

## The Ethics of Statistical Discrimination

Many sorts of social policies rely on statistical discrimination (hereafter, "SD"). By SD, I mean the following familiar practice: Some identifiable characteristic of persons (call it "A") is positively but imperfectly correlated with some other, perhaps underlying, characteristic B. We would prefer to discriminate among persons on the basis of B. However, for reasons of convenience, administrative cost, and so forth, we use A as a "surrogate" or "index" of B even though the A-to-B correlation is imperfect (that is, not strictly one-to-one). This paper concerns the ethics of such a practice. When is SD morally justified and when not?

Examples of SD abound. Some of the examples most relevant to the present discussion include: (1) "actuarial" insurance premia, where (for example) we discriminate against the class of young males even though the category we wish to target is unsafe drivers and not young males as such; (2) race-based affirmative action programs, where we use (say) "black" as a surrogate for "socioeconomically disadvantaged" or "victim of invidious discrimination"; (3) sex-based pension plans, where we discriminate against the class of women (by charging them higher payroll contributions) even though we want to control for longevity, not sex per se; finally, to take a less familiar example, (4) a bank policy which discriminates against local college students by restricting the amounts they may withdraw from automated teller machines: surely the relevant categories are not students and non-students but responsible and irresponsible customers; nevertheless the bank presumes all students to be bad risks, no doubt on the basis of an unhappy (if imperfect) correlation.

I do not mean to imply that these examples of SD are morally equivalent. Indeed, my experience has been that the average person reacts quite differently depending on which of the four is at issue. Nor do I believe that such a short list of examples

---

Copyright 1991 by *Social Theory and Practice*, Vol. 17, No. 1 (Spring 1991)

accurately reflects the broad range of policies which do (or might) rely on SD.<sup>1</sup>

This paper sketches a framework for the systematic analysis of SD which at the same time accounts for the complexity of the issues and our widely divergent intuitions about various policies which employ SD. I hope to provide a theory (based on comparing information-cost and net social utility) powerful enough to cover a wide array of policies without obscuring the very real complexities involved.

### **1. Legitimacy of Purpose**

For the sake of this discussion, let us start by assuming that a policy involving SD has a legitimate underlying purpose. That is, my theory will concern policies whose fundamental goals we have already determined to be defensible. To use examples (1)-(4), I take it that the underlying characteristic is in each case a legitimate target of social policy: in (1) we seek a match between premia and claims, in (2) we seek to identify and help disadvantaged victims, in (3) we try to adjust contributions to reflect benefits, and in (4) we try to restrict the privileges of irresponsible customers in order to minimize the costs of their irresponsibility. I think we can agree that, at bottom, such goals are worth pursuing and that in a perfect world (with no mismatch between our goals and our deeds) we would pursue them without a second thought. Indeed, for present purposes we might call a policy goal legitimate just in case we would choose to pursue the goal directly given cost-free and perfect information.

In our world, however, there is plenty of disagreement over the proper means of achieving such goals. Leave aside policies whose fundamental ends are indefensible, such as using race as an index of moral inferiority; policies like (1)-(4) are controversial quite apart from the (presumed) legitimacy of their goals. No doubt some persons will disagree with the very goals, and I will not try to change their minds here. Instead, I will suppose that the underlying purpose of a policy is legitimate; my theory will then provide analysis of the means used to achieve that purpose.

## **2. An Information-Cost Approach**

Information-cost explanations of SD have been available at least since the publication of Edmund Phelps's important paper "The Statistical Theory of Racism and Sexism."<sup>2</sup> Phelps refined the information-cost analysis by offering a statistical explanation of the tendency to discriminate instead of just assuming such a tendency as given. However, Phelps's paper does not pretend to offer a moral theory of SD. Rather, he analyzes SD solely in terms of the information-cost and utility of the discriminator: an instance of SD is rational when it "maximizes the expected utility" of the agent (Phelps, 659). Similarly, Richard Epstein defends certain kinds of SD on grounds of economic efficiency. Arguing on behalf of sex-based pension plans (example 3 above) Epstein writes, "Market processes select the most homogeneous subgroup on which it is possible to obtain aggregate statistics at reasonable cost."<sup>3</sup> Yet, like Phelps's account, Epstein falls short of an ethical defense of SD since among other things it fails to specify just what cost is "reasonable." Surely such specification requires reference to objective criteria which are external to the discriminator.

On the other side of the coin are theories which focus on the social cost of SD while inadequately addressing the information-cost of the discriminator. After a long and careful analysis, Robert Fullinwider ends up giving provisional approval to race-based preferential hiring on the basis of its social utility,<sup>4</sup> but he fails to recognize that systematic comparison of social utility and information-cost lies at the root of a sound moral theory of SD. Other writers, such as Alan Goldman, Louis Katzner, and James Nickel, acknowledge the importance of both factors in the justification of SD, but their comparisons of the factors seem to me neither explicit nor systematic.<sup>5</sup> I would like to provide a framework which better connects information-cost theory with judgments of SD based on social utility.

I believe a sound moral theory of SD must view information-cost and social utility as functions of one another. More precisely, the theory must plot the information-cost of a policy involving SD against the net social utility (exclusive of information-cost)

generated by the policy, and then examine any optima which emerge. Reflection on such a process will, I think, lead us to the following general conclusion: a social policy involving a given level of SD is justified if and only if the information-cost of further statistical refinement equals or exceeds the net social utility to be gained by such refinement.

