

ARTICLE

CONCEPTS OF DISCRIMINATION IN
"GENERAL ABILITY" JOB TESTING

Mark Kelman*

Federal civil rights laws prohibit employers from screening potential employees by means of general intelligence tests if the tests are not related to job performance and have a disparate impact on a protected group. Professor Kelman argues that the law takes too narrow a view of discrimination to capture all of the harms caused by such tests. Professor Kelman explores four concepts of discrimination in the context of general ability job testing and argues that in many instances the legal system's distinctions between different types of discrimination are unprincipled, because they are based on false intuitions about desert and unexamined assumptions about the desirability of private ordering. Professor Kelman argues that the use of general ability job tests that have a disparate impact on minorities should pose significant philosophical problems even for those who reject more expansive notions of discrimination. He criticizes testing proponents' claims that general ability job tests are highly predictive of job performance and argues that, even if such tests are both weakly predictive and culturally unbiased, they may fail to comport with the demands of liberal, individualistic meritocracy. This failure is of particular concern when these nonmeritocratic tests burden members of historically oppressed groups.

I. INTRODUCTION

As a group, blacks score significantly lower on "general ability" tests than do whites¹ and will thus obtain employment less frequently than white applicants if employers use such tests as a basis for hiring decisions. Proponents of testing believe that the tests accurately predict on-the-job performance and exclude only those people

* Professor of Law, Stanford University. I am grateful for comments on earlier drafts made by Joe Bankman, John Ely, Tom Grey, Toni Massaro, Mari Matsuda, and most especially Barbara Fried, as well as participants in workshops at both Stanford and the University of Toronto law schools. None of them is to blame for anything that went awry in this piece, and more generally they doubtless have all led exemplary lives. Hank Levin, who initially piqued my interest in this field, has been consistently helpful to me as well. Kim Leue, Patricia Greene, and Peter Savich were all wonderfully capable research assistants. I am grateful to the John M. Olin Program in Law and Economics, which provided research support.

¹ See FAIRNESS IN EMPLOYMENT TESTING 187 (J. Hartigan & A. Wigdor eds. 1989) [hereinafter FAIRNESS]; see also Gottfredson, *Societal Consequences of the g Factor in Employment*, 29 J. VOCATIONAL BEHAV. 379, 398-403 (1986) (indicating that patterns of occupational inequality between blacks and whites are likely to follow from the use of general intelligence measures in the hiring process).

who would perform less capably. Testing proponents therefore argue that the disadvantage blacks suffer when an employer uses tests is justified, not discriminatory.² Whether testing proponents convincingly prove this claim depends in part on the degree to which tests do in fact predict performance. The persuasiveness of their claim depends as well on how we answer a less familiar conceptual question — what do we mean by “discrimination”?

In this Article, I contend that there exist at least four distinct, legally and politically viable views of what constitutes illicit discrimination. Those who oppose a particular practice, such as the use of a screening device like a “general ability” test or the implementation of an allegedly inequitable compensation scheme, may have very different notions of why the practice is illegitimate. They may believe it is “statically” discriminatory,³ “dynamically” discriminatory,⁴ or “distributively” discriminatory,⁵ or they may believe that it draws distinctions that are illicit because they are not grounded in differences in individual output.⁶ Each allegation ultimately rests on a distinct normative model of proper income distribution. In each case, by failing to curtail the employer’s practice society has breached a differently conceived obligation to ensure a just distribution of rewards.

I will use these alternative frameworks largely, although not exclusively, to analyze the propriety of “general ability” job testing. I ultimately conclude not only that the use of such tests is suspect if one reads the antidiscrimination norm expansively, but also that their use is inconsistent with the sort of meritocratic principles that mainstream conservatives, who typically support job testing, voice in interpreting that norm. The legitimacy of such tests is of great current concern: the United States Employment Service (USES) has temporarily suspended the use of the General Ability Test Battery (GATB) pending Department of Labor determination whether the test is discriminatory and, if so, whether its discriminatory impact can be adequately dampened.⁷

In Part II of this Article, I elucidate the four conceptions of discrimination mentioned above and explore the ways in which cognitive tests might be considered discriminatory under each. Thus, I first argue that one might simply require the test to correlate with on-

² The most prominent proponents of job screening tests today are industrial psychologists John Hunter and Frank Schmidt. The most accessible summary of their policy perspective is in Schmidt, *The Problem of Group Differences in Ability Test Scores in Employment Selection*, 33 J. VOCATIONAL BEHAV. 272 (1988).

³ See *infra* pp. 1164–70.

⁴ See *infra* pp. 1170–83.

⁵ See *infra* pp. 1183–98.

⁶ See *infra* pp. 1198–1204.

⁷ See 55 Fed. Reg. 30,162 (1990) (proposed July 24, 1990); 55 Fed. Reg. 35,379 (1990) (proposed Aug. 29, 1990).

the-job performance under current modes of work organization. A test is "statically invalid" if it fails to predict applicants' performance on jobs as the jobs are currently conceived. The claim that an employer's practice discriminates illegitimately because of its "static invalidity" rests on the normative supposition that marketplace compensation disparities are not justified unless they are grounded in distinctions between workers' current marginal products. A statically invalid test, for instance, is illegitimate because it rejects certain workers and relegates them to less desirable jobs even though they would be just as productive as workers whom the test accepts. The claim of illegitimacy, however, does not challenge the assumption that distribution based on marginal productivity is either a normatively *desirable*⁸ or, at the least, a legally *acceptable* outcome of private markets, an outcome which should be altered only by tax and transfer programs.⁹

One might argue, however, that screening tests are also invalid if the minority applicants who would perform poorly on the job as it is currently constituted, just as the test predicts, could in fact perform on a par with nonminority workers if the employer reorganized the workplace. A predictor might then be considered "dynamically" invalid. The claim that a practice is illegitimate because "dynamically" discriminatory depends on the view that pay disparities are illegitimate when existing marginal productivity differences are solely a function of readily changeable practices (generally work organizational practices) that the employer should be obliged to alter.¹⁰ An obligation to

⁸ For the classic defense of the normative propriety of productivity-based distribution, grounded in part in the worker's desert and in larger part in the compatibility of such distribution with the wealth-maximizing decisions of consumers and firms, see J. CLARK, *THE DISTRIBUTION OF WEALTH* (1923). The claim that distribution will generally correspond legitimately to productivity could be grounded exclusively in an entitlement theory. Robert Nozick, for instance, asserts that people own the product of their labor and that any effort to alter the distribution that occurs when people retain the goods they produce or retain what others give them in exchange for the products they have produced is illegitimate. See R. NOZICK, *ANARCHY, STATE, AND UTOPIA* 149-64, 174-82 (1974). Of course, in Nozick's scheme a legitimate distribution would not accord with any discernable pattern, not even with productivity, because the unproductive could legitimately be given resources by those who themselves held them legitimately. Utilitarian economists, by contrast, argue that labor should be priced like any other good to avoid its misallocation; marginal productivity pricing avoids Pareto-inferior states in which potentially mutually beneficial trades are not made because prices misstate real resource costs. See, e.g., J. HICKS, *THE THEORY OF WAGES* 7-9, 14-18 (2d ed. 1966).

⁹ Economists frequently discourage interference with the pricing mechanisms of markets, yet they note that legislatures can legitimately redistribute the income shares earned in an unregulated market that has set prices for all "commodities," including labor. See, e.g., A. OKUN, *EQUALITY AND EFFICIENCY: THE BIG TRADEOFF* 88-120 (1975).

¹⁰ See, e.g., McGarity & Schroeder, *Risk-Oriented Employment Screening*, 59 *TEX. L. REV.* 999, 1020-21, 1027, 1031-32 (1981); Wegner, *The Antidiscrimination Model Reconsidered: Ensuring Equal Opportunity Without Respect to Handicap Under Section 504 of the Rehabilitation Act of 1973*, 69 *CORNELL L. REV.* 401, 442-51, 479-80 (1984).

alter practices might be seen as necessary either to correct historically imposed burdens, whether or not the employer herself imposed those burdens,¹¹ or to allow people to be treated in accordance with the potential that they cannot realize unless the employer restructures the work setting.¹² An employer who is ordered to stop using a dynamically invalid screening device would either have to face productivity losses or redesign aspects of her production methods. A judge embracing a dynamic validity theory of permissible screening mechanisms must face the thorny question of how extensive (and expensive, at least in transitional terms) such court-ordered restructuring should be.

More politically progressive commentators suggest, although often not explicitly, that a practice may be "distributively" illicit insofar as it distinguishes people on the basis of traits that, although they do influence productivity (even in a "reformed" workplace), *ought not* to influence access to either jobs or income.¹³ Claims of distributive illegitimacy deny not only the propriety of a post-transfer income distribution that reflects distinctions in productivity (at least in those cases in which distribution is skewed by social group), but also the legitimacy of allowing *private* employers to distribute in accord with either current or potential productivity. Many possible distributive invalidity claims exist. For instance, one might argue that productivity judgments unduly reflect the devaluation of the work of an oppressed subgroup or that markets may not be organized to recognize all the social contributions made by members of a subgroup. Most politically radical commentators believe strongly that, as an empirical matter, screening tests are both statically and dynamically invalid; thus, they have seen little need to detail a theory of distributive invalidity. Nonetheless, I believe it is important to elucidate such a theory, although I doubt that there is any short-run prospect either of interpreting title VII to embody this perspective in a wide variety of cases or of enacting legislation that embodies it more generally.

¹¹ See, e.g., Lawrence, "Justice" or "Just Us": *Racism and the Role of Ideology* (Book Review), 35 STAN. L. REV. 831, 849 (1983) (noting that facially neutral policies administered today may simply reproduce past illicit disparities).

¹² See, e.g., Frug, *Securing Job Equality for Women: Labor Market Hostility to Working Mothers*, 59 B.U.L. REV. 55, 55-61 (1979) (arguing that the structure of the labor market makes participation for women, who bear major responsibility for child care, difficult, thereby creating an unacceptable disparity between male and female workers' wages).

¹³ For the basic statement of the "distributive discrimination" view that comes closest to the one I explore below, see Littleton, *Reconstructing Sexual Equality*, 75 CALIF. L. REV. 1279, 1284-85 (1987). Littleton, however, uses the term "equality as acceptance" rather than "distributive discrimination." *Id.* at 1285. Littleton's conception is exemplified by her statement that "[t]he differences between human beings, whether perceived or real, and whether biologically or socially based, should not be permitted to make a difference in the lived-out equality of those persons." *Id.* at 1284-85 (emphasis in original). Below, I explore some of the practical and theoretical ambiguities of this formulation in the title VII setting. See *infra* pp. 1191-94.

Finally, those who interpret antidiscrimination law may also rely upon a fourth distributive principle, which I will call the "individual-output-centered" principle. Such a principle, which has not been invoked previously in discussing screening tests, would claim that a distribution is illegitimate if a worker capable of generating as much output as other workers is relegated to less desirable jobs because it is more costly to employ him. We must ultimately distinguish two sorts of cases. In one case, real social costs rise when the worker takes a particular job. For example, he is atypically prone to a disease exacerbated by the work setting so that either he must be specially protected from a hazard or someone in the society (perhaps the worker himself) must bear added health care costs and suffering from the disease. In the second scenario, the particular employer may prefer not to employ the worker because the employer would have to bear costs that would be borne elsewhere anyway. For example, the worker may be genetically predisposed to a disease that a self-insuring employer will have to pay to treat.

