

Proving the Unprovable

Proving the Unprovable: The Role of Law, Science, and Speculation in Adjudicating Culpability and Dangerousness

by Christopher Slobogin (Oxford University Press
(2007); \$55.)

Christopher Slobogin's slim, accessible, and engagingly written new book, *Proving the Unprovable: The Role of Law, Science, and Speculation in Adjudicating Culpability and Dangerousness*, tackles a fascinating subject and does so in a way that will make it of immediate value

to practitioners, judges, forensics experts, and academics. Slobogin's theme turns on this insight: Not all facts to be proved at criminal trials and hearings are alike. Some facts, particularly those involved in establishing that a suspect committed an alleged past criminal *act*, fit our commonsense notions of what is real and how we know so. Suppose John is charged with shooting and killing George, but John raises an alibi, claiming an alternative perpetrator (alt-perp) theory: Sally, not John, did it. If we could travel back in time, we could see for ourselves, one way or another, whether John or Sally indeed did the deadly deed. Because we cannot travel through time, however, we can turn to eyewitness testimony, expert testimony, and physical evidence to make our best reasoned judgment about what really happened.

Predicting the future. But, argues Slobogin, there are at least two important types of "facts"—those involving dangerousness and those involving culpability—for which there is no one, indisputably correct truth in the same sense as whether

John or Sally shot George. The first of these categories of unprovable facts is *predictions* of dangerousness. Such predictions are often important in the penalty phase of capital cases, more routine sentencing proceedings, sexual predator commitments, and mental health commitments, all of which often require proof of some sort that the defendant likely *will*, in the future, offend again. The testimony of forensic mental health specialists is often central to these determinations. Yet each such case involves proving something whose occurrence is not "provable" because it has not yet happened! Much like in the movie *Minority Report*, in which suspects are arrested and convicted by the police's "Future Crimes Unit" for crimes that the mutant fortune teller "pre-cogs" insist that the defendants *will* commit if not stopped, we approve harsher sentences or other deprivations of freedom for suspects' crimes that they may not yet have even conceived.

In *Minority Report*, the system collapses when it is revealed that sometimes one pre-cog "dissent" from the verdict of the other two, raising doubt about whether the future crime really will happen. But such a collapse of our current system is unlikely for two reasons. First, the real criminal justice system generally uses its crystal ball only to punish for future wrongs when an offender has already been convicted for some prior crime, likely lessening society's regret if it makes an error in enhancing his future penalties. After all, the defendant did already do something so wrong as to mark the person unworthy of freedom. Second, we have no mystical pre-cogs in which to place an almost religious faith in their prophetic prediction. We can conclusively prove future danger only by waiting to see if the offender indeed commits the predicted crime, a delay we will never tolerate because the whole point of preventive detention is to stop those future crimes before they happen. Accordingly, our criminal justice system routinely trusts ordinary humans, specifically, mental health professionals, to aid in proving that what is not now true will nevertheless soon be. We expect fact finders accurately to resolve disputes between these mental health experts. The substantive law in these cases mandates that judges become seers rather than historians, and Slobogin seeks at least to help courts do

Andrew E. Taslitz is a professor of law at the Howard University School of Law, the editor of this symposium on mental health and criminal justice, a member of the *Criminal Justice* editorial board, of the ABA Criminal Justice Standards Committee and of its Working Group on Transactional Surveillance, and codirector of the Communications Division of the ABA's Criminal Justice Section.

the best they can with this difficult task.

Culpability for the past. The second of the two “unprovable” types of facts is what academics call “culpability,” better known to practicing lawyers as *mens rea*. Past mental states are fundamentally different sorts of facts than are criminal acts. Even if we could travel back in time to see John shooting George, we cannot “see” what was in George’s mind at the time. His fear, anger, or confusion is invisible to us.

Nor would mind reading, were it possible, solve this problem. Humans are notoriously good at self-deception. They often do not consciously know the truth that explains why they committed a certain action. Furthermore, human thoughts and feelings are multiple, conflicting, and confusing. A person can act both from fear and hatred, not being clear which passion was stronger.

