

# Law and the Crystal Ball: Predicting Behavior with Statistical Inference and Individualized Judgment

Barbara D. Underwood\*

Important benefits and burdens are distributed in American society on the basis of predictions about individual behavior. Release from prison, places in schools, jobs, and retail credit are among the benefits distributed to those applicants who are found most likely to succeed. The effort to predict an applicant's behavior can be made in a variety of ways: by professional experts or ordinary laymen, by use of individualized judgment or formulas that assign fixed weights to predetermined characteristics of the applicant. No matter what method is used, it typically generates controversy.

This controversy is expressed in policy debates over the fairness or wisdom of choosing a particular method for selecting applicants. It also appears in litigation challenging a selection system on the ground that it violates some constitutional or statutory requirement. When the decisionmaker is a government agency, such as the parole authority or a public school, then the choice of a selection system is plainly a matter of public concern. As a political matter it involves the allocation of public resources, and as a legal matter it is subject to the requirements of fairness contained in the due process and equal protection clauses of the United States Constitution. But even when the decisionmaker is a private institution, such as a private employer or lender, its practices are often subject to public scrutiny and legal control. Many private decisionmakers are prohibited by law from discrimination on the basis of race, sex, and various other attributes. Enforcement of that prohibition requires the decisionmaker to respond to claims of illegal discrimination, by explaining his selection system, and thereby exposing it to public scrutiny. That exposure in turn generates a demand for new legal restraints on methods of selection.

In order to decide whether new legal restraints should be imposed, or even simply to choose a fair and effective selection system, it is

\* Professor of Law, Yale Law School.

## Prediction

necessary to identify clearly the advantages and disadvantages of the various available techniques for selection. The political and legal controversy over the various prediction methods in current use reflects considerable confusion over the differences among methods, and particular confusion over the nature of prediction methods based on statistically validated formulas or scoring schemes.

This article examines the tangled objections and ambivalent reactions to particular prediction methods, and to the process of prediction itself. It identifies the unarticulated values built into prediction schemes, and also the unarticulated values called into play by various objections. The resulting analysis demonstrates that important uses exist for predictive selection schemes, and that different circumstances may compel the use of one type of predictive scheme rather than another. The article presents criteria for choosing the system most appropriate to a particular factual context. Part I considers questions about the basic decision to predict: the decision to select applicants on the basis of their predicted behavior instead of their past merit, or their need, or their luck in a random draw. Part II considers questions about the choice of a predictive method: whether to rely on individualized judgment, or to combine objective data about applicants according to a statistically validated formula. Part III considers the questions that arise, if a statistically validated formula is used, concerning the choice of the particular items to be used in constructing that formula.

### I. The Decision to Predict

Two main themes dominate the controversy over predictive techniques of selection: questions of accuracy and questions of legitimacy. Critics contend that the techniques are not sufficiently accurate to justify their use, and that in any event their use is not legitimate in some particular context because it is inconsistent with respect for the autonomous individual.

One example of this controversy arises over the use of prediction to fix the period of imprisonment for a convicted criminal. The prevailing system for sentencing criminals relies on prediction at two different points. First, the sentencing judge typically fixes a period of imprisonment based partly on a judgment about the chance that the offender will commit further crimes. This criterion is seldom made explicit in the statutes, but it rests on the common understanding that one important function of imprisonment is incapacitation: the prevention of those crimes that would have been committed during the period of

imprisonment.<sup>1</sup> An offender who is thought likely to commit further crimes will for that reason be sentenced to a longer term than a similar offender who is thought less likely to do so. The effort at prediction enters the process again when the prisoner is considered for release to the supervised status of parole. Typically one of the most important prerequisites for parole is a finding that the prisoner is likely to avoid further crimes while at liberty.<sup>2</sup>

Critics have challenged both the accuracy and the legitimacy of prediction in the context of sentencing and parole. Most critics contend that predictions of criminal behavior are not accurate enough to use as a basis for a decision about a matter as important as liberty.<sup>3</sup> Indeed, existing techniques for parole prediction are highly inaccurate by any measure.<sup>4</sup> First, available methods are severely limited in their capacity to distinguish persons who will succeed from those who will fail.<sup>5</sup>

1. Sentencing is also influenced by other commonly recognized functions of punishment, such as deterrence, rehabilitation, and the expression of the moral condemnation of the community. For a discussion of the functions of criminal punishment, see, e.g., H. PACKER, *THE LIMITS OF THE CRIMINAL SANCTION* 35-61 (1968); Hart, *The Aims of the Criminal Law*, 23 *LAW & CONTEMP. PROB.* 401 (1958). Judges in general have great discretion in assigning weight to these various functions, and resolving conflicts among them. For an influential discussion of the resulting chaos, and a plea for reform, see M. FRANKEL, *CRIMINAL SENTENCES* (1972).

2. E.g., 18 U.S.C. § 4206(d) (1976) (prisoner may not be released if there is "reasonable probability that he will commit any Federal, State, or local crime"); CONN. GEN. STAT. § 54-125 (1977) (release requires "reasonable probability that such inmate will live and remain at liberty without violating the law"); 7 N.Y. CODES, RULES AND REGS., ch. XII, § 1910.5 (1974) (release requires "reasonable probability that the inmate will not violate the law"); see R. DAWSON, *SENTENCING* 263-78 (1969) (American Bar Foundation study of sentencing and parole practices in three midwestern states).

3. See N. MORRIS, *THE FUTURE OF IMPRISONMENT* 62-73 (1974); A. VON HIRSCH, *DOING JUSTICE* 19-26 (1976) (Report of Committee for the Study of Incarceration); Monahan & Monahan, *Prediction Research and the Role of Psychologists in Correctional Institutions*, 14 *SAN DIEGO L. REV.* 1028, 1029-30 (1977) (reporting controversy over accuracy of prediction of criminal behavior).

4. One way to describe the accuracy of a predictive device is to use the percentage of predictions that prove correct. That percentage will depend not only on the device, but on the true distribution of successes and failures in the population under study. For example, a device that always predicts success will be 100% accurate on a population that contains only people who will succeed; the accuracy of the device will decline with the proportion of true successes in the population.

Other measures of accuracy attempt to describe the discriminating power of the device in a manner that is independent of the population. Such measures are constructed from two component factors: the proportion of successes accurately identified by the device, and the proportion of failures accurately identified by it. If the device is more accurate in identifying successes than in identifying failures, then its total observed accuracy will increase and decrease with the number of true successes in the population. For discussion of various measures of accuracy, see F. SIMON, *PREDICTION METHODS IN CRIMINOLOGY* 16-29 (1971); Duncan, Ohlin, Reiss & Stanton, *Formal Devices for Making Selection Decisions*, 58 *AM. J. SOC.* 573 (1953); Ohlin & Duncan, *The Efficiency of Prediction in Criminology*, 54 *AM. J. SOC.* 441 (1949).

5. It is difficult to obtain the information necessary to characterize the accuracy of a predictive technique designed for parole decisions, because a prediction of failure usually

## Prediction

Second, under available methods, predictions of failure typically carry a much higher error rate than predictions of success.<sup>6</sup> Even when predictions of failure are made very conservatively, the high-risk group identifiable by present methods has a failure rate that may strike some as too low to use as a ground for extending incarceration.

The limited accuracy of existing parole prediction devices supports

leads to incarceration, preventing anyone from discovering whether the prediction was correct. Nevertheless, the information can sometimes be obtained or approximated, either by deliberate experimentation or by observations of people released on the basis of criteria different from those being tested. Even prisoners who are predicted to fail are released at the expiration of sentence, and hence they can be observed, like parolees, to see whether they commit new crimes.

In this way some data on the performance of various parole prediction techniques has been collected. See, e.g., F. SIMON, *supra* note 4 (reviewing prior studies and reporting new one); Hoffman, *Mandatory Release: A Measure of Type II Error*, 11 CRIMINOLOGY 541 (1974); Ohlin & Duncan, *supra* note 4. The studies suggest that available techniques often do less well than a policy of predicting success for everyone. This is in part because parole failure is a relatively rare event in many prisoner populations, and failure in the form of serious violent crime is even more rare. Therefore, a policy of predicting success for everyone would be fairly accurate, and can be bettered only by a quite powerful prediction technique. See generally Meehl & Rosen, *Antecedent Probability and the Efficiency of Psychometric Signs, Patterns, or Cutting Scores*, 52 PSYCH. BULL. 194 (1955), reprinted in P. MEEHL, *PSYCHODIAGNOSIS: SELECTED PAPERS* 32 (1973) (examining methods for evaluating validity of predictive psychometric devices) [hereinafter cited as *PSYCHODIAGNOSIS*].

The task of comparing predictive techniques is complicated by the fact that not all studies use the same definitions of success and failure. In the context of sentencing and parole, failure may be defined as any violation of parole conditions, any crime committed after release, or any violent crime committed after release. A failure may be counted if it occurs within six months of release, within one year, or within some longer period. A failure may be established by conviction, parole revocation, or independent evidence.

6. See, e.g., Kozol, Boucher & Garofalo, *The Diagnosis and Treatment of Dangerousness*, 18 CRIME & DELINQUENCY 371, 389-92 (1972) (61.3% error rate among predictions of failure, 8.6% error rate among predictions of success); Wenk, Robinson & Smith, *Can Violence Be Predicted?* 18 CRIME & DELINQUENCY 393, 394-96 (1972) (86% error rate among predictions of failure, 5% error rate among predictions of success). The higher error rate accompanying predictions of failure occurs partly because parole failure is less common than parole success. If a prediction device is equally accurate in identifying successes and failures, then its errors will be evenly distributed over the population. The erroneous predictions will contain succeeders in proportion to their numbers in the population. It follows, then, that in a population that contains mostly succeeders, erroneous predictions will involve mostly succeeders, and hence will be mostly erroneous predictions of failure.

A better accuracy rate can sometimes be obtained for the predictions of failure in these circumstances by changing the cut-off score on a selection device, tilting the decision toward prediction of success, and thereby increasing erroneous predictions of success. But even this possibility has not enabled anyone to predict parole failure with an accuracy rate higher than about 80%. That is, even among those identified as most likely to fail on parole, some 20% can be expected to succeed. See, e.g., Hoffman & Beck, *Salient Factor Score Validation—A 1972 Release Cohort*, 4 J. CRIM. JUST. 69, 71 (1976) (80% failure rate among people with score of zero on federal parole board scoring scheme in 1972, 75% failure rate among people with score of zero in 1970); Wolfgang, *Crime in a Birth Cohort*, 117 PROC. AM. PHILOSOPHICAL SOC'Y 404 (1973), reprinted in 3 CRIME AND JUSTICE 161 (L. Radzinowicz & M. Wolfgang 2d rev. ed. 1977) (70-80% recidivism rate among people who have three prior criminal offenses) [hereinafter cited as Wolfgang, with page references to 3 CRIME AND JUSTICE].

a telling argument against the use of any one of them. But it hardly counts against an effort to refine the methods of prediction, and to reduce the number and distribution of errors to some tolerable level. Nevertheless, several critics have explicitly rejected the effort to improve the accuracy of parole prediction,<sup>7</sup> and it seems likely that others, who now confine their attacks to the limitations of existing techniques, in fact object more fundamentally to the enterprise. In other fields, too, the challenge to the use of prediction seems to rest partly on a concern about accuracy and errors, and partly on some more fundamental ground.<sup>8</sup>

One reason to reject efforts to improve prediction may be grounded in the view that the limitations in existing techniques are inherent in the nature of prediction. The argument is that a prediction of future behavior is necessarily less accurate than any other determination that might be the basis for decision. Some may trace this inherent limitation to a distinction between future fact and present or past fact: the claim would be that for a question of past or present fact there is in principle a true answer, and any errors in factfinding are attributable only to inadequacies of evidence or of judgment. For a question of future fact, however, it may be that no amount of evidence or judgment can make the answer accessible.<sup>9</sup>

7. Professor Andrew von Hirsch explicitly considers the possibility of developing a predictive technique with an error rate that is no greater than the rate at which the existing system of criminal trials convicts innocent persons. Von Hirsch, *Prediction of Criminal Conduct and Preventive Confinement of Convicted Persons*, 21 BUFFALO L. REV. 717, 744-50 (1972). If such a technique were developed, he would oppose its use on the ground that it violates "basic concepts of individual liberty," *id.* at 745, and that it entails "unjustified risks of abuse," *id.* at 748. A similar point is made by Ronald Dworkin, though less elaborately; he argues that it is unjust to imprison someone on the basis of an actuarial prediction, however accurate, "because that denies his claim to equal respect as an individual." R. DWORKIN, *TAKING RIGHTS SERIOUSLY* 13 (1977).

8. See, e.g., J. GOLDSTEIN, A. FREUD & A. SOLNIT, *BEYOND THE BEST INTERESTS OF THE CHILD* 49-52 (1973) (criticizing efforts to predict successful parenting for child-custody determinations).

9. For a discussion of various philosophical views about the difference between past and future facts, see Sellars, *Time and the World Order*, in 3 MINNESOTA STUDIES IN THE PHILOSOPHY OF SCIENCE 527 (H. Feigl & G. Maxwell eds. 1962). See generally B. DE JOUVENEL, *THE ART OF CONJECTURE* (1976) (exploring nature of forecasting techniques).

Courts and legal commentators have noted in various contexts that prediction of future events is an unusually difficult form of factfinding. See, e.g., *Zablocki v. Redhail*, 434 U.S. 374, 402 n.4 (1978) (Powell, J., concurring) (prediction that children will not become public charges unacceptable as prerequisite for issuing marriage license to parent; statute authorizes standardless discretion and hence raises serious question of procedural due process "in light of the hazards of prediction in this area"); *State v. Turner*, 556 S.W.2d 563, 566 (Tex. 1977), *cert. denied*, 435 U.S. 929 (1978) (reasonable-doubt standard should not apply to civil commitment proceedings because prediction of dangerous conduct cannot be made with same certainty as determination of past fact); Black, *Due Process for Death: Jurek v. Texas and Companion Cases*, 26 CATH. U.L. REV. 1, 4-8 (1976) (prediction of future conduct less suitable than past fact as criterion for death penalty); Broun &

## Prediction

As a practical matter, however, the distinction between future facts and past or present facts is unsound. Some past or present facts are as elusive as any prediction, and some predictions can be made with as much confidence as most determinations of past fact.<sup>10</sup> If accuracy is the chief concern, it should be possible to specify the level of error that is tolerable in the decision to imprison an offender, and hold all fact-finding to that standard, whether it involves matters of past or future fact.<sup>11</sup> Perhaps no available predictive method is sufficiently accurate to satisfy the high standard of accuracy appropriate for the decision to incarcerate, but it may still be possible to develop one.

