Assessing the Capriciousness of Death Penalty Charging

Robert E. Weiss Richard A. Berk
Cathrine Y. Lee

We discuss capriciousness in decisions to charge homicide defendants with capital crimes. We propose using Shannon Information to assess capriciousness in a charging system and apply Shannon Information to analyze new data from San Francisco County, California. We show that about two-thirds of the potential systemic capriciousness is removed by the explanatory variables available. The one-third remaining is dependent on inherently unstable features of charging practices that necessarily produce capriciousness.

In *Furman v. Georgia* (1972:293), Justice Stewart noted that the existing system of charging and sentencing in death penalty cases was "cruel and unusual in the same way that being struck by lightning is cruel and unusual." Justice Brennan concurred that the existing procedures were "little more than a lottery system." In a recent article building on these concerns, published in *Law & Society Review*, Berk, Weiss, and Boger (1993a) develop the concept of an *as if* lottery for the role of chance in death penalty charging decisions (see also Paternoster 1993; Berk et al. 1993b). The authors argued that the decision to charge an offender with a capital crime takes on many of the characteristics of a lottery, although legal precedent and administrative practice seek deterministic outcomes. The issues must be addressed at the system level; the question is not whether a particular charging decision is capricious but, overall, how capricious charging practices are within a particular jurisdiction. The authors' focus was on the structure of charging practices.

We here extend the work of Berk, Weiss, and Boger (hereafter "BWB"). BWB argued that the statistical distribution of predicted probabilities from a model of the charging decision could be used to characterize systemic capriciousness. To take a simple illustration of a system with little capriciousness, imagine a charg-

Address correspondence to Richard A. Berk, Department of Statistics, University of California at Los Angeles, Los Angeles, CA 90095-1554.

Law & Society Review, Volume 30, Number 3 (1996)

ing process described as grouping offenders into a set of "good guys" and a set of "bad guys." The good guys have a predicted probability of 0.02 of being charged with a capital crime. The bad guys have a predicted probability of 0.97 of being charged with a capital crime. In contrast, imagine a charging system described as making few distinctions between offenders, so that all have a predicted probability of about 0.5 of receiving a capital charge. This system would be characterized as being very capricious.

BWB present a number of far more interesting distributions, but do not offer any way to summarize numerically the capriciousness in a distribution of predicted probabilities. Here, we add some precision to assessments of capriciousness. We suggest that Shannon Information can be used to productively characterize capriciousness in a charging system. We then illustrate its applicability with recent data on 427 death penalty charging decisions from the County of San Francisco.

I. Conceptions of Capriciousness

The outcome we consider is whether the defendant is *charged* with a crime for which capital punishment may be applied. Once the charge is determined, there is no uncertainty; the outcome is known with perfect accuracy by all individuals involved. However, before the charge is determined, there is prospective uncertainty insofar as the charging decision cannot be forecast with perfect accuracy.

Following BWB, we model this charging process as assigning defendants to *as if* lotteries, where the chance of a capital charge is a probability that depends on the nature of the crime committed and the biography of the offender. For example, a defendant who executes potential witnesses after committing a robbery might be, in effect, assigned to a lottery in which the chance of a capital charge is 0.9. It is as if a coin will later be flipped for which the probability of coming up heads is 0.9. If it comes up heads, a capital homicide will be charged.

To motivate this, consider a set of 100 defendants who execute potential witnesses after committing robberies; even if all backgrounds are identical and the crimes are identical, it is unlikely in our hypothetical that all 100 will be charged with a capital offense. In fact, only about 90 of them might be expected to receive a capital charge.

In contrast, a defendant who kills a clerk of a liquor store while committing a robbery might be, in effect, assigned to a lottery in which the chance of a capital charge is 0.3. It is as if a coin will later be flipped for which the probability of coming up heads is 0.3. If it comes up heads, a capital homicide will be charged.

Now, about 30 out of every 100 defendants will be charged with a capital offense.

The *as if* lottery is a way to formalize how prosecutors and defense attorneys may think about charging before charges are officially announced (Maynard 1984). It also readily translates into standard statistical procedures.

The same logic works retrospectively if the charging outcome, which has occurred, cannot be *backcast* or *retrodicted* from other information available. That is, if a new observer, who does not know the charges, cannot determine with perfect accuracy who was charged and who was not from the information available, there is *retrospective* uncertainty.

Capriciousness, as we use the term, refers to the degree of unpredictability or randomness in the output of any social system, *even if the same "inputs" are consistently applied.* We study here the social system producing prosecutors' charging decisions, where the "inputs" are characteristics of particular homicide defendants and their crimes. If there is any capriciousness, the charges made against each defendant cannot be predicted with perfect accuracy. One key implication is that if one could rerun history—if the particular set of defendants could be sentenced again "from scratch"—the outcome would almost certainly change. Not all defendants would receive the same charge in the second time around.

Two types of capriciousness in the charging system can be identified. One type of capriciousness occurs when differently situated offenders are treated identically for no apparent reason. Consider two hypothetical offenders. One commits a homicide with no aggravators specified by statute and the other commits a homicide with several aggravators specified by statute. Moreover, the first offender has no prior record, and the second has two prior homicide convictions. It would be surprising if neither, or if both, were charged with a capital crime. Alternatively, it would be surprising if both were assigned to lotteries with the same probability of being charged with a capital crime. We call this type of capriciousness, in which effectively different offenders are treated similarly, *chance homogeneity.*

The other type of capriciousness occurs when similarly situated offenders are treated differently for no apparent reason. This is the kind of capriciousness that perhaps most immediately comes to mind. Imagine two offenders with effectively identical prior records and who have committed effectively identical crimes. A failure to charge both with the same offense would be surprising. We call this *chance heterogeneity.*

Now, consider the implication of chance homogeneity at the systemic level. The BWB model of the charging system assigns offenders to lotteries, and lotteries then determine which offenders are charged with a capital crime. If charging practices are inconsistently applied, a charging system will tend to allocate differently situated offenders to similar lotteries. In effect, meaningful differences between defendants and between crimes wash out. In the extreme, the probabilities of all lotteries will cluster closely around some single value. In assigning offenders to lotteries, it makes intuitive sense that the more lotteries there are, and the more variable the probabilities associated with these lotteries, the more discriminating the charging system is. The charging system makes many nontrivial distinctions between offenders, which means that differently situated defendants will tend to face rather different probabilities of a capital charge. This, in turn, implies that it may be relatively rare for two offenders with very different crimes and backgrounds to be charged in exactly the same manner. Consequently, there is little chance homogeneity and capriciousness is low.

