The Prediction of "Dangerous" Behavior: A Continuing Legacy for *Barefoot v. Estelle* (1983)

Ehsan Bokhari and Lawrence Hubert

March 11, 2015

Abstract

The landmark U.S. Supreme Court case of *Barefoot v. Estelle* (1983) held that psychiatric testimony about a defendant's future dangerousness is admissible, irrespective of how unreliable or inaccurate that testimony may be. Under current Texas law, the execution of an individual found guilty of a capital offense may proceed when "there is a probability that the defendant would commit criminal acts of violence that would constitute a continuing threat to society." A jury's affirmative answer to this question of "future dangerousness" is routinely elicited by unstructured clinical testimony from psychiatrists retained by Texas prosecutors. The psychiatrists must respond solely to hypothetical questions about the defendant because any direct examination is effectively prohibited by another landmark U.S. Supreme Court case, *Miranda v. Arizona* (1966). This essay discusses the unfortunate legacy of *Barefoot v. Estelle*; it is still the controlling opinion whenever questions of "future dangerousness" arise, particularly for Texas death penalty cases.

Thomas Barefoot, an oilfield roughneck from New Iberia, Louisiana, fatally shot police officer Carl LeVin on August 7, 1978 in Bell County, Texas. Officer LeVin was in the midst of investigating a nightclub fire set by Barefoot to divert attention from a robbery he had planned to commit. An eyewitness

saw Barefoot emerge from some bushes near the nightclub and shoot the officer in the forehead at point blank range with a small caliber pistol. Barefoot returned to his home in a blood splattered shirt and told his roommates that he needed to leave immediately because he had just "wasted a cop." Later that night, Barefoot was arrested at a bus station, still carrying the gun used to kill officer LeVin.

The murder of a police officer is a capital offense in Texas, and one of five categories of homicide for which the state authorizes the imposition of the death penalty. Up to the present time, a capital murder trial in Texas is bifurcated. The first "guilt" phase establishes just that, guilt or innocence; Barefoot was convicted on November 14, 1978 in the same Bell County where the LeVin murder occurred. The second "sentencing" phase, heard by the same jury that determined "guilt," requires an affirmative response to two questions to trigger the automatic imposition of a death sentence. The first question comes with a foregone affirmative response determined from the "guilt" phase of the trial: whether the conduct causing death was "committed deliberately and with reasonable expectation that the death of the deceased or another would result." The second question is more problematic, particularly given the kind of evidence commonly introduced by the prosecution to elicit an affirmative response. The question asked is whether "there is a probability that the defendant would commit criminal acts of violence that would constitute a continuing threat to society."

As noted, the Texas statute controlling the trial bifurcation requires a jury to determine whether a defendant would pose "a continuing threat to society," and where by law, the term "society" includes the prison system itself. Therefore, a defendant

who could pose a risk to prison staff or other inmates would be eligible for the death penalty. The phrase "a probability" without specifying any further size seems odd to say the least, but Texas courts have steadfastly refused to delimit it any further. So, presumably a very small probability of future violence would be sufficient for the imposition of the death penalty, as long as this small probability could be proven "beyond a reasonable doubt."

In presenting evidence regarding a defendant's future "dangerousness," Texas prosecutors to this day routinely rely on the "expert" testimony of psychiatrists. In the "sentencing" phase of the Barefoot trial, for example, two psychiatrists, Drs. John Holbrook and James Grigson testified for the prosecution. Over defense objection, both were allowed to offer their clinical opinions regarding the ultimate issue of future "dangerousness." These opinions were offered without any direct psychiatric evaluation of the defendant, and were elicited from a hypothetical question posed to the psychiatrists by the prosecutor. The hypothetical question consisted mainly of a listing of the defendant's past anti-social behavior, criminal record, evidence introduced during the trial's "guilt" phase, and the defendant's behavior between the commission of the murder and police arrest.

On the basis of the "facts" stated in the hypothetical question, both Holbook and Grigson gave much the same testimony, which characterized Barefoot as a severe criminal sociopath. Grigson was the more confident and claimed an accuracy of prediction of "one hundred percent and absolute." He diagnosed the defendant as having "a fairly classical, typical, sociopathic personality disorder" of the "most severe category," with a complete "lack of a conscience." For Grigson, this special category

3

of sociopath includes "the ones that ... have complete disregard for another human being's life"; moreover, "there is not anything in medicine or psychiatry or any other field that will in any way at all modify or change the severe sociopath." When asked about future "dangerousness" explicitly, Grigson stated that he was "one hundred per cent" sure that Barefoot "most certainly would" commit future criminal acts of violence, and this would be the case whether Barefoot "was in the penitentiary or whether he was free."