This thesis needs several clarifications. The implementation of a policy which employs SD (an "SD policy," for short) will rely on statistics at some particular level of aggregation, for example, a statistical correlation between gender and longevity. That level of aggregation (i) imposes information-cost in some amount or other on the discriminator, and (ii) generates a degree of social disutility on account of the imperfect nature of the correlation. The information-cost is, roughly, the cost of obtaining reliable data at that level of aggregation. The disutility arises from the mismatch between the target characteristic and the characteristic used to aggregate the data.

To use the example of affirmative action, the disutility of the SD policy might include false positives (inclusion of blacks who are not disadvantaged), false negatives (exclusion of whites who are disadvantaged), and wasted resources (which might have been "spent" on genuinely disadvantaged blacks and whites). Granted, a mismatch between racial characteristics and genuine disadvantage does not pose problems for affirmative action programs designed to remedy underrepresentation pure and simple. If these programs intend solely to increase the numbers of minorities in professions until those numbers are proportioned to the minority population, then presumably race is the target characteristic. Clearly, such programs don't qualify as instances of SD. However, most affirmative action plans seek to remedy underrepresentation which is the result of racism, unfair disadvantage and the like. Indeed, this is crucial to the defense of affirmative action based on compensatory and other forms of justice. Thus, the plans usually seek to identify disadvantaged members of minority groups, and doing so has as a matter of fact entailed the use of SD. Moreover, SD will continue to figure in all other kinds of activities (such as insurance) which try to target an

underlying feature (such as risk) and not the just the index for that feature.

The foregoing example is merely illustrative and does not represent a careful application of the theory. Notice, however, the connection between information-cost and utility. Presumably, we could reduce the disutility in the example by using a more refined statistical aggregation, one which (say) grouped according to race and family income. But doing so would undoubtedly entail higher (if only slightly higher) information-cost for collection and analysis of the additional data. But increasing statistical refinement will also tend to reduce the disutility (or "social cost") arising from the imperfect correlations used in SD. As Alan Goldman notes, reducing the social cost of SD requires "narrowing of the specification of the group so as to maximize the correlation, but of course at the limit of such narrowing is a program administered on an individual basis."<sup>6</sup>

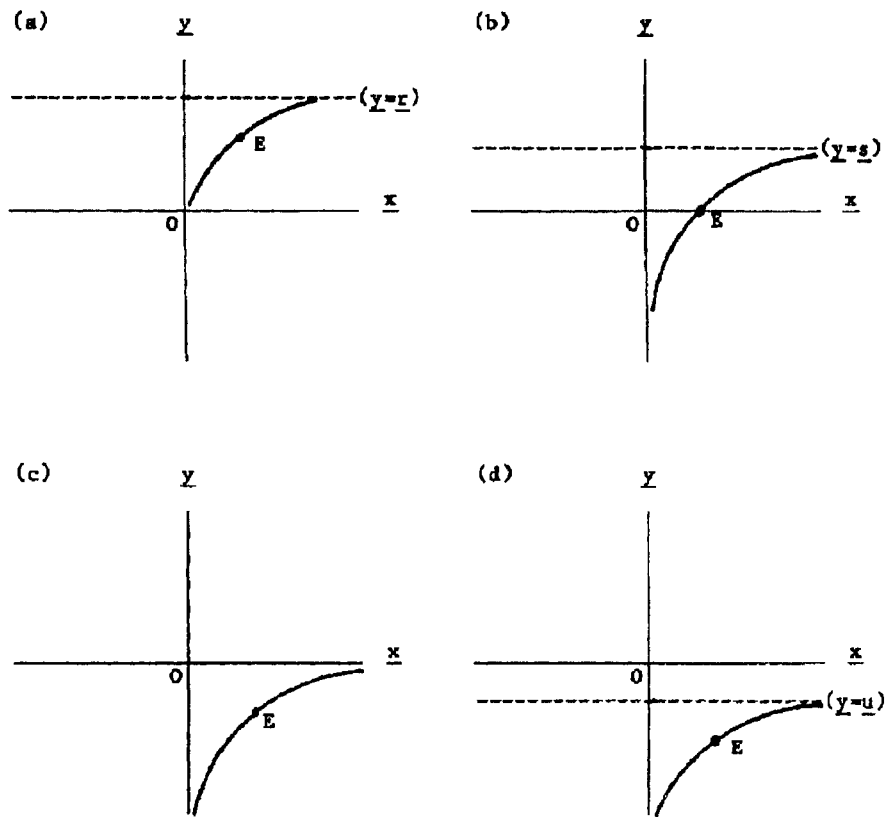
I take the "net social utility" of an SD policy to mean the sum of social costs and benefits generated by the policy: its positive social effects (which can include unintended benefits) less its negative effects (which are presumably not intended). Assuming it is possible to quantify the net social utility of an SD policy, we can plot that quantity against the associated information-cost, obtaining a curve. Figure 1 (on the following page) shows the curves for some sample policy-types.

