affect the first ballot jury's performance. But, as Table 4 does show, differences in ability do affect the probability of error of the optimally deliberating jury.

The upshot is that a jury basing its decision on a first ballot majority vote will be at a greater disadvantage compared with one that uses information optimally the greater is the variance in abilities of the jurors.

TABLE 4

PROBABILITIES OF ERROR WHEN THERE ARE DIFFERENCES IN ABILITY

IE	1	5	9	11	15	25	35
m=1	.40000	.25004	.17757	.15213	.11374	.05840	.03137
m=2	.40000	.26657	.19894	.17438	.13621	.07758	.04605
m=4	.40000	.27588	.21102	.18707	.14931	.08946	.05575
m=10	.40000	.28166	.21856	.19503	.15761	.09724	.06232
m= ∞	.40000	.28553	.22361	.20038	.16324	.10262	.06696
First Ballot	.40000	.31744	.26657	.24650	.21310	.15377	.11431

m=1   high variance

m=∞   no variance

55

VII. Concluding Remarks

In this paper, we have focused attention on the processing of information by juries. This is an aspect of the jury system that is central to almost all policy discussions but that has been neglected in the development of models that are intended to inform such discussions. Indeed, as we have shown, some of the more sophisticated models in the literature imply that the deliberative process can only yield decisions that are inferior to those produced by majority votes taken before any discussion among the jurors. The models we have presented here enable us to evaluate the role that can be played by jurors' sharing of their observations. Deliberation is important precisely because it enables differences in the strengths of jurors' views, differences in their decision standards, and differences in their abilities all to be taken into account in reaching a verdict. These are factors that a first-ballot, majority-rule decision cannot weigh and that previous models of the jury decision process have not encompassed.

There are, of course, other aspects of jury deliberation and its analogues that we have not captured in our model--for example, the possibility that different jurors see different kinds of information and the possibility (surely a reality) that jurors attach varying weights to the importance of different peers' views. In closing, we should reiterate a point we made very early in the paper: Ours is not a descriptive analysis of the value of actual jury deliberations; it is, instead, an effort to illustrate what deliberation can, in theory, achieve.

APPENDIX A

To see how the Kalven and Zeisel data can be used to estimate the parameters in the first stage of the Gelfand and Solomon model, let us introduce the following notation, where the concepts are those used by Gelfand and Solomon:

$f(k,N;G)$ = probability that on the first ballot k of N jurors vote to convict a guilty defendant;

$f(k,N;I)$ = probability that on the first ballot k of N jurors vote to convict an innocent defendant;

μ = probability that on the first ballot an individual juror will vote for the correct verdict;

θ = probability before the trial that the accused is guilty;

$h(k)$ = fraction of juries that register k votes for conviction on the first ballot.

The parameters in the first stage of the Gelfand-Solomon model are θ and μ. But note that

$$h(k) = \theta f(k,N;G) + (1-\theta)f(k,N;I).$$

If jurors vote independently on the first ballot, as Gelfand and Solomon assume they do, then

$$f(k,N;G) = \binom{N}{k} \mu^k (1-\mu)^{N-k}$$

and

$$f(k,N;I) = \binom{N}{k} (1-\mu)^k \mu^{N-k} .$$

57

Hence,

$$h(k) = \theta \binom{N}{k} \mu^k (1-\mu)^{N-k} + (1-\theta) \binom{N}{k} (1-\mu)^k \mu^{N-k}.$$

Assuming that θ and μ are both greater than .5, data on $h(k)$ can be used to obtain unique estimates of θ and μ. And, it is exactly observations on $h(k)$ that the Kalven-Zeisel data provide, although in a grouped form with k taking on only five possible first-ballot values instead of the full thirteen that it can actually assume in a 12-person jury. Using the Kalven-Zeisel data, Gelfand and Solomon estimate $\theta \approx .7$, $\mu \approx .9$.

Gelfand and Solomon also consider an alternative model in which they distinguish between $\mu_1 =$ the probability that on the first ballot an individual juror will vote to convict a guilty defendant and $\mu_2 =$ the probability that on the first ballot an individual juror will vote to acquit an innocent defendant. But their estimates lead them to believe that there is little difference between μ_1 and μ_2. Hence, they base their analysis on the two-parameter μ, θ model rather than on the three-parameter μ_1, μ_2, θ model.