Chance heterogeneity is introduced into the system by the lottery stage. If all probabilities for all lotteries are very near 0 or 1, then chance heterogeneity is small and the charging system is not capricious. Offenders assigned to the same lottery will indeed see the same charging outcome. In contrast, if many probabilities are near 0.5, then there is a set of offenders for whom chance heterogeneity is large—many offenders assigned to the same lottery with probability near 0.5 will be assigned a capital charge and many will not get a capital charge, yet the system has identified them as having similar enough backgrounds and crimes to assign them to the same or similar lotteries.

To summarize the systemic implications of chance homogeneity and chance heterogeneity, imagine a very simple charging system in which all of the offenders fall into four classes with known probabilities of a capital charge of 0.36, 0.38, 0.42, and 0.46. Note that there are only four different probabilities *and* they all have similar values. Consequently, capriciousness is high. In contrast, consider a charging system with eight known probabilities of 0.001, 0.02, 0.07, 0.15, 0.90, 0.91, 0.98, and 0.996. Since offenders are sorted into twice as many categories, since there are some rather dramatic distinctions between the probabilities of a capital charge, and since the probabilities are near 0 or 1, this charging system is much less capricious than the first.

II. Capriciousness in Capital Charging

What causes capriciousness specifically in capital charging systems? First, a substantial number of cases are judgment calls that could go either way. One implication is that unique, peripheral, and even formally irrelevant features of a case can determine the outcome. Prediction is then impossible for these cases.

Second, even when official charging guidelines exist, their application is always subject to interpretation. In California, for example, a homicide that is heinous may qualify for a capital charge. But where exactly is the line between heinous and nonheinous?

Third, charging practices necessarily evolve in response to changing circumstances. Recent "three-strikes" legislation in California, for instance, means that many more felony cases are going to trial; there is no reason to plead guilty on a third-strike offense when a life sentence automatically follows. As a result, prosecutors are increasingly hard pressed to pursue their usual mix of felony cases all the way to trial. And since capital cases will almost certainly mean a trial, prosecutors have to be more selective in which defendants they charge with capital crimes.

Finally, prosecutors are not immune to the volatile politics of capital punishment. The decision by the Los Angeles District Attorney not to charge O. J. Simpson with capital homicide is just one highly visible illustration. Our reading of the facts is that at least two statutory aggravators apply.

Many have argued that legally inadmissible variables often play a role in the charging process (Bowers, Pierce, & McDevitt 1984; Baldus, Woodworth, & Pulaski 1985, 1990; Paternoster & Kazyaka 1988; Gross & Mauro 1989; U.S. General Accounting Office 1990). This might be called unfairness in a charging system, a rather different issue. Our goal here is primarily to assess the capriciousness of charging systems.

To summarize, the capriciousness of a charging system depends on how effectively differences between offenders and their crimes translate *consistently* into the charges leveled. Differences have to be acted upon, and in a manner that does not vary from offender to offender. We now turn to a more formal representation of capriciousness.

III. Formalizing Capriciousness

We begin with some definitions and notation. Let A and B be discrete random variables taking on a finite number of outcomes A^k and B^j. There are K mutually exclusive possible events for random variable A, and J mutually exclusive B events, and $K * J$ total possible outcomes between the two random variables. Our notation and assumptions directly follow Khinchin (1957); the language surrounding the death penalty and capriciousness and the examples are ours.

Each outcome has probability π_k^A or π_j^B. In general, the random variables A and B do not need to have the same outcomes, probabilities, or possible number of outcomes. However, for death penalty charging decisions, we will consider random variables with only two outcomes; either the offender is charged with a capital crime or the offender is not charged with a capital crime. The argument *AB* denotes the joint outcomes $A^k B^j$ with corresponding probabilities π^{AB}_{kj}, which in general are not necessarily equal to $\pi^A_k * \pi^B_j$. If *A* and *B* are independent, then $\pi^{AB}_{kj} = \pi^A_k * \pi^B_j$.

A. Khinchin's Theorem

We now develop our capriciousness measure $H(A) = H(\pi^A_1, \pi^A_2, \ldots, \pi^A_K)$. The function $H(\cdot)$ will assess capriciousness of the random variable *A* or equivalently the associated set of probabilities $(\pi^A_1, \ldots, \pi^A_K)$. Suppose that given the event A^k occurs and the probabilities of the possible outcomes of random variable B_j change. We denote this situation by $H_A(B)$, the capriciousness of *B* given that the outcome of *A* has been observed. The capriciousness in *B* changes depending on the outcome of *A*. Let the function $H(\pi_1, \pi_2, \ldots, \pi_K)$ be continuous and satisfy the following three assumptions.

Assumption 1. For a given set of *K* mutually exclusive and exhaustive events with probabilities π_k so that $\Sigma^K_{k=1} \pi_k = 1$, we assume the function $H(\pi_1, \pi_2, \ldots, \pi_K)$ takes its largest value at $\pi_k = 1/K$, $k = 1, \ldots, K$, where *K* is the number of outcomes.

Assumption 2. For two discrete random variables *A* and *B* (such as two homicide defendants), $H(AB) = H(A) + H_A(B)$.

