

STATISTICS IN THE COURTS:
FROM THE HOWLAND WILL TO BRENNAN, ART, AND KIRBY

Murray Gerstenhaber
Department of Mathematics and School of Law
University of Pennsylvania
Philadelphia, PA 19104-6395
mgersten@math.upenn.edu

January 24, 1999

Pol Brennan, Kevin Barry John Artt, and Terence Damien Kirby escaped from Maze prison, just outside Belfast, in September, 1983, and were subsequently apprehended at different times in the United States. The United Kingdom requested their extradition and certifying orders were issued by Judge Charles A. Legge of the Northern District of California. They have now come before Circuit Judges Alfred T. Goodwin, Betty B. Fletcher, and Dorothy W. Nelson of the Court of Appeals for the Ninth Circuit in San Francisco. On October 8, 1998, with Judge Goodwin dissenting, the case is reversed and remanded.

Normally, extradition decisions are not appealable because they are not viewed as decisions of a court. They are made on request of the United States on behalf of a country with which we have a mutual extradition treaty. They come before a judge, justice, or magistrate sitting in chambers, not in open court. The decision is made under special authority granted by statute and normally is final. But there is something unusual about this case. The defendants point to the language of the treaty. In fact, there are two relevant treaties with the United Kingdom, the "1997 Treaty" and a "Supplementary Treaty of 1985". The latter contains an "Aquino Clause", after Benigno ("Ninoy") Aquino who had been extradited to the Philippines and was assassinated as he stepped off the plane in Manila.

The appellate court finds that despite the general non-appealability of orders certifying extradition, it does have the authority to review the interpretation of the clauses of the treaty under which they are made. It finds that the district judge did not clearly err in finding against the appellants with respect to the other clauses of the treaty, but Artt and Kirby cite the Aquino clause as a defense. It provides, in part,

Extradition shall not occur if the person sought establishes to the satisfaction of competent judicial authority by a preponderance of the evidence that the request for extradition has in fact been made with a view to try or punish him on account of his race, religion, nationality, or political opinions

Brennan, Artt, and Kirby had been convicted of various crimes and terrorist acts in a court system established under emergency legislation in the United Kingdom in 1973. These so-called Diplock courts (after the chairman of the commission creating them) had no juries, a single judge, reduced standards for evidence, and in general afforded considerably less protection to defendants than was common in the United Kingdom. The opinion states,

Artt and Kirby claim that their Northern Ireland convictions were based on false confessions obtained under coercive conditions. They argue that they would not have been convicted were it not for anti-Catholic and anti-Republican bias in the Northern Ireland justice system. Accordingly, they submit that the United States should deny extradition on the ground that the requests for extradition have been made in order to punish them on religious and political grounds.

Interpretation of the Aquino clause was a matter of first impression in the Ninth Circuit and no other appellate court had considered it. The district judge, writing on a blank slate, said that

As long as the courts of the United Kingdom applied a basic standard of fundamental fairness in their review of a coerced confession, and as long as those courts allowed a fair hearing on the issue, the court's inquiry must be limited to whether the confession was coerced because of the protected factors and not because of the offenses for which a respondent was arrested.

This, the appellate court agreed, correctly framed the question, but it split on Judge Legge's answer. He had issued an order severely limiting the scope of inquiry into whether Artt and Kirby's original convictions were to punish them because of protected conduct. All evidence, he said, had to be "directly related to the respondents" and could not involve "generalized enquiries into [the] Diplock court system." So the appellants were not allowed to develop any evidence of systemic bias. Judges Fletcher and Nelson found this too narrow. They said that statistical evidence of systemic bias should have been admitted. Judge Goodwin's dissent likened the present case to a death penalty case and said that "...I cannot see where generalized evidence of systemic bias could outweigh in this case the overwhelming particularized evidence of [the appellants'] guilt."

The split opinion reflects some long-standing tensions between deciding a case on the basis of what probably happened and deferring to public and social policy considerations. Courts and judges have limited political capital to spend.

Sometimes the problem seems unsolvable by courts. Charles R. Nesson, a professor of evidence at Harvard (cf. Jonathan Harr's "A Civil Action") poses the hypothetical "Case of the Blue Bus".

