PROBABILISTIC SENTENCING

KENNETH HALPERN

February 9, 2012

CONTENTS

1. ABSTRACT

We propose substituting a sentencing mechanism based on a probabilistic assessment of guilt or innocence. This allows jurists to better express their certainty or lack thereof than does our traditional all-or-nothing verdict. The natural place to reflect such an imputed degree of guilt is in the sentencing phase. We discuss the implications of such a system as well as certain issues with implementation.

2. INTRODUCTION

One of the major persistent attributes of legal systems throughout history is the use of a binary verdict; a defendant either is guilty or innocent. Implicit in this approach is the assumption of an absolute truth, though perhaps hidden from human inquiry. The idea that the innocent may be convicted and the guilty acquitted is understood and accepted. Depending on the relative acceptability of these two false outcomes, a greater or lesser burden may be placed on the defense. That jurists, professional or amateur, shall err is inevitable. This is true even if they are absolutely certain of their conclusion, and a wise jurist rarely is. The most perfect judge can only use the evidence presented. No human is omniscient, and mistakes will be made. The wise jurist knows this and can assess his own confidence in the verdict. By requiring a binary choice we rob ourselves of this important judgment, and in so doing miss an opportunity to make a necessarily imperfect system less so.

The definition of guilt is a fundamental question of judicial philosophy and differs between various theories. Its precise nature does not concern us here, and we require only that there exist a standard meaning in the relevant justice system. Nor do we consider the definition or nature of truth. Regardless of the philosophical viewpoint taken, there remains a binary choice: guilt or innocence. This is a characteristic of the trial process rather than any epistemological assumptions.

Similarly, our social perspective reflects a simple moral dichotomy. We tend to view the world in terms of good guys and bad guys, good and evil, guilt and innocence. Part of this may derive from religious traditions, but the root likely is deeper and represents a desired simplicity of thought. We see a convicted man as "guilty", a term of opprobrium – regardless of where or how he was tried, and irrespective of our certainty that some fraction of those so judged are innocent. The need for such

simplification transcends intellect. Regardless of our education and sophistication, this emotional perspective endures.

If we move beyond this prejudice, it becomes apparent that the existence of an absolute truth in each legal case is irrelevant. Instead of confounding truth and judgment – the fallacy implicit in a simple verdict – we realize that it is our certainty which is relevant to a trial with real human and social implications. Even were we to assume the existence of a particular binary truth, our determination of it need not mirror that form. We need no longer pretend that an absolute verdict must be rendered and that, insomuch as it represents our closest approximation of truth, it *is* truth from a legal standpoint.

In certain areas of law, a probabilistic or weighted assessment already is used. Civil lawsuits may involve an attribution of responsibility and apportionment of liability. Our proposal is to extend this to all trials. Psychologically, this may be difficult to accept. A defendant either is guilty or innocent, and it seems perverse to punish a man in the absence of certainty. But certainty is a convenient fiction. Consider a defendant who faces 10 years in prison if found guilty. A jury that finds him 80% guilty may assign an 8 year sentence. Depending on one's perspective that seems like 8 years too many or 2 years too few. It can appear ridiculous to punish a man in half measures because of uncertainty or the jury's laziness. But is it? It is we who are lazy, for we wish the whole matter to be neatly packaged and conveniently disposed of. We wish to be able to say that the man is guilty or innocent, condemned or vindicated. What does one say to a man who has been 80% convicted? How does one consider him? But all verdicts are probabilistic. We just force a binary choice from this probabilistic assessment. If a jury declares the man guilty, we need not worry ourselves; he deserves the full sentence. There is no need for discomfort or the acknowledgment of unpleasant realities on our part.

Our suggestion is to stop pretending, accept that an absolute verdict is a fiction, and directly employ a probabilistic assessment.

In addition to the greater flexibility inherent in a probabilistic system, there also are certain practical benefits. The cost of a mistake is reduced, as is the burden placed on jurists. Their responsibility remains, but they may directly vote their conscience rather than be forced to make an artificial choice. Deadlocked juries would cease to arise – and the corresponding pressure to make a choice, any choice, in order to go home is removed. A single strong personality may still sway a jury, but will not force it. The adjustment of sentences becomes easier as well. If new evidence is discovered, it may be incorporated in a manner that requires neither a full repudiation or a full ratification of the prior decision.