Notice several features of these curves. First, I assume a functional relation between the two variables; that is, I assume that a given level of information-cost determines a unique level of net social utility. Given that any SD policy can be implemented with greater or lesser statistical refinement, I assume that the level of utility generated will depend on the level of refinement.<sup>7</sup> The actual shape of the policy-curve may vary, but it will depict a function in any case. This assumption is justified by the expectation that, at a given level of efficiency, increasingly costly statistical aggregations will tend both to reduce the social cost of imperfect correlation and to increase the benefits of the SD policy by enabling it to reach a greater proportion of the target class.

The crucial feature of the curves in Figure 1 is the point, E, of diminishing marginal utility. E represents the point on each curve

Figure 1

Sample Curves for Policies Involving SD



$x$  = Total information-cost

$y$  = Net social utility

At E,  $dy/dx = 1$

Asymptotes:  $y = r$ ,  $y = s$ ,  $y = 0$ ,  $y = u$

## The Ethics of Statistical Discrimination

29

where the slope of the tangent (the derivative of  $y$  with respect to  $x$ ) equals 1. It is the point of maximum (or optimum) efficiency in the trade-off between information-cost and net social utility. In an important sense,  $E$  also represents an equilibrium point: at any point below  $E$  on the curve, an increase in information-cost produces more than an equivalent increase in net utility. However, at any point above  $E$  the marginal information-cost exceeds the marginal social utility to be gained by greater statistical refinement. Thus, in a sense yet to be defined, it's worthwhile below  $E$  to spend more on information, but not worthwhile above  $E$ .

Think of each point on the curve as representing a particular implementation of an SD policy. Each implementation incurs its own level of information-cost and produces its own level of net social utility. It will be useful to talk about the justice of an implementation as well as the justice of an entire SD policy. With these concepts in hand, we can give a refined version of the main thesis:

(T1) An implementation of an SD policy is justified if and only if marginal information-cost equals or exceeds marginal net social utility.

In other words, an implementation of some SD policy (that is, at a point on the policy-curve) is justified just in case the cost of improving the information (using more refined statistics, perhaps) equals or exceeds the net social utility to be gained by the improvement. If, on the other hand, an increase in information-cost creates more than its equivalent in increased net utility, failing to undertake that cost is unjustified. Additional features of the theory need clarification. To begin with, the policy-curves in Figure 1 are of course idealizations. Practically speaking, there are fewer possible implementations of a policy than there are points in the curve: changes in information-cost may be necessarily discontinuous. To take account of discontinuity, we could modify (T1) to state that an implementation is justified if and only if there is no feasible increase in information-cost that would buy more than its equivalent in extra social utility.

Second, curves (c) and (d) depict what might be called "degenerate cases," SD policies for which no amount of information-cost produces net positive social utility. I should think

we would demand that any implementation of SD generate net positive utility in order to be justified. Since according to (T1) any implementation at or above the equilibrium point E is *eo ipso* justified, we should add the requirement that the y-value at E be positive (that E lie in the first quadrant). Thus, we obtain two additional necessary conditions:

(T2) An implementation of an SD policy is justified only if the policy itself is justified.

(T3) An SD policy is justified only if the value of net social utility at the point of equilibrium is positive.

We can use these conditions to compare different policies, not just different instances of a single policy. We can rule out not just implementations which lie below E on the curve but entire policies for which the y-value at E is zero or less. Notice that each of the curves is asymptotic to a line perpendicular to the y-axis. I've made them asymptotic because it would seem there must be a limit to the social utility produced by any SD policy, regardless of the information-cost incurred to implement it. Unlike (T1)-(T3), however, the asymptotes are not essential to the theory.

Returning to the subject of social cost, in what sense is the disutility of an SD policy a cost to society? Example (3) (pension plan contributions) illustrates a typical case. Here, the disutility of SD will undoubtedly include misidentification and the inefficiencies which result from it: false negatives (men with high life-expectancy), false positives (women with low life-expectancy), and (thus) cases in which pension benefits are not proportioned to contributions. If SD is common in the design of pension plans, the inefficiencies will be social costs in at least the sense that lots of people will pay too much or too little for their benefits. Of course, the costs of occasional misidentification may be swamped by the overall efficiency of a pension system which discriminates according to sex. If the net effect of SD in pension plans is positive, it then becomes a case of comparing the cost of refining the data (for example, using, besides sex, information on medical history and lifestyle) with the benefit gained by the refinement.

In the remainder of this paper, I will expand and refine the basic theory in the course of addressing objections to it. I am especially



concerned to show how it qualifies as a normative theory of SD, how it can guide the morally justified implementation of SD policies.

### **3. Efficiency and Justice**

Our theory approves of instances of SD just in case they're efficient, just in case they maximize expected utility. However, I have stressed a broader sense of efficiency than the one employed by traditional information-cost theories of discrimination. A moral theory of SD requires comparing information-cost with social utility, not just the utility of the discriminator. But why is an SD policy morally justified if and only if it is efficient in the appropriate sense?