Assumption 3. The possibility of events of zero probability does not change the capriciousness function. That is, $H(\pi_1, \pi_2, \ldots, \pi_K) = H(\pi_1, \pi_2, \ldots, \pi_K, 0)$.

Assumption 1 requires that *K* equally likely events is the most capricious situation possible. Compare a fair coin toss with getting hit by a car when crossing the street. The outcome of the coin toss (random variable *A*) is much harder to predict than the outcome of crossing the street (random variable *B*), and $H(A) > H(B)$. If a head (or a tail) is predicted, that prediction will be correct about half the time. But since people are rarely struck by cars when crossing the street, a prediction of safe crossing will be correct most of the time.

Analogously, if offenders are assigned to lotteries with probability 0.5 of a capital charge, that lottery maximizes capriciousness, while a lottery with probabilities near 0 or 1 have minimal capriciousness. In general, the easier it is to predict the outcome of the random variable *A*, the less capricious *A* is.

Assumption 2 demands that the capriciousness of multiple random variables adds in a certain way. In particular, if *A* and *B* are independent, then $H(AB) = H(A) + H(B)$, and the capriciousness adds directly. For our homicide defendants, this means that the capriciousness of a situation with two independent defendants with the same probability of a capital charge (CC) is twice as

https://doi.org/10.2307/3054130 Published online by Cambridge University Press

capricious as the same situation with but a single defendant. With one defendant, there are two outcomes A^1 = CC and $A^2 \neq$ CC with probabilities π and $1 - \pi$. For two defendants, one can reformulate the problem using a single random variable with four possible outcomes; both CC; A only CC; B only CC; and neither CC with respective probabilities π^2, $\pi(1 - \pi)$, $(1 - \pi)\,\pi$ and $(1 - \pi)^2$. Assumption 2 forces this second random variable to have exactly twice the *total* capriciousness of the single-defendant random variable. Both situations have the same *average*, as opposed to *total*, capriciousness, which we might use for comparing different systems.

Assumption 3 states that including events of zero probability, which we avoid in practice, such as the prosecution calling a homicide victim to testify, do not affect the capriciousness of a system. In effect, events that are known with certainty cannot add capriciousness to the system.

Given assumptions 1, 2, and 3, we have

$$H(\pi_1, \pi_2 \ldots, \pi_K) = -\lambda \sum_{k=1}^{K} \pi_k \log \pi_k \qquad (1)$$

for some constant λ.

A proof for equation (1) can be found in Khinchin (1957). The H function is called alternately the entropy of a system, the uncertainty, the information, or the Shannon Information. The use of this function for summarizing uncertainty, unpredictability and information has a long history. See, for example, Shannon (1948) or Kullback (1958).

As a practical matter, the choice of λ does not matter. Choices for the base of the log also do not matter and are perfectly confounded with the choice of λ. For convenience, one can choose log base 2 or e or 10. We use base e and $\lambda = 1$.

B. Calculating Capital Charging Capriciousness

Now we narrow the discussion to a system of n binary and independent random variables $A_1, A_2, \ldots, A_n$, corresponding to n homicide defendants, with probabilities of a capital charge $P = \{\pi_1, \pi_2, \ldots, \pi_n\}$. For binary independent events, we define the function

$$C = C(P) = C(\pi_1, \ldots, \pi_n) = \sum_{i=1}^{n} H(\pi_i, 1 - \pi_i) \qquad (2)$$

as the capriciousness of the system of probabilities P. In general, the π_i's, the probabilities of a capital charge for each defendant, are unknown and must be estimated using data and a statistical model. Given some method of estimating the π_i's, the capriciousness $C(\pi_1, \ldots, \pi_n)$ of the system can be estimated. The next section discusses the interpretation of C.

IV. Understanding the Capriciousness Measure

First, we present two instructive inequalities that follow from the definitions (1) and (2) of capriciousness. They show that our measure of capriciousness performs in a fashion that is consistent with our earlier conceptual discussion of capriciousness.

Consider two sets of probabilities that have the same average. Capriciousness is greater for the set in which all the π_i's are the same rather than the set in which the π_i's differ. In particular, for any two probabilities π_1 and π_2, define

$\bar{\pi} = (\pi_1 + \pi_2)/2.$

Then

$$C(\pi_1, \pi_2) \leq C(\bar{\pi}, \bar{\pi}) \ . \quad (3)$$

In other words, a charging system with equal probabilities for a set of offenders is more capricious than another charging system whose average probability across defendants is the same, but whose individual probabilities are different.

To take a very simple example, consider a system in which the probability of a capital charge for each defendant is 0.20. Now consider another system in which the probability of a capital charge for half the defendants is 0.10 and 0.30 for the other half. The first system is more capricious, although both have the same average probability of a capital charge (i.e., 0.20).

Furthermore, for $0 < \alpha < 1$,

$$C(\pi_1, \pi_2) \leq C(\bar{\pi}_\alpha, \bar{\pi}_{1-\alpha}) \ , \quad (4)$$

where $\bar{\pi}_\alpha = \alpha\pi_1 + (1 - \alpha)\ \pi_2$. Inequality (4) says that if charging system 1 produces a set of probabilities that are more similar to one another than a set of probabilities produced by charging system 2, then charging system 1 will have higher capriciousness than charging system 2. We illustrated this point with a simple example earlier.

We are still left, however, with the need to interpret quantitatively our capriciousness measure. As is, the units do not have any simple meaning that would allow investigators to know how big is big, or when estimated differences in capriciousness between two charging system are large enough to be important.