Those who believe that antidiscrimination law should forbid only practices that are either economically irrational from the employer's perspective or motivated by her animus towards the protected group would deny there is any illegitimate discrimination in either of these cases.¹⁴ But some of those who either affirmatively favor or tolerate productivity-based distribution may nevertheless argue that refusals to hire in these cases are *not* in fact based on productivity.¹⁵ Although the argument is especially strong when the employer simply wants to avoid the private assumption of a social cost that will be borne anyway, commentators have not emphasized the distinction between this class of cases and the cases in which real social costs will increase if the applicant is employed.¹⁶ If real social costs would rise in a particular case, arguments opposing bars to employment must spring from one of two sources. They may come from those who deny the normative power of the productivity model — that is, our third group, who believe it is important to assess whether an employment practice is distributively discriminatory. But they could also come from those who espouse a more limited theory of distributive discrimination. Such people would permit employers to weigh a person's capacity to produce output — the employee's "talent" at work — in making a hiring decision, but would bar employers from considering the appli-

¹⁴ See, e.g., Gold, *Griggs' Folly: An Essay on the Theory, Problems, and Origins of the Adverse Impact Definition of Employment Discrimination and a Recommendation for Reform*, 7 *INDUS. REL. L.J.* 429 (1985); Recent Case, 85 *HARV. L. REV.* 1482, 1485 (1972) (discussing *Johnson v. Pike Corp.*, 332 F. Supp. 490 (C.D. Cal. 1971)).

¹⁵ See, e.g., Brodin, *Costs, Profits, and Equal Employment Opportunity*, 62 *NOTRE DAME L. REV.* 318, 320 n.14 (1987).

¹⁶ See, e.g., *id.*; Wegner, *supra* note 10, at 442-51.

cants' uncontrollable responsibility for increasing the employers' input costs, even if hiring the applicant decreases both firm profitability and net social product.

The arguments I briefly outline about the illegitimacy of statically invalid screening devices should be most well known to readers. Some of the arguments about dynamic discrimination ought to be fairly familiar as well. In contrast, although the arguments about distributive discrimination are by no means original, I deliberately state them in terms quite different from those in which they have heretofore been cast. The translations are *not* intended to be faithful to the sources, but rather represent a deliberate effort to recast aspects of the debate. Similarly, I attempt to alter considerably the formulation of arguments about screening out "costly" employees.

In Part III, I argue that general ability tests are illegitimate even if we both interpret the purposes of antidiscrimination law narrowly to exclude concerns about dynamic or distributive discrimination and make a series of factual assumptions quite favorable to the tests. Part IIIA argues that tests are neither universally valid across job types nor racially unbiased in the sense in which most psychometricians use the term. Nonetheless, in Part IIIB I assume, for argument's sake, the truth of what strong testing proponents such as John Schmidt and Frank Hunter believe: that the tests are equally valid in all settings and that they are equally valid and predictive for majority and minority group members (that is, they are culturally unbiased).¹⁷ I make the further assumption, shown to be plausible in Part IIIA, that tests screen out poorer workers better than would a random hiring system, although not very much better. Finally, I assume that tests predict performance better than would any alternative *ex ante* screening device with less racially disparate impact. Like the courts, which generally have not been explicit about how valid a test must be to be acceptable,¹⁸ I offer no hard-and-fast numerical guidelines.

Despite these assumptions, I argue that this sort of low validity, culturally unbiased general ability test poses far more severe title VII problems than either courts or commentators have acknowledged. Generally, commentators see the use of such tests as creating a conflict between liberal, individualistic, meritocratic fairness and the social interest in group advancement for blacks.¹⁹ I argue, however, that a hypothetically unbiased general ability test is problematic even within

¹⁷ For a fuller discussion of Hunter and Schmidt's position, see *infra* pp. 1207-21.

¹⁸ Courts that have addressed the issue directly have recognized that a low correlation test may be more acceptable if its adverse impact is less severe. See, e.g., *Robinson v. Lorillard Corp.*, 444 F.2d 791, 797-98 (4th Cir. 1971).

¹⁹ See, e.g., Fallon, *To Each According to His Ability, From None According to His Race: The Concept of Merit in the Law of Antidiscrimination*, 60 B.U.L. REV. 815, 816-20, 831-34 (1980); Newton, *Reverse Discrimination as Unjustified*, 83 ETHICS 308, 312 (1973).

the liberal, individualistic, meritocratic tradition, because the use of a low validity test does not necessarily comport with the demands of a rigorous definition of meritocracy. Although the test itself may not be discriminatory, its *use* is arguably biased, in that a black applicant who is as capable of performing a job as a white is less likely to get that job, and an incapable white is more likely to get a job than an incapable black.

One way to frame the question posed in Part III is to ask whether we should judge a system discriminatory when it treats factual equals unequally in situations in which equal treatment is costly. This important issue generally has been wished away by rhetorical fiat: political progressives simply assume that the cost of remedying discrimination is no defense, largely ignoring the argument that there *is* no discrimination to correct when an employer saves real resource costs by hiring in a way that effectively disadvantages blacks. Political conservatives, on the other hand, simply assume that an employer who is neither intentionally racist nor inadvertently racist (through unconscious use of unexamined, inaccurate stereotypes) is not discriminatory, and that decisions are *either* economically rational *or* impermissibly bigoted. Thus, conservatives pay inadequate attention to the claim that title VII demands "meritocratic" treatment, at least if protected groups are harmed by the breach of meritocratic principles, even when it is costly to ensure such treatment.

In Part II, I hope to press politically progressive commentators to think harder about the degree to which their view of antidiscrimination law ultimately constitutes a thinly disguised redistributive transfer program and to force them to consider whether they have chosen appropriate transferors and beneficiaries. In Part III, I ultimately hope to press politically conservative commentators to recognize that one can quite reasonably flip their ordinary account of the battle over testing. In the view I propound, the battle is not the one conservatives imagine between individualistic, meritocratic testing advocates and utilitarian social engineers seeking group pacification at the expense of individual justice. Rather, it is a battle between testing opponents concerned with injustice to the individuals in certain social groups and utilitarian testing proponents less interested in fairness than in maximizing social product.

II. MULTIPLE INTERPRETATIONS OF DISCRIMINATION

A. Static Discrimination

A broad consensus exists that federal law both should and does prohibit static discrimination. It is illegal to disfavor members of protected groups who are, without any shifts in the organization of employers' firms, as "market productive" as others who receive the

benefits, such as jobs, promotions, or pay, that members of the protected group seek. Antidiscrimination law does not mandate any general, productivity-based distributive principle; employers clearly remain legally privileged to hire a less productive member of an unprotected group rather than a more productive one. The law takes a far less agnostic stance, however, toward such spontaneous private ordering when members of protected groups are involved. There are a number of conceivable explanations for this position, none of which necessarily excludes any others.

First, antidiscrimination law might be based on the notion that members of protected groups are entitled to vindicate their affirmative meritocratic claims, because without legal protection these groups will face a systemic pattern of inadequate reward. A particular idiosyncratic employer may treat a member of a privileged group (for example, a white male) nonmeritocratically, but there is little reason to believe that a "privileged" person disfavored in one setting will be disfavored in others. Ultimately, we can expect such a worker to be rewarded for his productivity. Discrimination against members of protected groups, however, is systemic rather than idiosyncratic. There is a strong possibility that the productivity of individual workers in the protected class will remain unrewarded *everywhere* if the state does not intervene after the failure to reward productivity has been brought to light.

In addition to violating group members' affirmative claims to merit-based treatment, static discrimination may also breach a "negative rights" claim to be free from tort-like stigmatic injury. In this view, the racist employer inflicts a stigmatic injury beyond the deprivation of material benefits produced by his failure to hire, promote, or compensate adequately. The employer's failure to treat a protected group member as she would rationally be treated by an impersonal economic calculator — that is, solely in terms of her productivity — can only be understood as a function of animus or stereotyping, both of which are painful and demeaning to the person who encounters them.²⁰

Consider an employer who screens job applicants using a statically invalid test — a test that does not probabilistically predict how well applicants will perform on the job. Such a test will show a zero correlation between predictor (test performance) and criterion (job performance). In traditional title VII doctrine, the use of such a test is unquestionably impermissible whenever it has a disparate impact,²¹

²⁰ For a particularly compelling account of the direct injury caused by racist rejections, see Williams, *Spirit-Murdering the Messenger: The Discourse of Fingerprinting as the Law's Response to Racism*, 42 U. MIAMI L. REV. 127 (1987).

²¹ The statute directly precludes the use of tests that are not valid. See 42 U.S.C. § 2000e-2(h) (1988). The relevant question addressed in the text, however, is why this provision is

that is, whenever fewer blacks will be hired if the test is used than if hiring were done by lottery. By hypothesis, there is no business justification for the test.²² Because the unscreened work force will be no less capable than the work force that the employer selects by using the test, abandoning testing will not affect plant output and will eliminate the employer's costs of administering the test.²³

consistent with the more general theory of title VII. Even in the absence of the subsection, the invalid test would surely be barred under the disparate impact doctrine set forth in *Griggs v. Duke Power Co.*, 401 U.S. 424, 429-33 (1971). See *infra* note 23. This subsection was almost surely meant to *expand* the employer's scope of unregulated action in using tests beyond the level Congress believed courts might have allowed. See Brodin, *The Role of Fault and Motive in Defining Discrimination: The Seniority Question Under Title VII*, 62 N.C.L. REV. 943, 947-52 (1984); Wilson, *A Second Look at Griggs v. Duke Power Co.: Ruminations on Job Testing, Discrimination, and the Role of the Federal Courts*, 58 VA. L. REV. 844, 852-57 (1972).

²² There was once some question whether the use of a screening device had to be "necessary" or whether it could simply be predictive of performance in such a way as to be conducive to firm profitability. The dispute appears mindlessly formalistic to me; firms can always be more or less profitable and would always prefer to be more so. The question whether a test is "necessary" could be read to ask whether the firm could survive as an entity if it abandoned the test. But except to the degree that pushing marginal firms out of business itself causes transitional social costs that exceed the added social costs inherent in using any more expensive production technique, the issue of survival, as opposed to the issue of increasing costs, seems of no normative moment.

It now appears fairly well settled that employers may screen out employees who, although able to "do the job" in a broadly competent fashion — that is, without endangering fellow employees or customers — perform less capably than those who are hired. The screening device therefore need not be absolutely "necessary" to ensure that workers can perform the task. See *Watson v. Fort Worth Bank & Trust*, 487 U.S. 977, 978-79 (1988); *New York City Transit Auth. v. Beazer*, 440 U.S. 568, 587 n.31 (1979); *Contreras v. City of Los Angeles*, 656 F.2d 1267, 1283 (9th Cir. 1981), *cert. denied*, 455 U.S. 1021 (1982). In *Bernard v. Gulf Oil Corp.*, 841 F.2d 547 (5th Cir. 1988), the Fifth Circuit limited *Watkins v. Scott Paper Co.*, 530 F.2d 1159 (5th Cir. 1975), *cert. denied*, 429 U.S. 861 (1976), in accord with *Watson*. *Watkins* had held that a showing of business necessity could not be made unless the employer required the test to ensure safety or "efficiency." Because "efficiency" was never clearly defined, however, the functional distinction between the tests remains hard to fathom: the use of less productive employees is, perhaps, inefficient. The choice of judicial tests ultimately can be seen as signaling "moods" of more or less stringent scrutiny.

Increased efficiency cannot, however, justify the use of a test to segregate deliberately the work force. Deliberate segregation is a form of disparate treatment. See, e.g., *Knight v. Nassau County Civil Serv. Comm'n*, 649 F.2d 157, 161-62 (2d Cir.), *cert. denied*, 454 U.S. 818 (1981) (prohibiting relegation of a black employee to minority recruitment); EEOC Decision 70,350, 2 Fair Empl. Prac. Cas. (BNA) 498, 498-99 (1969) (forbidding the assignment of a black salesman to "all Negro" accounts even if it increases firm profitability). Such segregation cannot be justified by "business necessity" or profitability. See, e.g., *Parson v. Kaiser Aluminum & Chem. Corp.*, 727 F.2d 473, 477 (5th Cir. 1984) (holding that an employer cannot segregate its work force even if integration poses real threat of violent disruption of a plant). Further, an employer may not justify racial exclusion even under the more stringent "bona fide occupational qualification" (BFOQ) justification that the statute allows for other types of deliberate discrimination. See 42 U.S.C. § 2000(e)-3(b) (1988). Only religion, sex, and national origin are listed as possible bona fide occupational qualifications; race is not. See *id.* For a fuller discussion of BFOQs, see *infra* note 38.