Moreover, in contexts other than parole, the requisite standard of accuracy may be more easily met. For example, the attempt to predict school success or failure from test scores is notoriously limited in accuracy.<sup>12</sup> Nevertheless, the decision to reject applicants from school demands a lower standard of accuracy than the decision to deny pris-

Kelly, *Playing the Percentages and the Law of Evidence*, 1970 U. ILL. L.F. 23, 32-34 (probability estimates should be admissible for predictions, but not for questions of past fact, for which better evidence is available); Note, *Judicial Review of Facts in Informal Rulemaking: A Proposed Standard*, 84 YALE L.J. 1750, 1763-64 (1975) (judicial review of agency determinations should be more deferential with respect to predictions than with respect to findings of past fact).

10. For example, past states of mind are notoriously difficult to determine, and it is relatively easy to determine the amount of interest that will be paid by a bank on a deposit. A few courts and commentators have noticed that it is unsound to draw a distinction between the difficulties of predicting future events and the difficulties of discovering the truth about past events. See, e.g., *Murel v. Baltimore City Criminal Court*, 407 U.S. 355, 364 (1972) (Douglas, J., dissenting from dismissal of certiorari); Fiss, *The Jurisprudence of Busing*, 39 LAW & CONTEMP. PROB. 194, 211-12 (1975); Tribe, *Trial By Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1344-46 (1971); Note, *Evidential Use of Mathematically Determined Probability*, 28 HARV. L. REV. 693, 695-96 (1915).

11. This approach to the problem is difficult, however, if there are moral or other objections to making explicit the level of error that is tolerable in the decision to incarcerate. Cf. Tribe, *supra* note 10, at 1372-75 (moral objection to inviting juries to quantify probability of guilt and associated risk of error in criminal cases).

12. The Scholastic Aptitude Test (SAT) of the College Entrance Examination Board is designed for use in college admissions to predict college performance. The Board claims that SAT verbal scores give an applicant's "true score" with a "standard error of measurement" of 32 points. This means, *inter alia*, that applicants with the same "true score" are likely to show a difference in test scores of up to 64 points; applicants who differ by that amount may indeed perform equally well. Similarly, the measurement error for SAT mathematics scores is 35, yielding a probable range in test scores of up to 70 points. Furthermore, SAT scores do not predict college performance after the freshman year as well as they predict freshman-year grades. Although the probability is that the lower-scoring applicant will perform less well in college than his higher-scoring competitor who is still within his "standard error of measurement" range, there is significant chance that the true performance order will be reversed. Thus if some performance level is defined as "success" and used as an admissions criterion, then some applicants whose performance predicted by the SAT falls below that level will be erroneously excluded. See Reinhold, *What the Test Scores Do and Don't Say about a Child*, N.Y. Times, May 1, 1977, § 12, at 18, col. 1.

oners parole, and the limited accuracy of test predictions may satisfy that standard.

The use of predictive criteria for selection is subject to challenge not only on grounds of accuracy, however, but also on the ground that it conflicts with other important social values, involving respect for individual autonomy. The attempt to predict an individual's behavior seems to reduce him to a predictable object rather than treating him as an autonomous person. It is this apparent conflict, between predicting behavior and respecting autonomy, that leads some critics to argue that a prediction of individual behavior, however small the risk of error, is an inappropriate criterion for the selection of individuals for imprisonment, and for various other benefits and burdens as well. To imprison a person because of the crimes he is expected to commit denies him the opportunity to choose to avoid those crimes. Similarly, to reject a job applicant on the basis of a test score that predicts poor performance denies him the opportunity to outperform the prediction. The argument is that respect for individual autonomy requires recognition of the possibility that an individual can choose to refute any prediction about himself.

In principle, of course, predictive decisionmaking makes explicit allowance for the fact that some individuals would, if given the opportunity, prove a prediction wrong. Predictive judgments are generally tentative rather than certain,<sup>13</sup> because available evidence and present knowledge are limited, and perhaps also because individuals act autonomously notwithstanding predictions about them. But while predictive decisionmaking recognizes in principle the possibility of autonomous choice, and the possibility of refuting predictions, the use of a predictive system leaves little opportunity for the operation of that autonomy, and tends to minimize its importance for both applicants and decisionmakers.

For the applicant, the effect is to shift his attention from the predicted future behavior to the method used for making predictions. An applicant can improve his chances for selection, if at all, only by improving the prediction made about him, and not by attempting directly to improve his future performance. For example, an applicant for

13. That is, the predictive judgment claims only that the person is likely to succeed or to fail, not that he will. Thus the proponents of the requirement of proof beyond a reasonable doubt in civil commitment proceedings maintain that it is quite feasible to determine beyond a reasonable doubt that a person is mentally ill and *likely* to injure himself or others; that proposition can be true even if the person in fact would not injure anyone. See, e.g., *The Supreme Court, 1969 Term*, 84 HARV. L. REV. 1, 164 (1970) ("it would not be anomalous for a fact-finder to believe without any doubt that a person was 'dangerous' while still in doubt whether he would commit any specific harm if released") (footnote omitted).

## Prediction

credit may improve his chances by building a history of credit transactions.<sup>14</sup> An applicant for admission to college may improve his chances by participating actively in student organizations.<sup>15</sup> The applicant's efforts are thus directed toward improving his appearance as a good risk, without regard to any effect on actual future performance. Moreover, some predictive systems deny the applicant even this channel for the exercise of autonomous choice, for some systems render him powerless to affect the prediction about him. For example, if parole risk is assessed largely on the basis of race, gender, or age,<sup>16</sup> then the applicant for parole has no power by his own acts to improve his score. Similarly, if successful performance as a parent is predicted largely on the basis of one's experiences as a child,<sup>17</sup> then the applicant for child custody has no power by his own acts to improve the prediction about him. Some methods and criteria for making predictions give the applicant more opportunity than others to affect the prediction.<sup>18</sup> But predictive systems typically tend to encourage him to focus his attention on improving his predictive score, whether or not that also improves his future performance. A direct attempt to improve performance, if it does not also improve the predictive score, will have no effect on his chance of selection.<sup>19</sup>

When the predicted fact is not subject to individual control, then predicting that fact is less threatening to the value of respect for autonomy.<sup>20</sup> For example, prediction of violent behavior by the men-

14. See, e.g., R. COLE, CONSUMER AND COMMERCIAL CREDIT MANAGEMENT 245-322 (5th ed. 1976); A. SPENCE, MARKET SIGNALING 69-75 (1974); Hassler, Meyers & Seldin, *Payment History as a Predictor of Credit Risk*, 47 J. APPLIED PSYCH. 383 (1963).

15. Interview with Worth David, Dean of Undergraduate Admissions at Yale University, in New Haven, Conn. (Oct. 13, 1978).

16. Some studies suggest that race may be a reliable predictor of delinquency. See, e.g., Wolfgang, *supra* note 6, at 161. Wolfgang's study also suggests that age may be used as a predictor of crime. *Id.* at 172 (probability of arrest increases with age to 17, and then decreases with age); see E. VAN DEN HAAG, PUNISHING CRIMINALS 87 (1975) (violent crime is highest among young males).

17. Studies of child abuse suggest that the victims of child abuse in turn abuse their own children. See Spinetta & Rigler, *The Child-Abusing Parent: A Psychological Review*, 77 PSYCHOLOGICAL BULL. 296, 298 (1972) (reviewing literature).

18. This matter will be discussed in Parts II and III, *infra*.

19. Ideally, improving one's predictive score would also improve performance. But that will happen only if the factors that produce the score actually have some causal effect on performance. This causal relationship is not always present. See pp. 1441-42 *infra* (discussion of declining accuracy of prediction schemes relying on factors that applicants can control).

Moreover, when the focus of attention is on improving a predictive score, as it may be when the connection between score and performance is not obvious, then action taken to improve the score will not feel to the actors like action taken to improve performance.

20. Note that the fact predicted is different from the facts used to make the prediction. Predicting an individual's race, or gender, or age does not seem to threaten the individual's autonomous choicemaking. Using those factors to predict something else, however, would present a substantial threat to autonomy. See pp. 1434-36 *infra*.



tally ill, while widely challenged on the ground of inaccuracy, is seldom characterized as a threat to the autonomy of the mentally ill.<sup>21</sup> Similarly, there seems to be little challenge to the legitimacy of predicting behavior that is controlled by the individual only in a rather remote and indirect way. For example, the effort to predict death or illness for purposes of insurance does not seem to threaten the value of autonomy. To the extent that individuals control their future death or illness, that control is exercised indirectly, by a series of acts taken prior to the event. Those acts may well be proper subjects for inducements and penalties, which can be built into particular predictive systems,<sup>22</sup> but the possibility of influencing those choices does not convert the event of dying, or falling ill, into an autonomous act.<sup>23</sup> The conflict between prediction and respect for autonomy is most acute when the predicted behavior is strongly and directly subject to individual control. For example, the act of obeying or violating the criminal law is subject to individual control, and indeed that fact is central to the structure of the criminal law.<sup>24</sup> The strong tradition of respect for in-

21. *But see* Morse, *Crazy Behavior, Morals, and Science: An Analysis of Mental Health Law*, 51 S. CAL. L. REV. 527, 630-35 (1978) (commitment of mentally ill is objectionable not only because predictions of harmful behavior are inaccurate, but also because they incorrectly imply that harmful behavior of mentally ill is less freely chosen and controllable than that of anyone else). *See also id.* at 532-33 & n.11 (surveying commentary on mental health laws in general, and noting growing body of criticism urging that law should respect dignity of mentally ill persons by holding them responsible for choosing their behavior).

22. *See* pp. 1437-42 & note 83 *infra*.

23. *But see* Crawford, *Sickness as Sin*, HEALTH/PAC BULL., Jan./Feb. 1978, at 10 (analyzing and responding to emerging view that individuals should be regarded as responsible for their susceptibility to illness or accident).

24. Professor H.L.A. Hart has lucidly described and defended the view that criminal punishment is properly reserved for those who choose to commit offenses, and are for that reason blameworthy. *See* the collection of essays published as PUNISHMENT AND RESPONSIBILITY (1968), especially *Prolegomenon to the Principles of Punishment*, in *id.* at 1, and *Changing Conceptions of Responsibility*, in *id.* at 186. For other influential statements of this view, which dominates Anglo-American criminal law and scholarship, *see* H. PACKER, *supra* note 1, at 73-79; Hart, *supra* note 1; Packer, *Mens Rea and the Supreme Court*, 1962 SUP. CT. REV. 107.

While paying tribute to this ideal, Anglo-American criminal law has not always respected it. *See, e.g.,* United States v. Park, 421 U.S. 658 (1975) (corporation president criminally liable for storage of food in rat-infested warehouse, despite his lack of personal knowledge or intent). Moreover, some critics have urged the abandonment of the ideal of the voluntary and blameworthy offender. *See, e.g.,* B. WOOTTON, CRIME AND THE CRIMINAL LAW 52-57, 117-18 (1963). But the ideal seems unlikely to be abandoned, even if inconsistencies are tolerated. The ideal of the voluntary and blameworthy offender is essential to the retributive ideology that informs many recent proposals to reform the sentencing process. *See, e.g.,* N. MORRIS, *supra* note 3; P. O'DONNELL, M. CHURGIN & D. CURTIS, TOWARD A JUST AND EFFECTIVE SENTENCING SYSTEM (1977); REPORT OF THE TWENTIETH CENTURY FUND TASK FORCE ON CRIMINAL SENTENCING: FAIR AND CERTAIN PUNISHMENT (1976); E. VAN DEN HAAG, *supra* note 16; A. VON HIRSCH, *supra* note 3; J. WILSON, THINKING ABOUT CRIME (1975). For an attempt to explain the renaissance of retribution, *see* Coffee, *Repressed Issues of Sentencing: Accountability, Predictability, and Equality in the Era of the Sentencing Commission*, 66 GEO. L.J. 975, 1073-80 (1978).

## Prediction

dividual autonomy in criminal law theory may account for a large measure of the resistance to efforts to predict crime for purposes of sentencing and parole.

The conflict between predictive selection and respect for autonomy affects not only applicants but also decisionmakers. Just as applicants are discouraged by predictive selection from acting to improve their future performance, so is the decisionmaker discouraged from acting to improve the chances of the people identified as bad risks. A decisionmaker who selects and excludes individuals on the basis of their predicted behavior tends to view the prediction as a fixed attribute of the applicant, and tends not to consider ways of intervening to change the situation. If sentencing authorities have the power to extend the imprisonment of people identified as bad parole risks, then they may be less likely to seek ways of improving the prospects of those people through new forms of treatment during imprisonment or supportive services after release. If school officials have the power to exclude people identified as bad educational risks, then they may be less likely to seek ways of assisting those people to improve their prospects. If state officials can remove a child from the home of a parent who is expected to abuse that child, then they may be less likely to seek ways of intervening to reduce that risk.

There is no necessary connection between the use of predictive selection and the absence of effort to improve the prospects of the people identified as poor risks. In principle, a predictive selection system could be used to identify poor risks and provide special benefits to them rather than excluding them from benefits. Nevertheless, the practice of denying benefits to the bad risks is so widespread that it supports an objection to predictive selection that cannot be ignored.

In addition to the effect of predictive selection on applicants and decisionmakers, the general public may be affected in an important way by the symbolic qualities of the decision to predict individual behavior. The use of a predictive selection system makes the statement that prisoners or job applicants seldom act to refute the predictions about them, and it may cause others to act as if the prediction were true. The choice that would refute the prediction becomes simply a source of error in the predictive system, rather than an event to be encouraged or discouraged.