It is really *Smith v. Rapid Transit, Inc.*, 58 N.E.2d 755 (1945). Betty Smith was driving west on Main Street, Winthrop, Mass. at about 1 am on February 6, 1941. A speeding bus coming towards her forced her to swerve into a parked car. She sued the only bus company franchised to operate on Main Street; their time table showed that one of their buses would have passed the point of the accident at about the time that it occurred. The case was not allowed to go to the jury. On appeal the Massachusetts Supreme Judicial Court said,

The most that can be said of the evidence in current instance is that perhaps the mathematical chances somewhat favor the proposition that a bus of the defendant caused the accident. This was not enough.

Nesson supports the decision. Applying Savage's theory of subjective probability, let's suppose that at some future date we will learn precisely what happened on that February night, and that we are forced now to wager either that Betty Smith is right or wrong. What odds would it take for you to bet against her? If it is anything more than even money, then you are saying that she is more likely right than wrong, which is what "preponderance of evidence" normally means. But Nesson says, Cases of naked statistical proof present the most provocative example of probable verdicts that are unacceptable. The statistical nature of the evidence precludes both acceptance of the verdict against the defendant and internalization of the underlying norms.

Nesson accepts the decision in *Sindell v. Abbott Laboratories*, 26 Cal. 3d 588 (1980), a class action suit in which the plaintiff "DES daughters" got cervical cancer because their mothers had taken DES during pregnancy. There were many manufacturers and no one knew who took whose product. All manufacturers were held liable in proportion to their share of the DES market. Nesson distinguishes this because all the manufacturers were at fault. However he also accepts *Summers v. Trice*, 33 Cal.2d 80 (1948). There, two of three hunters mistook the third for game and fired at him, one pellet hitting him in the eye. Only one caused the injury and it was impossible to know which. The court was faced with the choice of holding neither liable or both. It made the wiser choice of holding both, on the theory that they were engaged in a common enterprise. In a sense so are all bus companies and even all who use the roads. In many states even those with perfect driving records must pay into a state fund which compensates victims for accidents caused by negligent uninsured motorists.

Before returning to Artt and Kirby, it may be useful to review some of the history of statistics in the courts. *Robinson v. Mandell* (C.C.D. Mass. 1868) is probably the first case in which probabilities are computed in a legal proceeding. It was suspected that Hetty Robinson (later Hetty Green the "Witch of

Wall Street") had forged her aunt Sylvia Ann Howland's signature on the third page of the latter's will. Benjamin Peirce testified to the extreme improbability of the signature being genuine. His binomial model may not have fit the data too well and his numerical computations are in error but his conclusion is almost certainly correct. (See Meier and Zabell, *Benjamin Peirce and the Howland Will*, 75 J. Am. Stat. Assoc. 497 (1980).) Peirce's statistical evidence was probably admitted only because he was a renowned Harvard professor and the greatest American mathematician of the time. The reception of statistics in American courts in subsequent years has been highly uneven.

Curiously, three decades earlier law had made a tremendous contribution to statistics. Siméon Denis Poisson (1781-1840) discovered the Poisson distribution while predicting the change in acquittal rates in the Cours d'Assizes that resulted from a recent change in its rules (*Récherches sur la Probabilité des Jugements*, 1837). It was many years, however, before the Poisson distribution played a role in any U. S. case, and in one of its most celebrated appearances, *People v. Collins*, 438 P.2d 33 (1968) it is misapplied.

In this country, the most important statistical issues in criminal law have centered on disparities in the selection of minorities for grand and petit juries. The Federal courts seem to have no problem when there is so clear a showing of purposeful discrimination that calculating probabilities is almost irrelevant. Total exclusion of black citizens as jurors by state statute was held unconstitutional in *Strauder v. West Virginia*, 100 U.S. 303 (1880). Notwithstanding, they continued to be excluded by selection processes which, although facially constitutional, at best permitted only token numbers to serve. An effective attack on tokenism did not begin until after World War II. In *Avery v. Georgia*, 345 U.S. 559 (1953) a black defendant was convicted by a jury selected from an all-white venire of sixty that had been drawn from a pool in which 5% were black. In overturning the conviction, Justice Frankfurter, speaking for the Supreme Court said that such an outcome could not be mere chance. (In fact, if the pool were exceedingly large, the probability of an all-white venire could be as large as .046. Nevertheless, the small number of blacks in the pool and the method of selection certainly indicate that blacks were being systematically excluded.)