²³ This is the basic structure of the inquiry in a traditional disparate impact case under title

An employer who uses such statically invalid tests may in fact *want* to exclude minorities. Because title VII prohibits him from openly following a discriminatory policy, the employer may conceal his bias through the use of substitute indicators of race.²⁴ Such blatant, conscious discrimination imposes a painful stigma on minority applicants and excludes them from benefits that they are qualified to receive. Moreover, the perpetrator is morally blameworthy, because the harm he inflicts is deliberate or intentional, rather than inadvertent or negligent.²⁵

VII. See, e.g., *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 425-35 (1975); *Griggs v. Duke Power Co.*, 401 U.S. 424, 429-33 (1971). In such a case, the basis of the claim is not that the particular employer used a consciously discriminatory practice, but that the employer engaged in a practice that produced an unjustified racially disparate outcome. In a disparate treatment case, on the other hand, the plaintiff alleges that "the employer simply treats some people less favorably than others because of their race." *International Bhd. of Teamsters v. United States*, 431 U.S. 324, 335 n.15 (1977). Generally speaking, employees can demonstrate disparate impact if the selection rate for a protected group is less than 80% of the rate for the group with the highest rate of selection. See *Uniform Guidelines on Employee Selection Procedures*, 29 C.F.R. § 1607.4(D) (1990). Once plaintiffs show disparate impact, the burden of production shifts to the employer to show that his practice is necessary. See *supra* note 22. For a basic description of the distinction between disparate treatment and disparate impact cases, see B. SCHLEI & P. GROSSMAN, *EMPLOYMENT DISCRIMINATION LAW* 13-15, 21-22, 82-92, 98-102 (2d ed. 1983); and Bartholet, *Application of Title VII to Jobs in High Places*, 95 HARV. L. REV. 945, 1004-06 (1982).

In the wake of *Wards Cove Packing Co. v. Atonio*, 490 U.S. 642 (1989), however, the status of disparate impact suits is unclear. In *Wards Cove*, the Supreme Court seemed to hold that it is inadequate for the plaintiff to show simply that a low proportion of workers in a particular job are minority workers, whether compared to the local population, as in *United States v. Ironworkers Local 86*, 443 F.2d 544, 551 (9th Cir. 1971) and *International Bhd. of Teamsters*, 431 U.S. at 337-38, to the particular pool of job applicants, as in *Hazelwood School District v. United States*, 433 U.S. 299, 305 (1977), or to workers in other job classifications in the same employment unit, as in *Parham v. Southwestern Bell Tel. Co.*, 433 F.2d 421, 424 (8th Cir. 1970). See *Wards Cove*, 490 U.S. at 650-55. Instead, *Wards Cove* implies that a plaintiff must show that a disproportionately low number of *qualified* minority workers have been hired. Clearly, to determine the size of the pool of qualified applicants one must answer what was traditionally thought to be the wholly distinct question of whether the screening device eliminates only those who will perform less satisfactorily.

Once the plaintiff establishes that a selection device has adversely impacted qualified minority applicants, *Wards Cove* requires the plaintiff to convince the factfinder by a preponderance of the evidence that the practice serves no valid business purpose. Placing the burden of proof on the plaintiff represents a substantial departure from previous understandings of title VII in many federal courts. See, e.g., *Nash v. Consol. City of Jacksonville*, 837 F.2d 1534, 1536 (11th Cir. 1988); *Bunch v. Bullard*, 795 F.2d 384, 395 (5th Cir. 1986) (placing the burden of proof on the employer).

²⁴ In this scenario, the employer resembles post-Reconstruction Southern voting registrars who attempted to find surrogates for race, such as property ownership, to serve as the basis for exclusion. See, e.g., J. FRANKLIN, *FROM SLAVERY TO FREEDOM* 328-33 (2d ed. 1956); B. MARSHALL, *FEDERALISM AND CIVIL RIGHTS* 13-15 (1963).

²⁵ In criminal law terms, a defendant who knowingly or purposely injures is almost invariably deemed more culpable than the inadvertent violator. See, e.g., G. WILLIAMS, *CRIMINAL LAW* 122-23 (2d ed. 1961); Fletcher, *The Theory of Criminal Negligence: A Comparative Analysis*,

Alternatively, ongoing use of the test may reflect cognitive dysfunctions that are the direct product of racist stereotyping.²⁶ It may appear "commonsensical" to a white employer that blacks do worse on tests, simply because "commonsensical" racist stereotyping supports the idea that blacks are typically inferior.²⁷ The failure of the employer to attempt to validate the test systematically results from the fact that bigoted common sense *implicitly* validates the test's outcome. Perversely, if a test excludes blacks its results may even be thought to confirm the accuracy of the testing mechanism.²⁸

119 U. PA. L. REV. 401 (1971); Hall, *Negligent Behavior Should Be Excluded from Penal Liability*, 63 COLUM. L. REV. 632 (1963). For an application of this principle to antidiscrimination law, see Note, *Discriminatory Purpose and Mens Rea: The Tortured Argument of Invidious Intent*, 93 YALE L.J. 111 (1983).

²⁶ For discussions of stereotyping, see Lawrence, *The Id, The Ego, and Equal Protection: Reckoning with Unconscious Racism*, 39 STAN. L. REV. 317, 332-34, 343, 372 (1987); and Snyder, *On the Self-Perpetuating Nature of Social Stereotypes*, in COGNITIVE PROCESSES IN STEREOTYPING AND INTERGROUP BEHAVIOR 183 (D. Hamilton ed. 1981).

²⁷ Such cognitive dysfunctions are frequent. White males often undervalue the achievements of women and non-whites, whether evaluating work performance or an applicant's credentials. See, e.g., Bigoness, *Effect of Applicant's Sex, Race, and Performance on Employers' Performance Ratings: Some Additional Findings*, 61 J. APPLIED PSYCHOLOGY 80, 80-81 (1976) (finding detrimental effect on basis of race, not gender); Deaux, *Sex and Gender*, 36 ANN. REV. PSYCHOLOGY 49, 65-69 (1985); Lott, *The Devaluation of Women's Competence*, 41 J. SOC. ISSUES No. 4, at 43, 54 (1985); Rosen & Jerdee, *Effects of Applicant's Sex and Difficulty of Job on Evaluations of Candidates for Managerial Positions*, 59 J. APPLIED PSYCHOLOGY 511, 511-12 (1974).

²⁸ A number of commentators have argued that racism can survive in unregulated competitive markets because of just such racist cognitive dysfunction. See, e.g., R. CHERRY, DISCRIMINATION 36-37 (1989); Arrow, *The Theory of Discrimination*, in DISCRIMINATION IN LABOR MARKETS 3, 23-32 (O. Ashenfelter & A. Rees eds. 1971). Racism can survive particularly when it is economically rational to stereotype, as it will be as long as it is more costly to ascertain individual traits than to screen out all group members. See Phelps, *The Statistical Theory of Racism and Sexism*, 62 AM. ECON. REV. 659, 659-61 (1972). The profit that an employer might make by hiring a black from the large pool of qualified available blacks while others foolishly bid up the wages of the more restricted pool of whites would make racist hiring practices unstable only if employers *recognized* the profit opportunity. Since racist cognitive dysfunction blocks this recognition, it is the precise phenomenon with which the legal system must interfere.

The affirmative argument for eliminating stereotype-based decision criteria is, once again, multifaceted. Such racist cognitive dysfunctions do not simply exclude blacks from access to desirable perquisites. They also inevitably make blacks aware that they are victims of demeaning stereotypes, a realization that may well be as degrading as being shunned or hated by the more overtly, self-consciously racist employer. See Lawrence, *supra* note 26, at 349-53.

In terms of the moral blameworthiness of the injurers, stereotyping employers by hypothesis are not as viciously racist as the first class of employers, but they are scarcely blameless either. They are not *just* inadvertently or negligently taking steps that have an adverse impact on blacks; rather, they are inadvertently indulging a set of bigoted perceptions. The moral distinctions may be reflected in traditional criminal jurisprudence: the law is much more tolerant of inadvertent factual mistakes than it is of inadvertent normative errors. Generally speaking, people are strictly liable for knowing the content of criminal norms. See MODEL PENAL CODE § 2.02(9) (1962). It is rare, however, for a defendant to be held strictly liable for his failure to

Opposition to such racial stereotyping, rather than opposition to racial exclusion, helps to explain why the law prohibits an employer from using an illegitimate screening device even when he employs a "bottom-line quota" to ensure that he hires as many blacks as he would have had the test not been employed.²⁹ Obviously, such an employer is not using the test as a pretext to avoid minority hiring, but he nevertheless inflicts stigmatic injuries on successful and unsuccessful minority applicants alike by perpetuating racist stereotypes. The Supreme Court's refusal to allow the "bottom-line" defense in such cases³⁰ frees each successful black applicant from the stigmatizing insult that she was hired "politically" rather than "meritocratically" and vindicates each unsuccessful black applicant's right to be evaluated through an unbigoted lens.³¹

This view necessarily incorporates group-based, social understandings into a quasi-tortious, individualistic analysis.³² Only our understanding of the gender and racial hierarchies within our society allows us to perceive white male distaste for non-whites or women as an attitude inflicting a remediable injury.³³ In a world of absolute group

know the correct facts when, had the facts been as the defendant believed them to be, he would not have committed an offense. *See id.* § 2.04(1). Drawing distinctions between knowing the content of a norm and knowing how it applies to particular circumstances may in fact be more difficult than the common distinction between mistakes of law and mistakes of fact suggests; see Kelman, *Interpretive Construction in the Substantive Criminal Law*, 33 STAN. L. REV. 592, 630-33 (1981), but there is little doubt that we have less sympathy for ignorance of general moral norms than we have for more particular errors of factual perception.

²⁹ *See, e.g.*, *Connecticut v. Teal*, 457 U.S. 440, 442 (1982).

³⁰ *See id.*

³¹ *See Welch, Superficially Neutral Classifications: Extending Disparate Impact Theory to Individuals*, 63 N.C.L. REV. 849, 860-63 (1985). Those who are particularly concerned that test validation is unduly expensive have argued that the use of quotas, alongside tests, enables employers both to meet reasonable societal demands to integrate the work force and to use reasonably good selection systems without incurring the prohibitive costs of validating a selection device. *See Fallon, supra* note 19, at 841. Validating tests locally can certainly be expensive, with costs running into the many hundreds of thousands of dollars. *See FAIRNESS, supra* note 1, at 43.

Others worry instead that the rejection of the bottom-line defense in *Teal* diminishes the incentive to engage in affirmative action, because employers who implement affirmative action policies will still be liable if they use unvalidated tests. *See Chamallas, Evolving Conceptions of Equality Under Title VII: Disparate Impact Theory and the Demise of the Bottom Line Principle*, 31 UCLA L. REV. 305, 370-82 (1983). Although I share this concern, it does not negate (though it may outweigh) the concern raised in the text that an invalid test imposes stigmatic injury on particular persons.

³² *See Gardner, Liberals and Unlawful Discrimination*, 9 OXFORD J. LEGAL STUD. 1, 21 (1989).

³³ *See id.* at 7; *see also C. MACKINNON, SEXUAL HARASSMENT OF WORKING WOMEN* 118-27 (1979) (arguing that, given the existence of social hierarchies, antidiscrimination law must eradicate dominance, not simply ensure that identically situated people receive equal treatment); Strauss, *The Myth of Color Blindness*, 1986 SUP. CT. REV. 99, 111-13, 118-32 (arguing that ordinary prohibitions against discrimination, not just schemes dubbed "affirmative action," are premised on improving the situation of blacks because they are an oppressed group).

equality, distaste might carry no connotation of inferiority. Stereotypes would merely demonstrate the stereotypers' cognitive deficiencies and would not reinforce the historically deflated self-images of the victims of discrimination.

In any event, there is so little controversy over the impropriety of static discrimination or statically invalid tests,³⁴ and so much familiar theorizing about why utterly invalid tests are illegitimate, that it is hardly necessary to rehash arguments about their impropriety. Courts and commentators, however, have not adequately dealt with cases in which employers may be acting rationally in excluding minorities, nor have they reached a consensus concerning the statistical validity of tests as predictors. I will return to these issues in Part III.