In spite of the persuasiveness of these criticisms of selection by prediction, the tension between predictive selection and respect for individual autonomy cannot be resolved in the abstract. It must be considered in light of the nature of the behavior that is being predicted, and the nature of the benefits or burdens to be distributed. Moreover

the assessment of predictive selection must take into account the nature of the plausible alternatives to predictive selection.

Criteria that purport to punish fault and reward merit frequently provide the chief alternative to predictive criteria for selection. In admitting students to school, the chief alternative to predicting future performance is rewarding past effort and industry. In selecting criminals for release from prison, the chief alternative to predicting future crime is punishing past crime in proportion to its seriousness.

But these alternatives have their own limitations with respect to both accuracy and legitimacy. Relative merit in the school context is difficult to define, and even more difficult to measure. If effort and industry constitute merit, then perhaps school behavior that merits reward is identified more effectively by grades and recommendations than by aptitude tests, but all available measures still fall far short of precise identification. Moreover, it is no more obvious that places in school should be distributed as rewards for the industrious than that they should be distributed to those with the best prospects for success.<sup>25</sup>

Similarly, relative blame in the criminal context is difficult to define and to measure. The seriousness of the crime is sometimes said to be the correct measure of blame, but there are major controversies over the proper determinants of seriousness: the actor's purposes for his crime, the magnitude of the harm intended, or the magnitude of the harm done.<sup>26</sup> Moreover, why should release from prison be given as a reward for low culpability rather than as an opportunity for those with the best chances of success? The person with relatively low culpability and a relatively high risk of failure poses a severe challenge for the opponents of prediction, a challenge which they generally resolve by incorporating some predictive criteria into their judgments of culpability. A prior criminal record is of dubious relevance to the seriousness of the present crime, but it is of substantial value as a predictor of future crime. By declaring that a prior record enhances culpability, the

25. It is tempting to deny the conflict between criteria for selection by supposing that they converge to select the same people. But in fact the people with the best prospects for success are not necessarily identical with the people who deserve reward for being industrious.

26. See, e.g., T. SELLIN & M. WOLFGANG, *THE MEASUREMENT OF DELINQUENCY* (1964) (when offenses ranked according to gravity by judges, students, and policemen, gravity appears to depend on harm done, harm threatened, intent, and other factors); A. VON HIRSCH, *supra* note 3, at 79-83 (gravity depends jointly on nature and likelihood of harm risked, and on blameworthiness of offender); Schulhofer, *Harm and Punishment: A Critique of Emphasis on the Results of Conduct in the Criminal Law*, 122 U. PA. L. REV. 1497 (1974) (gravity of offense should not be measured by harm done); Zimring, *The Medium is the Message: Firearm Caliber as a Determinant of Death from Assault*, 1 J. LEGAL STUD. 97, 114-18 (1972) (gravity in cases of assault and homicide by firearms should be measured by size of weapon, as indicator of harm risked).

## Prediction

critics of prediction retain the benefits of prediction while denying its legitimacy.<sup>27</sup>

The decisionmaker can avoid both predictions and evaluations of merit by declining to make comparative judgments at all. One way to avoid comparative judgments is to treat all applicants alike, accepting or rejecting all of them. Some employers and schools do not attempt to predict good performance, but instead accept all applicants who meet some minimum standard of qualification, and later exclude those who fail to demonstrate good performance.<sup>28</sup> Some insurers avoid selecting applicants by risk classification, and instead insure all applicants and spread the cost of varying risks over the whole insured population.<sup>29</sup>

When it is not feasible to accept or reject all applicants, the decisionmaker can avoid making comparative judgments by instead making random selections.<sup>30</sup> Random selection has been proposed as a method of admitting students to school,<sup>31</sup> selecting individuals for compulsory military service,<sup>32</sup> and selecting a parent to be the custodian of a child

27. The point is made by Professor Joseph Goldstein, criticizing the treatment of prior crimes in the Report of the Committee for the Study of Incarceration. *See Appendix: Additional Views of Individual Committee Members*, in A. VON HIRSCH, *supra* note 3, at 173-74. Von Hirsch apparently changed his view of the matter. As principal author of the Report of the Committee for the Study of Incarceration, he asserted that an offense that follows a conviction may be regarded as more culpable, because it demonstrates persistence, *id.* at 84-88, while in an earlier article he observed that counting prior convictions in current punishment is preventive confinement, not punishment. Von Hirsch, *supra* note 7, at 749.

28. One example is the open admissions program of the City University of New York (CUNY), which offers university admission to every high school graduate in New York City. *See The Open Admissions Story: 1970 at the City University of New York (1970)* (packet of internal documents, press releases, and news clippings on the program, available from CUNY, Office of University Relations) (on file with *Yale Law Journal*).

Acceptance of all applicants may be a deliberate choice to avoid prediction, as at CUNY, or it may be an interim measure, designed to aid in the development of a prediction device. *See, e.g.,* Gael, Grant & Ritchie, *Employment Test Validation for Minority and NonMinority Telephone Operators*, 60 J. APPLIED PSYCH. 411 (1975) (in order to study new employment test, participating companies hired without regard to test scores).

29. One example is Social Security, which tailors neither the individual's costs nor his benefits to an assessment of risk. Another example is group insurance provided to all employees of a particular employer, or all members of a fraternal organization.

30. *See* Greely, *The Equality of Allocation by Lot*, 12 HARV. C.R.-C.L. L. REV. 113, 123-26 (1977); *cf.* Divine, *Women in the Academy: Sex Discrimination in University Faculty Hiring and Promotion*, 5 J.L. & EDUC. 429, 443-44 (1977) (proposes random selection from pool of competent candidates for university professorships in order to suppress sex discrimination).

31. Professor Duncan Kennedy has talked with the Harvard Law School admissions office about using random selection to select entering students, but the proposal has never reached a formal stage. Telephone Interview with Professor Duncan Kennedy (Oct. 13, 1978).

32. *Random Selection for Military Service*, Proc. No. 3945, 34 Fed. Reg. 19,017 (1969), reprinted in 50 U.S.C. § 455 app. (1970).

after divorce.<sup>33</sup> But the advantage of random selection is at the same time its chief flaw: its obvious function as a technique to avoid making comparative judgments. The decisionmaker who resorts to random selection is vulnerable to the charge that he has abdicated responsibility for decision in an important matter.

Alternatives to predictive selection, then, often have no stronger claim to legitimacy than predictive selection. Thus, despite substantial questions about the legitimacy and accuracy of predicting individual behavior, predictive selection will sometimes be preferable to any other available means of selection. For this reason, the various methods of making predictions must be considered more carefully.

## II. The Choice Between Statistical and Clinical Methods

Techniques for predicting individual behavior generally use one of two competing approaches to the problem. One approach relies on the subjective judgment of experienced decisionmakers, who evaluate each applicant on an individual basis in light of the experience accumulated by the decisionmaker and his profession. Psychiatrists use this method to predict violent behavior for purposes of involuntary hospitalization, and in that context it is often called the exercise of clinical judgment, or clinical prediction. The term clinical prediction can be generalized to refer to the subjective evaluations performed in other fields as well. For example, sentencing judges and some parole boards evaluate individuals in a subjective manner in order to predict future crime, and similar methods are sometimes used by admissions officers to predict academic success, by personnel officers to predict job success, and by loan officers to predict loan repayment. In all these fields, clinical prediction occurs when a decisionmaker attempts to assess applicants as individuals, taking into account the characteristics that distinguish each applicant.

The alternative method for making predictions evaluates each applicant according to a predetermined rule for counting and weighting key characteristics. The relevant characteristics are specified in advance, and so is the rule for combining them to produce a score for each applicant. This score must be convertible into an estimate of the applicant's expected performance.<sup>34</sup> This method of making predictions

33. J. GOLDSTEIN, A. FREUD & A. SOLNIT, *supra* note 8, at 63, 153 (judicial drawing of lots may be best process for resolving hard choice in child custody proceeding).

34. Performance is sometimes treated as a dichotomous variable: each individual either succeeds or fails. In that case, it is common to use a scoring device to divide the population into groups, and to estimate the proportion of individuals who succeed at each score. If, at a certain score, 80% of the individuals succeed, then the score is interpreted as

## Prediction

is often called statistical prediction, because statistical techniques are generally used to generate the rule from an analysis of prior cases,<sup>35</sup> to measure the accuracy of the rule in describing those prior cases,<sup>36</sup> and to decide whether the rule should be used to predict results in future cases.<sup>37</sup>

Statistical scoring schemes have been used to predict parole success from such items as number of prior convictions, type of crime, em-

identifying persons each of whom has an 80% chance of success. Selection is then made by setting a minimum acceptable chance of success.

Alternatively, performance can be treated as a variable that is many-valued or even continuous: each individual's performance can be assigned to a place on an ordered scale. In that case, the scoring device can be used to estimate a performance level, and selection is made by setting the minimum acceptable performance level.

35. Various methods exist for identifying the predictive factors that seem most strongly associated with performance, and assigning weights to them. Among the most widely used are multiple regression analysis, analysis of variance, and factor analysis. *See generally* H. BLALOCK, *SOCIAL STATISTICS* (2d ed. 1972) (describing range of elementary statistical techniques, their underlying mathematical assumptions, and the problems of statistical inference associated with each technique); F. SIMON, *supra* note 4, at 72-117 (statistical techniques described and applied to sample of probationers in United Kingdom). These various methods do not necessarily identify the same factors or assign the same weights to them; indeed, even a single technique, multiple regression analysis, can produce different results depending on which factors are investigated first and on other matters of judgment. *See* F. MOSTELLER & J. TUKEY, *DATA ANALYSIS AND REGRESSION* 258-332, 381-406 (1977).

36. The rule for combining predictors produces an estimate of performance. For the prior cases, both that estimate and actual performance are known. It is therefore possible to examine the accuracy of the rule in predicting performance. *See* note 4 *supra*. More generally, the accuracy of the device can be described by a measure of the association between estimated and actual performance. Measures of association include correlation coefficients and other indices. The choice among these measures depends on the number, form, and distribution of the variables whose association is being tested. *See generally* H. BLALOCK, *supra* note 35, at 219-502. For a given set of predictive factors, predictive accuracy is maximized by using the weights assigned by multiple regression; that technique cannot however guarantee the choice of those factors that would combine to produce the rule with the greatest possible accuracy.

37. The observed association between actual performance and the estimates produced by the rule might be attributed to chance, in which case the rule is useless as a guide to future decisionmaking. If the relationship is not attributed to chance, then the assumption is often made that it can be attributed to some cause that will persist, and that it is therefore a useful guide to predicting the future. The technique of determining whether an observed association should be attributed to chance is often called significance testing. The relationship is accepted as likely to persist only if there is a sufficiently small probability that the observed pattern of association could have occurred by chance. *See generally* H. BLALOCK, *supra* note 35, at 109-216; F. MOSTELLER, R. ROURKE, & G. THOMAS, *PROBABILITY WITH STATISTICAL APPLICATIONS* 494 (2d ed. 1970).

The assumption that the observed relationship is attributable to some persistent cause rather than to chance does not require the identification of any particular causal relationship between the variables being examined. To identify a particular causal relationship would require substantially more statistical and theoretical analysis than does the assumption, for predictive purposes, that the forces producing a correlation will continue to operate in a similar way, regardless of what those forces are. *See* H. BLALOCK, *CAUSAL INFERENCES IN NONEXPERIMENTAL RESEARCH* 38-44 (1961). For a discussion of the role of causal theory in prediction, see pp. 1444-47 *infra*.

ployment history, and family ties.<sup>38</sup> They have been used to predict the repayment of loans from such items as age, marital status, location of residence, income, and assets.<sup>39</sup> Educational testing is a more elaborate version of the same technique. Instead of a small number of predictive items, a test used to select applicants for a school typically rests on a large number of items—the answers to the questions that constitute the test. These answers are scored and weighted according to a predetermined rule, and applicants are then ranked on the basis of their scores.<sup>40</sup>

The choice between clinical and statistical methods of prediction has generated substantial controversy in each field where the choice has been made. Although some of the controversy relates to the details of particular clinical or statistical methods, a number of important and recurring issues are raised by the choice between these two general approaches to the problem of prediction. In general, proponents of

38. One such scheme has been adopted by the United States Parole Commission as an aid in predicting parole performance. 28 C.F.R. § 2.20 (1978). Called a Salient Factor Score, it uses seven predictive factors: prior convictions, prior incarcerations, age at first commitment, offense involvement of auto theft or checks, prior parole revocations, drug history, and employment history. Under each of these categories the individual receives points. For example, if the individual has no prior convictions, he receives three points; if one prior conviction, he receives two points; if two or three, one point; if four or more, no points. The higher the total score, the better the parole prognosis. The Salient Factor Score is one important determinant of the date of parole release; the other important determinant is the seriousness of the offense.

For a more detailed discussion of the Salient Factor Score by its originators, see Hoffman & Beck, *Parole Decision-Making: A Salient Factor Score* (Apr. 1974) (United States Board of Parole Research Unit Report Two), *adapted for publication*, 2 J. CRIM. JUST. 195 (1974). For the history of its development, including various modifications made since its first experimental use in 1972, see Coffee, *supra* note 24, at 991-93, 1017-32; Project, *Parole Release Decisionmaking and the Sentencing Process*, 84 YALE L.J. 810, 820-41, 861-78 (1975).

39. See, e.g., Boggess, *Screen-test your credit risks*, 45 HARV. BUS. REV., Nov.-Dec. 1967, at 113; Wells, *New Customer Credit Pointing System*, in NUMERICAL POINTING PLANS FOR EVALUATING CONSUMER CREDIT RISKS 4-21 (1963) (report of Second Consumer Credit Symposium, University of Pennsylvania); Note, *Credit Screening and the ECOA: Applying the Effects Test*, 88 YALE L.J. 1450, 1453-58 (1979).