It should be obvious by now that I adopt a broadly utilitarian approach to social policy. But I will not attempt a full-blown defense of utilitarianism here, since my theory does not require one. The view that the morality of SD involves questions of utility-maximization does not depend on the global success of utilitarian ethics. One can apply utilitarian criteria to the analysis of SD without believing that they resolve every other question of justice as well. Just the same, I do believe in a general consonance between net social utility and (for instance) fundamental rights in the context of SD. This consonance will become clearer as I discuss specific challenges to the utilitarian framework of the theory. Moreover, I should stress that nearly all analyses of SD (whether favorable or not) rely to some extent on considerations of utility, whatever else they employ. Indeed, some of the most careful analyses end up judging SD primarily on the basis of its social utility (for example, Fullinwider).<sup>8</sup>

Before moving on, I should try to prevent a potential misinterpretation of the theory. As I say, we ought to judge SD at least partly with regard to its marginal information-cost. Why marginal? Why start the discussion in terms of improving an implementation of SD? Does this reflect a basic presumption against the use of SD? No. It simply reflects a basic presumption against inefficiency in social policy. The "marginalist" approach

## 32 Social Theory and Practice

of the theory reflects the belief that it is always rational to spend more on refinement when marginal information-cost is less than marginal social utility—and always irrational to do so when the inequality is reversed. Indeed, given the utilitarian bent of the theory, any initial presumption against SD would have to rely on a general fact about the consequences of SD, namely that SD generally creates negative net utility. But the general effect of SD is an empirical question to which no one yet knows the answer, so it surely cannot form the basis of an initial presumption about SD at this stage of the game.

### 4. Whose Information-Cost?

Exactly whose information-cost is relevant to the justification of SD, and why is it relevant? Surely it is the information-cost of the statistical discriminator that is peculiarly relevant, for the following simple reason. Assume that a given SD policy is reasonably efficient in translating additional information-cost into additional social utility. This doesn't mean that the policy won't suffer from the diminishing marginal utility of information; indeed, we have assumed that it must (starting at the equilibrium point, E). Rather, I assume that the diminishing marginal success of the policy is attributable to the diminishing marginal utility of information and not to some perverse fluctuation in the efficiency with which the policy uses information to achieve its ends. From this simple assumption it seems clear that the net social utility of SD depends on the information-cost incurred by the discriminator (but see note 11, below). In the case of insurance rates, for example, the relevant information-cost is that borne by the insurance company (or the entire insurance industry if the data are pooled). With state-mandated affirmative action, presumably the government must incur the cost of information. And so on. But this analysis might be considered vulnerable to the following objection.

There are said to be "morally arbitrary" aspects to the actual cost and availability of information. Any normative theory relying heavily on information-cost is said to incorporate that arbitrariness. The availability of aggregate statistical data depends on what data

## The Ethics of Statistical Discrimination

33

society chooses to collect. There may be many statistically useful questions that we neglect or explicitly choose not to ask. Some kinds of data-collection might be considered too intrusive or just not worth the effort. How, then, can a normative, information-cost theory succeed when we might not possess information which in some sense we ought to possess? I will say more about intrusiveness when I discuss the right to privacy (section 6). Meanwhile, let me address the question of moral arbitrariness.

Granting that the availability and cost of information really are to some degree arbitrary, it's not clear that any moral theory which starts from that reality is thereby arbitrary. Take an extreme (but parallel) example. Most theories of distributive justice would agree that, all other things equal, it would be nice if everyone's reasonable, noninjurious desires were at once satisfied. But resources are surely too scarce to satisfy that aggregate desire; they are scarce even when we restrict ourselves to the provision of basic needs. Surely the sad fact about scarce resources is in some sense morally arbitrary. Just the same, few theories of justice would require endless redistribution in the vain attempt to satisfy every desire, even every reasonable desire. This is, in part, because good moral theories recognize at least some of the constraints imposed by initial conditions. They recognize, for example, whatever constraints on utopia might result from the scarcity of economic goods. Or they recognize the importance of actual information-cost in the moral assessment of SD, even though that cost has arbitrary dimensions. Moreover, I should stress that our theory of SD does view information-cost from an ethical standpoint, namely from the standpoint of its marginal social utility. Thus, in many instances the theory will demand statistical refinement for the sake of the social good. In some cases that refinement may require expanding the amount of data we gather, keeping in mind the social cost (for example, the intrusiveness) of gathering it. The theory brings moral pressure to bear on the very collection of information.

### 5. Equilibrium and Incentive

Let us examine further the notion of "equilibrium" to which I've frequently referred. Our theory assumes, uncontroversially, that the optimal implementation of an SD policy occurs at E, the point on the curve for which  $dy/dx = 1$ . We can think of E as the point of equilibrium because there is pressure to approach E from above and from below. Any implementation above E is justified but more costly than necessary; so there is economic pressure to approach E from above. Any implementation below E is certainly cheaper, but it is unjustified on grounds of marginal efficiency, so there is moral pressure to approach E from below. The two kinds of pressure, economic and moral, ensure that implementation tends toward equilibrium. In order to exert the latter kind of pressure, our theory need not provide for degrees of justification; all we need is a means of deciding one way or the other about an instance of SD. Put another way, the "yes or no" test of justified implementation counterbalances the economic incentive the discriminator already has to make the policy cheaper. It provides moral incentive to spend more, to complement the economic incentive (already there) to spend less.<sup>9</sup>

I need to say more about that moral incentive and about the relation between marginal information-cost and marginal social utility. So far the discussion has assumed that we measure cost and utility in commensurable units, an assumption which underlies the comparisons required by our theory. However, I expect the objection that we typically measure information-cost in units of money and utility in "utils" (whatever those are), and I have said nothing to indicate how they are commensurable. At one level, this objection doesn't matter to the theory, but at another level it matters deeply. Let me explain why.