More importantly, understanding our own inner world requires interpretation more than observation. If our heart beats rapidly, our hands sweat, our mouth is dry, do these signs reflect fear, lust, or hate? We draw on contextual cues—what triggered these reactions—prior experience, and social expectations to decide just what it is we are feeling. We may still have doubts about our interpretation’s accuracy, change our minds, and even anguish over it, as might a young lover wondering whether he is merely infatuated or truly in love. Yet the interpretive nature of mental state determination is even stronger when a third party, such as a judge or jury—that cannot hear our internal conversations, remember our life experiences, or know our life’s pressures—stands in judgment.

The fact finders’ limited information base will therefore lead them to draw on social conventions, stereotypes, biases, analogies, and their own personal experiences, rather than the defendant’s, in seeking to enter into the latter’s mind. But the *mens rea* categories of the criminal law often do not match everyday understandings of the human mind. Terms like “malice aforethought,” “depraved heart,” and “heat of passion” invite moral judgments as much as historical re-creation of events. The result, therefore, is that fact finders, whether judges or juries, construct, rather than “discover,” mental states by crafting values-infused stories to explain a defendant’s actions. This sort of “fact” is a far cry from deciding whether John shot George.

Yet, once again, the substantive criminal law requires fact finding concerning the interpretive, narrative fact called *mens rea*. Although moral

judgments are part of this process, they are not all of it, and, in any event, turn on an understanding of human nature. Where lay preconceptions may be uninformed ones, mental health professionals’ testimony again seems important.

The courts have agreed. Most often psychologists and other behavioral experts testify about routine diagnoses common in their professions, such as whether the defendant suffers from paranoid schizophrenia or some other diagnosis from the *Diagnostic and Statistical Manual of Mental Disorders* of the American Psychiatric Association. But a wide array of expert testimony is more infrequently offered on less everyday subjects like passive-dependent personality disorder, urban war syndrome, and a host of other “abuses excuses.” Mental health professionals’ testimony about mental state can be relevant to insanity, diminished capacity, and partial responsibility defenses, as well as to self-defense, entrapment, and witness credibility. Many commentators flatly reject *any* such expert testimony about mental state or would limit it to merely reciting observations of a patient’s behavior precisely because of its scientifically “unprovable” nature. Slobogin concedes that most such testimony fails traditional standards of scientific trustworthiness, yet he argues, contrary to these other commentators, that mental health professionals’ expert opinions should sometimes be permitted. He also objects to the attitude of most courts, which freely admit expert testimony on traditional mental health issues and apply only marginally more scrutiny to novel or unusual ones.

Ultimately, therefore, Slobogin is less interested in whether mental health experts’ opinions about mental state and future dangerousness should be admitted at criminal trials—he enthusiastically believes that they should be—than in when, why, and how they should be used at trial. All this theory about the nature of facts thus leads up to a nuts-and-bolts defense of a four-factor evidentiary test that Slobogin finds implicit in the Federal Rules of Evidence and adapts to the problem of “proving the unprovable.” Those five factors are:

1. *necessity*: the degree of the proponent’s need for the evidence given the unavailability of equally effective alternatives.
2. *materiality*: whether “testimony” logically relates to the relevant substantive law.
3. *probative value*: a measure “of the accuracy of expert testimony, with respect to

both its general principles and its specific applications to the case at hand.”

4. *helpfulness*: the “extent to which the testimony adds something to what the fact finder can accurately figure out for itself.”
5. *prejudicial impact*: whether, even if the above three factors favor admission, the testimony should “nonetheless be . . . excluded because it will most likely be misused by the fact finder.”

(*Id.* at 15.)

The rest of Slobogin’s book is largely an elaboration of these factors and their application to first, culpability, second, future dangerousness, determinations.

Culpability and the mental health expert

Slobogin’s bottom line position is that, under his five-factor test, there are at least sometimes good reasons to admit clinical or interpretative social science mental health testimony to prove culpability. Such testimony is necessary, material, probative, and helpful, with a relatively small risk of being unfairly prejudicial.