We propose to anchor the capriciousness measure C at a high end and a low end. Essentially, this produces a ruler with which to measure capriciousness. A mathematical upper bound for the capriciousness C of any system P of n probabilities is C associated with a vector of n probabilities equal to 0.5. This would lead to a C_{max} of $0.693n$; where $0.693 = 0.5 * \log 0.5 + 0.5 * \log 0.5$ is the mean capriciousness. However, $0.693n$ is a misleadingly high upper bound for systems where the fraction of defendants receiving a capital charge is low; the 0.5 probability of a capital charge per se overstates maximum capriciousness. The fraction of defendants who are charged with a capital crime is certainly an important issue, but one that needs to be distinguished from whether important distinctions are being made *between* defendants based on their crime, prior record, and other factors.

We propose, therefore, that the upper bound take the observed fraction of defendants charged with a capital crime as given. In particular, we propose fixing all the predicted probabilities at the value of the overall sample proportion of defendants charged with a capital crime. Thus, no distinctions are made between defendants, and all face the same probability of a capital charge equal to what is empirically observed.

Zero is always available as a lower bound for *C;* it occurs when the probabilities are all 0 or 1, as in a deterministic system. This suggests a goal of perfection, which seems unreasonable for social systems. A slightly larger and convenient, if ad hoc, minimum amount of capriciousness can be defined when the predicted probabilities are fixed at $1/n$ and at $1 - (1/n)$. The $1/n$ and $1 - (1/n)$ are smallest and largest noncertain sample frequencies empirically possible from a sample of size n. This lower bound we propose is more demanding when more defendants are processed because n is larger. Actually, since $H(\pi, 1 - \pi) = H(1 - \pi, \pi)$, there is no need to be concerned about setting up the proportion of $1/n$'s and the $1 - (1/n)$'s so that the mean predicted probability is maintained.

To help fix these ideas consider the following example. There are 50 defendants and 5, or 10%, are charged with a capital crime. It follows that the upper bound to capriciousness is $50 * (0.1 \log 0.1 + 0.9 \log 0.9) = 16.25$. Our proposed lower bound for capriciousness is $4.9 = 50 * (0.02 \log 0.02 + 0.98 \log 0.98)$.

Suppose now that the estimated distribution of predicted probabilities has a mean of 0.10 but is highly skewed with a long right tail. Half of the 50 defendants have a predicted probability of a capital charge of 0.01, 20 have predicted probabilities of 0.1, 4 have probabilities of 0.4, and 1 of the 50 defendants has a probability of a capital charge of 0.90. The capriciousness of this distribution is 10.9.

On a ruler marked at the low end by 4.9 and at the high end by 16.25, a score of 10.9 is a little less than half of the distance from the upper bound of maximum capriciousness (given the mean) to our lower bound of minimum capriciousness. More usefully, 47% of the system's total possible capriciousness has been removed by how the charging is done. Alternatively, 53% of the total possible capriciousness remains. Note that this 53% is just as real as the systematic 47%. It is a characteristic of the charging system and the mix of cases that can and should be estimated. Capriciousness occurs when differences between defendants and their crimes are not consistently translated into charges.

V. Comparing Two Sets of Offenders

As a computational matter, it is relatively easy to compare the capriciousness of two charging systems. However, one cannot directly compare the estimated C_1 and C_2, since the sizes n_1 and n_2 of the two sets of defendants probably differ; relative assessments of capriciousness can vary solely because the sample sizes differ. Therefore, we propose to compare the average capriciousness measures $\bar{C}_1 = C_1/n_1$ and $\bar{C}_2$ as estimates of the average capriciousness that each set of defendants faces.

More difficult is the decision about how to estimate average capriciousness for the two charging systems. One option is to fit a single model to the pooled data. Another option is to fit the two sets of data separately, but with the same model. Still another option is to fit two different models to the two sets of data.

While the strategy of fitting two different models to the two different datasets may be preferable, there is always the worry that estimated differences in capriciousness are in part a function of differences in the quality of the models applied. We recommend, therefore, fitting the same model separately to both datasets. That is, the set of explanatory variables used in both models should be the same. If, in fact, different sets of explanatory variables are important in the different models, we suggest pooling the explanatory variables. For example, if the first model has *A* and *B* as explanatory variables, and the second model has *X* and *Y* as explanatory variables, we recommend using *A, B, X,* and *Y* as explanatory variables for both analyses. When the same explanatory variables are not available in both datasets, we suggest building separately the best models possible for each dataset and then qualifying any capriciousness measures accordingly, especially if one model is arguably worse than the other because of, for example, obvious omitted variables.

VI. Some Caveats

Capriciousness depends on the data and the model. Thus, estimated capriciousness is no different from other parameter estimates in that poor data or model misspecification can undermine the entire enterprise. However, we suspect that estimated capriciousness is somewhat less vulnerable to these problems than, say, estimated regression coefficients. The capriciousness C depends on the distribution of the model's predicted probabilities, and many different plausible combinations of the explanatory variables can produce the same distribution of predicted probabilities. Thus, if two rather different models generate approximately the same distribution of predicted probabilities, estimated capriciousness will be approximately the same.

Moreover, the important predictors of a death penalty charge are generally known (e.g., U.S. General Accounting Office 1990). We argue later that the likelihood of finding new and powerful predictors is very small. Such predictors could not be among the hundreds that have been tested and discarded, and would have to substantially alter estimated capriciousness beyond that explained by the other predictors routinely employed.

Another specification concern is that the standard models do not properly represent the distribution of y given the selected x's. We will later use standard statistical tools for modeling, and do not offer any statistical innovations guaranteeing the appropriateness of our models. We note, however, that no substantive research has suggested methods other than those we apply. While some have considered Classification and Regression Tree (CART) models (Breiman et al. 1984), this approach also implies a substantial amount of capriciousness.

A deeper concern is that, in effect, each case is unique; that there is a specific covariate uniquely attached to each case determining the charge. Were these known, there would be no capriciousness. While we acknowledge that this could be true in some sense, this position necessarily and completely rejects the possibility of statistical or other kind of modeling. Furthermore, this implies a total inability for *anyone* to predict the outcome of the charging process given *any* amount of historical information—which is exactly our thesis here. If each case is unique, it is impossible to predict the future from the past. Readers interested in a detailed dissection of the uniqueness argument should consult BWB.