At one level, the theory just suggests evaluating an SD policy not simply by determining its administrative cost or its effect on society but by (marginally) comparing both features of the policy. Moreover, this methodology is supposed to account for our widely divergent attitudes toward SD and to provide a way of forming the right judgment about particular policies. At these tasks the unembellished theory partly succeeds. It identifies (broadly) the

## The Ethics of Statistical Discrimination

35

crucial comparison; it suggests (see below) that the intricacy of utilitarian calculation explains why people differ about the ethics of SD; and it proposes that, in spite of their intricacy, utilitarian calculations are key to the correct moral assessment of SD policies.

At a more detailed level, however, the theory needs considerable elaboration. What kind of metric will make the units of cost and utility truly commensurable? Marginal comparisons are crucial to fixing the point of equilibrium: how can we calculate the marginal social utility of an SD policy? Finally, how can we translate social utility into real incentive for the statistical discriminator?

Let me sketch tentative answers to these tough questions, taking my cue from Richard Posner, whose ethical theory prefers the maximization of wealth to the maximization of utility.<sup>10</sup> Posner's economic approach does appear to solve some of the notorious problems of utilitarianism. Moreover, by considering, say, dollars instead of utiles we can measure both variables in the same units. Nevertheless, the pure economic view seems insensitive to some costs of SD which we should not ignore, so I adopt Posner's approach only with substantial modification.

We can measure both marginal information-cost and marginal social utility in dollars. Marginal information-cost is borne by the statistical discriminator, as described earlier. Marginal social utility is defined as the amount society would be willing to pay, in the aggregate, for a given increase in the statistical precision of an SD policy. Because high transaction-costs prohibit an actual exchange between the discriminator and society, we resort to a "hypothetical market" to determine the values of the policy-curve. I suggest below a mechanism for conducting the hypothetical exchanges.

Suppose that an improvement in the precision of his statistics costs the discriminator an amount  $X$ . Suppose that society, in the aggregate, would be willing to pay him  $Y$  to undertake the improvement. Then wealth maximization favors the improvement just in case  $X < Y$ . Since prohibitive transaction-costs preclude an actual market exchange, some other mechanism must enforce wealth-maximizing behavior. One mechanism that Posner suggests consists of imposing efficient liability rules.<sup>11</sup> Thus, for example, we might entertain a class-action lawsuit by appropriately situated plaintiffs: the statistical discriminator

would be liable if and only if  $X$  (the marginal information-cost) is less than  $Y$ , where  $Y$  represents the expected marginal value (to the plaintiffs) of the refined implementation. Ideally, enforcing this rule would have the effect of encouraging all and only wealth-maximizing expenditures on information.

This approach has its problems, however. First, Posner's theory has the "put up or shut up" flavor of most economic analysis: roughly speaking, nothing counts as a social cost unless society (in the aggregate) would be willing to pay money to avoid it. Thus, it is not clear that Posner's analysis captures the noneconomic costs (if any) of SD, such as injustice, demoralization or negative symbolism. But it might, provided society were willing to pay to avoid or reduce those effects of SD.

Second, and more important, it's not clear how to identify the right buyers in our hypothetical market. Suppose that the class of young, cautious male drivers considers itself unjustly harmed by SD which unfairly benefits the class of young, unsafe female drivers. The auto insurance rates in most states would seem to justify this feeling of unfairness. Suppose, further, that these men would collectively pay for statistical refinement which made their rates more consistent with the risk they impose. Say the refinement costs the insurance industry  $X$ . Let  $Y$  equal the expected value (to the men) of the refinement: the aggregate difference between their current rates and the (lower) rates they'd expect to pay a competitive industry which possessed the refined data. As rational wealth-maximizers, the men would collectively pay any amount less than  $Y$  for the refinement. We find the industry liable if  $X < Y$  (since, if so, the men would pay any amount between  $X$  and  $Y$ ). However, on the pure economic view, the difference between the improved situation and the status quo is just "distributional": imposing liability (in the amount of  $Y$ ) simply transfers  $Y$  from one class to another, with no necessary effect on the total wealth of society. Even if the transfer leads to more efficient behavior by both classes of drivers, the improved efficiency bears no necessary relation to the magnitude  $Y$ . Thus, if the goal is wealth maximization, we've chosen the wrong plaintiffs.

I answer these objections by departing from the pure economic view, in the following way. Presumably the goal of SD in

insurance is to match premia with objective risks. One's risk as a driver depends on one's objective propensity to cause collisions, and insurers use statistics as a means of estimating that propensity. Increasingly refined statistics can improve the estimate, up to the point at which the match between premium and risk becomes as close as we can make it. But it may be socially inefficient to reach the point of maximum statistical accuracy: if that point occurs well above E, then the marginal cost of reaching it exceeds the marginal benefit of reaching it. Still, a match between rates and risk remains the policy goal (to be pursued efficiently). Given that goal, then, I want to insist that Y is the relevant cost in the above example. In terms of the goal, the practice misallocates by an amount equal to Y, even though the effect of the misallocation on GNP has little or nothing to do with Y. Thus, in our idealized example we've chosen the right plaintiffs, and they recover iff the misallocation of Y could be prevented by a smaller expenditure of X.