Necessity. Slobogin argues that there is a strong social need for expert evidence on culpability precisely because of the interpretive nature of mental states. Ultimately, mental state determination involves values-infused narrative thinking. A mental health expert can help jurors to consider plausible mental state alternatives to “normal,” confront their own unconsidered assumptions about human nature, and otherwise open their minds to hypotheses that they might miss unaided. Moreover, by helping defendants, who are otherwise incapable of doing so, to accurately tell their tales, mental health experts give those on trial voice, a sense of participation, and an opportunity to bring diverse views and experiences to the jury’s attention so that the moral, normative portion of mental state “fact finding” is consistent with the American republic’s democratic values and with equal respect. Additionally, by making the process appear fair, it promotes defendants’ acceptance of verdict legitimacy and provides the other benefits of procedural justice, such as an increased willingness to obey the law.

Materiality. Materiality is a simple question of logical relevancy, yet this evidentiary minimum of admissibility is too often ignored when mental

health experts are involved. One of Slobogin’s most interesting examples is the prosecution of Damian Williams and Henry Watson for aggravated assault, felony murder, and attempted murder of Reginald Denny and others during the Los Angeles riots following the acquittal of the officers who beat Rodney King. There, the defendants offered social science testimony, which was quite successful, supporting the theory that they were caught up in a “group contagion” of anger and frustration caused by the King riots. Accordingly, they insisted, they did not have the intent to harm. But, under California law, such “diminished capacity” testimony is relevant only to negate specific intent, yet assault and mayhem are general intent crimes in that state, rendering the testimony immaterial to those charges. Attempted murder is a specific intent crime in California, requiring a purpose to kill. Premeditation is not necessary. But, explains Slobogin, proving that the defendants were “out of control” is immaterial too, for killing impulsively from anger and frustration still involves a purpose to kill.

Probative value. The probative value of mental health experts’ testimony, Slobogin maintains, is best measured by “generally accepted content validity.” There are several types of validity concepts used by social scientists. But, argues Slobogin, most of these measures are impracticable in this context because judgments about past mental states are “not . . . easily reducible to objectively verifiable components in the same way abuse and eyewitness identifications [of physical acts] are.” The only way to measure matters relevant to the concept of blameworthiness is, therefore, “by studying the extent to which they are accepted by judges, juries, or panels of experts in mental health law, and then determining what factors (type of expert, length of evaluation, clinical tests used) correlate with those cases when the court/panel agrees with the expert.” (SLOBOGIN, *supra*, at 64.)

Few studies of this nature are currently available, so Slobogin recommends the creation of a computerized database of the best expert case studies, subjected to peer review and readily accessible to future experts. Expert panels would create the database, which would be published in journals and on the Internet. These case studies would illustrate the “types of inquiries, instruments, tests, and analytical processes that experienced forensic clinicians have developed in evaluating various legal issues that have proven useful and persuasive to the

courts.” (*Id.* at 64.) Slobogin offers the example of a murder trial in which a psychologist seeks to opine, in support of an insanity defense, that the defendant suffered from post-traumatic stress disorder (PTSD). Slobogin describes the practical value to the court of a database of properly formatted case studies in making the admissibility decision:

The role of such a format in determining admissibility would be straightforward. If future evaluators faced with potential PTSD cases failed to consider the various clinical factors it incorporates—the diagnostic criteria, their forensic implications, sources of information—the admissibility of their opinion would be called into doubt, depending on the degree of failure and whether reasonable alternatives were used. Ultimately, a clinician who wanted to testify about PTSD-induced insanity but who paid little heed to various factors that case law (in both the legal and pragmatic psychology sense of the term) has identified as significant to that inquiry would be prohibited from taking the stand. In this sense, at least, *Daubert*’s demand for verification that an expert’s opinion be reliable *can* be implemented in the culpability context.

(SLOBOGIN, *supra*, at 66.)

Helpfulness. Slobogin addresses expert helpfulness by building on the concept of “incremental validity.” Evidence offers incremental validity if it *improves* accuracy, here meaning that the proffered expert evidence improves the jury’s accuracy in assessing mental state. But how can improved accuracy be measured for such an amorphous concept as culpability? Slobogin’s answer is “factor-based incremental validity,” ideally relying on experimental studies of the factors juries use in making mental state judgments when informed by an expert and comparing that to the factors they use when left to their own devices. If an informed jury used more material factors than an uninformed one, that would establish incremental validity. Similar research has already been carried out in other areas, though not yet concerning culpability.