B. Dynamic Discrimination

Let us assume that an employer tests applicants and that test performance correlates reasonably well with job performance. Let us further assume, however, that this correlation would decline markedly if the employer altered the work environment in which newly hired employees are placed. Such a relationship between job organizational reform and diminished test validity might exist for a variety of reasons. In this section, I discuss the effects of several possible barriers to equal performance in order to develop a fuller account of the concepts of illicit dynamic discrimination and dynamic invalidity in the hiring process.

1. *Illegitimate Discrimination Hampers Performance.* — In the first scenario, applicants who do badly on the screening test are victims of productivity-limiting practices in the workplace that are themselves illegitimately discriminatory. For instance, women workers might fare badly on a screening test *and* be victimized by sexual

³⁴ Prohibitions of private discrimination may in fact be difficult to justify, given the usual deference to unfettered private choice in our liberal legal tradition. See Gardner, *supra* note 32, at 3. While private acts of discrimination may seem to harm others in a conventional "tort-like" way, the basis of the harm may well be the broader social practice of exclusion and stigmatization for which the "tortfeasor" is not individually responsible. See Goldman, *The Justification of Equal Opportunity*, in EQUAL OPPORTUNITY 88, 99-100 (1987). Although prominent libertarian legal theorist Richard Epstein does not argue that the 1964 Civil Rights Act constituted a "taking" by depriving the owner of a public accommodation of his traditional right to exclude, see 42 U.S.C. § 2000a(a) (1988), his failure to do so seems wholly tactical, not principled. See R. EPSTEIN, *TAKINGS: PRIVATE PROPERTY AND THE POWER OF EMINENT DOMAIN* (1985). Given Epstein's supposition that a shift in common law entitlements is a *per se* taking, see *id.* at 96-98, and his more specific statement that the right to exclude is a defining aspect of property ownership, see *id.* at 58-66, logically his conclusion must be that the 1964 Act was indeed unconstitutional unless it provided owners with compensation for the loss of the right to exclude. Similarly, given that at common law no private individual could compel a person to enter a labor contract with another, Epstein should logically believe that the state has no legitimate authority to compel labor contracts through title VII. See *id.* at 12-13, 318-19.

harassment at work that diminishes their productivity. The test might then appear criterion validated, in that it predicts performance — the poor test performers do a poor job. The test would not, however, be construct or content valid — that is, it would not measure traits actually needed to perform the job.³⁵ Instead, it would simply reflect the fact that as a result of on-the-job discrimination, productivity on a particular work site is predominantly a function of gender.³⁶

In these cases, the employer is independently required by law to remedy the productivity-blunting practice by protecting women workers from sexual harassment.³⁷ But it should be equally uncontroversial to say that as long as the employer fails to remedy the on-the-job violations, he cannot rely on the test to screen workers. Women

³⁵ A test is content validated if the content of the examination closely matches the content of the job, as determined by a careful job analysis. The classic example is a typing test given to determine eligibility for a typing pool. A test is construct validated if a professional job analysis shows that the job requires a series of traits that the test is known to measure. A test is criterion validated (also described as predictively validated or empirically validated) if there is a substantial correlation between measured performance on the predictor (for example, a screening test) and performance on the job, as measured by supervisor ratings, job samples, or other objective measures. See Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. §§ 1607.5(B), 1607.16(D)–(F) (1990) (describing these validation procedures in employment discrimination law); see also M. MINER & J. MINER, *EMPLOYEE SELECTION WITHIN THE LAW* 75–106 (1979) (discussing the design of validation research and the development of criterion measures); Bartholet, *supra* note 23, at 1016–23 (discussing commonly accepted methods of test validation). In my view, the conceptual core of validation is criterion validation. Tests are valid if, and only if, they predict performance, although content or construct validation may be cheaper and more convenient techniques to use. Content or construct validity are simply plausible predictions that true criterion validity would in fact exist if tested for.

³⁶ Some may be skeptical that such a test could prove to be criterion valid. But if women actually performed substantially worse on the test, and if women performed substantially worse as workers as a result of harassment, the fact that correlations might be near zero among male applicants would not preclude a significant, positive validity rating for the whole group. If, however, the test had to be separately validated for each subgroup, such a test would prove invalid for men and would probably be invalid for predicting performance *differentials* among women.

The more serious problem arises when a test is culturally biased in favor of the majority group, but the cultural bias is hidden by the fact that those who are most prone to perform poorly on the test (for example, minority workers) perform poorly on the job as a result of discriminatory in-plant practices (such as racial harassment, noncooperation from fellow workers, or hostile or unsupportive supervision). Such a test may well appear valid for the whole group *and* valid for the larger majority group. The *aggregate* predictions for minority group performance will also seem fair. Intraminority validities will probably be low if the test is culturally biased (that is, intragroup differences in test performance will not correlate with intragroup productivity differentials). However, to the degree that the capacity to do well on the test is related to traits that diminish one's subjection to productivity-limiting racism, even intragroup validities may appear substantial. More importantly, as a practical matter, even low intragroup validities for minorities may be correctly dismissed as inconsequential if the minority sample is too small to generate statistically significant results.

³⁷ See *Meritor Sav. Bank v. Vinson*, 477 U.S. 57, 73 (1986) (holding that a claim of "hostile environment" sex discrimination is actionable under title VII).

should not be made to suffer the consequences of the employer's inability to protect them. Rather, merit must be measured in a setting in which one's basic rights are protected.

To support this rather commonsensical result, one might argue that the test is not in fact properly statically validated. Although it might conceivably be criterion validated, any correlation between predictor and criterion is spurious, because there is no functional relationship between the screening device and performance. If the employer understood that the test "worked" only because it selected women, he could then use a directly discriminatory hiring principle. Such a decision would clearly violate title VII unless gender were a bona fide occupational qualification (BFOQ).³⁸ When one can predict poor productivity based upon membership in a protected group (other than race), one can exclude members of the group only if the prediction is grounded on the intrinsic demands of the job.³⁹

In a sense, however, this argument seems circular and question-begging. As long as employers are not required to treat women as if they performed as they would in the absence of rights violations and as long as the violations themselves are not eliminated or reduced, maleness will appear to be a BFOQ. In other words, the illegitimacy of the employer's screening device depends on our normative judgment obliging the employer to assign socially desired benefits, such as jobs and pay, in accord with the workers' potential rather than actual productivity. If such a duty to assign benefits in accord with potential productivity is not placed on employers, employees surely have no cognizable claim in the more difficult cases discussed below.

In this first scenario, however, the worker's claim to be treated as if she were as productive as she would be in a reformed, nonsexist environment is particularly strong. First, the employer has an *independent* legal obligation to establish the more ideal environment. We might find it difficult to impose workplace reform duties on an employer in the absence of direct statutory authority obligating reform. Demanding that an employer hire someone who is statically less productive penalizes the employer and implicitly imposes an obligation to reform the workplace. Unlike a fine, imprisonment, or a damage judgment, the "penalty" for the breach of the obligation is not directly

³⁸ Employers may discriminate against members of a protected group when membership in the non-protected group is a BFOQ: the applicant is unqualified for the job if she lacks the ability to perform the job, if the privacy interests of those with whom she will come into contact will be compromised if she is hired, or if customers "legitimately" prefer to deal with non-protected group members (for example, if theatergoers want male actors to play male characters). See, e.g., A. LARSON & L. LARSON, EMPLOYMENT DISCRIMINATION §§ 14.00-17.00, at 4-4 to 4-71 (12th ed. 1990).

³⁹ See, e.g., I C. SULLIVAN, M. ZIMMER & R. RICHARDS, EMPLOYMENT DISCRIMINATION § 3.6, at 105-26 (2d ed. 1988).

imposed by the state, but instead exerts its force in the form of declining output and profitability — results that are the direct and intended outcome of the state's intervention in the screening process. Here, the "indirect" imposition of a penalty is readily legitimated because the obligation is "direct"; the state could order the employer to end harassment or face civil penalties.

Second, the employer who permits sexual harassment on his work site is like one attempting to profit from his own wrongdoing,⁴⁰ or at least attempting to avoid one legal obligation by breaching another. But for the illegitimate tolerance of sexual harassment, women would be as productive as men and clearly would be non-excludable.⁴¹ Thus, the case for proscribing dynamic discrimination is most compelling when the barrier to equal performance by the protected group is an independently unacceptable violation of the group members' rights.

2. *Inadequate Training Hampers Performance.* — In the second scenario, I assume that applicants who do poorly on the screening device will do poorly on the job unless they receive some supplementary preliminary training that is not now offered. The training effectively overcomes genuine obstacles to performance, permitting those who would otherwise perform poorly to perform adequately. The training will not, however, raise the productivity of those who currently qualify for the job as much as it will improve the productivity of those who have been rejected. By severing the connection between test and on-the-job performance, such training will blunt the screening device's criterion validity. It will not simply increase the level of everyone's performance and reproduce the same relative rankings. Courts have, in fact, ordered employers to train applicants rather than simply screen them out in just such circumstances.⁴²

⁴⁰ There is no universally applicable rule that one can never profit from one's own wrongdoing, but it is surely difficult to explain why we resolve certain ethical or legal controversies as we do unless we accept the notion that profiting from one's own wrongdoing is morally objectionable. See Wellington, *Common Law Rules and Constitutional Double Standards: Some Notes on Adjudication*, 83 YALE L.J. 221, 222-24 (1973).

⁴¹ In roughly analogous criminal law settings, defendants typically cannot claim a defense of necessity or justification if they have placed themselves in the position of having to perform proscribed acts. See, e.g., MODEL PENAL CODE § 3.02(2) (1985). Similarly, the initial aggressor who must ultimately defend himself will not be able to claim self-defense successfully. See, e.g., *United States v. Peterson*, 483 F.2d 1222, 1231 (D.C. Cir.), cert. denied, 414 U.S. 1007 (1973).

⁴² See, e.g., *Kinsey v. First Regional Sec., Inc.*, 557 F.2d 830, 837 (D.C. Cir. 1977); *Pennsylvania v. Local Union 542, Int'l Union of Operating Eng'rs*, 488 F. Supp. 988, 992 (E.D. Pa. 1980); *Buckner v. Goodyear Tire & Rubber Co.*, 339 F. Supp. 1108, 1124-25 (N.D. Ala. 1972), *aff'd*, 476 F.2d 1287 (5th Cir. 1973); *United States v. Local No. 86, Int'l Ass'n of Bridge, Structural, Ornamental & Reinforcing Ironworkers*, 315 F. Supp. 1202, 1247 (W.D. Wash. 1970); see also *Uniform Guidelines on Employee Selection Procedures*, 29 C.F.R. § 1607.5(F) (1990) (stating that employers who screen applicants should avoid making employment decisions

In this case, the employer has no independent duty to establish training programs, as he has an independent duty to curtail harassment. Indeed, if his work force is not unduly segregated in the absence of a training program, he need not establish one even if it would bolster minority performance. If, however, the employer places untrained applicants on the job immediately *and* uses screening tests that have a disparate impact on black applicants, courts may either order training or forbid the use of the tests.⁴³ Although there is no across-the-board affirmative duty to adopt any particular affirmative workplace reform in the same way that there are duties to prevent discriminatory practices such as sexual harassment, there is always a *residual* duty under title VII to adopt any cost-effective alternative that reduces racially disparate impact.⁴⁴ This is true even though the employer is not causally or morally responsible for the poor performance of untrained applicants in the way that the employer who tolerates sexual harassment is responsible for the relative failure of women applicants.

In examining the basis of the employer's duty to train, it is useful to hypothesize two more distinct scenarios. In the first, the training program is virtually costless. The adoption of a special training program could have a net cost approximating zero if it increased the productivity of the trained workers sufficiently beyond the level achieved by screened workers to offset the costs of administering the program. In the second scenario, the program is costly and could reduce firm profitability if enacted. A training program would be costly if training costs exceeded screening costs and if the program increased trainees' productivity only to the level of untrained but screened workers.