The situation gets more complicated as the number of injured parties increases, but the approach is the same throughout. A class of persons is injured by an SD policy just in case those persons (acting as rational wealth-maximizers) would pay to improve the statistical aggregation which grounds the policy, that is, just in case they would pay for stricter adherence to the goals of the policy. Whether they recover depends on the relative cost of the improved aggregation. On this view, the cautious male drivers would be legitimate buyers in our hypothetical market. But the unsafe female drivers would not be, even though they might be willing to pay to maintain the status quo.

Thus, in evaluating an SD policy, our modified economic approach makes essential reference to the (presumably legitimate) goals of the policy. This has two consequences. First, it means that we seek not maximization of total wealth but efficiency in the pursuit of individual policy goals. The penalties for inefficient conduct aren't measured by the effect of that conduct on GNP; they're measured by the gap between the goal of the policy and the actual practice. Our approach says, in short, if you use SD to pursue a policy goal you had better do it efficiently. Second, the evaluations get complicated when a policy involving SD has multiple goals. No doubt the purpose of many affirmative action

programs, for example, is to improve the lot of the disadvantaged. The pursuit of this goal involves SD insofar as a trait such as race is used as an index of disadvantage. But other affirmative action programs might use SD to pursue a different end (say, diversity among students). Some goals, such as increasing minority representation, needn't involve SD at all. Consequently, our theory won't provide the same judgment for every possible instance of affirmative action. Assessing the efficiency of an SD policy requires taking account of its goals.

I have just sketched a hypothetical-market mechanism for comparing marginal cost and utility in commensurable units, for determining the marginal utility of an SD policy, and for encouraging the statistical discriminator to behave efficiently in the pursuit of policy goals. As a solution, the mechanism strikes me as a first approximation whose consistency with the rest of the theory is debatable. In the remainder of this paper, I shall return to broader questions of cost and utility without further reference to the market.

## 6. Fundamental Rights

The question of fundamental rights and SD really comprises a cluster of issues. Those issues include, if I may use shorthand labels, the following: (i) immutable traits, the view that SD based on immutable traits (such as race) is almost always unjustified, since people have a basic right not to be discriminated against on account of such traits; (ii) lifestyle discrimination, the question whether we should permit SD based on the personal choices people make, for example, about marriage, sexual orientation (assuming that's a matter of choice), smoking, and so on; (iii) constitutive rights, including the view that some rights are so fundamental as to bar any discrimination which infringes them; and (iv) the public/private distinction, the belief that we shouldn't interfere with discrimination (including SD) when it involves the "private" choices of persons even though similar discrimination in the public arena would be intolerable. This list is not exhaustive; nor have I stated every view in all its subtlety. But the list is representative,



## The Ethics of Statistical Discrimination

39

and it will suffice to show our theory's wide range of application to questions of rights.

I want to suggest the same approach to fundamental rights as the theory takes to other issues. That is, I want to suggest an analysis of rights in terms of efficiency and net social utility. Opponents of such a strategy will insist that my utilitarian calculus misses something fundamental about our intuitions concerning rights. I realize that the issues between utilitarians and rights-theorists are deep and enormously difficult. I do not pretend to have found a resolution of those issues, much less one that I can offer here. Therefore, the following account is tentative and suggestive. Still, I consider it a promising start on an analysis of the relation between SD, rights, and utility.<sup>12</sup> Again, I do not defend utilitarianism in this paper, but I think that considerations of social utility are particularly well suited to the normative analysis of SD.

No rights are so fundamental as to be absolutely inviolable. For example, rights of ownership, habeas corpus, free speech, and even the right to life can be compromised under appropriate circumstances (such as war or emergencies requiring the imposition of martial law). I don't mean to suggest that every historical instance of their compromise has been appropriate or justifiable. That's tragically far from true. But I think even most libertarians can conceive of circumstances which would justify abridging some basic right or other. Sometimes rights may be overridden. Another way of putting this, I suggest, is that there are no rights whose violation creates infinite disutility.

I'm not suggesting that we should take rights lightly. I merely propose that even basic rights can be analyzed in terms of finite utility and disutility. Moreover, it is not conceptually necessary that SD conflict with the most basic kinds of rights. Whether it does seems to me an empirical question whose answer depends on the type of SD.

Take an SD policy which discriminates according to an immutable trait such as race. We may believe that people have a right not to be discriminated against on account of traits they cannot control. On the approach I am suggesting, since no right is absolutely inviolable we assign a finite (but high) disutility to the

violation of that right. Consequently, an SD policy which violates it incurs that disutility. Although the numbers will differ, the same general analysis applies to discrimination based on lifestyle choices (marriage, smoking) and the like. What might the curve look like for an SD policy which unavoidably infringed a basic right? I can think of three possible answers consistent with our theory: (a) The policy might fail to identify a legitimate goal underlying the SD (for example, a policy which denied education to blacks because they were deemed uneducable); (b) The policy might fail to generate net utility at equilibrium, thus failing to meet condition (T3) for justification; or (c) The policy might satisfy (T3) but nonetheless require extremely high information-cost in order to be justified under (T1).