40. Tests produced by the Educational Testing Service are widely used to select students for admission to college, graduate, and professional schools. These tests are designed to predict an applicant's performance as defined by first-term or first-year average grades. See Schrader, *The Predictive Validity of College Board Admission Tests*, in COLLEGE ENTRANCE EXAMINATION BOARD, *THE COLLEGE BOARD ADMISSIONS TESTING PROGRAM* 117 (W. Angoff ed. 1971) [volume hereinafter cited as *ADMISSIONS TESTING*]. When new items are proposed for addition to these examinations, the items are tested to determine how well their results correlate with results for old questions (inter-test reliability), and also how well their results correlate with performance as defined by grades (predictive validity). Donlon & Angoff, *The Scholastic Aptitude Test*, in *ADMISSIONS TESTING*, *supra*, at 27-29; Fremer & Chandler, *Special Studies*, in *id.* at 156-65; McPeck, Pitcher & Carlson, *The Predictive Effectiveness of Several Experimental Item Types and the Operational Item Types in the Law School Admission Test in 1970-71*, in 2 LAW SCHOOL ADMISSION COUNCIL, *REPORTS OF LSAC SPONSORED RESEARCH: 1970-74*, at 499-503 (1976).

## Prediction

statistical method claim advantages of aggregate accuracy and efficiency, while proponents of clinical method claim advantages of accuracy and fairness in individual cases. These competing claims require closer examination, but first it is necessary to consider the possibility that the conflict is illusory.

It has sometimes been argued that the dichotomy between methods of prediction is false, because clinical methods are simply inchoate and inarticulate forms of statistical decisionmaking. The argument is that a clinical decisionmaker can only respond to a finite number of characteristics, and he inevitably weights them in accordance with the importance they have shown in other cases, either in his own experience or in a larger body of received experience. Statistical analysis, on this view, makes explicit and articulates the process that would occur in any event, if less carefully and candidly.

The argument, however, misses a critical distinction. A clinical decisionmaker is not committed in advance of decision to the factors that will be considered and the rule for combining them. He is free to respond to individual differences whose relevance was not anticipated by any rule. This power to respond to individual differences may render clinical judgments either better or worse than statistical decisions, with respect to either accuracy or legitimacy. Thus the distinction between the methods cannot be dismissed as unimportant.<sup>41</sup>

A number of studies have attempted to test the relative accuracy of clinical and statistical methods. Paul Meehl, in an important book on the subject, examined this literature up to 1954 and found that, in the reported studies, statistical methods always performed as well as or better than clinical methods;<sup>42</sup> in papers since then, he has presented additional evidence for the superior accuracy of statistical methods.<sup>43</sup> Of course the comparison of particular clinical and statistical methods does not necessarily provide a trustworthy test of either general

41. The distinction arises not only in the context of prediction, but also in any other factfinding process. It describes two different ways of approaching the task of moving from evidence to facts, and that task presents similar problems whether the facts to be found are past or future.

42. P. MEEHL, *CLINICAL VERSUS STATISTICAL PREDICTION* 90-119 (1954) (in 20 studies comparing clinical with statistical methods of making predictions in education, corrections, psychiatry, and organic medicine, about half show statistical method superior, and about half show no difference between methods).

43. P. MEEHL, *When Shall We Use Our Heads Instead of the Formula?* in *PSYCHODIAGNOSIS*, *supra* note 5, at 88 (1957 review giving new total of 27 studies, with 17 showing statistical methods superior and 10 showing no difference); P. MEEHL, *What Can the Clinician Do Well?* in *id.* at 171-72 (1967 review giving new total of 35 studies, with 23 showing statistical methods superior and 12 showing no difference).



methodology. There is no reason to suppose that either methodology is represented in these studies by typical examples.<sup>44</sup>

Meehl suggests that clinical methods might be more accurate than statistical methods if the available data are too diverse, and the understanding of the phenomenon too weak, to support the formulation of a useful rule for combining factors to make predictions.<sup>45</sup> That is, if the designers of the statistical rule have been unable to identify and assign weights to all the important predictive items, then perhaps their rule will be less accurate than the clinical judgment of an experienced observer who relies on perceptions that he cannot articulate as explicit rules. Meehl has conceded the apparent superiority of clinical methods in one such reported comparison,<sup>46</sup> and there may be others. But in general, the limited available evidence suggests that statistical methods tend to make fewer errors than do clinical methods that allow the decisionmaker to consider unanticipated factors in the individual case.<sup>47</sup>

Moreover, any such difference in accuracy is likely to be enhanced by the effects of the comparative operating costs of the two types of systems. A clinical system requires the careful exercise of skilled judg-

44. As Meehl observes, he uses statistical methods of analysis to assess the performance of the two methodologies, in the sense that he generalizes from the cases studied to the general methodologies they purport to represent. P. MEEHL, *supra* note 42, at 83-128, 136-38. But doubts about statistical method are based on doubts about the ability of individuals to represent or be represented by a class. A skeptic about the advisability of using statistical methods for selecting applicants questions the ability to select individual applicants based on inferences from their membership in a class. Such a skeptic would hardly be convinced by Meehl's demonstration, because he would also doubt the ability to choose a selection technique based on inferences from its membership in a class of techniques.

45. P. MEEHL, *What Can the Clinician Do Well?* *supra* note 43, at 169-70. Meehl identifies six factors which favor clinical prediction: 1) *open-endedness*, or ambiguity about what is to be predicted; 2) *unanalyzed stimulus-equivalences*, or inability to articulate the rules that guide subjective judgment; 3) *empty cells*, or the failure of relevant factors to appear in the cases from which the statistical rule is derived; 4) *theory mediation*, or the availability of general theory to generate predictions without statistical support; 5) *insufficient time* for devising or applying a suitable statistical rule; 6) *highly configured functions*, or relationships that cannot be expressed in general mathematical form. *Id.*

46. Meehl, *Seer Over Sign: The First Good Example*, 1 J. EXPERIMENTAL RESEARCH PERSONALITY 27 (1965) (discussing, with approval, Lindzey, *Seer versus Sign*, 1 J. EXPERIMENTAL RESEARCH PERSONALITY 17 (1965)). Lindzey found that clinical predictions of homosexual behavior among prisoners were significantly more accurate than statistical predictions. Lindzey, *supra*, at 18. But see Goldberg, *Seer over Sign: The First "Good" Example?* 3 J. EXPERIMENTAL RESEARCH PERSONALITY 168 (1968) (arguing that Lindzey's study does not show superiority of clinical methods).

47. See Dawes, *A Case Study of Graduate Admissions: Application of Three Principles of Human Decision Making*, 26 AM. PSYCHOLOGIST 180 (1971), reprinted in STATISTICS AND PUBLIC POLICY 295 (W. Fairley & F. Mosteller eds. 1977) (statistical method superior to clinical method in predicting academic performance; reviewing literature on methodological controversy); Mosteller, *Assessing Unknown Numbers*, in STATISTICS AND PUBLIC POLICY, *supra*, at 168-70 (relative strengths of statistical method).

## Prediction

ment for the evaluation of each individual applicant, and therefore the marginal cost of evaluating additional applicants tends to be high. A statistical system, on the other hand, requires less skill and less time from its decisionmakers; the major costs of a statistical system arise at the level of collecting and analyzing data and designing the system, rather than administering it.

As the number of applicants increases, the costs of a clinical prediction device increase much more rapidly than those of a statistical device. Thus, even if a clinical device and a statistical device originally were designed to attain roughly the same level of accuracy, failure to meet rising costs would in practice undermine the accuracy of the clinical device much more seriously than it would undermine the accuracy of the statistical device.

Often, the two methods are equivalent in accuracy, or nearly so. In that case, the choice between them is properly dominated by other considerations, which may be characterized as issues of legitimacy.

One set of legitimacy issues concerns the extent to which each method pays attention to the uniqueness of individuals. There is a pervasive sense that in some circumstances justice requires individualized decisionmaking, and that statistical methods fail to satisfy that requirement. Thus, for example, sentencing and parole decisions made according to predetermined rules have been criticized as inappropriately "fixed and mechanical,"<sup>48</sup> and insufficiently sensitive to individual differences. Indeed, for capital sentencing, at least, this crit-

48. *E.g.*, *United States v. Schwarz*, 500 F.2d 1350, 1352 (2d Cir. 1974) (invalidating "fixed and mechanical approach" that denied Youth Corrections sentence to privileged defendant); *Woosley v. United States*, 478 F.2d 139, 144 (8th Cir. 1973) (invalidating "mechanical" imposition of maximum sentence on selective-service violators). Several courts have suggested that strict adherence to the federal Parole Commission guidelines would violate a constitutional or statutory requirement of individualized determinations. *See, e.g.*, *Geraghty v. United States Parole Comm'n*, 579 F.2d 238, 259-63 (3d Cir. 1978), *cert. granted*, 47 U.S.L.W. 3586 (1979) (No. 78-572); *United States v. Cruz*, 544 F.2d 1162, 1164 n.6 (2d Cir. 1976); *United States v. Norcome*, 375 F. Supp. 270, 274 n.3 (D.D.C.), *aff'd*, 497 F.2d 686 (D.C. Cir. 1974). These courts have rejected only strict adherence to the guidelines, and not all reference to them; one court concluded that guidelines can be used "as a tool but not as a rule," *Page v. United States*, 428 F. Supp. 1007, 1009 (S.D. Fla. 1977). But other courts have held that there is no inconsistency between individualized sentencing and strict adherence to the parole guidelines, because each defendant is individually evaluated to get his guideline score. *See, e.g.*, *Daniels v. United States Parole Comm'n*, 429 F. Supp. 518 (W.D. Pa. 1977); *Barr v. United States*, 415 F. Supp. 990 (W.D. Okla. 1976).

In fact the parole guidelines include a provision for decisions that depart from the guideline terms, so long as reasons are given for the departure. Whether or not such departures often occur, the possibility of discretionary departure may be sufficient to meet objections to a rigid rule. In practice, statistical rules are seldom adopted without provision for clinical decisions that depart from the rules. Although there is no reason to think that this technique improves aggregate predictive accuracy, it seems to make a statistical device more acceptable to decisionmakers and their constituencies.

icism has the status of a constitutional principle: the Supreme Court has held that the Constitution prohibits capital sentencing according to fixed rules, and requires individual discretionary judgments.<sup>49</sup> Another example can be found in school admissions. Many institutions refuse to follow strictly the predictions made from test scores and grades, on the ground that rigid reliance on numbers would be insufficiently sensitive to individual differences.<sup>50</sup>

49. *Lockett v. Ohio*, 98 S. Ct. 2954, 2965 (1978) (plurality opinion of Burger, C.J., joined by Stewart, Powell, and Stevens, JJ.) (statute unconstitutional because it fails to allow consideration of defendant's character and record and circumstances of offense, thus violating principle that "an individualized decision is essential in capital cases"); *Roberts v. Louisiana*, 428 U.S. 325, 333-34 (1976) (opinion of Stewart, Powell, and Stevens, JJ., announcing judgment of Court) (same); *Woodson v. North Carolina*, 428 U.S. 280, 303-05 (1976) (opinion of Stewart, Powell, and Stevens, JJ., announcing judgment of Court) (same). This requirement of individualized consideration, although consistently articulated by only three (or four) justices, represents controlling law, because any capital statute violating the requirement will be held unconstitutional by a majority consisting of this group and Justices Brennan and Marshall, who find any death sentence unconstitutional.

The history of capital punishment in the Supreme Court provides a vivid illustration of the conflict between the fundamental value of decisionmaking in accordance with general standards and the equally fundamental value of individualized decisionmaking. In 1971 the Court rejected the claim that a death penalty statute was constitutionally required to provide standards for the decision to impose the death penalty. *McGautha v. California*, 402 U.S. 183, 196-208 (1971). The eloquent dissent of Justice Brennan, *id.* at 248, became in 1972 the implicit holding of the Court in *Furman v. Georgia*, 408 U.S. 238 (1972). In *Furman*, Justices Brennan and Marshall took the position that any death penalty statute is unconstitutional. *Id.* at 304-05 (Brennan, J., concurring); *id.* at 358 (Marshall, J., concurring). The three other Justices in the majority invalidated the statutes at hand because they authorized sentencing authorities to exercise standardless discretion. *Id.* at 256-57 (Douglas, J., concurring); *id.* at 309-10 (Stewart, J., concurring); *id.* at 313 (White, J., concurring).

In response to *Furman*, many states enacted new capital sentencing laws either specifying criteria for mandatory death sentences or specifying criteria to guide the discretion of sentencing authorities. See Note, *Discretion and the Constitutionality of the New Death Penalty Statutes*, 87 HARV. L. REV. 1690, 1691 & n.6, 1699-1712 (1974). The Court has invalidated several of these statutes on the ground that they failed to provide for sufficiently individualized decisionmaking. *Lockett v. Ohio*, 98 S. Ct. 2954, 2965 (1978) (plurality opinion of Burger, C.J., joined by Stewart, Powell, and Stevens, JJ.); *Roberts v. Louisiana*, 428 U.S. 325, 333 (1976) (opinion of Stewart, Powell, and Stevens, JJ., announcing judgment of Court); *Woodson v. North Carolina*, 428 U.S. 280, 304 (1976) (opinion of Stewart, Powell, and Stevens, JJ., announcing judgment of court). Other new statutes have been upheld as constitutional, partly because they allow consideration of particularized mitigating factors. *Jurek v. Texas*, 428 U.S. 262, 271-74 (1976) (opinion of Stewart, Powell, and Stevens, JJ., announcing judgment of Court); *Proffitt v. Florida*, 428 U.S. 242, 254-58 (1976) (opinion of Stewart, Powell, and Stevens, JJ., announcing judgment of Court); *Gregg v. Georgia*, 428 U.S. 153, 196-204 (1976) (opinion of Stewart, Powell, and Stevens, JJ., announcing judgment of Court). It has been suggested that the new statutes, prescribing criteria for sentencing that include consideration of individualized factors, give decisionmakers the same unbridled discretion as they had under the pre-*Furman* statutes. See Black, *supra* note 9, at 9-14.