The first explicit calculation of binomial probabilities in a jury case seems to have been in *Whitus v. Georgia*, 385 US 545 (1967). Blacks listed in the tax digest constituted 27.1% of the taxpayers, but only 9.1% of those on the grand jury venire. Only 7 blacks appeared on a venire of 90 (and none in the petit jury). Citing a law journal article, the court took the probability to

be .000006, but purposeful discrimination would have been obvious even to the innumerate. The court did some binomial calculations again in *Alexander v. Louisiana*, 405 U.S. 625 (1972), getting a probability of "one in 20,000" that the observed event was due to chance.

Finally, in the landmark case of *Castaneda v. Partida*, 430 U.S. 483 (1976), the Supreme Court admitted a showing of statistical disparity using standard deviations. Rodrigo Partida had been indicted in March, 1972, in Hidalgo County, a border county of southern Texas, for burglary of a private residence at night with intent to rape. The grand jury that indicted him had a majority of Spanish surnamed members, as did the petit jury that convicted him and the presiding judge. But over the 11-year period from 1962 through 1972 the average percentage of Spanish surnames amongst the total of 870 summoned to serve as grand jurors was only 39% while the 1972 census showed that 79% of the county were Mexican Americans. The court, in a footnote, enunciated the "Castaneda rule".

As a general rule for such large samples, if the difference between the observed number and the expected value is greater than two or three standard deviations, then the hypothesis that the jury drawing is random would be suspect to a social scientist.

The actual difference was 29 standard deviations, and the probability of chance occurrence less than 1 in 10^{140} . The dissenters' argument was that in the previous line of cases tokenism was obvious, while here many Mexican Americans served as grand jurors. Distinguishing *Alexander*, Chief Justice Burger, joined by Justice Rehnquist, said "... the challenger's venire [in *Alexander*] included only one member of the identifiable class and the grand jury that indicted him had none." Partida had not shown, they claimed, that the statistical disparity was not simply due to a difference in eligibility rates. That issue had in fact been addressed in a footnote to Justice Blackmun's majority opinion. He argued that the selection system left openings for prejudice and that the challenger's statistics had been sufficiently strong to shift to the state the burden of proof that there had been no prejudice. Here is the beginning of another source of tension over the use of statistics in the courts. A standard counter to a statistical study is to point to a neglected variable. It is virtually impossible for the side presenting the statistics to show that all variables have been taken into account. At what point must the burden shift?

The Fifth Circuit and Eleventh Circuits (split from the Fifth) seem to have effectively negated *Castaneda* within their borders by deliberately focusing on absolute disparity in percentages rather than on their statistical significance. The index case appears to be *U.S. v. Maskeny*, 609 F.2d 183 (1980), where the

court said

It is true that the Supreme Court in *Castaneda* discussed disparity in standard deviations but it seems clear to us that the court in that case based its holding on an absolute disparity of 40 percent between the population of Mexican-Americans in the community and the percentage summoned for jury service[.]

The *Maskeny* court claimed its authority from *Duren v. Missouri*, 439 U.S. 357 (1979). There a jury was successfully challenged on the grounds that women served only if they volunteered to do so, as a result of which women constituted only 15% of jurors but more than half the population. Rehnquist dissented. *Duren* hardly supports *Maskeny* and its Fifth and Eleventh Circuit progeny, e.g. *U.S. v. Bautista*, 776 F.2d 1509 (Eleventh Circuit, 1985). There the proportion of blacks eligible for jury service in the population of the Miami Division was taken as 18.82% and the percentage on the qualified wheel as 12.146%. Since the absolute disparity was only 6.674%, the court did not find it necessary to make the State of Florida explain why more than one out of every three eligible blacks was denied an opportunity to serve.