Among rights, the right to privacy is especially interesting for an information-cost theory of SD. Presumably that right places constraints on the intrusiveness of data collection. In other words, the search for more information may create its own disutility even as it attempts to reduce the disutility of statistical imprecision. Undoubtedly this tension will affect the marginal utility of refinement and hence the overall utility of SD. Nevertheless (I am suggesting) the right to privacy, like other rights, admits of analysis in terms of net utility.

Finally, let me address the public/private distinction. Like the other issues in this section, that distinction can be viewed in terms of social utility. In the literature on this topic, the archetypal example of (legitimate) private discrimination is a woman's choosing among suitors for her own reasons (whatever those reasons may be). As David Thomasson notes,<sup>13</sup> there may be good reasons for regarding the woman's choice as morally objectionable; suppose she capriciously abandons her fiancé, Dick, at the altar after promising to marry him, thus dashing his "legitimate expectations." Nevertheless, we would not consider ourselves justified in interfering with her choice. But, Thomasson argues, the morality of both private and public discrimination depends on their social utility. There is only a quantitative difference between Mary's capricious treatment of Dick and an employer's immoral discrimination among applicants:

## The Ethics of Statistical Discrimination

41

Thus the sole criterion of [justifiability] is social utility.... Why is Mary's choice of a particular spouse wrong? Because of the resulting disutility. Why is the employer's discrimination wrong? Because of the resulting disutility. What permits the enforcement of a rule against discrimination in employment? The relatively high disutility that would accrue from recognizing the employer's right to discriminate. And what bar is there to enforcing the rule [against Mary]? The resulting disutilities are insufficiently extensive.<sup>14</sup>

Thomasson's analysis is more subtle and complex than I have made it seem. What is important here, though, is his suggestion that the familiar public/private distinction ultimately turns on questions of social utility. I think that's correct. Indeed, consequentialist considerations figure prominently in the jurisprudence of public and private discrimination: in decisions outlawing racial discrimination in public accommodations, for instance, the Supreme Court has frequently emphasized the social disutility of the practice.<sup>15</sup>

What's the upshot of all this for the theory? For one thing, it means that we might at once account for and transcend the public/private distinction in matters of SD. The theory accounts for the distinction by tracing it to quantitative differences in social utility. It transcends the distinction by comparing the information-cost of an SD policy with the net utility of the policy, regardless of the public or private nature of the discriminator. Hence, some instances of "public" SD (for example, affirmative action at a state college) might turn out to be justified while an instance of so-called "private" SD (for example, a banker's discrimination against students) turns out to be unjustified. It's a matter of the empirical relation between marginal information-cost and marginal social utility.

### 7. Conclusion

Several considerations recommend the theory of SD I have sketched in this paper. The theory provides a principled, systematic framework for assessing the variety of social policies which employ SD. It covers a wide array of rationales both for and against SD, and it can serve to arbitrate among them. It connects information-cost and social utility, and it avoids the

defects of moral theories which rely entirely on one concept or the other. It shifts the argument from an arena in which conflicting moral perspectives talk past one another to an arena in which questions might admit of empirical resolution. The empirical questions are complex, to be sure; I think it's that complexity which accounts for our widely divergent moral intuitions about SD. We should not fault the theory if, pending the resolution of knotty empirical questions, it allows reasonable people to disagree.

The empirical problems raised by the theory are not, in principle, insoluble. After all, in deciding what statistics to collect insurance companies already make detailed comparisons of marginal information-cost and marginal revenue. But the unsolved problems are substantial and interesting. What are the relevant empirical questions concerning SD, to the extent the theory hasn't already specified them? What are the answers to those questions? What do the policy-curves look like for different SD policies, that is, what are the functions? I suspect they'll resemble those in Figure 1. Where does a given implementation lie on the curve? Are the equilibria of SD policies stable? If instituted, does the enforcement mechanism sketched above work? And so on. We don't know all the answers, and in that light I should stress that the curves in Figure 1 are chiefly heuristic devices for facilitating the discussion. I've given reasons for suspecting they're approximately right, but they're not crucial to the theory. The central tenets are, instead, the normative principles (T1)-(T3).

My theory attempts to analyze a wide array of issues in terms of information-cost and social utility. Yet it is not so broad that it lacks distinctive consequences. The theory rejects purely "epistemic" justifications of SD: the statistics used to identify the target characteristic must be appropriately refined. It shows that strategies which attack SD because it relies on imperfect correlations, or because it violates rights, don't automatically succeed: the appropriate marginal comparisons must be made. And it also shows how some SD policies might be flawed at two separate levels. Recall example (4), the bank which discriminates against college students: first, the criteria actually used to identify "students" (age and length of residence) are imperfect; second, the statistical correlation between students and irresponsible

customers is presumably imperfect. These imperfections compound the problem of (for example) false positives, making the implementation, in my view, highly suspect. I doubt that it can be justified on (T1). Posner (p. 369) notes similar features in the implementation of affirmative action: the means of identifying members of racial minorities are imperfect, as is the correlation between race and (for example) disadvantage. Of course, as Posner recognizes, this observation alone does not condemn affirmative action. The normative analysis of SD is a comparative task.