VII. An Application

BWB used data from the County of San Francisco including all nonvehicular homicides (363) from 1978 through 1988. The data were coded from official records and forms filled out by police and prosecutors. The outcome of interest was the decision by prosecutors to charge defendants with *special circumstances.* In California, a charge of special circumstances means that associated with the case are certain aggravating factors specified by statute, which, if found by plea or trial, make the defendant eligible for the death penalty.

For this illustration, we return to San Francisco County with new data on all nonvehicular homicides from 1986 to 1993. We report parameter and capriciousness estimates from several logistic regression models fit to the data. We display traditional output, although interpretation of, for example, the standard errors is debatable since this is a population. Following Berk, Western, and Weiss (1995), one might treat the inference from a Bayesian perspective.

There are 427 cases coded using the same sorts of primary sources as the earlier study and effectively the same coding sheets. Changes in the coding sheets reflect primarily clearer definitions and technical refinements. For example, explicit instructions were given to distinguish between no mention of any weapon and a clear indication that a gun was not the murder weapon. As before, the unit of analysis is the defendant. Multiple murders were treated as a single case unless the defendant(s) was (were) tried in separate cases. There were nine multiple murder cases in the dataset.

As in the earlier study, the response variable is the decision by prosecutors to charge special circumstances. Of the 427 homicide cases, 29 (6.8%) were so charged. In the earlier study, 27 homicide defendants out of 363 (7.4%) were charged with special circumstances. Once again, prosecutors seek the death penalty in only a small fraction of homicide cases.

As before, there are well over 100 possible explanatory variables describing the biography and background of the defendant (e.g., age, ethnicity, race, gender, prior record), biography and background of the victim (e.g., age, ethnicity, race, gender, relationship with offender), and nature of the crime (e.g., weapon used, aggravating or mitigating circumstances, location of crime). We use logistic regression to model the probability of a capital charge as a function of the covariates.

We began with a model that was identical to the final model used in BWB and then tried a number of other models in an effort to increase the number of useful explanatory variables. We were anxious to include as many explanatory variables as could be sensibly justified; we were anxious to avoid the charge that later findings of substantial capriciousness were the result of omitted explanatory variables. Table 1 probably pushes the data too hard, and we will present a simpler model later.

We lose about 100 cases in Table 1 because of missing data. No one variable is at fault. Rather, deletion of all cases missing any covariate add up. Simple methods of recovering the missing cases are to code "missing" as just another dummy variable or by treating missing as "0" (the absence of the attribute in question) along with the nonmissing 0s. Neither of these strategies changed the substantive story in Table 1.

Table 1 shows the results; standard errors, odds multipliers and one-sided *p*-values are based on the usual output of a logistic regression. Our analysis is nominally Bayesian with flat priors, and inference is based on the standard maximum likelihood approximation to the posterior.

The findings are much like those reported by BWB. The features of defendants and cases that prosecutors claim to affect charging decisions surface strongly. The odds of a death-eligible charge increase dramatically when there is more than one victim,

Table 1. Logistic Regression Results of Model 1 ($N = 312$)

Predictors	Coefficient	Standard Error	One-sided *p*-value	Odds Multiplier
(Intercept)	−11.305	2.345		
Victim white or Asian	1.203	0.738	0.0521	3.332
Multiple victims	10.072	2.804	0.0002	23,678.280
Gender: victim female	0.929	0.858	0.1401	2.531
Defendant committed prior homicides	3.844	1.305	0.0017	46.735
Relationship: victim familial, acquaintance, or other known to defendant	1.692	0.810	0.0188	5.428
Manner of killing was by shot (firearms) or stabbed	2.020	0.968	0.0189	7.538
Contemporaneous felony involved (robbery, burglary, sexual assault)	2.570	0.990	0.0050	13.063
Other contemporaneous crime involved	3.177	0.997	0.0008	23.971
Victim raped and/or bound/gagged	3.426	1.326	0.0051	30.765
Robbery, robbery-narcotics, sex/prostitution, burglary, sex-rape, child abuse, attack on police, assassination	2.509	0.974	0.0053	12.290
Defendant had accomplice(s)	1.004	0.814	0.1091	2.730
Victim under influence of drugs	−2.179	1.841	0.1188	0.113
Victim was gay	−3.427	1.849	0.0324	0.032
Defendant has less than high school diploma or equivalent education	−2.442	1.066	0.0113	0.086
Defendant has no occupation or unskilled	1.636	1.051	0.0603	5.135
Defendant's education or occupation unknown	3.187	1.268	0.0062	24.219

Null deviance 193.203 on 312 degrees of freedom.
Residual deviance 68.981 on 296 degrees of freedom.

when the defendant has been previously convicted of a prior homicide, and when there is a contemporaneous felony or some other aggravator.

The earlier study did not find clear race effects and did not replicate the common finding that the odds of a capital charge increase if the victim is white rather than a member of a minority group. But there is some evidence in Table 1 that if the victim is white or Asian (compared to African American or Latino), the odds of a capital charge are about four times larger. Finally, some other biographical variables such as the defendant's education and occupation may play some role, but the precise effects are difficult to pin down because of large amounts of missing data on those variables. Missing data for education and occupation was here coded as a separate dummy variable.