50. See Turnbull, McKee & Galloway, *Law School Admissions: A Descriptive Study*, in 2 LAW SCHOOL ADMISSION COUNCIL, *supra* note 40, at 319-21 (summarizing use of non-objective criteria at five schools). The departure from strict reliance on scores and grades occurs largely after the high scorers have been accepted and the low scorers have been rejected, so that subjective judgments are used only to override relatively small differences

## Prediction

This general argument against the statistical method concedes that the method may produce decisions that in the aggregate are correct often enough to meet some requisite standard of accuracy. The argument denies, however, that aggregate results provide the sole measure of the legitimacy of the method. The claim is that in some situations the individual applicant is entitled to a decision that not only contributes to a useful aggregate, but also reflects consideration of the distinctive aspects of his individual case.

This concern for individualized decisionmaking is sometimes expressed in rhetoric that seems to protest all classification or categorical judgments. Taken to that extreme, however, the argument seems to reject clinical decisionmaking as well, because that technique too employs a form of categorical judgment. Although the clinician need not identify in advance the characteristics he will regard as salient, he must nevertheless evaluate the applicant on the basis of a finite number of salient characteristics, and thus, like the statistical decisionmaker, he treats the applicant as a member of a class defined by those characteristics.<sup>51</sup>

If the clinical method pays more attention to individuals, it does so by giving each applicant the opportunity to call attention to his own special characteristics. Even if special pleading seldom affects the decision, and even if it may slightly impair the accuracy of aggregate decisionmaking, it may nevertheless enhance the perceived legitimacy of the process.

A clinical decision process allows the applicant to present the evidence and arguments he regards as most favorable, rather than restrict-

in scores. Small differences in scores tend to predict only small differences in performance, and tend to do so with little accuracy. See Schrader, *supra* note 40, at 124, 130; Turnbull, McKee & Galloway, *supra*, at 307. While there is no evidence that the use of clinical judgment improves on the limited predictive power of tests in a restricted range, at least the use of clinical judgment is unlikely to cause much loss of predictive power in such circumstances.

School admissions officers sometimes depart from reliance on grades and test scores not in an effort to improve the prediction of school performance, but in an effort to predict later professional performance, or to achieve heterogeneity among students for the benefit of the educational process, or for other reasons. See *id.* at 319-21. The effort to predict professional performance has been singularly unsuccessful; this is at least in part because of the difficulty in finding satisfactory definitions and measures of professional performance. See H. PACKER, T. EHRLICH & S. PEPPER, *NEW DIRECTIONS IN LEGAL EDUCATION* 22 (1972) (Carnegie Commission on Higher Education); Boyer & Cramton, *American Legal Education: An Agenda for Research and Reform*, 59 CORNELL L. REV. 221, 270, 275 (1974). The effort to achieve heterogeneity is described and approved in Regents of Univ. of Cal. v. Bakke, 438 U.S. 265, 311-20 (1978) (Powell, J., announcing judgment of Court) (opinion necessary to disposition of challenge to minority admissions program).

51. For example, when a clinical decisionmaker selects a hard-working even-tempered applicant as a good risk for a job, he is implicitly treating that person as a member of the class of hard-working and even-tempered people.

ing him to the items defined as relevant by a predetermined statistical scoring scheme. The opportunity to participate actively in the decision process is widely thought to enhance the participant's sense of legitimacy and fairness.<sup>52</sup> Moreover, the decisionmaker who must confront applicants as individuals is more likely to experience the disappointment of rejected applicants, and to agonize over the selection process.<sup>53</sup> By giving the applicant the opportunity to make a claim on the personal attention of the decisionmaker, a clinical decision process demonstrates a certain respect for the personal dignity of each applicant.

Statistical methods not only lack this particular advantage, but they pose an additional threat to the perceived legitimacy of the process. A growing body of psychological literature suggests that the intuitive individualized judgments of most people are systematically inconsistent with statistical judgments in several respects. The studies show that in making individualized judgments people rely primarily on information about the case at hand, paying relatively little attention to background information about other cases.<sup>54</sup> This gap between prevailing intui-

52. See *Joint Anti-Fascist Refugee Comm. v. McGrath*, 341 U.S. 123, 171-72 (1951) (Frankfurter, J., concurring); J. THIBAUT & L. WALKER, *PROCEDURAL JUSTICE* 2, 77, 95, 118 (1975); Mashaw, *The Supreme Court's Due Process Calculus for Administrative Adjudication* in Mathews v. Eldridge: *Three Factors in Search of a Theory of Value*, 44 U. CHI. L. REV. 28, 45 (1976) (difficult judgments best legitimated by fullest possible participation in decisional process). This is a recurring theme in the literature about the functions of hearings, and of the adversary system. See, e.g., G. HAZARD, *ETHICS IN THE PRACTICE OF LAW* 120-35 (1978); F. JAMES & G. HAZARD, *CIVIL PROCEDURE* § 1.1 (2d ed. 1977); Michelman, *Formal and Associational Aims in Procedural Due Process*, in *DUE PROCESS* 127-28 (NOMOS XVIII J. Pennock & J. Chapman eds. 1977); Verkuil, *The Emerging Concept of Administrative Procedure*, 78 COLUM. L. REV. 258, 279-80 (1978).

The opportunity for participation is not completely inconsistent with a statistical decision rule: the applicant may be given an opportunity to challenge his score, or, more rarely, to urge a change in the rule. Although such opportunities are conceivable, they are unlikely to exist in practice. One of the chief advantages of a statistical rule is its ease of administration. Therefore, a decisionmaker who chooses to use such a rule is unlikely to accord individual applicants the opportunity to participate in the decision process and thereby impede the administration of the rule.

53. See J. NOONAN, *PERSONS AND MASKS OF THE LAW* xi (1976) (decisionmaker's sense of responsibility is "greater the more one is conscious that he or she—not some imagined entity—is acting, and the more one is conscious that the action affects not a hypothetical A but a real Helen Palsgraf").

54. For a review of the empirical results, see Tversky & Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCI. 1124 (1974) (identifying and classifying some ways in which intuitive judgment systematically differs from statistical inference). The authors report, for example, that if people are asked to judge the probability that a given case belongs to a certain class, they tend to pay little attention to the base rate or prior probability of that class in the population, and to give controlling weight to any resemblance between the given case and the typical class member.

For experiments supporting this and other generalizations about the divergence between intuitive judgment and statistical inference, see, e.g., Chapman & Chapman, *Genesis of Popular but Erroneous Psychodiagnostic Observations*, 72 J. ABNORMAL PSYCH. 193 (1967); Kahneman & Tversky, *On the Psychology of Prediction*, 80 PSYCH. REV. 237 (1973); Kahneman & Tversky, *Subjective Probability: A Judgment of Representativeness*, 3 COGNITIVE

## Prediction

tions and statistical inference may explain a pervasive lack of confidence in the legitimacy of statistical methods. A method that is profoundly counterintuitive inspires suspicion and distrust, and detracts from an applicant's sense that he has been evaluated by a legitimate process.

The gap between intuitive individualized judgment and statistical inference may in time be narrowed. As education in statistics becomes more pervasive, statistical reasoning will become more familiar to all decisionmakers, and perhaps it will exert an influence on individualized decisionmaking. To the extent the gap persists, however, it leaves room for questions of comparative accuracy and comparative legitimacy. It might seem that the gap demonstrates the error of intuitive individualized judgment as a general methodology. On the other hand, individualized judgment techniques have sometimes performed as accurately as statistical techniques in the available studies.<sup>55</sup> And the available studies of comparative accuracy have hardly exhausted the field. Moreover, whatever the verdict on comparative accuracy, the persistence of this gap may confer greater perceived legitimacy on the clinical method.<sup>56</sup>

Clinical and statistical methods differ not only in the extent to which they pay attention to individuals, but also in a related respect that also has important implications for the legitimacy of the process. A statistical method makes explicit and visible the precise criteria by which applicants are selected or rejected, while a clinical method leaves the decision process somewhat more obscure. The explicit statement of criteria protects against the danger that a decisionmaker may rely on illegal criteria, such as race. It also provides the opportunity for public

PSYCH. 430 (1972); Tversky & Kahneman, *Availability: A Heuristic for Judging Frequency and Probability*, 5 COGNITIVE PSYCH. 207 (1973); Tversky & Kahneman, *Belief in the Law of Small Numbers*, 76 PSYCHOLOGICAL BULL. 105 (1971).

In one experiment, subjects were asked to give the probability that the person described in a brief character sketch was a lawyer. Each subject was told that the sketch had been drawn from a population of 70 lawyers and 30 engineers. Some of the sketches described people who fit the stereotype of a lawyer or engineer, some contained information unrelated to either stereotype, and one gave no information beyond the fact of the random drawing. When the sketch fit one stereotype or the other, the subjects evaluated according to stereotype, even if that entailed also guessing a high probability that the person was an engineer. When the sketch fit neither stereotype, they thought the chance he was a lawyer was 50%, ignoring the base rates in the population. Only when they had no individualized information at all did they use the base rates and conclude that the chance that he was a lawyer was 70%. Kahneman & Tversky, *On the Psychology of Prediction*, *supra*, at 241-43.

55. See pp. 1423-24 *supra*.

56. Cf. Tribe, *supra* note 10, at 1375-77 (counterintuitive statistical inferences should not be used in evidence at trial because they may impair perceived legitimacy of trial).

discourse about the desirability of the criteria that are used.<sup>57</sup> Both features of a statistical system tend to foster public confidence in the legitimacy of the system.

The advantages of obscurity are somewhat less obvious. First, the less explicit system is probably in practice more open to adjustment in the face of new information about individuals or about the society in which they live. In principle, a statistical system can be regularly revised to take account of new information, but in practice frequent revision is unlikely. One reason is that the decisionmaker who operates the system and thereby acquires the new information is usually different from the person who formulates the rule. Another is that statistical methods are often chosen for their low operating costs, and a practice of frequent revision would undermine that advantage. A second advantage of obscurity is that it avoids labeling certain characteristics as predictors of failure. Because there is disagreement about the goals to be served by the choice of predictors, the choice of any particular predictors will inevitably provoke controversy. A decision-making scheme that avoids the choice avoids that controversy.<sup>58</sup>

A final aspect of obscurity can operate either as an advantage or a disadvantage of the clinical method. When obscurity surrounds the criteria for a particular type of decision, it encourages the development of a specialized group of people entrusted with the power to decide. To the extent that these people inspire general respect and confidence, they enhance the perceived legitimacy of the system. Indeed public confidence in their unarticulated judgment may surpass public confidence in any explicit rule. But by the same token, to the extent there is no such respected group of clinical judges, the clinical method will be perceived as illegitimate.

Moreover, the connection between clinical methods and an entrenched group of clinical judges may have a more sinister effect as well. The incumbent specialists often have a territorial interest in preserving their discretionary powers, and the respect associated with

57. The requirement that important decisions be governed by explicit standards occupies a high place in the American legal tradition. For a masterful examination of the sources and functions of that requirement, see *McGautha v. California*, 402 U.S. 183, 248-312 (1971), discussed in note 49 *supra* (Brennan, J., dissenting) (arguing that Constitution requires explicit standards for capital punishment).

58. For the advantages of obscurity, see *Regents of Univ. of Cal. v. Bakke*, 438 U.S. 265, 315-20 (1978) (Powell, J., announcing judgment of Court). In the view of Justice Powell, whose vote controlled the disposition of the case, a school's effort to achieve a heterogeneous student body justifies attention to the race of applicants among other factors, but does not justify an explicit racial preference. This distinction, for Justice Powell, is supported both by the Constitution and by the desirability of creating the appearance of fairness to individual applicants. *Id.* at 319 n.53.

## Prediction

that discretion. Thus any such group, once established, may resist efforts over time to make explicit their criteria for decision. On some occasions, the established clinicians may yield to the promise of assistance from an efficient statistical rule. The United States Parole Board, for example, was initially reluctant to experiment with a statistical method for predicting dangerous conduct. The Board feared that a statistical method would cast doubt on the value of their efforts at discretionary decisionmaking. Eventually the Board was persuaded to allow the development of a statistical rule designed in part to predict dangerous conduct, but designed primarily to predict the decisions that the Board itself would make. The success of that effort convinced the Board that the statistical rule codified its collective wisdom rather than devaluing it, and the rule was adopted as the basis of parole prediction in the federal system.<sup>59</sup>

But sometimes the clinical judges refuse to surrender their discretionary power. Hospital authorities entrusted with the decision to discharge involuntary mental patients have resisted the use of statistical predictions, even though their decision closely resembles the prediction made in the parole context. An involuntary mental patient is typically hospitalized on the ground that he has a mental illness that renders him dangerous to himself or others; he is entitled to release when those conditions are no longer met.<sup>60</sup> Hospital authorities purport to make a discharge decision on the basis of a complex clinical evaluation of the patient. Observers have found that among the important predictors of a release decision are the availability of a job for the released patient, and some suitable housing arrangements, but hospital authorities have opposed efforts to make explicit use of such predictors.<sup>61</sup> As a result, the statistical prediction of violent behavior has been resisted in the context of involuntary hospitalization while it has been embraced in the context of parole.

One explanation may be that hospital psychiatrists, unlike parole boards, do not specialize in the release decision. Instead they view it as part of a broader process of providing service and treatment. They may thus be more concerned with preserving the personal and individu-

59. Gottfredson, Hoffman, Sigler & Wilkins, *Making Parole Policy Explicit*, 21 CRIME & DELINQUENCY 34, 37-41 (1975). The Parole Board was replaced by the United States Parole Commission in 1976. Parole Commission and Reorganization Act, Pub. L. No. 94-233, § 2, 90 Stat. 219 (1976) (codified at 18 U.S.C. §§ 4201-4218 (1976)). This article generally refers to the agency that administers the federal parole system as the Commission but occasionally uses the old name in reference to actions taken before 1976.

60. See *Developments in the Law—Civil Commitment of the Mentally Ill*, 87 HARV. L. REV. 1190, 1201-07, 1384-89 (1974).

61. See Weinstein, *Real & Ideal Discharge Criteria*, 1964 MENTAL HOSPITALS 680.



alized character of that whole process, and less interested in improving the efficiency of the release decision, which constitutes only a segment of that process. They seem to find it unacceptably restricting and demeaning to replace their clinical evaluations with a statistical device.