Based on Grigson's testimony regarding future "dangerousness," Thomas Barefoot was sentenced to death by an affirmative response from the jury to the "a probability" question. Barefoot was executed by lethal injection on October 30, 1984, but not before his case for appeal reached the U.S. Supreme Court, and decided in the landmark (and eponymous) ruling in *Barefoot v. Estelle* (1983). This unfortunate opinion let stand the imposition of a death penalty sentence based on highly unreliable and inaccurate psychiatric behavioral prediction; it is still the controlling opinion whenever the prediction of dangerous behavior is at issue, as it often is, particularly in Texas death penalty cases.

The explicit holding in the Supreme Court decision in *Barefoot v. Estelle* (1983) on the prediction of "dangerous" behavior is as follows:

> There is no merit to petitioner's [that is, Barefoot's] argument that psychiatrists, individually and as a group, are incompetent to predict with an acceptable degree of reliability that a particular criminal will commit other crimes in the future, and so represent a danger to the community.

The Court held, in effect, that no matter what the data might show, and for both clinical and actuarial (that is, statistical) prediction, such predictions of future crime can be made at an acceptable level to be of value in the criminal justice system (and in the Texas context, to permit an execution to proceed, as it did for Thomas Barefoot). As we have noted earlier, *Barefoot v. Estelle* is still the controlling opinion whenever issues of behavioral prediction of dangerous or violent behavior come before the court.

Two extensive redactions are given in appendices located at an external web site named at the end of this essay: one is the majority opinion in *Barefoot v. Estelle*, along with an eloquent Justice Blackmun dissent; the second is an *amicus curiae* brief in this same case from the American Psychiatric Association on the accuracy of clinical prediction of future violence. These documents are detailed, self-explanatory, and highly informative about our current lack of ability to make assessments that lead to accurate and reliable predictions of future dangerous behavior.

The psychiatrist featured so prominently in the opinions for *Barefoot v. Estelle* and the corresponding American Psychiatric Association *amicus* brief, James Grigson, played the same role repeatedly in the Texas legal system. For over three decades before his retirement in 2003, he testified when requested at death sentence hearings to a high certainty as to "whether there is a probability that the defendant would commit criminal acts of violence that would constitute a continuing threat to society." An affirmative answer by the sentencing jury imposed the death penalty automatically, as it was on Thomas Barefoot—he was executed in October of 1984, as noted earlier. When asked if he had a last statement to make, he replied:

> Yes, I do. I hope that one day we can look back on the evil that we're doing right now like the witches we burned at the stake. I want everybody to know that I hold nothing against them. I forgive them all. I hope everybody I've done anything to will forgive me. I've been praying all day for Carl LeVin's wife to drive the bitterness from her heart because that bitterness that's in her heart will send her to Hell just as surely as any other sin. I'm sorry for everything I've ever done to anybody. I hope they'll forgive me.

James Grigson was expelled in 1995 from the American Psychiatric Association and the Texas Association of Psychiatric Physicians for two chronic ethics violations: making statements in testimony about defendants he had not actually examined, and for predicting violence with 100% certainty. The press gave him the nickname of "Dr. Death." Most famously, Grigson was culpable in the famous exoneration depicted in the award winning film by Errol Morris, *The Thin Blue Line* (1988). This documentary tells the story of Randall Dale Adams, wrongly convicted of the murder of a Texas police officer; it includes a confession from the real killer, David Harris. Grigson served in his usual role as an expert witness for the original death penalty verdict for Adams. Ron Rosenbaum in his collection of essays, *Travels with Dr. Death* (1991), reports that eventhough Adams was exonerated a year earlier, [Grigson] told him that, despite everything, he *still* [original emphasis] has no doubt about it; the confession by [David Harris] was a sham ... [Grigson] later testified not only that he believed Randall Dale Adams did the killing, but that he was certain Adams "will kill again." As reported in a belated obituary from the *New York Times* (June

25, 2011), Adams lived a quiet life until his death in 2010.

Barefoot v. Estelle has another connection to the distinction between actuarial and clinical prediction, and where the former is commonly better than the latter. There is evidence mentioned in the APA brief that actuarial predictions of violence carried out by statistically informed laymen might be better than those of a clinician. This may be due to a bias that psychiatrists might (unsuspectingly) have in overpredicting violence because of the clients they typically see, or for other reasons related to their practice. There is a pertinent passage from the court opinion (not otherwise given in our web-posted redactions):

> That psychiatrists actually may be less accurate predictors of future violence than laymen, may be due to personal biases in favor of predicting violence arising from the fear of being responsible for the erroneous release of a violent individual. ... It also may be due to a tendency to generalize from experiences with past offenders on bases that have no empirical relationship to future violence, a tendency that may be present in Grigson's and Holbrook's testimony. Statistical prediction is clearly more reliable than clinical prediction ... and prediction based on statistics alone may be done by anyone.