Pending such studies, Slobogin offers a rougher, more commonsense test:

[A] rougher test of helpfulness, derived from the law itself, can be suggested. Stated most

simply, testimony about acts or mental states that is meant to rebut presumptions, overcome statements or innuendo from the opposing side, or in some other way provide counterintuitive or corrective information should be considered helpful.

(*Id.* at 78.)

For example, notes Slobogin, both the criminal law and laypersons assume that individuals intend the natural and probable consequences of their acts. Expert testimony that this is not so in a particular case is helpful in an insanity, provocation, or lack of mens rea defenses because it rebuts lay preconceptions. Likewise, battered woman’s syndrome evidence can overcome a presumption that a truly battered woman would have left her husband, thus again rendering the testimony helpful, as would evidence that the defendant had an extraordinarily passive personality, supporting the counterintuitive notion that she did not commit the act charged, despite having what would appear to be motive and opportunity for a more typical person in her situation.

Unfair prejudice. Finally, Slobogin argues that it will be the rare case where necessary, material, probative, helpful testimony will also unfairly prejudice the jury more than it will improve its ability to assess culpability accurately. He nevertheless concedes that there are some cases where the contrary may seem true. Psychological evidence based on a new theory not yet making its way into the proposed forensic database might overawe a jury, which may be presented with insufficient critiques or alternative theories or lack a judge’s practice in evaluating them. A theory that reaffirms strong preconceptions might also worsen the risk of jury error. But exclusion should still be a last resort, insists Slobogin, if effective adversarial correctives are in place. Opposing lawyers should, through cross-examination and rebuttal evidence, be able to point out statistical flaws, relative implausibility, and unresolved ambiguity, and research suggests that juries will be able to use this counterevidence effectively, especially because they tend to *under-* rather than overvalue social science testimony. To ensure that adversarial safeguards are in place, Slobogin recommends in limine motion hearings to determine whether this is likely. If not, the trial judge can appoint his or her own expert under Federal Rule of Evidence 706 or a state equivalent. Jurors can also be authorized to ask experts questions and can even

be allowed to read briefs prepared for them on the scientific issues, while judges can enforce ethical rules more vigorously against incompetent counsel.

Predicting dangerousness

Background. Slobogin devotes only one-third as many pages to the analysis of predicting dangerousness as he uses to address culpability assessments, but his conclusions are no less important. Remember that predicting dangerousness involves predicting an *act*, not, as with culpability assessment, a mental state. The existence of a criminal act—stabbing, shooting, or beating someone—is a much more objective, less normative, and more easily measurable matter than is mental state. The problem that arises, however, is precisely that we are predicting that which has never happened and thus can make only a best guess, a probability assessment. Usually the guess is made in significant part by relying on what other, similarly situated, defendants have done and comparing that to the current defendant's characteristics and behavior. Future dangerousness assessments of this sort are common in capital case penalty phases, other case sentencing hearings, commitment proceedings, and sexual predator hearings.

After weighing all the competing considerations, Slobogin's bottom line is this: Traditional *clinical* testimony about future behavior is often so unhelpful and unfairly prejudicial, despite being probative and material, that it should generally be inadmissible unless its subject chooses to offer such testimony first to prove his or her nondangerousness. Otherwise, the state can seek to prove future dangerousness only by using alternative means, namely "properly normed actuarial instruments or structured interviews that are tied to explicit probability estimates."

Slobogin's conclusions turn on the distinctions among four types of predictive methods: clinical, actuarial, adjusted actuarial, and structured professional judgment. Clinical testimony consists of the relatively subjective judgment of a clinician based on the factors deemed necessary from the clinician's training and experience. Actuarial prediction works much like insurance actuarial tables. For example, the Violence Risk Appraisal Guide (VRAG) uses 12 empirically derived, narrow variables, and nothing else, to produce a score estimating the probability of recidivism. An adjusted actuarial approach uses such tests as a baseline but raises or lowers the score based on

other observations known to relate to recidivism but not included in the test's computation, such as this particular offender's doing well in treatment (lowering the risk) or threatening upon release to join a gang (raising the risk). Structured professional judgment assigns ratings to a number of specified items in three categories: historical, clinical, and risk management. Unlike with actuarial prediction, however, no algorithm is used to produce a score statistically correlated with recidivism probabilities. Instead, the ratings simply inform the offender's clinical judgment, though the ratings have each been correlated with risk.