The courts have been willing to impose an obligation to adopt the costless or nearly costless program.⁴⁵ They defend this obligation

based on test performance when tests have an adverse impact on minorities and when job skills can be learned through a brief training program).

⁴³ See, e.g., *Long v. Ford Motor Co.*, 496 F.2d 500, 505-06 (6th Cir. 1974); *Buckner v. Goodyear Tire & Rubber Co.*, 339 F. Supp. 1108, 1124-25 (N.D. Ala. 1972), *aff'd*, 476 F.2d 1287 (5th Cir. 1973).

⁴⁴ See sources cited *infra* note 46. At times, courts have recognized that an employer will have to bear training costs high enough to have significant impact on firm profitability but have argued that the employer must nevertheless provide the necessary training, at least as long as the trained employees will be effective. See, e.g., *Chrapliwy v. Uniroyal, Inc.*, 458 F. Supp. 252, 271 (N.D. Ind. 1977) (estimating the cost of a mandated training program as roughly \$600,000 but finding that "dollar cost alone is an immaterial consideration under the business necessity doctrine"). In other cases mandating training programs, courts have simply ignored the issue of cost. See, e.g., *United States v. St.L.-S.F. Ry.*, 464 F.2d 301, 308, 310-11 (8th Cir. 1972), *cert. denied*, 409 U.S. 1116 (1973).

⁴⁵ See *Kinsey*, 557 F.2d at 837. Arguably, courts that impose costs on the employer have avoided the need to defend that controversial choice by interpreting ambiguous facts so as to find training program reforms nearly costless. Courts may simply treat claims that high training

without regard to the weight of the applicants' claims to a particular outcome and interpret the failure to adopt an equally profitable system with less disparate impact on a protected group as a form of impermissible animus discrimination.⁴⁶ The implicit argument is that, although the initial adoption of the program may have been benign, the failure to switch to an equally cost-effective alternative can best be explained as an outgrowth of animus. The employer may claim that she is simply indifferent to the goal of increasing minority representation and that indifference can scarcely be condemned as active discrimination. But the structure of the law's demands ultimately precludes such a defense. Once the law requires the use of an equally efficient, more integrated system, one would not refuse if one truly were indifferent between the systems. The refusal to make costless reforms belies the claim of indifference the employer needs to overcome the law's initial demand to reform.

The problem with relying exclusively on this "active perpetration" model becomes clear, however, when one examines cases in which reform is costly. In such cases, an obligation to reform cannot derive from the supposition that the employer's failure to change unequivocally bespeaks impermissible animus. Instead, any duty imposed on the employer must be correlative to some sort of distributive right of potential workers. The individual applicant might have a right to receive "long-term potential marginal product" rewards or a right to a remedy for illegitimately imposed handicaps to productivity. Alternatively, the right might belong collectively to a protected group. This group-oriented right must be the right of each group to full participation in the polity and the economy.

Whatever its nature or source, the scope of the employer's duty — how much she must spend to meet that duty — remains difficult to specify normatively. It is not surprising, however, that employers should be singled out to shoulder *some* duty. Employers may well be specially able to tailor programs that overcome illegitimate handicaps or that enable applicants to reach their full potential in the particular work setting.⁴⁷

costs will be incurred if practices with adverse impact are ended as pretextual. *See, e.g.,* United States v. N.L. Indus., 479 F.2d 354, 366 (8th Cir. 1973); Jones v. Lee Way Motor Freight, Inc., 431 F.2d 245, 250 (10th Cir. 1970), *cert. denied*, 401 U.S. 954 (1971).

⁴⁶ *See, e.g.,* Guardians Ass'n of New York City Police Dep't, Inc. v. Civil Serv. Comm'n, 630 F.2d 79, 100-06 (2d Cir. 1980), *cert. denied*, 452 U.S. 940 (1981); Allen v. City of Mobile, 464 F. Supp. 433, 439-40 (S.D. Ala. 1978). *See generally* B. SCHLEI & P. GROSSMAN, *supra* note 23, at 156-57 (discussing the obligation to adopt schemes having less disparate impact).

⁴⁷ To arrive at this conclusion, one must not only consider the propriety of having employers shoulder such burdens, but also consider whether Congress in fact intended employers to shoulder them. Courts commonly interpret title VII to mean that an employer may *not* raise a cost defense once it has been established that he discriminated. *See, e.g.,* Kinsey, 557 F.2d at 837. This position is not enormously helpful, however, because the very question in many cases

Any determination of the employers' duties will in large part depend on our normative views of merit.⁴⁸ Claims of merit are always buffeted between essentialist and contingent views of persons. It is always possible to adopt a completely static and formal view of merit — an individual merits a particular benefit as long as he actually possesses the specified qualifications for the benefit.⁴⁹ Even after one discards such a hyper-formalistic formulation and acknowledges that qualifications for receiving benefits should be relevant to attaining a sensible social goal,⁵⁰ one can still maintain an essentialist focus that assumes that people either do or do not possess the requisite qualities for receiving a benefit. On the other hand, a contingent view of personhood and merit suggests that an individual merits a desired reward as long as she *could* meet all relevant qualifications once the illegitimate barriers to becoming a more qualified person are removed.⁵¹

The counterfactualism that the contingent meritocrats demand cannot be pressed too far. The question of who an individual might have been had any number of background conditions shifted is so unanswerable that its moral relevance may be of little practical moment, except as a more general argument against applying meritocratic principles.⁵² But at the legal operational level, we can demand reforms in current practices to change outcomes without determining what it would take to meet all the demands of contingent meritocrats. If a

is whether an employer has in fact discriminated if increasing minority representation requires costly alterations in employment practices. It seems best, in my view, to focus directly on the question of what sorts of costs an employer should be required to bear to integrate a work force. For example, an employer's legal obligation to spend the money to build new restroom facilities for once-excluded women workers has little bearing on whether that employer should bear the costs of training someone who would be much less productive if not given elaborate training. Yet the slogan that the cost of the remedy is not an excuse for ongoing discrimination arguably applies in both cases.

⁴⁸ In discussing meritocracy more fully in Part III, I consider why an individualistic legislator might believe that the employee ought to be able to make an affirmative claim to meritocratic treatment by claiming reward in accord with his potential product from a particular employer. For now, however, I simply assume that, *if* the employee had a claim of merit here, the employer would have a correlative duty to meet the claim.

⁴⁹ See Fallon, *supra* note 19, at 827-29.

⁵⁰ See *id.*; see also Karst & Horowitz, *Affirmative Action and Equal Protection*, 60 VA. L. REV. 955, 961-63 (1974) (arguing that qualities are recognized as meritorious when they contribute to meeting ever-shifting social needs).

⁵¹ See, e.g., Jencks, *Whom Must We Treat Equally for Educational Opportunity to Be Equal?*, 98 ETHICS 518, 519-20 (1988) (noting that advocates of "strong" humane justice argue that people do not receive adequate opportunities unless inputs are provided that permit them to develop as they would have had they not been limited by environmental or genetic factors while advocates of "weak" humane justice demand that all environmentally based handicaps to performance be corrected).

⁵² See, e.g., B. ACKERMAN, *SOCIAL JUSTICE IN THE LIBERAL STATE* 53-59, 185-86, 242-49 (1980); J. RAWLS, *A THEORY OF JUSTICE* 516 (1971).

training program in fact alters productivity, it follows that past educational deficiencies blunted productivity. Assuming that we adopt a more nuanced, contingent view of merit, we still may approach the question of the scope of the duty to remove barriers in two rather different ways. First, we can focus on the remedy and argue that any barrier to productivity equalization that is readily remedied is illegitimate simply because nothing so cheaply and easily fixed should stand in the way of people's achievements. Alternatively, we might focus on the injury and establish some substantive baseline of opportunities to measure injury. Regardless of how high the cost, protected group members would be entitled to claim this baseline level of benefits even against particular employers who were not uniquely responsible for their deprivation. We might require no greater improvements in opportunity beyond this baseline, however, no matter how low the cost. Whenever it is possible to predict how an individual would have performed had she been provided the normative baseline of opportunities, we might thus require an employer to do whatever is necessary to bring her up to the level she would have achieved without the handicap of earlier deprivations.

Title VII does not currently impose extensive remedial obligations on individual employers. At the same time, members of protected groups excluded by tests are conclusively presumed to have faced unjust, sub-baseline opportunities. Thus, the employer cannot avoid the obligation to improve the productivity of protected workers through a training program by arguing that such workers have already had an adequate chance to develop their skills. Courts apparently assume that minority group members are entitled to a little more opportunity than currently provided; they do not establish some absolute baseline of opportunity and determine whether it has been provided or order that it be provided regardless of cost.⁵³

The scope of the employer's duty could be further limited. Rather than limit the absolute expense of the training program, one could

⁵³ In cases in which the employer is ordered to implement a remedial program, the employer himself is typically found to have engaged in racist practices that hamper minority group employees' capacity to develop skills. See, e.g., *United States v. N.L. Indus., Inc.* 479 F.2d 354, 361-64 (8th Cir. 1973); *United States v. St.L.-S.F. Ry.*, 464 F.2d 301, 307-08 (8th Cir. 1972), *cert. denied*, 409 U.S. 1116 (1973); *Jones v. Lee Way Motor Freight, Inc.*, 431 F.2d 245, 247-49 (10th Cir. 1970), *cert. denied*, 401 U.S. 954 (1971). The relevance of the finding of past discrimination to the order that the "cheap" training program must be offered is, however, unclear unless we treat training as a forced in-kind compensation payment for the past illicit acts. If the order were simply directed at remedying current discrimination, one might think the employer could argue against such a remedy by showing that black employees have had the opportunity to develop the skills needed for the job without the employer's intervention and are not entitled to training once the particular exclusionary practices of the employer have ceased. That employers cannot raise such a defense suggests that the courts assume there is always an ongoing lack of opportunity.

limit the program functionally by constricting the type of training an employer is obliged to offer. Similarly, one might require an employer to raise protected group members to some normative baseline of opportunity only to the extent that the employer is uniquely capable of remedying the injury of sub-baseline opportunity. To the degree possible, then, we might require employers to establish training programs focused only on highly plant-specific skills but insist that the employer has no special obligation to engage in any general remedial education. To the extent that we justify placing the affirmative burden on the employer largely in terms of his special capacity to tailor a particular narrow remedy to a more general problem, this functional approach appears sensible. In some situations, however, an employer might reduce a test's static validity and its disparate impact more cheaply by offering more general skills training rather than job-specific training. In such cases employers could be ordered to hire screened-out workers and could be allowed to condition permanent employment on the workers' completion of state-funded job training or educational programs. Such a remedy would indeed impose costs, but it would limit the employer's costs to the administrative burdens of hiring probationary employees, including the costs inevitably borne due to increasing uncertainty about the size of the available work force. At the functional level, these appear to be social costs that the employer is uniquely capable of minimizing.

If courts focused more explicitly on the affirmative obligation issue in establishing training programs, the limits they would establish on the obligation's scope would probably incorporate both functional and absolute expense concerns. In plans to equalize performance, employers would probably be asked to bear only those social costs that they are uniquely equipped to minimize and would be asked to bear only costs that appear "reasonably low" in comparison with other costs employers incur in screening, hiring, and placing other employees.