3. THE BASIC PROPOSAL

The mechanism that we propose is independent of the trial process, though we shall discuss certain practical issues of implementation with judges and juries. The idea is simple. A trial is conducted as before, but instead of a binary choice of guilt or innocence a probability of guilt is rendered. However this is arrived at, we adjust the sentence proportionately. In order to do this, the sentencing phase must be completely separate. Otherwise the choice of a sentence from an allowed range may subvert the use of a probabilities. Let us consider an example. A man is tried for crime X. He is found guilty by a jury, with 70% probability. The legally prescribed sentence for crime X is 10-15 years in prison, at the discretion of the judge. If the judge feels that the man is guilty and knows that the sentence will be reduced by 30%, he may be inclined to issue a harsher sentence. For our idea to work, the judge would have to reach a decision independently. The solution is for the judge to decide on a sentence before the jury verdict is read. Or a third party could determine whether a sentencing phase is necessary, only revealing the actual probability after the defendant has made his pre-sentencing statement and the judge has assigned a sentence. Either way, the judge must decide

on a punishment without knowledge of the verdict probability. And he must do so on the assumption that the man is completely guilty.

Once the sentence has been determined, the jury's verdict is read and the sentence reduced appropriately. In our example, if the judge decides on a sentence of 12 years (for complete guilt) and the jury then reveals a 70% certainty, the final sentence would be 8.4 years in prison.

Note that although the present ability to levy multiple charges and allow a verdict on each count has certain similarities to our system, the principle behind it is quite different. In that case a choice of the actual *crime* committed is being made, rather than an assessment of the probability that *a* particular crime has been committed.

4. ADJUSTMENTS

A practical application of probabilistic sentencing must make allowances for very small or very large probabilities. Although jurists may choose to do so implicitly themselves, explicit controls must be introduced as well. In theory, there can never be certainty of guilt or innocence. We do not wish to punish everyone who is accused, and the difference between no jail time and *any* jail time is far greater than small differences in the duration of a sentence. Among other things, such a state of affairs could lead to prosecutorial abuse and a condition of social paranoia. A de minimus level must be set, below which no sentence is imposed. Such a level could be based on either the probability or sentence, and the former is the correct approach. That is, we would acquit if the probability is below a certain value.

If we used the punishment itself rather than the probability, we would run into problems concerning small and large sentences. Small sentences may be appropriate for certain offenses. To avoid excluding these, our de minimus level would have to depend on the base sentence. The reasonable way to do this would be through a proportional adjustment – which is tantamount to the use of probabilities. Large sentences could lead to anomalous outcomes that are unacceptable. For example, a 1% certainty of guilt for a crime with a 100 year sentence could lead to a year in prison.

What is an appropriate choice of de minimus probability? There are two approaches to this. We could make it small on the theory that it represents a limit of the jurists' ability to discern. In that case, a large number of people would end up in prison for short terms. If we make it large instead, then fewer innocent people will be punished but more guilty ones would go free (rather than merely receiving a lighter sentence).

The suitable choice depends on the prevailing social values and anticipated behavior of the prosecutorial organs. If the prosecution is expected to bring charges only in the presence of substantial evidence, and jurists are expected to underestimate their certainty or impose their own de minimus standard, then it may make sense to set a low statutory minimum. On the other hand, if one does not wish to rely on the discretion of prosecutors and jurists or if society wishes to minimize the harm to its citizenry or costs of incarceration, then a higher de minimus level is called for. A 50% standard could serve as a reasonable choice in that case. Given our nation's emphasis on individual freedom and minimizing unnecessary harm to the citizenry, we likely would adopt something in this range.