There are advantages and disadvantages with each method, but what I want to focus on here is their predictive accuracy. Gauging accuracy requires knowing both the false positive rate—how often the predictions are wrong in finding dangerousness—and the base rate—how often the population studied reoffends overall. Thus, if a predictor's false positive rate is 50 percent, but the population, say, of sex offenders, reoffends at a 50 percent rate, the predictor is no better than chance. One measure that takes into account both these factors is the Receiver Operating Characteristics Curve (ROC), the computation of which I will not review here. Numerous recent studies reveal ROCs better than chance for each of the four predictive methods, averaging .67 (a 67 percent chance of accuracy) for clinical prediction and .71 for actuarial prediction on a cross-validation group, with some actuarial instruments scoring as high as .89 on certain populations. Where a person's freedom is at stake, these error rates may still be high, but accuracy is high enough to conclude that predictive dangerousness assessments can potentially offer fact finders much of value.

The factors applied. Slobogin does not much discuss the "necessity" of predictive dangerousness testimony, suggesting that the need for it is great, for few, if any, alternatives beyond the unguided judgment of the fact finder come to mind. Indeed, I took a bit of liberty in describing "necessity" as a factor in Slobogin's approach because he does not so label it, though his extensive use of it as a matter to be weighed in the admissibility judgment reveals its role to be no different than those things he does label "factors." With this background, Slobogin applies his remaining four-part admissibility test, beginning with materiality.

Materiality. There are three likely objections that expert dangerousness testimony is immateri-

al: first, that the studies are based on group average behavior when it is *this individual* who matters; second, that the methodology of the studies used was based on a population excluding others like this alleged offender, thus being inapplicable to this individual; and third, that the predictions rely on factors, such as race, that are not legally cognizable.

Slobogin summarily rejects the first objection, partly because group behavior “can still be directly linked to the individual who is the subject of prediction. It is *John*’s age, prior record, marital status, psychopathic personality traits, and so on that place him in the 76 percent recidivism category.” Moreover, what is offered is merely a probability assessment, not a guarantee, thus being in principle “no different in kind from an assertion that John probably committed the criminal act in the past (even though he either did or did not). . . .”

Slobogin concedes that the second and third objections are weightier. The generalizability of, for example, VRAG, which has been normed only for white subjects, to black subjects may be debatable. VRAG also used a sample population that included not only violent felons but those who engaged in minor misdemeanor assaults, so it is unclear how effective VRAG is in predicting serious violence. But Slobogin does not see these sorts of concerns as meriting exclusion, for adversarial combat can easily expose such weaknesses.

Concerning the predictive value of allegedly legally irrelevant factors, such as race, Slobogin offers a mixed assessment. On the one hand, he is unworried because race specifically is empirically not a good predictor. On the other hand, he points out that compelling state interests—which should include future violence—can justify using otherwise legally suspect categories, and it is behavior, not blameworthiness, that is in question, so no wrongful moral assessment based on such categories will easily tempt the jury toward highly emotional, irrational judgments. Therefore, if there is a category, say, gender, that proves to be a useful part of a future dangerousness prediction, the law should not bar consideration of that factor.

Probative value. Given the studies showing both clinical and actuarial methodologies doing substantially better than chance, Slobogin sees proof of their probative value as easy. They are clearly relevant under Federal Rule of Evidence 401 because they raise the probability of dangerousness beyond what it would be without such evidence (that is, beyond the base rate). Similarly, he