3. *Organizational Reform.* — In the third hypothetical, applicants who do poorly on the screening device will do poorly on the job if they must perform the precise tasks now demanded of those who currently occupy the position for which they have applied. Although the test adequately measures capacity to perform the job as currently constituted, the rejected applicants could in fact produce the same output as accepted workers if the job were modified so that certain capacities, concededly well measured by the test, were no longer required. To take a simple example, tests for upper-body strength (which we will assume correlates with gender) might correlate well with the capacity to drive a bus *unless* the employer henceforth purchases buses with power steering. More realistically, it is a BFOQ for those who guard male prisoners to be male, because guards perform body searches that invade personal privacy. If, however, guard duties are reassigned so that only a subset of guards perform all body

searches, women could be guards in the reorganized workplace.⁵⁴ Similarly, lower level jobs may require facility with the English language as long as the employer hires few supervisors who can communicate with non-English speakers. If non-English speaking supervisors are hired, however, the correlation between English language proficiency and job performance may decline dramatically.⁵⁵

Other claims about dynamic discrimination may conceivably involve not workplace organizational reform, but more general changes in social behavior. Assume an employer refuses to employ a clerk with a foreign accent, correctly believing that many customers will have difficulty comprehending him. Assume also that difficulty in understanding those with less common accents is socially contingent, and that customers will ultimately adjust to the accented speaker if our civil rights law insists that accented speakers be allowed to hold positions in which they regularly communicate with the public. Statically, the failure to hire is productivity validated, and the system that deprives accented speakers of a job is nondiscriminatory; dynamically, it remains discriminatory.⁵⁶

Finally, assume that all written tests in part measure a certain "docility," a certain unquestioning cooperativeness with the test givers that may indeed prove useful in workplaces that demand docility from employees. If the workplace were less hierarchical, docility might be useless or affirmatively harmful.⁵⁷ Again, the sorts of plant-wide reorganizations that would blunt the validity of the screening mechanisms could be costless or costly to implement; if costly, the costs could be one-shot transitional costs or more permanent efficiency costs.

Obviously, the bulk of the arguments I have offered concerning whether an employer should have an affirmative obligation to train

⁵⁴ See, e.g., *Hardin v. Stynchcomb*, 691 F.2d 1364, 1373-74 (11th Cir. 1982); *Gunther v. Iowa State Men's Reformatory*, 612 F.2d 1079, 1086-87 (8th Cir.), cert. denied, 446 U.S. 966 (1980).

⁵⁵ For a general discussion of "accent" discrimination that also deals with claims that non-English speakers might make to participate in a multicultural polity, see Matsuda, *Voices of America: Accent, Antidiscrimination Law, and a Jurisprudence for the Last Reconstruction*, 100 YALE L.J. 1329 (1991). For an analogous case outside the "accent" discrimination area, see *EEOC v. Murphy Motor Freight Lines, Inc.*, 488 F. Supp. 381, 390 (D. Minn. 1980) (ordering the employer to educate supervisors so that they will be more "sensitized" to the nature of racial discrimination and to requirements of equal employment conditions for all employees).

⁵⁶ Compare *Carino v. University of Oklahoma*, 25 Fair Empl. Prac. Cas. (BNA) 1332, 1337 (W.D. Okla. 1981) (holding that the plaintiff's accent did not sufficiently impede communication with workers the plaintiff supervised to protect the employer from a title VII claim) with *Duong Nhat Tran v. City of Houston*, 35 Fair Empl. Prac. Cas. (BNA) 471, 472 (S.D. Tex. 1983) (holding that an employer's rejection of a Vietnamese employee for a position of inspector was reasonable and non-discriminatory given the employee's inability to communicate with persons "unaccustomed to the peculiarities of [his] speech").

⁵⁷ See generally S. BOWLES & H. GINTIS, *SCHOOLING IN CAPITALIST AMERICA* 53-54, 131-41 (1976) (arguing that schools prepare workers for the capitalist economy in part by making them submissive to authority).

workers who perform poorly on screening tests are germane here as well. The more difficult question, however, is whether these cases are sufficiently *different* from the training cases to justify excusing an employer from reorganizing production even though he should have an obligation to train.

The most obvious preliminary distinction is that judicial orders to train do not purport to question the employer's general managerial competence or her prerogative to organize production. In this view, title VII does not question the rationality of work design; it simply precludes employers from selecting people illegitimately to fulfill whatever design the employer might have. This distinction, however, does not necessarily have any normative force. It is hard to justify absolute collective deference to private work design decisions and simultaneously to question the employer's control over the sorts of training programs, if any, she will provide. A judicial order designed to facilitate the employment of women by requiring an employer to purchase buses with power steering is less intrusive, more readily supervisable by the court, and more readily tested for its subsequent efficacy in reducing unwarranted disparate impact than is an order requiring remedial training.

Because the distinction between ordering training and ordering workplace reform cannot be justified in terms of the institutional competence of the courts, it must be based on the perceived need to draw a bright conceptual line between justifiable claims that an employee merits a particular *existing* position and claims that positions themselves should be altered if they are not necessary to maximize productive efficiency. In a sense, training is different not because it is less costly to the employer nor because the court can assure itself that training practices will reduce unwarranted disparate impact, but because training operates most directly on the employee, fitting her to the job rather than fitting the job to her. The perceived need to draw such a distinction may arise from a sense that title VII should focus on an employee's eligibility for an already defined job rather than on her abstract merit, whether merit is conceived as ultimate productivity (as proposed here) or some other form of virtue. Focusing on abstract measures of merit misinterprets the limited scope of collective intervention the statute does and should demand. Title VII cannot, in this view, purport to ensure that each employee will be rewarded according to abstract "merit," if only because there are far too many conceptions of what constitutes merit (for example, market productivity, wit, empathy, athletic prowess, effort, etc.).⁵⁸ The statute merely purports to eliminate illegitimate barriers that private institutions

⁵⁸ See, e.g., E. PAUL, *EQUITY AND GENDER* 40-42, 111-16 (1989) (attacking the idea that market earnings should be deemed measures of "merit").

place in the way of members of protected groups who can in fact help meet the goals of those organizations.

Although it is true that title VII is not generally interpreted as legitimating or delegitimizing certain organizational ends,⁵⁹ this generalization may not justify differentiating between training cases and organizational reform cases. Intervention in work redesign cases does not necessarily require questioning the employer's announced commitment to profit maximization. The applicant claims to "fit the job," if we understand "the job" to mean producing the highest level of output in excess of input costs.

Thus, any claim that the court should be especially deferential to production design decisions seems to turn not on legitimate concern for maintaining productivity, but only on an unjustifiable presumption that judicial inquiry into production design will inevitably be more intrusive than inquiry into the impact of pre-employment training programs. It may be true that courts should refrain from acting when they cannot ascertain whether effective job redesign is feasible. Due to the complexity of the factual inquiry and the burden of creating, policing, and evaluating the efficacy of an ongoing order, a court would never intervene to demand, for example, a less hierarchical workplace designed to blunt the hypothesized impact of docility on productivity. However, if a court could magically run a short-term experiment in less hierarchical production that resulted in equal aggregate product and screened out fewer members of protected groups, the court should not hesitate to order the less hierarchical system. When implementing and evaluating the effects of job redesign in fact proves easy, redesign should be ordered.

I strongly suspect that courts that might adopt the dynamic discrimination perspective would be more disposed to demand that the employer bear one-time transitional costs than ongoing increased operational costs, even though each cost could be readily expressed in comparable present value terms. In the power steering case, for example, one could imagine three scenarios: (a) power steering costs no more to purchase or operate than manual steering, and women drivers can be hired to operate power steering units that must be purchased in the ordinary course of operation (job reform has zero cost); (b) power steering costs no more to purchase or operate than manual steering, but in order to hire women *now* the employer must incur a loss from prematurely retiring some manual steering buses (job reform has a one-time transitional cost); or (c) power steering costs more to purchase or operate (job reform is costly on an ongoing basis). My intuition is that even if the discounted present value of the cost in (b) were *higher* than the cost in (c), courts would more

⁵⁹ For a discussion of a possible exception to this broad proposition, see *infra* p. 1189.

readily order step (b), seeing it as a one-time cost of remedying past discrimination. In contrast, courts would be likely to see the costs in (c) as indicating an ongoing real gap in the net productivity (worker output minus input costs) of women and men workers.⁶⁰ In the prison guard cases, an employer's costs in planning and assigning a new set of work tasks, even if substantial, are not likely to preclude a title VII remedy, although proof that the new scheme would remain slightly more costly than the old one might defeat the plaintiffs' claim.

The intuition behind this distinction between one-time costs and ongoing costs must be that when an employer has initially had equally profitable alternative paths open to him, his arbitrary choice of a system that limits job access to members of a certain group ought not to be allowed to penalize that group. Had the bus company *always* bought buses with power steering or had guard work *initially* been assigned so as to concentrate in a subgroup of guards certain tasks that must be performed by males, no transitional costs would now be necessary. This intuition might be defended as a matter of corrective justice or as an incentive against making discriminatory decisions today that might be especially costly to remedy later. The intuition that employers should more readily be ordered to bear transitional costs might also be present, albeit less justifiably, in cases in which these rationales are less applicable. If, for instance, the decision not to buy buses with power steering were once justified because such buses were once more costly than buses without power steering, a court might still ask the employer to bear the transitional expenses associated with replacing the buses prematurely, even though the transitional costs are not the result of an earlier *unjustified* decision.

4. *Summary.* — It seems relatively uncontroversial to argue that dynamically invalid screening devices are illegitimate when the employer has an independent duty to alter work conditions in ways that would reduce the predictive validity of the screening device. It also appears fairly unexceptionable to require an employer to adopt costless changes in work organization or job training programs that will reduce the validity of statically valid devices. When altering production techniques or providing training entails either transitional costs or long-term costs, however, the applicant's claim that she is being treated illegitimately is harder to sustain. In such cases, the applicant's claim must be that her entitlement to gain rewards in accord with her readily attainable, potential merit is strong enough to impose a duty on the employer to shoulder costs in order to vindicate the applicant's entitlement. Courts ought to be most willing to impose such costs on an employer when they are relatively insubstantial, most

⁶⁰ Such cases ultimately raise issues of when applicants are entitled to protection from what I call discrimination against those with equal individual output. See *infra* pp. 1198-1203.

rationality borne by the employer, who can most cheaply develop the employee's potential merit, and relatively easily ascertained by a court, which can order a particular organizational reform without supervising the entire production process. However, the assertion that employers owe employees any costly affirmative duties to meet meritocratic claims, either as a matter of current positive law or as a matter of ideal theory, is highly controversial. I return to this issue in Part III.

C. Distributive Discrimination

The goal of antidiscrimination law, as conventionally understood, is simply to alter the identity of the persons who will occupy particular slots in a fundamentally unaltered employment and compensation hierarchy.⁶¹ In this mainstream view, title VII is not designed to alter the shape of the income distribution curve. Rather, it rearranges the people who occupy particular points on the curve, permitting members of protected groups to occupy more desirable points from which they had been wrongfully excluded. In the most commonplace, zero-sum conception of title VII, more qualified members of the legislatively protected groups displace less qualified members of the socially privileged group.⁶² A more optimistic version of this position argues that title VII serves, at least in part, to eliminate the underemployment of protected group members, thus spurring overall growth.⁶³ In this

⁶¹ See, e.g., Maltz, *Title VII and Upper Level Employment — A Response to Professor Bartholet*, 77 NW. U.L. REV. 776, 784 (1983) (arguing that title VII is intended to guarantee minorities the opportunity to perform jobs previously reserved for whites, not to guarantee equal outcomes either by guaranteeing minority access to a given set of jobs or by altering the job structure); Note, *Title VII Today: The Shift Away From Equality*, 20 J. MARSHALL L. REV. 525, 528 (1987) (same). Even those commentators who argue that title VII could be understood to mandate equal outcomes for blacks and whites as groups see the statute as requiring the integration of blacks into existing positions held by whites, not the flattening of existing economic hierarchies. See, e.g., Fiss, *A Theory of Fair Employment Laws*, 38 U. CHI. L. REV. 235, 245-46 (1971).

⁶² See, e.g., Fiss, *supra* note 61, at 245-46.

⁶³ See Caldwell, *Reaffirming the Disproportionate Effects Standard of Liability in Title VII Litigation*, 46 U. PITT. L. REV. 555, 579-83 (1985); see also Cooper & Sobol, *Seniority and Testing Under Fair Employment Laws: A General Approach to Objective Criteria of Hiring and Promotion*, 82 HARV. L. REV. 1598, 1678 & n.37 (1969) (discussing empirical evidence of success of programs in activating employees' untapped potential). This view was surely held by many of those who initially supported the passage of title VII. Senator Humphrey's statement is typical: "There is considerable evidence to demonstrate that permitting people to be hired on the basis of their qualifications not only helps business, but also improves the total national economy." 110 CONG. REC. 13,088 (1964). For a different, but no less favorable, account of the efficiency implications of title VII, see Donohue, *Is Title VII Efficient?*, 134 U. PA. L. REV. 1411, 1423-31 (1986) (focusing on the role of title VII in facilitating the inevitable movement of a competitive economy toward the elimination of employment discrimination by driving discriminatory firms' costs up and profits down).

view, income may be distributed more equally once antidiscrimination laws are aggressively enforced, not because the statute has been used to either raise pay for some existing jobs or lower it for others, but simply because low-paying jobs that underutilize workers' skills will have disappeared.