In discussions of SD, I seem to encounter people from two general camps: those who argue from the neoclassical economic perspective ("the economists"), and those who judge SD primarily in terms of such concepts as fairness, justice and rights ("the deontologists"). The economists approve of SD just in case it maximizes total wealth. Charges of injustice and the like don't normally move them unless the injustice can be shown to diminish GNP (a nontrivial task since there's no necessary connection between the two). The deontologists, by contrast, believe that right and justice are paramount and almost always trump considerations of utility and efficiency. Deontologists, for example, often tell me that SD in insurance is "just plain wrong" regardless of its supposed efficiency. Maybe I've drawn the contrast too sharply (and there are plenty of people who fall in between), but the basic dispute between the two camps is live and inescapable.

As I've suggested before, my theory seeks to combine the insights of both sides. In general, I adopt the utilitarian approach of the economists. But the enforcement mechanism I propose ranks other values above (though it needn't conflict with) the maximization of total wealth. And to the deontologists, I say: assuming some fundamental goal is worth pursuing, an SD policy may provide the most efficient way to pursue it. I acknowledge the centrality of goals in the moral assessment of SD, but I also recognize the importance of efficiency to the social good.<sup>16</sup>

Notes

1. I take it that all four examples might legitimately be called "social policies": they all have social impact to one degree or another. As for "SD," we might broaden the definition of that term by calling anything SD which exhibits the functional relation between information-cost and social utility that I describe.
2. Edmund Phelps, "The Statistical Theory of Racism and Sexism," *American Economic Review* 62 (1972): 659-61.
3. Richard Epstein, *Takings: Private Property and the Power of Eminent Domain* (Cambridge: Harvard University Press, 1985), p. 313, n. 15.
4. Robert Fullinwider, *The Reverse Discrimination Controversy* (Totowa, New Jersey: Rowman and Littlefield, 1980), pp. 249-50. Fullinwider adopts a rule-utilitarian justification of race-based preferential hiring, but—I should stress—only after concluding that preferential hiring violates no genuine rights.
5. See, for example, Alan H. Goldman, "Reparations to Individuals or Groups?" and James W. Nickel, "Preferential Policies in Hiring and Admissions" (esp. p. 341), in Barry R. Gross, ed., *Reverse Discrimination* (New York: Prometheus Books, 1977); and Louis Katzner, "Is the Favoring of Women and Blacks in Employment and Educational Opportunities Justified?" in Joel Feinberg and Hyman Gross (eds.) *The Philosophy of Law*, 3d ed. (Belmont, CA: Wadsworth, 1986), esp. p. 360.
6. Goldman, "Reparations to Individuals or Groups?," p. 322.
7. I assume the two variables are functionally covariant, and I'm skeptical about alleged counterexamples, for example, a low-cost, over-inclusive instance of SD which creates greater net utility than a more costly and refined version: the cheaper version is wasteful and also liable to under-include (which would reduce its net utility).
8. Fullinwider, *The Reverse Discrimination Controversy*, Chapter 15.
9. Here one might accuse me of ignoring the behavior of government bureaucracies, which (on some theories) have every incentive to spend as much as possible in order to increase their annual budget allocations. No doubt many bureaucracies behave this way, and many are in charge of implementing SD policies. Nevertheless, I assume a degree of classical economic rationality on the part of government agencies in the conduct of individual SD programs (though they might expand the number of such programs to soak up money). Even government agencies are (increasingly) constrained by the budget ax. Moreover, this perverse incentive won't affect the SD policies of profit-motivated private firms.
10. Posner develops the wealth maximization theory of justice in *The Economics of Justice* (Cambridge: Harvard University Press, 1983), esp. pp. 48-115.

## The Ethics of Statistical Discrimination

45

11. Posner, *The Economics of Justice*, p. 70. In fashioning a liability rule for SD, I have assumed that the information-cost of refinement is borne most cheaply by the statistical discriminator. However, I concede that in some cases it might be cheaper, in the aggregate, for the class of plaintiffs (potential victims of SD) to conduct the desired refinement than it would be for the discriminator to conduct it. In such cases, the efficient liability rule must exculpate the discriminator, and so must the liability rule that I propose.
12. Amartya Sen, I should note, has already defended an intriguing combination of rights-theory and (non-utilitarian) consequentialism. See his "Rights and Agency," *Philosophy & Public Affairs* 11 (1981): 3-39.
13. David Thomasson, "Rights, Wrongs, and Discrimination," *Harvard Journal of Law & Public Policy* 11 (1988): 831. Thomasson distinguishes throughout between "justifiable" and "permissible" acts, so his use of those terms no doubt differs from my own. (He believes that a woman's morally offensive discrimination among suitors is permissible without being morally justifiable.)
14. Thomasson, "Rights, Wrongs, and Discrimination," p. 831.
15. See, for example, *Heart of Atlanta Motel v. U.S.*, 379 U.S. 241 (1964), and *Katzenbach v. McClung*, 379 U.S. 294 (1964).
16. For helpful comments on an earlier version of this paper, I wish to thank Anthony Appiah, Sterling Harwood, George Hay, and referees from this journal.

**Stephen Maitzen**  
Sage School of Philosophy  
Cornell University