58

FOOTNOTES

1. When Kalven and Zeisel compared the first-ballot votes of these 225 juries with their final verdicts, they found that "in the instances where there is an initial majority either for conviction or for acquittal, the jury in roughly nine out of ten cases decides in the direction of the initial majority. Only with extreme infrequency does the minority succeed in persuading the majority to change its mind during the deliberation... [W]ith very few exceptions the first ballot decides the outcome of the verdict... The upshot is a radical hunch about the function of the deliberation process. Perhaps it does not so much decide the case as bring about the consensus, the outcome of which has been made highly likely by the distribution of first ballot votes." Kalven and Zeisel (1966), pp. 488-489.

2. Gelfand and Solomon modify the Davis scheme for two reasons. First, they want the resulting social decison scheme to fit reasonably well the transitions from first-ballot vote to final verdict that were actually found by Kalven and Zeisel in the 225 cases mentioned earlier. Second, when the resulting social decision scheme is taken together with their estimated values of $\theta \approx .7$ and $\mu \approx .9$ (see Appendix A), Gelfand and Solomon want it to yield an overall frequency distribution of convictions, acquittals, and hung juries that fits the empirical distribution found by Kalven and Zeisel in their overall study.

3. Although Gelfand and Solomon state their results in terms of the probability of convicting an innocent person and the probability of acquitting a guilty one, these are not the probabilities that they calculate and present in their paper. Instead, to measure jury errors, they compute, and their tables display, the conditional probability that a defendant is innocent given that he/she is convicted and the conditional probability that a defendant is guilty given that he/she is acquitted. Both sets of conditional probabilities--those which condition on the true state of nature (guilt or innocence) and those which condition on the jury's verdict--are of interest in a discussion of the accuracy of the jury decision process. But the error probabilities that condition on the true state of nature are the relevant ones for a decision-theoretic analysis of jury structure and jury decisions rules. It is this set of error probabilities that we display in Table 1, and we indicate that they are calculated from inrormation in the Gelfand and Solomon article because the latter did not directly provide these numbers. The qualitative comparisons that Gelfand and Solomon draw and the ones that we draw for their results are invariant with respect to which set of error probabilities one uses.

4. This implication of the Gelfand and Solomon results was observed independently by Grofman (1979, pp. 24-25; July 1980, pp. 298-299), though his discussion is based solely on the error probabilities that Gelfand and Solomon present. He uses the implication as part of his "case for majority verdicts," but he does not suggest any explanation of the observed phenomenon. In addition, he does not provide a formal argument about conditions, beyond those in the Gelfand and Solomon study, in which one could expect to observe predeliberation, first-ballot majority verdicts that are superior to full deliberation, unanimous verdicts.

Penrod and Hastie (1979, pp. 480-483) observe a similar phenomenon in their illustrative calculations comparing the performance of twelve-person juries operating under unanimous and nonunanimous (two-thirds majority rule) decision rules. In these calculations, they use an initial vote distribution that produces a good fit to the Chicago Jury Project data on the 225 criminal cases mentioned earlier, and they use a pair of social decision schemes that are roughly comparable to the ones Gelfand and Solomon use. Penrod and Hastie's calculations yield error rates for nonunanimous juries that are lower than those in the corresponding unanimous juries. They interpret this outcome by examining what happens when the jury's predeliberation vote would exactly meet the two-thirds majority requirement of the nonunanimous jury, though of course their interpretation applies as well when the predeliberation vote exceeds the nonunanimous decision criterion. Hence, the situations for which they interpret their result are exactly analogous to those in which (a) there is a majority on the first ballot and (b) one is comparing the simple majority decision rule with the unanimity requirement. The explanation that Penrod and Hastie offer is that the decision schemes they use "do not assume that deliberation serves to correct the errors made in first ballot votes" but rather "these decision schemes reflect relatively unfavorable views of the deliberation process insofar as they assume that both correct and errorful initial majorities are equally likely to be reversed during deliberation." (p. 481).

Let us look at the results in Table 1 in yet another way. Suppose we accept the error probabilities based on Gelfand and Solomon's refined estimates, those in columns (2) and (4), as the basis for choosing between six-person and twelve-person juries when both kinds of juries are subject to a unanimity rule. Then it appears that allowing deliberation subject to a unanimity requirement introduces much larger errors than does the change from twelve-person to six-person juries. Using their estimates, Gelfand and Solomon would seem to have made a stronger case for having juries decide by a majority vote without any deliberation than they have made for having twelve-person rather than six-person juries.