More politically radical commentators argue that the antidiscrimination principle demands more than the removal of illegitimate barriers to the attainment of rewards in the current system. In this view, the reward structure itself may be discriminatory, particularly insofar as it favors traits that legislatively protected groups tend to lack and ignores the traits protected groups possess in relative abundance.⁶⁴

For a more concrete example, consider the following hypothetical. Imagine a psychological test battery that measures aggression. Imagine further that those with higher aggression scores perform better as litigators and that male applicants are, on average, markedly more aggressive than female applicants. Could a law firm legitimately use the aggressiveness test as a screening device in hiring attorneys? The test appears statically valid. It also might be considered dynamically valid, at least to the extent that the firm could not reasonably be expected to alter the relationship between aggression and success in litigation.⁶⁵

It is often difficult to discern what pure theories of distributive invalidity would look like, because the more politically radical commentators who have sketched them would generally deny the plausibility of *any* hypothetical like the one offered above. Although they emphasize the importance of differences between cultural groups, these commentators generally discuss only the ways in which the special strengths of historically subordinated groups have been inadequately appreciated.⁶⁶ Few discuss directly what the appropriate

⁶⁴ This perspective has been enunciated largely in the context of gender discrimination rather than racial discrimination. See, e.g., Abrams, *Gender Discrimination and the Transformation of Workplace Norms*, 42 VAND. L. REV. 1183, 1189-92 (1989); Becker, *Barriers Facing Women in the Wage-Labor Market and the Need for Additional Remedies: A Reply to Fischel and Lazear*, 53 U. CHI. L. REV. 934, 936 (1986); Littleton, *supra* note 13, at 1280, 1321-23; Poole, *On Merit*, 1 LAW & INEQUALITY 155, 157 (1986). As Littleton notes, lawyers remain committed to the idea that blacks are not different from whites in ways that would be relevant to antidiscrimination law. See Littleton, *supra* note 13, at 1288-41. By contrast, women's most obvious differences from men — particularly their ability to become pregnant and the disproportionate share of household and child rearing responsibilities they have historically borne — have led legal feminists to confront problems of what equal treatment of non-identical people should mean. See *id.*

⁶⁵ The relationship between aggression and litigation skill, or the relationship between gender and aggression, may not be culturally immutable over the longer run. I assume, however, that an employer breaches a duty not to discriminate dynamically only by failing to make changes within her short-run control.

⁶⁶ See M. DALY, *GYN/ECOLOGY* 17-22 (1978); Matsuda, *Affirmative Action and Legal Knowl-*

response should be in the event that differences *handicap* the members of a title VII protected group in the performance of a significant task;⁶⁷ yet, it is this situation that most directly raises distributional issues.

In a case such as the one I have hypothesized, most politically radical commentators would argue that the aggressiveness test is almost certain to be statically invalid. They would argue that, due in large part to illegitimate and unexamined stereotypes, the employer's measures of output irrationally emphasize those aspects of output that men perform best. The evaluator may simply overvalue the skills men possess, or overvalue particular skills *because* men possess them and it seems commonsensical to those in power that they have measured productivity accurately if they discover that men are more productive on whatever scale they have formulated.⁶⁸ Even if aggressiveness does help in certain litigation settings, it may hurt in negotiation, counseling, and a variety of other tasks an attorney must perform.⁶⁹ If employers irrationally underestimate the productive value of the tasks that members of protected groups best perform, especially if they do so in part precisely because those tasks are associated with members of the protected but socially devalued group, the static validity of screening devices with disparate impact will be systematically overestimated. The criterion (job performance) will be defined too narrowly. It will stress those elements of performance that the predictor (screening test) values, which are in turn the strengths of the socially dominant group.

edge, 11 HARV. WOMEN'S L.J. 1, 5 (1987) [hereinafter Matsuda, *Affirmative Action*]; Matsuda, *Looking to the Bottom: Critical Legal Studies and Reparations*, 22 HARV. C.R.-C.L. L. REV. 323, 324, 346 (1987) [hereinafter Matsuda, *Looking to the Bottom*]. But see Williams, *Deconstructing Gender*, 87 MICH. L. REV. 797, 813-21 (1989) (arguing that it is tactically disadvantageous to laud women's differences because empowered men will continue to devalue traits identified as female).

⁶⁷ Some scholars have argued that subgroups do have distinct cognitive or linguistic styles — distinct and valuable “ways of knowing.” See, e.g., J. HALE-BENSON, *BLACK CHILDREN: THEIR ROOTS, CULTURE, AND LEARNING STYLE* 4-8 (1986); Lang, *A Language Theory of Discrimination*, 101 Q.J. ECON. 363, 367 (1986). The most plausible theory that sees cultural separateness as a productive handicap argues that linguistic barriers (broadly defined) between majority and minority group members dampens the productivity of linguistically integrated firms. See Lang, *supra*, at 363-64, 369-77. In this view, there is nothing intrinsically wrong with “Black English,” but if whites do not understand it well, the presence of black workers in a majority-dominated workplace will hamper productivity. It is, of course, difficult to sort out whether the use of “Black English” genuinely handicaps communication or simply causes majority group members to devalue the users' competence. Moreover, the genuine difficulties that those in hierarchically superior positions have in understanding social subordinates may be culturally contingent and mutable as well. See Matsuda, *supra* note 55, at 1352 & nn.82-84. If, nevertheless, communication handicaps do exist, one would expect rational employers to segregate work forces by common language, not by race per se, and one would anticipate that minority workers would face constricted demand for their labor.

⁶⁸ See *supra* p. 1168.

⁶⁹ See, e.g., Menkel-Meadow, *Portia in a Different Voice: Speculations on a Woman's Lawyering Process*, 1 BERKELEY WOMEN'S L.J. 39, 55-58 (1985).

These more radical theories of antidiscrimination law necessarily rely on something more than liberal skepticism about the legitimacy of a productivity-based distributive system. Welfare-state liberals believe that the state can and should consciously redistribute income from those who are paid in accord with their higher marginal product to those with lower productivity.⁷⁰ At times, mainstream legal scholars, following Rawls, deny the propriety of any merit-based distribution.⁷¹ These scholars argue that a person does not deserve his productive capacity because it is a function of two factors — genetic predisposition and social learning — that the individual does not deserve in any moral sense.⁷² But such scholars generally assume that if productivity-based distribution is unjust, it is to be corrected not at the private or firm level, but through collective tax and transfer programs.⁷³ The more politically radical commentators reject the implicit assumption that collectivities intervene in distribution only when they most self-consciously redistribute. Instead, these commentators emphasize the extent to which all collective decisions about entitlement structure, including decisions that define illegal discrimination, frame a distinct distribution.⁷⁴ These commentators make no

⁷⁰ See J. RAWLS, *supra* note 52, at 60–75 (arguing that the difference principle requires that inequalities be permitted only insofar as they redound to the benefit of the representative worst-off person); see also R. MUSGRAVE, *THE THEORY OF PUBLIC FINANCE* 17–22 (1959) (describing one of the three main roles of budget policy as distributing income more equitably than the market).

⁷¹ See, e.g., Michelman, *In Pursuit of Constitutional Welfare Rights: One View of Rawls' Theory of Justice*, 121 U. PA. L. REV. 962, 972–76 & n.42 (1973) (noting that tensions between liberty of property disposition and the egalitarian difference principle find resolution in the observation that because accidents of birth and natural endowment are morally irrelevant, social institutions should allow nature's endowment of ability to work for the good of all, not just for those who are gifted). *But see* Grey, *Property and Need: The Welfare State and Theories of Distributive Justice*, 28 STAN. L. REV. 877, 883–86 (1976) (arguing against Rawls's notion that one's distributive share must be deserved).

⁷² See J. RAWLS, *supra* note 52, at 73–75, 100–04, 310–15; see also Nagel, *Equal Treatment and Compensatory Discrimination*, 2 PHIL. & PUB. AFF. 348, 353–55 (1973) (arguing that differential abilities usually do not determine whether individuals deserve economic and social benefits). Persuasive criticisms of Rawls's view of desert can be found in R. NOZICK, *supra* note 8, at 213–31; and M. SANDEL, *LIBERALISM AND THE LIMITS OF JUSTICE* 72–103 (1982).

⁷³ See R. MUSGRAVE, *supra* note 70, at 17–22; Shavell, *A Note on Efficiency vs. Distributional Equity in Legal Rulemaking*, 71 AM. ECON. A. PAPERS & PROC. 414, 414 (1981) (arguing that legal rules should be designed to promote only efficiency, not equity, even in a world of optimal tax and transfer policy, because redistributive rules will generate the same work disincentives as non-optimal explicit taxes). Some limits on the reach of this view in mainstream thought are discussed in Kronman, *Contract Law and Distributive Justice*, 89 YALE L.J. 472 (1980).

⁷⁴ Cf. M. KELMAN, *A GUIDE TO CRITICAL LEGAL STUDIES* 103–04, 165–66 (1987) (observing that government enforcement of property rights also distributes legal and economic burdens between owners and non-owners); Hale, *Coercion and Distribution in a Supposedly Non-Coercive State*, 38 POL. SCI. Q. 470, 471–72 (1923) (arguing that the initial distribution of property in any system determines that system's subsequent production of goods and services).

normative presumptions in favor of the tax and transfer method of shaping distributive outcomes but believe it perfectly appropriate to reorder private hiring practices in order to guarantee a more desirable distributive outcome.⁷⁵

Although distributive discrimination theories have hardly been at the center of title VII jurisprudence or commentary, I discuss below some possible theories of distributive discrimination that might justify state-mandated shifts in hiring patterns in a case such as my hypothetical about aggressive litigators. I will discuss in turn the three themes on which most arguments focus: skepticism about productivity, subsidy theory, and reparations.

1. *Skepticism About Productivity.* — The first theory expresses a normative skepticism about the concept of productivity. In essence, proponents of this view argue that the ends sought by a particular organization are inevitably poorly defined, particularly when the organization is not a private firm committed to profit maximization. When an organization has vaguely defined goals, measures of its "output" are necessarily contestable, because it is impossible to say with certainty what goals define its useful output.⁷⁶ If the goals and the related measures of their achievement by particular workers are contested, a multiplicity of cultural perspectives should be used to determine definitively the organization's proper objectives.⁷⁷ There may be no single "right" answer to the question of proper organizational objectives, just a series of culturally distinct perspectives on what the right answer should be. To the degree that those who control an organization exclude members of protected groups ostensibly because they do not contribute as much to the achievement of the organization's aims, those in control will continue to define the organization's aims too narrowly.

Arguments of this form are most readily made when the employer is a nonprofit organization, particularly one that produces goods rarely thought to be valued properly by markets. Thus, if universities had to validate *any* sort of screening device used to hire academics (for example, paper credentials or past scholarly production), they would need what is surely unavailable: a universally acceptable *output* measure for academic excellence.⁷⁸ Politically radical commentators sug-

⁷⁵ See, e.g., Abrams, *supra* note 64, at 1226-33; Littleton, *supra* note 13, at 1324-36.

⁷⁶ See, e.g., Bartholet, *supra* note 23, at 1019-23, 1026-27. The problem perhaps most visibly arises in the case law when courts discuss the meaning of "academic merit" in tenure review cases. See, e.g., Lynn v. Regents of the Univ. of Cal., 656 F.2d 1337, 1342-45 (9th Cir. 1981), *cert. denied*, 459 U.S. 823 (1982).

⁷⁷ See Bartholet, *supra* note 23, at 1026-27; Poole, *supra* note 64, at 157-58.