Elsewhere, when a specific group, even though quite small, has been excluded in violation of the Jury Selection and Service Act of 1968, 18 U.S.C Sections 1861-78, convictions were overturned. In *U.S. v. Calabrese, et al.* 942 F.2d 218 (Third Circuit, 1991), all potential jurors who in their returned questionnaires simply indicated that they knew any defendant were automatically excused, without any questioning that might have revealed actual bias. The result was the wrongful exclusion of 12 from a pool of approximately 300. The court held, despite the small percentage excluded, that this was not a mere technical violation of the act but "...the de facto creation of a new category of exclusions." The court also noted that the Fifth Circuit upheld a conviction in a similar case, but was the only circuit to do so.

The cases cited so far concern flawed systems which generated wholesale exclusions. Shouldn't it be sufficient for the defendant to show that the process was flawed in his specific case, or must he actually show that the system is flawed? When fighting peremptory challenges, showing that the system is flawed is almost impossible because prosecutors do not memorialize their prejudices. Nevertheless it was demanded in *Swain v. Alabama* 380 U.S. 202 (1965). This was partially reversed in *Batson v. Kentucky* 476 U.S. 79 (1986). At *Batson's* trial for second-degree burglary and receipt of stolen goods, the prosecutor used his peremptory challenges to strike all four black persons from the venire, leaving an all-white jury. *Batson's* defense successfully moved to discharge the jury by showing that of two jurors who seemed to have comparable qualifications and attitudes, the black one was struck and the white one seated. Burger and Rehnquist

dissented. Later, in a Ninth Circuit case with Judge Nelson writing for a unanimous panel including Judges Fletcher and Harry Pregerson, a new trial was ordered where the appellant simply showed that of two prospective jurors with comparable qualifications and attitudes, the black one was peremptorily struck and the white one seated, *Turner v. Marshall*, 121 F.3d 1248 (1997).

The increasing safeguards in the jury selection process have been publicly accepted because it is essential to a justice system that one be able to trust the verdicts. When the question is not whether the verdict is fair but whether the punishment fits the crime there is great disagreement. In this arena, the question of whether the burden has shifted in a statistical battle, and therefore, in effect, of who prevails, has been determined more by matters of policy than by mathematics. Warren McClesky, a black man, was convicted of murdering a white police officer who responded to a silent alarm during the robbery of an Atlanta furniture store. He was sentenced to death. He challenged the sentence on the grounds that blacks were unfairly sentenced to death in Georgia much more frequently than whites. (The raw numbers showed that killers of white victims were sentenced to death in 11% of the cases but killers of black victims in only 1%.) The District Court rejected the massive Baldus statistical study, *McClesky v. Zant*, 580 F. Supp. 338 (N.D. Georgia 1984). It claimed, amongst other things, that there were substantial flaws in the data base, a lack of sufficient predictive value in the experts' models to justify a claim of discrimination, and that being a white victim might be a surrogate for unaccounted-for aggravation in Baldus' multiple regression model. The Court of Appeals did not deal with the asserted flaws but said that even if the results were valid, they were insufficient to justify upsetting the system, *McClesky v. Kemp*, 753 F.2d 877 (11th Circuit, 1985). It argued that "The very exercise of discretion means that persons exercising discretion may reach different results from exact duplicates." The Supreme Court affirmed by a 5-4 decision. The nub of the court's answer to McClesky's statistics is contained in the following excerpt from Justice Powell's opinion.

Thus, the application of inference drawn from the general statistics to a specific decision is not comparable to the application of an inference from general statistics to a specific venire-selection or Title VII case. In those cases, the statistics relate to fewer entities, and fewer variables are relevant to the challenged decisions.

In footnote 15, Powell explains his use of "entities" and "variables".

We refer here not to the number of entities involved in any particular decision but to the number of entities whose decisions necessarily are reflected in a statistical display such as the Baldus study. The decisions of

a jury commission or of an employer over time are fairly attributable to the commission or employer. Therefore, an unexplained statistical discrepancy can be said to indicate a consistent policy of the decisionmaker. The Baldus study seeks to deduce a state "policy" by studying the combined effects of the decisions of hundreds of juries that are unique in their composition.

Reading this one may ask, Do the effects of systemic prejudice pass constitutional muster simply because its agents vary?