Whatever the reason for this resistance, it tends to obstruct one very desirable path of progress. For in a field where knowledge is scarce, it seems reasonable to begin with a decision process that depends on the clinical judgment of wise people, and then move to a statistical rule that codifies their collective wisdom and enables less skilled decision-makers to apply the rule with uniformity and efficiency. But the clinical judges who would provide the foundation for such a move may often be its strongest opponents.

One way to describe the difference between clinical and statistical methods is to say that clinical methods pay more attention to individual applicants, and statistical methods pay more attention to the rules for selecting them. Thus the tension between clinical and statistical methods mirrors in large part the familiar tension between discretion and rules.<sup>62</sup> Clinical methods have the advantages of individualized discretionary decisionmaking, while statistical methods have the advantages of a system that depends on the uniform application of well-defined rules. Thus, clinical methods protect most effectively against failure to consider unanticipated individual differences, and statistical methods protect most effectively against the implicit use of illegal or otherwise unacceptable criteria for decision. The choice of a method must depend in part on which of these dangers seems most threatening. It must also depend on the extent to which there is available a group of clinical judges who command respect and confidence. A decisionmaker using a statistical rule will be perceived as less legitimate than a highly respected clinical decisionmaker, but more legitimate than a clinical decisionmaker who commands less community respect. Because in many circumstances these considerations will lead to the choice of a statistical rule, it is necessary to consider in some detail the varieties of statistical rules, and the grounds for choosing among them.

### III. Criteria for Choosing Particular Predictive Items

The decision to engage in predictive decisionmaking through a statistical method does not command the use of any particular statistical

62. See, e.g., K. DAVIS, *DISCRETIONARY JUSTICE* 15-21 (1969); R. DWORKIN, *supra* note 7, at 31-39; H.L.A. HART, *THE CONCEPT OF LAW* 127-29 (1961); Kennedy, *Form and Substance in Private Law Adjudication*, 89 *HARV. L. REV.* 1685, 1688, 1699 (1976).

## Prediction

prediction device. It remains to consider precisely what items to use as predictors, and what weights to assign to each item.

To some extent, the choice of predictors can be guided by concern for accuracy. If the device is meant to predict success, then the group selected by the device should contain at least as many people who actually succeed as would be found in a group selected by some other means.<sup>63</sup> But accuracy cannot be the sole criterion for choosing predictors. First, there may be no device that uniquely satisfies this criterion; many alternatives may be equally accurate.<sup>64</sup> For that reason alone, additional criteria are needed for choosing among possible items and devices. Moreover, the values that inform the choice of one from a group of equally accurate predictors or devices may be so important that they require the choice of a predictive scheme that sacrifices some degree of accuracy for other goals.

The choice of particular predictors reopens many of the issues that accompany the decision to engage in prediction. The decision to predict may tend to minimize the applicant's opportunity to control his prospects for selection, and it may tend to reject the notion that selection should reward meritorious behavior and punish fault. The choice of particular predictive items may either reinforce those tendencies, or restore to some extent the possibility of individual control and the role of reward and punishment. Like the decision to predict, then, the choice of predictors can raise questions concerning respect for the autonomy of applicants.

Similarly the choice of particular predictors reopens issues that accompany the choice between statistical and clinical methods. The decision to use a statistical method rejects flexibility and attention to individual claims in favor of ease of administration and explicit criteria for decision. The choice of particular predictive items may either reinforce that decision, or restore some measure of flexibility and individualized judgment. Like the decision to use a statistical method, the choice of predictors can present the conflict between the

63. As noted in note 4 *supra*, the actual performance of a device depends not only on its discriminating power but also on the distribution of success and failure in the population. To satisfy criteria of accuracy in performance, then, it is necessary to choose a device tailored to perform within specified limits on a population of known characteristics.

64. When several items with predictive power are themselves interdependent, their collective predictive power can be arbitrarily allocated among them. One item can be used to the exclusion of the others, or each item can be given a place in the predictive device, without changing the accuracy of the device. This problem is called collinearity. See F. MOSTELLER & J. TUKEY, *supra* note 35, at 280-87; Farrar & Glauber, *Multicollinearity in Regression Analysis: The Problem Revisited*, 49 REV. ECON. & STATISTICS 92 (1967); Tukey & Wilk, *Data Analysis and Statistics: Techniques and Approaches*, in *THE QUANTITATIVE ANALYSIS OF SOCIAL PROBLEMS* 377 (E. Tufte ed. 1970).

uniform application of general rules, and the flexible application of individualized judgment. This section will examine the issues that arise in choosing particular statistical predictors and suggest the values that should inform that choice.

## A. *Autonomy*

### 1. *Nonvolitional Predictors: Race and Others*

Of all possible predictive items, race may have received the most sustained scrutiny and censure. Both constitutional and statutory authority impose stringent restrictions on the use of race as a ground for predicting success or failure. The constitutional restriction lies chiefly in the command of the Fourteenth Amendment that no state shall "deny to any person within its jurisdiction the equal protection of the laws." This provision controls decisionmaking by governmental bodies, such as parole authorities, public employers, and public schools. For such decisionmakers, the Constitution prohibits selection based on race unless the justification is extremely strong.<sup>65</sup> Private decisionmakers, while not governed by the constitutional requirement of equal protection, are in many circumstances similarly restricted by statute; legislatures have prohibited or sharply limited the use of race as a criterion for selection by private employers, schools, lenders, and other institutions.<sup>66</sup>

Several reasons support the hostility in American law to the use of race in selecting from a pool of applicants. First, there is the desire to protect the accuracy of the decisionmaking process by purging it of any unwarranted bias against a particular group. Historical experience suggests that in the absence of a special prohibition, decisionmakers may be influenced by negative views about minority racial groups to make

65. For a review of the doctrine that racial classifications require "strict scrutiny," see Gunther, *The Supreme Court, 1971 Term—Foreword: In Search of Evolving Doctrine on a Changing Court: A Model for a Newer Equal Protection*, 86 HARV. L. REV. 1 (1972); *Developments in the Law—Equal Protection*, 82 HARV. L. REV. 1065, 1081-1137 (1969).

The Court has upheld racial exclusion or preference on grounds of threat to national security in wartime, *Korematsu v. United States*, 323 U.S. 214, 215-18 (1944) (exclusion of Japanese from much of West Coast); threat to prison security, *Lee v. Washington*, 390 U.S. 333 (1968) (racial segregation of prisons) (dictum); need to remedy prior illegal racial discrimination, *Swann v. Charlotte-Mecklenburg Bd. of Educ.*, 402 U.S. 1, 22-25 (1971) (race-conscious assignment of students to schools); and need to "prepare students to live in a pluralistic society," *id.* at 16 (dictum). Cf. *Morton v. Mancari*, 417 U.S. 535, 551-55 (1974) (special constitutional status of Indian tribes justifies employment preference for Indians in Bureau of Indian Affairs).

66. *E.g.*, Equal Credit Opportunity Act of 1974, 15 U.S.C. § 1691(a)(1) (1976) (lenders); Title VI of the Civil Rights Act of 1964, 42 U.S.C. § 2000d (1976) (programs, including schools, that receive federal funds); Title VII of the Civil Rights Act of 1964, *id.* § 2000e-2(a) (employers).

## Prediction

negative predictions even when the predictive power of race is non-existent.

Limitations on the use of race as a predictor also stem from goals independent of concerns for accuracy. The repeated use of negative predictions based on race has a cumulative effect. By repeatedly excluding from various benefits the members of the same well-defined group, the practice contributes to the formation of a discrete disadvantaged class, whose members share a massive sense of injustice.<sup>67</sup> A prohibition on the use of such predictions can have both symbolic and real consequences in reducing the separation and stigmatization of such a group.<sup>68</sup>

While this second goal is distinct from a concern for accurate prediction, it need not conflict with that concern. In some cases, the predictive power of race is very small, and alternative predictors are available with at least as much predictive power. In such circumstances, a prohibition on the use of race causes no loss of accuracy. Indeed, if the predictive power of race is so small that it would ordinarily be disregarded, the decision to select on the basis of race suggests a bias on the part of the decisionmaker that is inconsistent with the constitutional and statutory concern for equal treatment.<sup>69</sup>

When the predictive power of race is large, the attempt to avoid racial criteria may create a greater conflict with the concern for accurate prediction. Even then, it is usually possible to find alternative predictors that are at least as powerful as race.<sup>70</sup> In these circumstances the alternative predictors will correlate strongly with race, and thus have the effect of excluding disproportionately large numbers of one racial group. A nonracial predictor that correlates highly with race has some of the effects of selection based explicitly on race.<sup>71</sup> For all

67. Moreover, the use of race as a criterion of selection is likely to induce behavior by applicants that increases the apparent validity of race as a predictor. See A. SPENCE, *supra* note 14, at 98.

68. See P. BREST, *PROCESSES OF CONSTITUTIONAL DECISIONMAKING* 481-82 (1975); Fiss, *Groups and the Equal Protection Clause*, 5 *PHILOSOPHY & PUB. AFF.* 107, 157-68 (1976).

69. This idea is particularly well put by Professor Brest, who calls it "racially selective indifference." Brest, *The Supreme Court, 1975 Term—Foreword: In Defense of the Anti-discrimination Principle*, 90 *HARV. L. REV.* 1, 6-8, 14-15 (1976).

70. Compare Wolfgang's view that when race is used to predict criminal recidivism, school achievement and other background variables add little predictive power, see Wolfgang, *supra* note 6, at 163, with the view of Peter Hoffman, principal architect of the federal parole guidelines. Hoffman asserts that when criminal record and other background factors are used to predict criminal recidivism, race adds little predictive power. See Project, *supra* note 38, at 877 n.329.

71. Indeed, under Title VII of the 1964 Civil Rights Act, 42 U.S.C. § 2000e-2(a) (1976), the use of a selection criterion that has a disproportionate exclusionary impact on a protected group can amount to impermissible discrimination against that group. See *Griggs v. Duke Power Co.*, 401 U.S. 424, 436 (1971). See generally B. SCHLEI & P. GROSSMAN, *EMPLOYMENT DISCRIMINATION LAW* 65-184 (1976).

its disadvantages, however, it avoids the extreme effects of a racial predictor, which makes a symbolic statement about the inferiority of one racial group.

A final reason for restricting selection based on race is that race is beyond the power of anyone to change, and hence the use of predictions based on race deprives individuals of the power to improve their chances. The idea that each individual should have the opportunity to seek his own destiny occupies a high place in the hierarchy of American values.<sup>72</sup> Although any decision to rely on predictive judgments poses some threat to the autonomy of individuals, that threat is greatest when the predictions are based on factors over which an individual applicant has no control. The threat is reduced when the applicant has the power to improve his predicted chances by changing his behavior. The law does not consistently prohibit or restrict the use of factors beyond the power of the individual to control; the law often permits the use of age, physical strength, and verbal facility to predict success.<sup>73</sup> But it may be important that an individual is not bound to one age forever, and strength and verbal facility may be influenced by effort. It is extremely difficult, however, for a person to avoid the racial category to which he is assigned by society, and partly for that reason the use of race as a predictor is regarded with special skepticism by the law.

Factors other than race, such as gender or illegitimate birth, are similarly resistant to change by the applicant. Although the use of such factors as predictors differs somewhat from the use of race in its history and consequences, nevertheless some of the same concerns are implicated. Of all the factors that might be used for predictive purposes, those beyond the individual's control present the greatest threat to individual autonomy. Use of such factors in a statistical prediction device is particularly undesirable if the device is to be used in a context in which autonomy is highly valued.<sup>74</sup>

72. *E.g.*, *Regents of Univ. of Cal. v. Bakke*, 438 U.S. 265, 319 n.53 (1978) (Powell, J., announcing judgment of Court) ("Fairness in individual competition for opportunities, especially those provided by the State, is a widely cherished American ethic."); *Wisconsin v. Yoder*, 406 U.S. 205, 245 (1972) (Douglas, J., dissenting in part) (footnote omitted) ("It is the student's judgment, not his parents', that is essential [to evaluate the Amish community's restrictions on schooling] if we are to give full meaning to what we have said about the Bill of Rights and the right of students to be masters of their own destiny.").

73. *See, e.g.*, *Massachusetts Bd. of Retirement v. Murgia*, 427 U.S. 307 (1976) (age, strength); *Washington v. Davis*, 426 U.S. 229 (1976) (verbal facility).

74. In the context of decisions about incarceration, individual autonomy has traditionally commanded great respect. It assumes less importance in other contexts. The success of an organ transplant, under present technology, can be predicted to some degree from an analysis of the tissues of the potential donor and recipient. A person who seeks

## Prediction

### 2. *Predictors that Induce, Reward, and Punish*

Predictive items that fall to some degree within the control of an applicant include personal habits, such as drinking, and life choices, such as years of formal education. The use of controllable factors to predict success will tend to induce an applicant to behave in a way that improves his predictive score. Indeed, this inducement effect is precisely what makes it possible to use a prediction scheme while respecting the individual's autonomous choice: the use of controllable factors provides an opportunity for the individual who really wants to be selected to choose conduct that improves his predictive score and hence his prospects for selection.

But in some cases, the inducement of behavior may seem almost as great a threat to autonomy as the use of factors beyond the individual's control. That happens if the induced behavior falls in a zone in which there is a strong social commitment to protecting an individual's private and personal choices. For example, objections based on that commitment may have contributed to defeat a parole board's use of marriage to predict parole success,<sup>75</sup> and the use by employers of pregnancy or maternity to predict job failure.<sup>76</sup>

When the predictive factor is beyond the applicant's control, his autonomy is denied in that he has no opportunity to influence his

to receive a transplant may be selected or rejected on the basis of a prediction of success or failure, which in turn is based on biological factors wholly beyond the applicant's control. Yet the disadvantages of selecting on the basis of uncontrollable factors seem outweighed by the strong interest in avoiding the waste that would accompany erroneous selection, by the relative accuracy of prediction, and by the absence of suitable alternative criteria for selection.