5. Klevorick and Rothschild (1979).

6. There is an analogy in the competitive bidding literature, more specifically, in the phenomenon referred to as "the winner's curse." Potential bidders with private information should recognize that they will win only if their observations are extreme. Hence, when evaluating the information in their observations, they should condition on the observations being extreme.

7. An alternative interpretation of the entries in Table 2 is that $\theta = 1 - \theta = \frac{1}{2}$ so that $K = \frac{\Gamma(G|I)}{\Gamma(I|G)}$, the relative social costs of the two kinds of misclassification. In this case, the entries in Table 2 equal the expected losses due to misclassification multiplied by $\frac{2}{\Gamma_I + \Gamma_G}$.

8. Kendall and Stuart (1958), p. 329.

9. To establish this relationship between m and $\text{Var}\left[\frac{1}{\sigma_1^2}\right]$, let $t_{w,p}$ be defined by $\Pr\{t_w < t_{w,p}\} = p$; that is, $t_{w,p}$ is simply the critical value for tail probability p of a t distribution with w degrees of freedom. But from the properties of the t distribution, it follows that for $p < \frac{1}{2}$, as w increases, $t_{w,p}$ increases—that is, $|t_{w,p}|$ decreases. Hence, as m increases, if the probability in (6) is to remain constant, we must have $\sqrt{\gamma m}$ decrease. But then it follows that for p to stay fixed, γ must decrease more than proportionately to the increase in m so that $u'(m) < -\frac{u(m)}{m}$. Substituting $u(m)$ for γ in the expression for $\text{Var}\left[\frac{1}{\sigma_1^2}\right]$ in (3), differentiating with respect to m, and using this result about $u'(m)$, yields the desired result: as m increases, the variance of the individual jurors' abilities decreases.

10. The upper bound, $m = 1$, is imposed only by our desire to have the probability computed in (6) be a t distribution. If $1/\sigma_1^2$ is distributed as a gamma distribution with parameters ν and ξ, then the same logic as was used to derive (6) shows that the probability that a juror decides incorrectly is just $\Pr[y < \delta]$ where y is a standardized normal variate divided by the square root of an independent gamma distribution and δ is a function of ν and ξ. This is, in principle, easy to calculate. It is not, as is the t distribution, extensively tabulated.

REFERENCES

Anderson, T.W. , An Introduction to Multivariate Statistical Analysis.
John Wiley and Sons, Inc. , New York, 1958.

Davis, J.H. , "Group Decision and Social Interaction: A Theory of Social
Decision Schemes," 80 Psychological Review 97 (1973).

Davis, J.H. , R.M. Bray, and R.W. Holt, "The Empirical Study of Social
Decision Processes in Juries" in J.L. Tapp and F.J. Levine, editors,
Law. Justice. and the Individual in Society: Psychological and Legal
Perspectives. Holt, Rinehart, and Winston, Inc. , New York, 1977.

Gelfand, A.E. and H. Solomon, "An Argument in Favor of 12-Member Juries,"
in S. Nagel, editor, Modelling the Criminal Justice System. Volume 7
of Justice Systems Analysis. Sage Publications, Beverly Hills,
1977(a).

_____, "Analyzing the Decision-Making Process of the
American Jury," 70 Journal of the American Statistical Association 305
(1975).

_____, "Considerations in Building Jury Behavior
Models and in Comparing Jury Schemes: An Argument in Favor of
Twelve-Member Juries," 17 Jurimetrics Journal 292, 1977(b).

_____, "Modeling Jury Verdicts in the American Legal
System," 69 Journal of the American Statistical Association 21 (1974).

_____, "A Study of Poisson's Models for Jury
Verdicts in Criminal and Civil Trials," 68 Journal of the American
Statistical Association 271 (1973).

Grofman, B., "The Case for Majority Verdicts," TRIAL Magazine 23, December
1979.

_____, "The Slippery Slope: Jury Size and Jury Verdict Requirements
--Legal and Social Science Approaches," 2 Law and Policy Quarterly 285
(1980).

Kalven, H. , Jr. and H. Zeisel, The American Jury. Little, Brown, and Co. ,
Boston, 1966.

Kendall, M.G. and A. Stuart, The Advanced Theory of Statistics. Volume 1,
Charles Griffin and Company, Ltd. , London, 1958.

Klevorick, A.K. and M. Rothschild, "A Model of the Jury Decision Process,"
8 Journal of Legal Studies 141 (1979).

Penrod, S. and R. Hastie, "Models of Jury Decision Making: A Critical
Review," 86 Psychological Bulletin 462 (1979).