More important for us is the question of capriciousness. Figure 1 shows for model 1 a histogram of the predicted probabilities of a capital charge. It is clear that the vast majority of defendants have predicted probabilities of less than 0.1, but there is a small group of defendants with predicted probabilities of over 0.9. In addition, there are a number of defendants with predicted probabilities between 0.2 and 0.8. These are cases that are most likely, roughly speaking, to go either way, and it is also these cases that make the largest contributions to the overall capri-

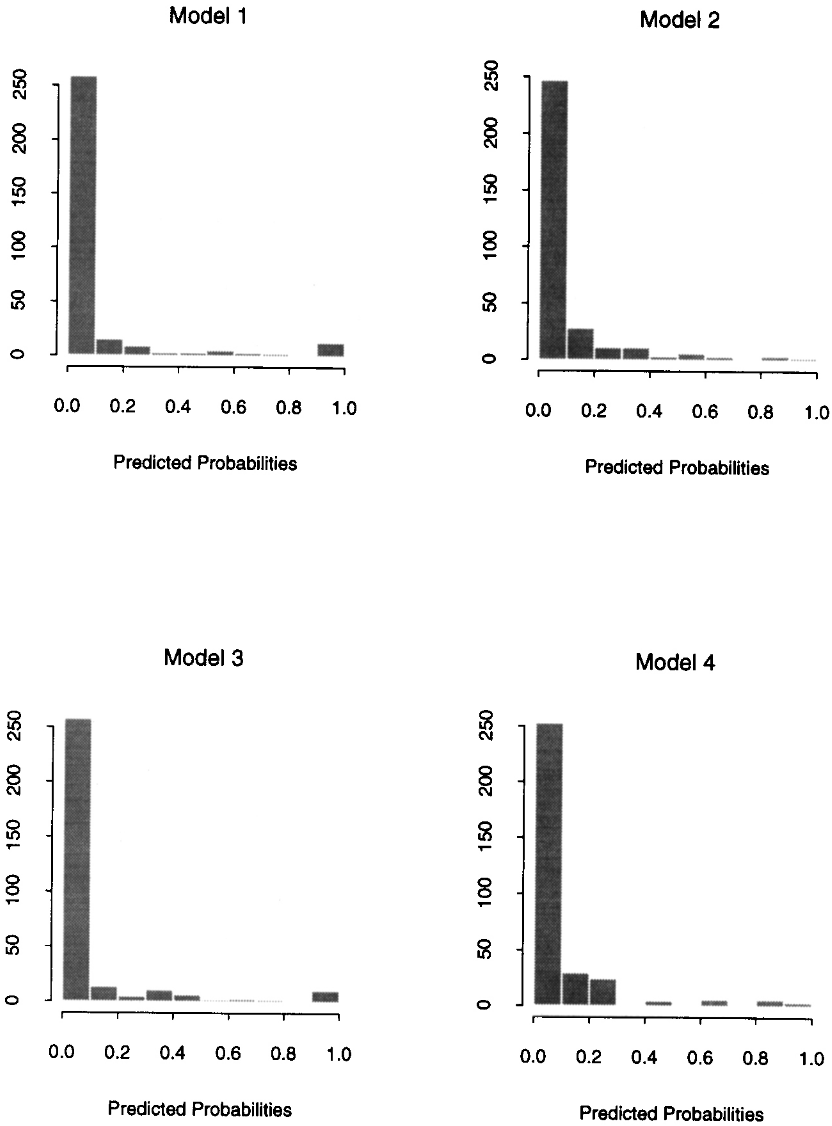

Fig. 1. Histograms of fitted probabilities $\hat{p}_i$'s for models 1–4

ciousness of the charging system. In short, the histogram has the long right tail researchers have come to expect, and the charging system contains substantial capriciousness.

The maximum amount of capriciousness occurs for this sample when all of the predicted probabilities are fixed at the overall proportion of cases charged with a capital crime. Since for the analysis shown in Table 1 there are 313 cases with complete data, the overall proportion is $29/313 = 0.093$. The maximum capriciousness is estimated at 96.85. Our suggested minimum amount of capriciousness occurs when there are two values for the predicted probability of a capital charge: $1/n$ and $1 - (1/n)$. For

https://doi.org/10.2307/3054130 Published online by Cambridge University Press

these data, the minimum capriciousness is 6.75. Finally, the estimated capriciousness from model 1 is 34.49.

From these figures, the capriciousness possibilities range from 6.75 to 96.85, or about 90 units. The procedures used by the local prosecutors manage to remove 64.36 (96.85 – 34.49) of those units, or about 69%. In other words, of the total capriciousness possible for these data, our model suggests that the charging system is able to remove about two-thirds of the capriciousness and leave one-third intact. Thus, about two-thirds of the pattern in charging outcomes is signal, depending on characteristics of the defendant, victim, and crime. And about one-third of the pattern in charging outcomes is noise.

One might object that our model grossly overestimates the amount of capriciousness because important variables that affect capital charging have been overlooked. One way to better appreciate the merits of those argument is to show the consequences for our measure of capriciousness when we *knowingly* omit one or more important predictors.

From Table 1 it is apparent that killing more than one person substantially increases the odds of a capital charge. A cross-tabulation of a capital charge by whether there was more than one victim shows that nearly a quarter (7 out of 29) of the defendants who were charged with a capital crime had multiple victims, and every defendant but one with multiple victims (7 out of 8) was so charged. Clearly, killing more than one person makes it a very good bet that a capital charge will be filed, but capital charges can be filed for other reasons too.

When the model shown in Table 1 is reestimated with multiple victims no longer included as an explanatory variable, capriciousness increases to 60.67. This is model 2. Given the ruler established earlier, the charging system now removes about two-fifths, or 40%, of the capriciousness from the system rather than two-thirds; now 60% of the charging system appears to be noise. One can see from the figure that the right tail of the distribution has been shifted to the left; there are few defendants with very high predicted probabilities and more with middling predicted probabilities. It follows that capriciousness should be higher since there are now more defendants clustered around the gray area of 0.5.