This brings us back to Artt and Kirby. By all accounts, they were not peaceable people. Brennan had been arrested on a street in Belfast carrying a revolver, but his companion, Ann Marie Quinn was carrying a bomb containing 23 pounds of explosives wired to a detonator, battery, and watch. She testified that they had been on a bombing mission for the Irish Republican Army. Kirby had been convicted of possession of an explosive device, of a submachine gun, assault, imprisonment, and felony murder. Artt was convicted of murdering a prison official. Nevertheless, it was clear that the Diplock system that had incarcerated them had substantial procedural flaws. Brennan admitted his guilt, so the issue with respect to Artt and Kirby became, does this case more closely resemble *McClesky* or *Castaneda*? The dissenting Judge Goodwin, citing the quote from Powell, says,

I believe that Artt and Kirby's cases resemble a death penalty case, where generalized statistical evidence has been deemed inadmissible, more than they resemble venire-selection or Title VII cases.

By contrast, Judges Fletcher and Nelson, quoting employment discrimination cases and *Turner* (on which they sat) say The United States Supreme Court has made it "unmistakably clear" that more generalized statistical analyses "play an important role" when the existence of discrimination is in dispute.

They send the case back to District Judge Legge, in whose hands it rests at this writing.

In concluding, I want to mention briefly four cases, two presided over by Judge Jack B. Weinstein, and two coming from salary and tenure disputes in academia. The first Weinstein case is actually a series of four "Shonubi" cases, *U.S. v. Shonubi I*, 802 F. Supp 859 (E.D.N.Y. 1992), *Shonubi II*, 998 F.2d 84 (2d Cir. 1993). *Shonubi III*, 895 F. Supp. 460 (E.D.N.Y. 1995), and *Shonubi IV*. 103 F.3d 1085 (2d Circuit, 1997). Charles O. Shonubi, a 34-year old citizen of Nigeria, was arrested on December 10, 1991, at John F. Kennedy airport in New York. An X-ray revealed foreign bodies in his intestine. He subsequently passed 103 balloons containing 53% pure heroin totaling 427 grams. This was his eighth trip from Nigeria. The statistical issue going back and forth between Judge Weinstein and the Second Circuit was how much heroin, in total, Shonubi had likely brought into the United States in all his trips; it would determine his sentence. For an

excellent analysis I can recommend a term paper by a former student, Darren Tucker: *The Appointment of Expert Statisticians under Rule 706: A Case Study of United States v. Shonubi*.

The second Weinstein case is *Geressy et al. v. Digital Equipment Corporation* 980 F.Supp. 640 (E.D.N.Y. 1997). Judge Weinstein applies *Castaneda* to determine whether damages previously awarded to women suffering repetitive stress injuries from using Digital's keyboards were excessive. He makes an extensive review of past damage awards and finds that the present awards are within two standard deviations of the mean of comparable preceding ones. This, he says, satisfies *Castaneda*'s "two or three standard deviations" rule.

The academic cases are *Presseisen v. Swarthmore College*, 442 F. Supp. 593 (E.D. Pa. 1977) and *Ottaviani et. al. v. State University of New York at New Paltz et al.*, 875 F.2d 365 (1989). Dr. John deCani, former chair of the Statistics Department at the University of Pennsylvania appeared for Barbara Presseisen, and Dr. Mary Gray, organizer of this symposium, appeared for the plaintiffs in *Ottaviani*. In both cases the issue was one of discrimination against women in matters of promotion and salary. Despite impressive evidence, none prevailed. The opinions are couched mainly in challenges to the statistical proof, but as in *McClesky*, I believe that policy considerations overrode the evidence. Had Barbara Presseisen won a tenured position, Swarthmore, a small liberal arts college, would have been obligated to employ her for life. If Swarthmore did discriminate against women, was Barbara Presseisen the appropriate person to reward with tenure because she was the one who brought this suit? The case might have been settled for monetary damages, but I think that this question, which was never explicitly raised, was always present. Similar policy questions, complicated by the fact that the institution being sued was a state university, may have led to the outcome in *Ottaviani*. Their analysis is best left to Dr. Gray.