A similar consideration applies at the upper end of the probability range. In any judicial system where guilt is the presumption rather than innocence, the aforementioned considerations should be applied to a de maximus probability instead of a de minimus one. In most real systems, we likely would wish to employ both de minimus and de maximus levels. In our prior example, 50% could be the de minimus and 80% could be the de maximus. Any verdict with over 80% probability of guilt would receive the full sentence, and any with under 50% would result in acquittal. Note that if the de minimus and de maximus levels are equal (they need not be 50%) , we simply have a formalized

version of the usual binary verdict. The farther apart the two lay, the greater the discretion of the jurists.

Although we assumed a proportionate adjustment of sentences, it also is possible to apply a nonlinear scale[1]. Such a map could serve to smooth out the de minimus or de maximus step functions, or it could better reflect the moral view we take of penalty vs certainty[2].

Independent of the statutory de minimus and de maximus probabilities, any practical system also would require a means of immediately excluding cases which do not meet a minimum standard of evidence. This would avoid excess trials, prevent prosecutorial excess, and limit unnecessary damage to citizens' lives.

5. PLEA BARGAINS

The treatment of plea bargains is a little tricky. It may pay to do away with them altogether, though this could deprive prosecutors of a valuable tool and hobble efforts to elicit cooperation. The justification for use of probabilities does not extend to plea bargains. Presumably, a defendant knows whether he is guilty or not. If we allow probability pleas – that is, offers to plead some percent guilt – then the process is reduced to a game of chance. However, the present ability to negotiate a choice from among charges of varying seriousness lends plea bargaining this character anyway. It is reasonable to assume that with any form of plea bargaining, the sentence is the negotiated quantity. The probability or choice of crime merely serve as parameters that can be adjusted to legally obtain the agreed upon sentence. From this standpoint, probability pleas are reasonable.

6. MULTIPLE DEFENDANTS

Multiple individuals may be tried for the same crime under our current system, but only if they are defendants under a single prosecutorial theory. With a probabilistic system, individuals could simultaneously be tried under different theories. In practice, this would require a joint trial and the requirement that the probabilities of guilt under different theories not add to more than 100%. Of course, the admission of multiple theories may give prosecutors the incentive to employ a shotgun approach. We do not recommend allowing multiple theories in general; however, there are situations where it is not unreasonable to do so.

In a trial, there typically is a single prosecutorial theory, according to which the defendants are guilty as charged. However this theory often incorporates implicit subordinate theories. While it is the individual defendants who are in jeopardy and upon whose persons punishment would be exercised, it is the theories which really are on trial. A guilty verdict involves the conjunction of two determinations: (1) the prosecutorial theory is correct as a whole and (2) the defendant is guilty as imputed by that theory. These two components may be indistinguishable or quite distinct. For example, the theory that person X and two masked associates robbed a liquor store may be accepted by a jury, yet the subordinate theory that defendant Y was one of those associates may be rejected. On the other hand, if person X was on trial then the theories would coincide for all practical purposes.

Returning to the case of person Y, would it be unreasonable for the prosecution to additionally offer persons A and B as candidates? If there is evidence that each could be guilty, then why not allow multiple prosecutorial theories. In this example, only two could be guilty – so we could require that

[1]Mathematically, we require a monotonically non-decreasing function $f : [0, 1] \to [0, 1]$ such that $f[0] = 0$ and $f[1] = 1$. However f need not be continuous or have an image with nonzero measure.

[2]For example, we could decide that everything between the de minimus and de maximus levels should receive a 50% penalty.

the probabilities sum to no more than 200%. They likely would be less, because the primary theory also carries uncertainty. Such an approach may or may not be palatable, but it deserves mention.

7. Determination of Probabilities

The means by which a probability of guilt is arrived at would depend on the structure of the trial. With a single judge, the assessment entirely is personal. It is possible that a professional jurist would view the probabilistic system as a nuisance, preferring simply to dispose of the entire case in one action. For this reason among others, it would be best if a single judge did not govern all aspects of the trial.

Aggregation of the opinions of jurors or multiple judges could be accomplished in many ways. The simplest is an average of the individual jurist probabilities. Additionally, we could require that a certain number of jurors exceed a threshold in one or the other direction. There is no problem with deliberation per se, though it likely would lead to highly correlated results. However the individual probabilities will be highly correlated anyway, because the jurists all saw the same evidence.