argues that the testimony survives *Daubert* because the error rates are known and can be revealed to the jury, the methodologies are generally accepted in the field of mental health, there are professional standards governing their use, and there has been ample peer review and publication, these being the five *Daubert* factors for judging the reliability of the principles and methods involved in proffered expert testimony. The trial judge’s major task concerning probative value, therefore, is to ensure that the mental health professional is, in fact, relying on those generally accepted assessment techniques that have been found to be highly reliable and to ensure that the fact finder receives clear, complete, and accurate error rate information. Some critics object that, since expert testimony is often the only evidence of dangerousness, it cannot alone be sufficient in criminal cases to meet the beyond a reasonable doubt burden of persuasion. Slobogin responds that in many instances dangerousness predictions occur where the burden is clear and convincing evidence, and methods with accuracy rates in the range of 75 percent should meet this standard. Furthermore, many statutes requiring dangerousness assessments, such as many state commitment laws, define someone as dangerous if he or she is “likely” or “substantially likely” to harm another. If “likely” means roughly 51 percent probability, “then proving beyond a reasonable doubt that a person is dangerous under this definition would require only a 46 percent likelihood (.90 x .51) that the person will harm another.” In any event, Slobogin cautions that this objection wrongly confuses admissibility with sufficiency.

Helpfulness. Nor does Slobogin see helpfulness as a major concern, though he expresses some cautions about clinical testimony. Ideally, a sufficient number of recent studies would be done comparing the accuracy of lay and clinical predictions to determine whether the latter are better than the former by offering evidence that attunes judges and juries to risk factors they would not otherwise have contemplated. If it does, it should be helpful, provided it is based on an intimate knowledge of the relevant research literature. This qualification is no minor one, for mental health practitioners rarely have the necessary research familiarity, though academic social scientists in the relevant research field will. If the expert is properly qualified, his or her testimony likely will be helpful, for jurors are unlikely to be familiar, for example, with such matters as how stress affects the risk of dangerousness for different individuals.

Slobogin has still less of a problem with actuarial and structured professional judgment assessments' helpfulness, for they are clearly rooted in the risk literature and offer probability assessments beyond the ken of laypersons. Slobogin cautions, however, that the testimony should concern only risks and may not be permitted to declare that a person *is* dangerous, committable, or a threat to society, all conclusions expressing a level of certainty that the data do not support.

Unfair prejudice. But Slobogin does have a problem with the risks of unfair prejudice from *clinical* predictions of dangerousness. Unlike at trial, dangerousness predictions, such as at sentencing, occur after a suspect has been convicted of a crime. The suspect loses the presumption of innocence, and fact finders, knowing the suspect has committed one recent crime of violence, are strongly tempted to assume that he or she will do so again. Furthermore, because the dangerousness expert testifies about acts, rather than the more amorphous and normative notion of culpability, fact finders may also give the expert's testimony great weight. Nor are adversarial safeguards likely to work, for research demonstrates that jurors are far more likely to give clinical testimony significant weight, more than it deserves and more than they will give actuarial testimony, even though the latter is clearly more accurate. One study shows that even strong cross-examination and use of an opposing expert "do not shake the influence of a state [clinical] expert willing to pronounce the defendant dangerous," yet indigent offenders often cannot afford to hire their own experts, exacerbat-

ing the problem. A survey of Maryland criminal commitment hearings, almost all of which were contested, found that judges accepted staff predictions of dangerousness in 86 percent of the cases.

By contrast, fact finders are likely, if anything, to undervalue the more accurate forms of prediction testimony, such as actuarial studies and structured professional judgment. At least one study shows that cross-examination of actuarial experts, unlike clinical experts, does raise fact finder skepticism about the testimony. Accordingly, Slobogin would bar prosecutors' first use of clinical dangerousness testimony but not of actuarial and structured professional judgment assessments because adversarial safeguards are only likely to work well with the latter. However, fairness demands that if subjects open the door by offering clinical testimony in their favor, then the state should be free to respond with its own clinical experts, though Slobogin would require defendants to give prosecutors prior notice of the defense intention to rely on a clinician.

Conclusion

I have likely done Slobogin's new book a disservice because I have but scratched the surface of his rich analysis. What I hope I have done is to illustrate how nicely he has married theory and practice. His arguments are persuasive and practical, and the law is likely to move in the direction he suggests over time. Defense attorneys, prosecutors, judges, and legislators who ignore Slobogin's sage advice do so at the peril of being left behind the wave of sensible and incremental law reform that this meaty yet concise volume will slowly help to build. ■