More specifically, there are two ways in which dropping explanatory variables can inflate estimated capriciousness. First, the number of different values for the predicted probabilities may decline. Where there were once 30 "bins," for example, there are now 25. Second, some extreme values for predicted probabilities become less extreme. For instance, 0.9s may become 0.7s, and 0.05s may become 0.3s. Explanatory variables with large regression coefficients and/or large variances are more able to powerfully affect how far the predicted probabilities move. Since the variable multiple victims has relatively little variance, it is having its impact because of a very large regression coefficient. A few very high predicted probabilities are shifted dramatically downward and many low predicted probabilities shift upwards slightly when multiple victims is deleted from the model.

Clearly, if there is an omitted variable or set of omitted variables with effects similar to that of multiple victims, we are grossly overestimating capriciousness using the results reported in Table 1. But how plausible is this?

In California, aggravating factors that may lead to a capital charge are specified by statute, and we have in the dataset measures of each. So, whatever it is that has been overlooked is not an aggravator specified by law. Second, one must keep in mind that the omitted variable(s) would have to exert their effect after adjusting for the effect of the other variables shown in Table 1 and not be one of the about 90 other variables that had little demonstrable impact for these data. Finally, this variable would have to account for a substantial fraction of the 29 capital charges filed after multiple victims has had its effect and not occur in cases where no capital charge occurred. We have reviewed the literature on capital charging and sentencing and have found no such variable mentioned.

Consider now a less powerful explanatory variable: whether the victim knew the defendant. From Table 1, the regression coefficient of 1.69 translates into an odds multiplier of 5.4. When the victim is a friend or acquaintance of the defendant or a member of the defendant's family, the odds of a capital charge are more than five times greater. While this is a nontrivial effect, dropping the relationship variable from model 1, giving model 3, changes the histogram of predicted probabilities vary little; compare the histogram for model 3 to the histogram from model 1. In fact, estimated capriciousness increases from 34.5 to only 42.2. It would take about a half-dozen such variables with odds multipliers of about 5 to have the same impact on estimated capriciousness as the variable multiple victims. Keep in mind that such variables would have to not be already included in our list of about 90 and would have to prove important after partialing for all the explanatory variables shown in Table 1. Again, there is nothing we have found in the relevant literature suggesting what those variables might be.

The model shown in Table 1 probably has too many explanatory variables for the number of defendants charged with a capital crime. Thus, capriciousness is probably underestimated by model 1. One result of too many predictors is that some of the estimated coefficients may be unstable and small changes in the model or modest amounts of measurement error could affect the story told. To explore these concerns, we examined several smaller models more like those reported by BWB. Table 2 shows one set of results, called model 4, in which we include the race variable and other explanatory variables with the largest odds multipliers.

Table 2. Logistic Regression Results of Model 4 ($N = 329$)

Predictors	Coefficient	Standard Error	One-sided p-value	Odds Multiplier
(Intercept)	−4.297	0.519		
Victim white or Asian	0.577	0.502	0.13	1.782
Multiple victims	5.949	1.182	0.0000003	383.208
Defendant committed prior homicides	1.902	0.788	0.0082	6.670
Contemporaneous felony involved (robbery, burglary, sexual assault)	2.417	0.525	0.000003	11.211
Other contemporaneous crime involved	1.796	0.650	0.003	6.028

Null deviance 201.1 on 329 degrees of freedom.
Residual deviance 121.8 on 324 degrees of freedom.

The story from Table 2 is straightforward. The odds of being charged with a capital crime increase dramatically if there is more than one victim, if the defendant has a prior homicide conviction, and if there was either a contemporaneous felony or another associated felony. Compared with the figures in Table 1, the race of victim effect is dramatically reduced. Finally, the estimated capriciousness is now 60.1, roughly double the estimated capriciousness for the first model.

The main point is that if in fact a number of useful predictors are excluded, estimated capriciousness can increase substantially. Compared with the histogram from model 1, the histogram for model 4 shows that the long right tail is shortened and that there are many more cases with predicted probabilities from 0.1 to 0.3.

Whether one prefers model 1 or model 4 depends on the substantive questions being asked. Since model 1 probably has too many predictors for this particular dataset, the coefficient estimates may be unstable because of collinearity not fully represented by inflated standard errors. If interest lies in estimating capriciousness, the larger model at least provides a plausible lower bound. Of course, it may turn out in practice that a single model will suffice whether the interest is in the regression coefficients or in capriciousness.

VIII. Discussion and Conclusions

We have here extended the earlier work of Berk, Weiss, & Boger (1993a), in which the concept of an *as if* lottery was applied to death penalty charging decisions. Our contributions are both methodological and substantive. We have suggested that Shannon Information (definition (1)), when considered in the context of plausible minima and maxima can be used to quantitatively characterize the systemic capriciousness of decisions to charge homicide defendants with a capital crime. Our measure of capriciousness responds to chance homogeneity and heterogeneity in outcomes; estimated capriciousness is increased when observed differences between homicide defendants and their crimes do not consistently translate into differences in the probability of a capital charge.

Our most optimistic estimate is that compared with a charging system with maximum possible capriciousness where no distinctions are made between defendants, the current procedures in San Francisco wring out about two-thirds of the potential capriciousness. We doubt that it can be shown that the systematic component is substantially larger. First of all, the relevant literature is silent on variables we have overlooked that would dramatically alter the model's predicted probabilities, especially after adjustment for the other variables already included. In addition, model 1 already contains 16 explanatory variables when there are only 29 capital charges in the data. With added explanatory variables, one would begin to closely approximate one explanatory variable for each death penalty charge. At that point, the model would be pushing against the uniqueness problem discussed briefly above, and at great length in BWB. Finally, it seems inevitable that there will always be a substantial number of *close calls* for which the charging decision could go either way. In fact, homicide cases vary on a variety of dimensions, and many cases will fall in a *gray area* in which it is unclear whether a capital charge is appropriate.

Several conclusions follow. First, we have suggested a way to quantify capriciousness that may be usefully applied not just to death penalty charging decisions but to any people-processing institution with dichotomous outcomes. The measure seems to map well onto common-sense notions of capriciousness and is easy to compute.