75. The United States Parole Commission's Salient Factor Score originally gave a point toward a prediction of success for a "[r]elease plan to live with spouse and/or children." 41 Fed. Reg. 22,344 (June 3, 1976) (Item I) (correcting omission in *id.* at 19,326 (May 12, 1976)). Subsequently, the Commission dropped that factor from the scoring scheme, and over the course of the next two years provided several different explanations for the change. First, the Commission stated that the factor was "sometimes difficult to score reliably," and no loss in predictive power would result from revising the device (the revision also deleted educational achievement, and added weight to prior criminal history). *Id.* at 52,890 (Dec. 2, 1976) (revision proposed). The Commission later stated that it was deleting the factor of "plan to live with spouse" in part because that factor was likely to be perceived as discriminatory, because it was a matter of social status over which the prisoner had relatively little control. 42 Fed. Reg. 12,043 (Mar. 21, 1977) (revision adopted). According to the Research Director of the Parole Commission, however, an important disadvantage of the factor was that it was too easily controlled by the prisoner; the Director stated that inmates were contriving living arrangements in order to earn a favorable point. See Coffee, *supra* note 24, at 1027 n.159 (reporting interview with Dr. Peter Hoffman). If the use of marital status or living arrangements is perceived as unfair, it seems likely that the unfairness lies not so much in the idea that change in marital status is beyond the individual's control, but rather in the idea that such a change ought not to be induced by governmental decisions about parole.

76. *E.g.*, *Cleveland Bd. of Educ. v. LaFleur*, 414 U.S. 632 (1974); *Phillips v. Martin Marietta Corp.*, 400 U.S. 542 (1971).

chance of selection. When the predictor is within his control, he can influence his chances, but his autonomy is impaired in a different way: not with respect to the results of the selection process, but with respect to the behavior that is used as a predictor.

Certain kinds of behavior are widely regarded as private, and for that reason unsuitable as a target for inducements. For this reason, some statutes prohibit private employers, lenders, and other institutions from selecting applicants on the basis of marital status or other related behavior.<sup>77</sup> Public institutions may be restricted not only by such statutes but also by the Constitution. For if the choice of marital or other family relationships is a fundamental right,<sup>78</sup> then selection based on those factors may unconstitutionally deny liberty without due process, and it may deny equal protection by impinging on a fundamental right without adequate justification.

Respect for autonomy thus counsels not only against the use of uncontrollable factors, but also against the use of those controllable factors that involve behavior generally regarded as private and protected against official interference. That leaves, as the most satisfactory predictors, other factors that lie within the individual's control. Even these pose some threat to autonomy, for their incorporation in a prediction device will tend to induce behavior, and the inducement may be so powerful that it leaves the individual effectively without choice.<sup>79</sup> Yet the only possibilities are to use a predictor that denies people the chance to change their scores, or to use a predictor that gives them the chance but thereby influences their behavior. Given that choice, the controllable factor—so long as it is nonprivate—is preferable because it leaves more room for individual autonomy.

Considerations unrelated to autonomy may also offer guidance in choosing from among these controllable, nonprivate factors. A good reason to use a particular behavior as a predictor of success appears

77. See, e.g., Equal Credit Opportunity Act, 15 U.S.C. § 1691(a)(1) (1976); CONN. GEN. STAT. § 31-126 (Supp. 1978) (employment); N.Y. EXEC. LAW § 296(5)(a)(1) (McKinney Supp. 1978) (sale or rental of housing). For convenient tables listing state antidiscrimination laws concerning credit, housing, and public accommodations, and displaying their coverage of sex, marital status, and related factors, see B. BROWN, A. FREEDMAN, H. KATZ & A. PRICE, *WOMEN'S RIGHTS AND THE LAW* 270-94 (1977).

78. See *Moore v. City of East Cleveland*, 431 U.S. 494 (1977); *Boddie v. Connecticut*, 401 U.S. 371 (1971); *Loving v. Virginia*, 388 U.S. 1, 12 (1967).

79. A similar problem arises when Congress uses federal grants to induce state enforcement of federal policy. The inducement provided by a federal grant may be so powerful that it leaves the states effectively without choice, and thereby intrudes excessively on their autonomy. See *EPA v. Brown*, 431 U.S. 99 (1977); Stewart, *Pyramids of Sacrifice? Problems of Federalism in Mandating State Implementation of National Environmental Policy*, 86 YALE L.J. 1196, 1250-62 (1977).

## Prediction

when society seeks to encourage or reward that behavior for reasons apart from its predictive power.

For example, when a high school diploma is used to predict job success, and the prediction leads to a decision to hire an applicant, that decision can alternatively be viewed as a decision to reward the effort that led to the diploma. When prior convictions are used to predict parole failure, and that prediction leads to a decision to extend incarceration, the decision can alternatively be viewed as a decision to impose additional punishment for the prior blameworthy acts. Recent revisions of the federal parole scoring system have shown a clear preference for predicting failure with predictors that can also be viewed as blameworthy acts. Each time a controversial item has been deleted from the list of predictors, it has been replaced by a factor that gives additional weight to the fact of prior convictions.<sup>80</sup> Perhaps the most important reason for the change is that of all the factors that predict parole failure, a record of prior convictions is the one that most clearly represents blameworthy behavior.

The preference for blameworthy or meritorious predictors appears also in the debates about suitable criteria for risk classification in insurance. In the field of automobile liability insurance, many insurers use risk groups defined by gender, age, marital status, and place of residence.<sup>81</sup> Increasing public opposition to the use of these categories has led insurers, regulatory agencies, and legislatures to consider proposals to substitute categories defined by an individual's voluntary acts.<sup>82</sup> The reformers show a strong preference for predicting high risk from blameworthy acts, and predicting low risk from meritorious acts, such as participation in driver training.

The same issues arise in connection with insurance that requires a prediction of life expectancy. Those who oppose the use of gender to predict long life argue, *inter alia*, that unfavorable predictions should be based instead on faulty behavior, such as heavy drinking or poor

80. See note 75 *supra*.

81. For a survey of history and current practice, see STANFORD RESEARCH INSTITUTE, THE ROLE OF RISK CLASSIFICATION IN PROPERTY AND CASUALTY INSURANCE 1-78 (Supp. 1976). See also MASSACHUSETTS COMM'R OF INSURANCE, OPINION, FINDINGS, AND DECISION ON 1978 AUTOMOBILE INSURANCE RATES 113-23 (1977); Shayer, *Driver Classification in Automobile Insurance*, in MASSACHUSETTS DIVISION OF INSURANCE, AUTOMOBILE INSURANCE RISK CLASSIFICATION: EQUITY & ACCURACY 6-16 (1978).

82. See, e.g., *Rights and Remedies of Insurance Policyholders*, Hearings Before the Subcomm. on Citizens and Shareholders Rights and Remedies of the Senate Comm. on the Judiciary, 95th Cong., 2d Sess. 37-52 (1978) (Kenneth Hahn, Supervisor of Los Angeles County) [hereinafter cited as *Hearings*]; *id.* at 214-22 (J. Robert Hunter, Deputy Federal Insurance Administrator, HUD); Shayer, *supra* note 81, at 4-6, 16-22; Stuart, *Big Shift May be Ahead in Auto Insurance Rates*, N.Y. Times, Dec. 5, 1978, at D1, col. 4.



driving habits, and that favorable predictions should be based on meritorious behavior such as preventive health care.<sup>83</sup>

In predicting life expectancy it may seem difficult to follow the rule of predicting favorable consequences from meritorious behavior and unfavorable consequences from blameworthy behavior. For a short life expectancy has unfavorable consequences in the form of higher premiums with respect to some kinds of insurance, and favorable consequences, in the form of lower premiums, with respect to other kinds. In the context of death benefits, short life is a disadvantage leading to higher rates. In the context of lifetime annuities, on the other hand, short life is an advantage leading to lower rates.<sup>84</sup> The solution is to use different predictions for each kind of insurance. When short life is a disadvantage, it can be predicted by blameworthy behavior like alcohol abuse or bad driving habits. When short life is an advantage, it can be predicted by meritorious or socially valued behavior like mili-

83. The legislatures and the courts have increasingly made it clear that sex-based distinctions cannot be used in determining pension contributions or payments. See, e.g., *City of Los Angeles v. Manhart*, 435 U.S. 702 (1978); Bernstein & Williams, *Title VII and the Problem of Sex Classifications in Pension Programs*, 74 COLUM. L. REV. 1203, 1213 (1974) (arguing that Title VII requires employers to provide pension plans based on unisex actuarial tables estimating longevity). There is evidence to suggest that predictions based on blameworthy and meritorious factors may be at least as accurate as predictions based on gender. See Waldron, *Why Do Women Live Longer than Men?* 10 Soc. SCI. & MED. 349, 353 (1976) (major causes of death related to "behaviors which are encouraged or accepted more in males: using guns, drinking alcohol, smoking, working at hazardous jobs and being adventurous and acting unafraid"); cf. Haner, *Prediction of Automobile Claims by Psychological Methods—A Case Study in Automobile Insurance*, 35 J. RISK & INS. 49 (1968) (attitudes and personality traits pervasive enough to provide workable base for insurance).

Prediction for insurance appears to differ from other predictive selection in important respects. These depend, however, on a critical assumption that may often be false. The assumption is that the result of an unfavorable prediction is to make insurance benefits more costly, and not to exclude the applicant from benefits altogether. On that assumption, the individual may be said to be harmed less by an unfavorable insurance rating than by an unfavorable prediction of performance on parole, or in school, or in a job. And the group that receives an unfavorable insurance prediction may be said to suffer no net loss at all, because while all members of that group pay more for insurance, their collective returns are also commensurately greater; indeed, the net difference between premium paid and benefits received should be the same for a high-risk group as for a low-risk group if the actuarial predictions are sound.

The flaw in this reasoning comes from the fact that an unfavorable prediction of risk sometimes excludes people from insurance altogether. This can happen either because the premium is beyond the financial means of the applicant or because the insurance company elects not to sell insurance at all to selected high-risk groups. In either case, the result is to disadvantage both the individual and the group by denying the benefit of risk-spreading.

84. This fact can hardly be said to neutralize the vice of using gender to predict life expectancy. First, not everyone buys both death benefits and lifetime annuities. Second, even when both are available, they often use inconsistent gender differentials, so that the cost imposed in one context is not compensated by a benefit in the other context.

## Prediction

tary combat performance or employment as a coal miner. In neither case is it necessary to predict short life from male gender.

When a favorable prediction is based on socially valued, meritorious behavior, then there can be little objection to the fact that the prediction scheme may induce applicants to perform the acts that improve their scores. That inducement should be seen not as an unjustified infringement of autonomy, but rather as an independent good, on the theory that refraining from crime, or acquiring a high school education, or obtaining good health care has some social value apart from its power to predict success. Indeed, the independent value of the behavior serves to respond to several other criticisms of the predictive scheme as well. A system that uses praiseworthy behavior to predict success may satisfy even those who object in principle to the whole enterprise of predictive selection. They can ignore the predictive aspect of the scheme and accept it as a system for punishing and rewarding behavior according to its deserts, or for inducing desirable behavior and discouraging undesirable behavior.

Moreover this alternative way of justifying the selection system also decreases the importance of the possibility that a prediction system based on voluntary behavior is likely to decline in accuracy over time. This concern is one of the chief objections to the use of voluntary behavior as a predictor. The fear is that when applicants adjust their behavior to improve their scores, the behavior will lose its predictive power. For example, if high school diplomas are found to predict job success, and people are thereby induced to acquire diplomas, then diplomas might cease to distinguish between success and failure on certain jobs. That could happen if the education that leads to a diploma does not contribute to success, but instead simply correlates with success under present conditions.<sup>85</sup> On that hypothesis, the attempt to use education to predict job success would ultimately be unsuccessful.

85. The Educational Testing Service is greatly concerned about the possibility that review courses can teach applicants how to improve their test scores, and that the result might be to reduce the predictive power of the tests. Turnbull, McKee & Galloway, *supra* note 50, at 318.

The United States Parole Commission has also been concerned with the problem of an inducement effect that may weaken the power of a predictive factor. In the context of high school diplomas, the Commission dealt with the problem by counting only a diploma earned prior to conviction, and not a diploma earned subsequently in prison. An alternate explanation of the practice was proffered by Peter Hoffman, the Parole Commission's research director, who claimed that post-incarceration achievement did not correlate significantly with parole success. Coffee, *supra* note 24, at 1019 n.125 (interview with P. Hoffman, Mar. 22, 1978). Subsequently, the factor of pre-incarceration educational achievement was rejected as well, for the stated reason that the item was not highly correlated with parole success, 42 Fed. Reg. 12,043 (Mar. 21, 1977); moreover, it might be perceived as an unfair item of social status beyond the prisoner's control, *id.* at 12,044.

Nevertheless, such a prediction scheme would induce many people to obtain high school diplomas. If there is no independent value in the behavior that leads to a diploma, then the investment in diploma-gathering is pointless. If, on the other hand, education has some social value or merit apart from its power to predict job success, then the predictive effort that induces and rewards education has some redeeming value even if its predictive accuracy ultimately fails.

Concern for autonomy, then, suggests a hierarchy of preference in choosing predictors for use in a statistical prediction scheme. Non-volitional status is the least desirable predictive item from this point of view, because it denies the applicant any possibility of affecting the result. Only slightly less objectionable is voluntary behavior that falls within the zone of protected private decisionmaking, because it allows the applicant to affect the result only if he submits to strong influences on his autonomous choicemaking in the zone of private, personal behavior. Other kinds of voluntary behavior can be used as predictors with much less threat to autonomy. The most satisfactory predictor of all is voluntary behavior that deserves independently to be encouraged or discouraged, both because that independent value justifies the inducement effect of the predictive system, and because it helps to justify the system in the event that its predictive power should decline.<sup>86</sup>

## B. *Uniformity and Neutrality*

The choice of a statistical method rejects the flexibility and individualized judgment of a clinical method in favor of general rules, explicit criteria for decision, lower administrative costs, and aggregate accuracy. Nevertheless, the selection of items and the construction of the device permit considerable variation in these characteristics. The same considerations that influence the choice of a method thus play an important role in the choice of particular predictive items.