The *Federal Rules of Evidence* and the admissibility of expert witnesses and scientific data was influenced heavily by the case of *Daubert v. Merrell Dow Pharmaceuticals* (1993) that promulgates the *Daubert* standard for admitting expert testimony in federal courts. The majority opinion in *Daubert* was written by Justice Blackman, the same justice who wrote the dissent in

Barefoot v. Estelle. The court stated that Rule 702 of the *Federal Rules of Evidence* was the governing standard for admitting scientific evidence in trials held in federal court (and now in most state courts as well). Rule 702, Testimony by Experts, states:

> If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.

The point of much of the current discussion has been to emphasize that actuarial evidence about future violence involving variables such as age, race, sex, or prior behavior and arrest is all there really is in making such predictions. More pointedly, the assignment of a clinical label, such as "sociopath," adds nothing to an ability to predict, and to suggest that it does is to use the worst "junk science," even though it may be routinely assumed true in the larger society. All we have to rely on is the usual psychological adage that the best predictor of future behavior is past behavior. Thus, the best predictor of criminal recidivism is a history of such behavior, and past violence suggests future violence. The greater the amount of past criminal behavior or violence, the more likely that such future behavior or violence will occur (a behavioral form of a "dose-response" relationship). At its basis, this is statistical evidence of such a likely occurrence and no medical or psychological diagnosis is needed or useful.

Besides the specious application of a sociopath diagnosis to predict future violence, after the Supreme Court decision in *Estelle v. Smith* (1981), such a diagnosis had to be made on the basis of a hypothetical question and not on an actual psychological examination of the defendant. In addition to a 100% incontrovertible assurance of future violence, offering testimony without actually examining a defendant proved to be Grigson's eventual downfall and one reason for the expulsion from his professional psychiatric societies. This prevention of an actual examination of a defendant by the Supreme Court case, *Estelle v. Smith*, also involved James Grigson. Ernest Smith, indicted for murder, had been examined by Grigson in jail, and who determined he was competent to stand trial. In the psychiatric report on Smith, Grigson termed him "a severe sociopath" but gave no other statements as to future dangerousness. Smith was sentenced to death based on the sociopath label given by Grigson. In *Estelle v. Smith* the Supreme Court held that because of the well-known case of *Miranda v. Arizona* (1966), the state could not force a defendant to submit to a psychiatric examination for the purposes of sentencing because it violated a defendant's Fifth Amendment rights against self-incrimination and the Sixth Amendment right to counsel. Thus, the examination of Ernest Smith was inadmissible at sentencing. From that point on, predictions of violence were made solely on hypothetical questions and Grigson's belief that a labeling as a sociopath was sufficient to guarantee future violence on the part of a defendant, and therefore, the defendant should be put to death.

The offering of a professional psychiatric opinion about an individual without direct examination is an ethical violation of the Goldwater Rule, named for the Arizona Senator who ran

for President in 1964 as a Republican. Promulgated by the American Psychiatric Association in 1971, it delineated a set of requirements for communication with the media about the state of mind of individuals. The Goldwater Rule was the result of a special September/October 1964 issue of *Fact:* magazine, published by the highly provocative Ralph Ginzburg. The issue title was "The Unconscious of a Conservative: Special Issue on the Mind of Barry Goldwater," and reported on a mail survey of 12,356 psychiatrists, of whom 2,417 responded: 27% said he was mentally fit; 49% said he was not; and 24% said they did not know enough about Goldwater to answer the question. Much was made of Goldwater's "two nervous breakdowns," because such a person should obviously never be President because of a risk of recurrence under stress that might then lead to pressing the nuclear button.

Goldwater brought a $2 million libel suit against *Fact:* and its publisher, Ginzburg. In 1970, the United States Supreme Court decided in Goldwater's favor giving him $1 in compensatory damages and $75,000 in punitive damages. More importantly, it set a legal precedent that changed medical ethics forever. For an updated discussion of the Goldwater Rule, this time because of the many psychiatrists commenting on the psychological makeup of the former chief of the International Monetary Fund, Dominique Strauss-Kahn, after his arrest on sexual assault charges in New York, see Richard A. Friedman's article, "How a Telescopic Lens Muddles Psychiatric Insights" (*New York Times*, May 23, 2011).