Second, estimated capriciousness is model and data dependent. Like any summary statistic, it can be challenged with a showing of significant model misspecification or substantial measurement error. In our view, this raises the issue of where the scientific burden of proof lies. Following a good faith and credible showing that substantial capriciousness exists, sweeping claims of errors should not carry any weight unless solid empirical evidence can be brought to bear.

Third, death penalty charging decisions in San Francisco, and almost certainly elsewhere, would seem to be marked by substantial capriciousness. In effect, large numbers of defendants are assigned to lotteries in which the probabilities are some distance from 0.0 and 1.0. This means that even at the empirical extremes, there will be defendants who are charged with capital crimes by no manifest rationale. There will also be defendants who are not charged with a capital crime by no manifest rationale. Insofar as there are significant numbers of defendants who are not easily assigned to the extremes, unpredictable and inexplicable charging decisions will be relatively common. Without clearly articulated and exhaustive rules for assigning capital charges, systemic capriciousness is inevitable, and the real question is how much capriciousness the courts and legislatures of the land are prepared to accept in capital cases. We have provided some tools to help quantify the question of *how much* capriciousness.

Finally, there are additional issues we are pursuing that are not considered here. On the technical side, capriciousness is a summary statistic and when applied to a sample, needs an estimated posterior density. We have employed several different procedures to this end, but have yet to settle on the best approach. In the special case of flat priors, logistic regression, and estimation based on posterior modal estimates, our capriciousness measure is numerically equal to one half of the deviance (McCullagh & Nelder 1989). We are providing, therefore, a useful generalization that can be easily used to characterize capriciousness in situations where the deviance does not apply.

Our technical contributions are, therefore, fivefold. First, we provide an interpretation of the deviance as Shannon Information for the models we employ. Second, we show how Shannon Information can be made more useful by providing ways to compute minimum and maximum capriciousness. This is *not* the same grounding used for the deviance. Third, we give an interpretation to the deviance that can be employed by Bayesians. Fourth, we advocate the calculation of Shannon Information in models besides linear logistic regression with flat priors and estimation techniques beyond maximum likelihood. In particular, we suggest that the Shannon Information is an interesting quantity to estimate in its own right. Finally, we apply capriciousness to a real world problem.

Our implicit loss function is also an issue. Capriciousness, as we have defined it, is *unbiased* in the sense that each predicted probability is weighted the same in the calculations. Alternatively, one might choose to weight the predicted probabilities differently. For example, one might want to weight more heavily the small predicted probabilities because they represent the lightning strikes that so concerned Justice Stewart. The high predicted probabilities, in contrast, may represent random mercy, which is presumably less problematic.

On the substantive side, we argue capriciousness as an inherent quality of social systems. A sensible initial goal, therefore, measure the amount of capriciousness. Having established the amount of capriciousness, one may consider whether the amount is intolerable and whether an effort should be made to reduce it.

These issues are especially pressing when, as in capital cases, the stakes are very high.

We hope to accumulate more experience with estimated capriciousness using data from a number of different jurisdictions. Given consistent findings of substantial capriciousness in many jurisdictions, we will then be faced with the problem of developing social systems that eliminate or substantially decrease the capriciousness.

References

Baldus, David C., George Woodworth, & Charles A. Pulaski, Jr. (1985) "Monitoring and Evaluating Contemporary Death Penalty Sentencing Systems: Lessons from Georgia," 18 *Univ. of California–Davis Law Rev.* 1375.

——— (1990) *Equal Justice and the Death Penalty: A Legal and Empirical Analysis.* Boston: Northeastern Univ. Press.

Berk, Richard A., Robert E. Weiss, & Jack Boger (1993a) "Chance and the Death Penalty," 27 *Law & Society Rev.* 89.

——— (1993b) "Rejoinder," 27 *Law & Society Rev.* 125.

Berk, Richard A., Bruce Western, & Robert E. Weiss (1995) "Statistical Inference from Apparent Populations" (with discussion), 25 *Sociological Methodology* 421.

Bowers, William J., Glenn L. Pierce, & John F. McDevitt (1984) *Legal Homicide: Death as Punishment in America, 1864–1982.* Boston: Northeastern Univ. Press.

Breiman, Leo, Jerome H. Friedman, Richard A. Olshen, & Charles J. Stone (1984) *Classification and Regression Trees.* Belmont, CA: Wadsworth.

Gross, Samuel R., & Robert Mauro (1989) *Death and Discrimination: Racial Disparities in Capital Sentencing.* Boston: Northeastern Univ. Press.

Khinchin, A. I. (1957) *Mathematical Foundations of Information Theory.* New York: Dover.

Kullback, Solomon (1958) *Information Theory and Statistics.* Gloucester, MA: Peter Smith.

Maynard, Douglas W. (1984) *Inside Plea Bargaining: The Language of Plea Bargaining.* New York: Plenum Press.

McCullagh, Peter, & John A. Nelder (1989) *Generalized Linear Models.* 2d ed. London: Chapman & Hall.

Paternoster, Raymond (1993) "Comment: Assessing Capriciousness in Capital Cases," 27 *Law & Society Rev.* 111.

Paternoster, Raymond, & Annmarie Kazyaka (1988) "Racial Considerations in Capital Punishment: The Failure of Evenhanded Justice," in K. C. Haas & J. A. Inciardi, eds., *Challenging Capital Punishment: Legal and Social Science Approaches.* Newbury Park, CA: Sage Publications.

Shannon, Claude E. (1948) "The Mathematical Theory of Communication," 27 *Bell System Technical J.* 379, 623.

U.S. General Accounting Office (1990) *Death Penalty Sentencing: Research Indicates Patterns of Racial Disparities.* GAO/GGD-90-57. Washington: General Accounting Office.

Case

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