### 1. *Predictors Subject to Objective and Reliable Scoring*

Under a statistical prediction method, decisions depend more on a general rule than on the judgment of a decisionmaker. A preference

86. It can happen that socially preferred behavior predicts a bad result, and blameworthy behavior predicts a good result. For example, people who commit homicide are in general good parole risks, whereas check forgers (whose crime is less serious, if not exactly preferred) are worse parole risks. See, e.g., L. OHLIN, *SELECTION FOR PAROLE* 52 (1951). The United States Parole Commission uses this information as the basis for giving negative credit for auto theft and check forgery. That item has provoked criticism on the ground that it singles out for harsher treatment crimes that are not especially serious. See 42 Fed. Reg. 12,043 (Mar. 21, 1977). Perhaps one explanation for the failure of the Parole Commission to refine its scoring scheme further by crime category is that it would exacerbate this problem of appearing to reward fault and punish relative merit.

## Prediction

for that feature may also dominate the choice of predictive items, and lead to a preference for those that may be objectively and reliably scored. When items are easy to score, then an applicant's score is unlikely to depend on the person who does the scoring. Early parole scoring schemes were severely criticized for using items whose scoring required a large degree of subjective judgment, and therefore seemed likely to produce results that would vary with the scorer.<sup>87</sup> Similarly, when the United States Parole Commission deleted from its scoring scheme the applicant's plans to live with a spouse after release, the reason given for the change was the difficulty in reliably ascertaining those plans.<sup>88</sup>

Ease and objectivity of scoring are attractive partly because they seem likely to increase the overall accuracy of the device. But in principle subjective judgments could form the basis for a highly accurate statistical scoring scheme. Indeed, despite the subjective and potentially unreliable character of information about a prisoner's release plans, that item showed considerable predictive power in the studies that led to the construction of the federal parole prediction device. The difficulties of scoring that item failed to destroy its predictive accuracy. Perhaps errors in scoring were too few to reduce the accuracy of aggregate results. Or perhaps the errors in scoring were systematic in some way that actually enhanced accuracy, so that parole success correlated even better with recorded release plans than with actual release plans. That might happen if, for example, prisoners who tell persuasive lies about their release plans also have the skills needed to succeed on parole.

The more important objection to subjectively scored factors, whether or not the scores are reliably reproduced by various decisionmakers, is the opportunity they provide for the decisionmaker to cloak his personal biases in the mantle of a scientific judgment. The presence of subjectively scored factors makes it easier for a decisionmaker to manipulate an applicant's score in order to confirm a decision made on grounds independent of the scoring system. Even if it happens too seldom to damage the accuracy of the system as a whole, the possibility of this kind of abuse may generate suspicion and mistrust on the part of applicants. The use of such factors destroys one of the principal ad-

87. For one such scheme, see L. OHLIN, *supra* note 86, at 124-29 (evaluator required to assign parole applicant to one of following categories: erring citizen, marginally delinquent, socially inadequate, farmer, ne'er-do-well, floater, socially maladjusted, drunkard, drug addict, sex deviant). See F. SIMON, *supra* note 4, at 30-57, 145-48 (reviewing scoring schemes).

88. See note 75 *supra*.

vantages of statistical prediction over other methods: the apparent neutrality of a fixed rule, uniformly applied.

In some circumstances, however, the arguments in favor of clinical prediction are advanced to support the use of subjective scoring in a statistical system. Deliberate use of some subjectively scored items might serve as a hybrid technique that preserves a role for clinical judgment and that gives some scope for recognition of individual differences. In practice, statistical prediction devices are seldom adopted without some such hybridization: either some subjectively scored items are incorporated into the statistical device, or the subjective judgment of the decisionmaker is permitted to override the results of the objectively scored statistical scheme.

## 2. *Predictors Supported by a Plausible Causal Theory*

Critics of statistical prediction schemes have been particularly vehement in their criticisms of factors that lack intuitive support in the form of a plausible theory of causation. There is widespread agreement, among both opponents and proponents of the statistical method, that something more than mere statistical correlation should be required as the basis for selecting applicants.<sup>89</sup>

The argument sometimes arises in response to the use of tests of vocabulary and general knowledge to predict job and school performance. Successful performance may be correlated with knowledge of certain words and facts that have no obvious relevance to the demands of the job or school. Test questions based on such knowledge have been criticized on the ground that they measure knowledge that lacks any plausible causal relationship to successful performance.<sup>90</sup>

A similar argument has sometimes been made in the controversy over the use of gender to predict life expectancy for insurance pur-

89. See, e.g., F. MOSTELLER & J. TUKEY, *supra* note 35, at 270-71 (given several formulas that will be about equally good at forecasting, decisionmaker should choose one that is "reasonably presentable and defensible" and will avoid "hilarious newspaper columns or . . . the appearance of injustice").

90. One example is the test for the selection of police officers upheld in *Washington v. Davis*, 426 U.S. 229 (1976), reprinted as an appendix to the lower-court opinion, 512 F.2d 956, 967-76 (D.C. Cir. 1975); another is the test for the selection of firefighters, at issue in *Association Against Discrimination in Employment v. City of Bridgeport*, 454 F. Supp. 751, 756 (D. Conn. 1978), *remanded*, No. 78-7400 (2d Cir. Feb. 23, 1979).

Courts have long been plagued by the difficulty of evaluating a test whose claim to usefulness rests on a statistical correlation, and not on a theoretical construct that renders each test question a plausible measure of the quality being tested. Cf. *United States v. Alexander*, 471 F.2d 923, 953-55 (D.C. Cir.), *cert. denied*, 409 U.S. 1044 (1972) (Bazelon, C.J., dissenting in part) (discussion of psychological tests in connection with insanity defense to criminal responsibility); *United States v. Schappel*, 445 F.2d 716, 718 n.4 (D.C. Cir. 1971) (same).

## Prediction

poses. Life expectancy is used to decide whether, and at what price, to sell death benefits and lifetime annuities to applicants. If female gender is used to predict long life, then lifetime annuities will be more expensive or unavailable to women, and death benefits will be more expensive or unavailable to men. Other attributes could be used instead, to partition the population into different groups with statistically stable life expectancies. Among the possibilities are groups defined by habits of work or nutrition, history of illness, or ancestors' ages at death. It is a matter of actuarial indifference whether the cost of lifetime benefits for the long-lived woman is borne by all other women, or by people who share with her some other characteristics, such as similar habits of work or nutrition. Similarly, the cost of death benefits for the short-lived man can be borne either by all other men, or by people who share with him some other characteristic.<sup>91</sup>

One objection to the use of gender to predict life expectancy arises because gender is a factor beyond the individual's control.<sup>92</sup> A further objection to the use of gender to predict life expectancy rests on the claim that gender is not causally related to life expectancy. In fact, available evidence does not answer the question whether biological sex differences cause the observed difference between the sexes in average length of life, or whether that difference is attributable to other factors whose association with gender is fortuitous.<sup>93</sup> Many people have thought the answer to that question relevant to legal or policy decisions about the use of gender to predict life expectancy. The argument seems to be that if gender does not cause the difference, then its use as a predictor is especially unfair, and that if it does cause the difference, that fact tends somehow to justify its use.

Part of the difficulty with this argument is that the notion of causation is notoriously difficult to define with precision.<sup>94</sup> Nevertheless,

91. See *City of Los Angeles v. Manhart*, 435 U.S. 702, 710 (1978) (footnotes omitted) ("To insure the flabby and the fit as though they were equivalent risks may be more common than treating men and women alike; but nothing more than habit makes one 'subsidy' seem less fair than the other.")

92. See note 83 *supra*.

93. See Waldron, *supra* note 83. The male mortality rate in the United States is higher than the rate for females. Much of this sex differential can be attributed to the incidence of arteriosclerotic heart disease among males, and the incidence of this disease is related to competitive, aggressive behavior, and cigarette smoking—characteristics more common among men than among women in this culture. *Id.* at 349-50, 357. Other factors contributing to the high rate of male mortality include suicide rate, fatal motor vehicle accidents, cirrhosis of the liver, and respiratory diseases. *Id.* at 353.

94. The task of defining causal explanations, other kinds of explanations, and chance occurrences has been an important and perplexing concern for philosophers of science. See, e.g., C. HEMPEL, *ASPECTS OF SCIENTIFIC EXPLANATION* 347-54 (1945) (statement of causality may be viewed as working hypothesis); K. POPPER, *THE LOGIC OF SCIENTIFIC*

the demand for a plausible causal theory seems to have some identifiable features. In large part it is a demand for a story explaining how and why the predictor leads to successful performance. And that story should have support in evidence aside from the correlation at hand. Lack of such a story is perhaps one of the more troubling features of the claim of a correlation between knowledge of obscure facts and successful performance as a police officer.

The preference for causal factors does not necessarily serve the goal of accurate prediction. Observed correlations may lead to highly accurate predictions without any causal theory, or even without any explanatory theory at all. Moreover, correlations supported by a causal theory may cease to provide accurate predictions under changed circumstances.

The importance of a causal theory is not that it guarantees the continuing effectiveness of the predictive scheme, but that it suggests the circumstances under which the scheme will remain effective. A statistical correlation in data about one group of people may not hold when used as a basis for predictions about another group of people. A causal theory helps to identify any relevant differences between the two groups, or differences in the surrounding circumstances. Changes in the job market could remove a previously valid connection between lack of education and parole failure; changes in the typical employment patterns of men and women may remove the connection between male gender and short life.

If a causal theory identifies such possibilities for change, it may accomplish more than warning the decisionmaker of the limitations of the prediction device. For the theory, if publicized, may suggest to applicants or decisionmakers a course of action that would change both the prediction and the performance. For example, a statistical device employed in setting rates for fire insurance may use the location of buildings as a predictor of risk.<sup>95</sup> Under this scheme, buildings located in cities receive lower scores than buildings located in suburbs, and therefore coverage in urban areas is more expensive or even unavailable. The causal theory behind the prediction device might take

DISCOVERY 56-62 (1968) (discussing varied meanings attributed to "causal explanation"). Legal commentators have struggled to clarify the meaning of causal language in law. See, e.g., G. CALABRESI, *THE COSTS OF ACCIDENTS* 295 (1970); H.L.A. HART & A. HONORE, *CAUSATION IN THE LAW* 1-7 (1959).

95. The practice of using location as a predictor of risk is sometimes called "redlining," especially when it entails the identification of inner-city areas as high risk, and the refusal to insure property in those areas. For a survey of the extent of the practice and its consequences, see Federal Insurance Administrator, U.S. Dep't of HUD, *Insurance Crisis in Urban America*, reprinted in *Hearings*, *supra* note 82, at 223-351 app.

## Prediction

any of several forms. The city might have an unusually high concentration of old buildings that are susceptible to fire damage, but for new urban buildings the risks might be no greater than for new suburban buildings. Or the greater risk of fire damage may arise because the city lacks adequate firefighting forces, or because the area is home to several arsonists. Each of these theories has implications for action that might reduce the risk of fire and render the prediction false: the individual owner could choose to occupy a new building rather than an old one; the city could improve fire protection or prosecute its arsonists. Thus a causal theory may operate not to improve prediction but to improve performance.

Another advantage of a causal theory is that it enhances the perceived legitimacy of the predictive scheme. Part of the function of a selection process is to reconcile unsuccessful applicants to their failure to obtain parole, the job, the loan, or the other benefit at stake. In establishing the legitimacy of a selection system, an appeal to scientific authority may have contradictory effects: it suggests orderly decision rather than arbitrariness, but it also suggests decision by experts who have unchecked power to manipulate their system. A prediction based on unexplained correlations requires the applicant simply to trust the technical experts who designed the system. If the predictors are supported by a plausible causal theory, then the unsuccessful applicant can have an intelligible explanation along with the invocation of scientific authority.<sup>96</sup> By using predictors that are both statistically valid and intuitively plausible, the decisionmaker may be able to appropriate the benefits of the appeal to science without the costs.

## Conclusion

The choice of factors for use in a predictive scheme implicates values that sometimes point in different directions. Respect for autonomy counsels the choice of controllable factors, involving nonprivate behavior. Respect for uniformity and neutrality counsels the choice of easily scored factors, and those that are supported by a plausible causal theory. But a factor that is easy to score will often be beyond the applicant's control, and one that is easily controlled will often be difficult to score. Moreover, a plausible causal theory may be found on either side of the conflict, or on neither. The choice of factors, then,

96. Decisionmakers as well as applicants may prefer a scheme that has intuitive plausibility. *See, e.g.*, Wells, *supra* note 39, at 56 (bank executive reluctant to adopt credit-scoring scheme that would cause implementing personnel to "think we are crazy and . . . have a little less faith in the system").



requires a choice between competing values in much the same way that a choice between competing values was required to reach the decision to predict, or the decision to use a statistical method of prediction.

The decisionmaker who would choose a prediction device for some particular use must find a way to balance these competing values. He must also decide how much weight to give to predictive accuracy. For the conflict in values can be posed not only by factors with equal predictive power, but also by factors that are unequal in predictive power. And in that case, it is not clear whether accuracy should triumph over all other values. In order to make the choice, it is necessary to know what is being allocated, what sort of behavior is being predicted, what alternative selection techniques are available, and what advantages and disadvantages inhere in the various techniques.

The advantages and disadvantages depend in large measure on one's vision of applicants and decisionmakers. Statistical method and easily scored factors are most appealing if wisdom is thought to be concentrated in a few wise rulemakers, who can anticipate most important contingencies, while the everyday decisionmakers are thought to have inadequacies that require control by rule. Discretionary decisionmaking is most appealing if wisdom is thought to reside in the people who confront applicants and make decisions, rather than in the policymakers who make the rules.

If the nature of decisionmakers helps to resolve the conflict between a system of rules and a system of discretion, it is the nature of applicants that is critical in deciding whether to predict at all, and whether to prefer predictive factors that lie within the applicant's control. If applicants are regarded as fixed on a life course, then it is both feasible and just to predict their future behavior, and to base the prediction on factors beyond the individual's control. If, instead, applicants are regarded as able to change, and especially if certain changes are desirable, then it is preferable to use nonpredictive criteria of selection that tend to induce those changes by reward and punishment, or to use predictive methods that base predictions on factors within the individual's control.