Because of the doctrine of *stare decisis* (Latin for "to stand by things decided"), where earlier Supreme Court decisions are respected and hold precedent, the legacy of *Barefoot v. Estelle*

continues to this day. For example, the American Psychological Association (APA) in 2011 filed a "friend of the court brief" with the U.S. Supreme Court in the case of *Billy Wayne Coble v. State of Texas*. The brief asked the Supreme Court to grant Mr. Coble's petition for *certiorari* (that is, for a review of the lower court decision) based partly on the "unreliability" of testimony about the risk of "future dangerousness" given by a forensic psychiatrist, Dr. Richard Coons. The Supreme Court denied *certiorari* in June of 2011, citing *Barefoot v. Estelle* as justification for the denial. Billy Wayne Coble is currently an inmate on the Texas death row, waiting for his execution date to be set.

The APA brief makes three assertions about predicting future dangerousness, two of which we would affirm and one with which we would disagree and instead assert that the empirical facts would suggest otherwise. The two we would affirm:

> Unstructured clinical testimony like that at issue is not based on science and should not be relied upon to establish future dangerousness.

> Unstructured clinical risk-assessment testimony is unduly persuasive to juries.

The one that we would question:

> In contrast to Dr. Coon's unstructured approach, structured risk-assessment methods are scientifically based and can reliably inform assessments of future dangerousness in a variety of contexts.

What is at issue here is the term "reliably," which we would interpret as meaning "accurately." Given the empirical literature on structured risk assessment instruments (some of which

is cited at the end of this article), and the usual values characterizing the "reliability" of an instrument (that is, the familiar diagnostic test statistics of "sensitivity"' and "specificity") plus the low prior base rate of "dangerous" behavior, a prediction of "dangerous" is generally more likely to be wrong than right; that is, when a prediction of "dangerous" is made, it is *more* likely that the person proves "not dangerous" than "dangerous." This is such an odd situation that it has been called the "false positive paradox." To us, and contrary to the assertion in the APA brief, this indicates an *unreliable* means to "inform assessment of future dangerousness in a variety of contexts."

In choosing between clinical and actuarial prediction strategies, preference should be tilted toward actuarial strategies, if only to avoid the capricious judgment of individuals such as Coons, Grigson, and Holbrook who offer no scientifically justified bases at all for their predictions. But that is not then to say that the various actuarial mechanisms available are at a level of accuracy that could automatically justify their routine use in imposing severe restrictions on personal liberty (or for allowing executions to proceed, as in Texas). One problem is in the ambiguity as to what constitutes "dangerous" behavior and how encompassing the definition is (for example, verbal threats made with some makeshift weapon in hand but with no attendant physical injury or touching, are routinely considered to be "violent"). Possibly, it is always best to remember Sir William Blackstone's adage from the *Commentaries on the Laws of England* (1765): "It is better that ten guilty escape than one innocent suffer."

For a recent and thorough review of the literature on the prediction of dangerous or violent behavior as it relates to the death

penalty, see Michael L. Perlin, *Mental Disability and the Death Penalty: The Shame of the States* (2013; Rowman & Littlefield); Chapter 3 is particularly relevant: "Future Dangerousness and the Death Penalty."[1] A good resource generally for material on the prediction of dangerous behavior and related forensic matters is the Texas Defender Service, and the publications it has freely available at its web site (`www.texasdefender.org`):

A State of Denial: Texas Justice and the Death Penalty (2000)

Deadly Speculation: Misleading Texas Capital Juries with False Predictions of Future Dangerousness (2004)

Minimizing Risk: A Blueprint for Death Penalty Reform in Texas (2005)

The redactions of the majority opinion, the Blackmun dissent, and the APA brief in *Barefoot v. Estelle* are at:

`http://cda.psych.uiuc.edu/barefoot_majority_opinion_`
`blackmun_dissent`

`http://cda.psych.uiuc.edu/barefoot_apa_brief`

There is a large meta-analytic literature on the use of risk assessment instruments that would support our disappointing overall conclusion about the prediction of future "dangerous" behavior, either clinically or actuarially. One of the most comprehensive such studies appeared recently in the Open Access *British Medical Journal (BJM)* on July 24, 2012, by Seena Fazel,

---

[1]Some of the background material on *Barefoot v. Estelle* included in the present essay is adapted from Lawrence Hubert and Howard Wainer, *A Statistical Guide for the Ethically Perplexed* (CRC Press, 2013); in particular, Section 6.6 (The (Un)reliability of Clinical Prediction; pp. 159–171).

Jay P. Singh, Helen Doll, and Martin Grann ("Use of Risk Assessment Instruments to Predict Violence and Antisocial Behaviour in 73 Samples Involving 24,827 People: Systematic Review and Meta-analysis"). The overall conclusion of the study is stated as follows:

> Although risk assessment tools are widely used in clinical and criminal justice settings, their predictive accuracy varies depending on how they are used. They seem to identify low risk individuals with high levels of accuracy, but their use as sole determinants of detention, sentencing, and release is not supported by the current evidence. Further research is needed to examine their contribution to treatment and management.