⁷⁸ Professor Randall Kennedy argues that those who claim the input of black scholars on race-related issues is understated must demonstrate that *good* work is ignored. See Kennedy, *Racial Critiques of Legal Academia*, 102 HARV. L. REV. 1745, 1770-78 (1989). However, his view cannot persuade those who would claim that judgments about what work is good are

gest that in cases in which no output measure can be unanimously agreed upon, organizations must ensure representation of all protected groups.⁷⁹ One can justify proportional representation simply as the default solution for an employer whose selection process yields a disparate impact that the employer cannot validate by appealing to uncontroversial principles. Alternatively, a more affirmative claim might be that the organization will, loosely speaking, maximize output by getting contributions from people with diverse understandings of good output.⁸⁰ Yet another affirmative claim is that proportional representation is not just a convenient baseline but an outright good that promotes the leveling of social hierarchies and ensures all social groups a part in the polity and civil society.⁸¹

There is some support in existing case law for views skeptical of organizations' capacity to define *legitimate* goals. Courts have allowed race-conscious hiring designed to take advantage of the special skills minority group members bring to a job. Courts have held, for instance, that a black applicant may be more qualified for a job in a police department than an otherwise similar white applicant because a black officer can function more effectively in black neighborhoods.⁸² Although it is possible that the black policeman's advantage can be quantified, it seems clear that part of what lies behind such a decision is the assumption that a policeman's output cannot readily be measured in an uncontroversial way. The existing output measures that racially exclusive departments use downplay the unique contributions and perspectives that minority group members bring to the job.

themselves illicit. For discussions of the illegitimacy of mainstream judgments of merit, see Espinoza, *Masks and Other Disguises: Exposing Legal Academia*, 103 HARV. L. REV. 1878, 1882-83 (1990); and Poole, *supra* note 64, at 155-58.

⁷⁹ See, e.g., Kennedy, *Comment: A Cultural Pluralist Case for Affirmative Action in Legal Academia*, 1990 DUKE L.J. 705.

⁸⁰ As Professor Bartholet puts it:

Is the best school principal one who possesses advanced academic degrees, or one who will seek to learn about and fulfill the needs of the community the school serves? Is the best doctor one who will excel at state-of-the-art surgical procedures, or one who will provide badly needed health care to poor families in unglamorous communities? To the extent that such value judgments are inevitable . . . it is hard to justify selection devices with a racially exclusionary impact. Selection on a racially proportionate basis seems an appropriate solution. Ethnic minorities would then be represented in positions in which they could make the essentially political judgments about . . . how jobs carrying significant power to influence society should look.

Bartholet, *supra* note 23, at 1027.

⁸¹ See Kennedy, *supra* note 79, at 712-21.

⁸² See, e.g., *Detroit Police Officers' Ass'n v. Young*, 608 F.2d 671, 695-96 (6th Cir. 1979), *cert. denied*, 452 U.S. 938 (1981); *Bridgeport Guardians, Inc. v. Members of the Bridgeport Civil Serv. Comm'n*, 482 F.2d 1333, 1341 (2d Cir. 1973), *cert. denied*, 421 U.S. 991 (1975); *cf. Regents of the Univ. of Cal. v. Bakke*, 438 U.S. 265, 314 (1978) (Powell, J.) (arguing that a diverse student body in a medical school may enrich the training of students and better equip graduates to render their service to humanity with compassion).

Most mainstream interpreters of the antidiscrimination principle would doubtless agree that an organization cannot defend its screening practices as necessary for the achievement of just *any* declared set of ends, especially if a court believes these stated ends are actually pretexts for group exclusion.⁸³ Nonetheless, deference to an organization's statement of its goals would be substantially greater under the conventional view than under the politically radical one. Unless a statement of ends appears implausible or offered in bad faith, the conventional commentator would permit each private entity to declare its own ends, without allowing a court to question whether that entity might reasonably have adopted other objectives.

Thus, assume our aggressiveness-tested lawyers seek employment, not at a profit-maximizing firm, but at a government agency or a nonprofit organization. The conventional analyst would surely think it legitimate for such agencies to select applicants on the basis of litigation skill, even if he acknowledged that the focus on litigation might be debatable. Another employer's legitimate desire to hire lawyers who had better negotiation skills, or who were more able to collaborate with clients to empower them to solve their own problems, would be of no moment, even if members of protected groups would less frequently be excluded from an entity with this latter set of aims.

The more politically radical analyst, however, would almost surely deny the implicit substantive premise of the mainstream position: the belief that the diversity of private organizations will provide adequate opportunities for members of protected groups if their preferred ends really are worth seeking.⁸⁴ Radicals would argue instead that orga-

⁸³ See, e.g., *Hardin v. Stynchcomb*, 691 F.2d 1364, 1372-74 (11th Cir. 1982) (holding that there is no substantial reason to defer to the sheriff's office's judgment that deputy sheriffs must spend six months working in the jail before taking on other duties); *Diaz v. Pan Am. World Airways, Inc.*, 442 F.2d 385, 388-89 (5th Cir.) (holding that the primary purpose of flight attendants is to help deliver passengers safely, so that the airline's determination that one job of a flight attendant was to reassure passengers, a function the court conceded for argument's sake might be better performed by women, is not dispositive), *cert. denied*, 404 U.S. 950 (1971); *Wilson v. Southwest Airlines Co.*, 517 F. Supp. 292, 304 (N.D. Tex. 1981) (rejecting airline's determination that one of a flight attendant's roles is exuding feminine spirit and sex appeal, at least in the absence of a convincing demonstration from the airline that profitability was increased as a result of such factors). Courts will not even allow hiring exclusions that increase profitability if the increase results simply because the employer's customers do not want to be served by members of a protected group. See, e.g., *Fernandez v. Wynn Oil Co.*, 653 F.2d 1273, 1276-77 (9th Cir. 1981).

⁸⁴ In the analogous area of diversity of expression, the notion that cultural uniformity is a product solely of the state monopolization of production heavily influences opponents of state regulation of the press, even when the state simply mandates access. See, e.g., B. SCHMIDT, *FREEDOM OF THE PRESS VS. PUBLIC ACCESS* 29 (1976). But see J. BROWN, *FREEDOM OF THE PRESS FOR WHOM? THE RIGHT OF ACCESS TO MASS MEDIA* 4-5 (1973) (arguing that voices may go unheard even when the private sphere is free from governmental interference); Yackle, *Confessions of a Horizontalist: A Dialogue on the First Amendment*, 27 KAN. L. REV. 541, 546-47 (1979) (same).

nizations dominated by members of socially privileged groups will tend to choose organizational ends that require the sorts of skills that the socially privileged more frequently have, even where the choice is unlikely to be a pretext for exclusion. A combination of disguised self-interest⁸⁵ and a genuine belief in the universal significance of the aims deemed important within one's own subculture will bias decisionmaking. Biased decisionmaking will result in the ongoing exclusion of those who would otherwise challenge the tacit consensus about proper ends. Radicals might not directly question the other key bulwark of the mainstream position — the *procedural* skepticism of mainstream scholars concerning the competence of courts to review the legitimacy of an organization's chosen ends. Radicals might resolve this lack of faith in judicial inquiry differently by directing courts to refuse to certify *any* contestable ends that perpetuate disparate impact in hiring practices.

2. *Subsidy Arguments.* — Arguments based on the contestability of organizational goals will be less cogent if the employer claims that she seeks solely to maximize profits. In this situation, those who challenge the distributive validity of a test that selects in a profit-maximizing fashion might demand that the firm take the socially valuable contributions of excluded workers into account in its hiring decisions, even if these contributions do not benefit the firm in particular. Commentators rarely make this argument in its pure form. Rather, they generally link it to an extreme factual skepticism about the claim that some types of employees better promote a firm's profitability than others. Thus, in the radical view our hypothetical law firm overestimates not only the social importance of litigation skills, but also the revenue-generating power of aggressive litigators.⁸⁶ Similarly, in the radical view firms would typically undervalue the skills of an excluded manager who might be better able to motivate workers, even though such skills would ultimately increase profitability.⁸⁷

If one sets aside skepticism that firms do not invariably maximize profits, the radical position could be put, without grotesque distortion, in terms of relatively traditional subsidy arguments. Take another hypothetical law firm that deliberately selects employees whose psychological tests or interview statements show that the employees will readily accept separation from their children. Assume too that the reluctance to separate from children is more pronounced in women and, finally, that the incapacity to separate reduces a lawyer's in-firm productivity (perhaps because being responsible for children makes it

⁸⁵ See Bell, *The Supreme Court, 1984 Term — Foreword: The Civil Rights Chronicles*, 99 HARV. L. REV. 4, 9-10, 23, 75-77 (1985).

⁸⁶ Cf. Frug, *supra* note 12, at 94-103 (discussing alternate methods of organizing work that would aid women without reducing a firm's profitability).

⁸⁷ See Barthelet, *supra* note 23, at 1027.

harder to work late or to be available whenever clients require). At the same time, assume that this inability to separate is concededly of great social importance, because it helps to ensure that children are properly nurtured. In such a case, subsidies might help induce socially optimal levels of child care by reducing socially undesirable pressures on women to conform to the firm's values — that is, to abandon nurturing for less socially valuable lawyering. Because markets in nurturing are inadequately developed, the actual "demand" for nurturing will not be manifest unless explicit subsidies are provided to cure the market failure.⁸⁸

One might consider subsidies appropriate even if women would nurture for free. Such a conclusion would rest on a distributive principle that required society to reward all socially valuable contributions made by members of traditionally devalued groups.⁸⁹ Interference in employers' privately rational decisions might well be appropriate if one insisted that group income should correspond to moral worth and if one recognized that there is no way to gauge the relative worth of different "ways of being." Presumably, such a claim would be based on a theory of group, rather than individual rights; there appears to be no corresponding argument for equal distribution based on the moral equality of individuals.⁹⁰ Under this view, when we

⁸⁸ This seems true even under the most conventional economic assumptions about the conditions under which subsidies are needed. Numerous transaction cost barriers hamper the organization of a market for those who otherwise would willingly pay for "intimate" child care providers. Free rider and contract enforcement barriers preclude third parties interested in happier children — and eventually more productive or at least less sociopathic adults — from manifesting their preferences. For straightforward discussions of transaction costs, see R. POSNER, *ECONOMIC ANALYSIS OF LAW* 54–55 (3d ed. 1986); and Calabresi, *Transaction Costs, Resource Allocation, and Liability Rules — A Comment*, 11 J.L. & ECON. 67 (1968). The difficulty of specifying what sort of care is owed, capital market imperfections, and cognitive incapacity would all preclude children themselves from bargaining for optimal care. See R. POSNER, *ECONOMIC ANALYSIS OF LAW* 111–12 (2d ed. 1977) (explaining child protective laws on the basis of children's inability to bargain). Economistic values aside, heightened care might be socially desirable. The relative lack of market power in the hands of those who would redirect social resources in a particular way hardly demonstrates the impropriety of such a redirection. First, current social patterns of demand are obviously a function of the existing distribution of resources. A change in the distribution of resources might well produce a different pattern of demand. See, e.g., R. POSNER, *ECONOMIC ANALYSIS OF LAW* 4–5, 10 (3d ed. 1986). Moreover, some social practices, like the neglect of children, can be judged without any regard to their actual or potential economic efficiency: as Robin West notes, rape is wrong not because, as Posner suggests, rapists circumvent the market in sex (prostitution). See Posner, *An Economic Theory of the Criminal Law*, 85 COLUM. L. REV. 1193, 1197, 1198–99 (1985). Rather, prostitution and rape are each wrong because they victimize women. See West, *Submission, Choice and Ethics: A Rejoinder to Judge Posner*, 99 HARV. L. REV. 1449, 1449–50 (1986). Analogously, neglect is wrong because it causes children to suffer and may harm them in later life as well.

⁸⁹ See Littleton, *supra* note 13, at 1284–85; Poole, *supra* note 64, at 155–58.

⁹⁰ Although it is perfectly reasonable to oppose any distinction in material reward based on productivity, such a view would be grounded in some variety of general egalitarianism, not in a specific focus on cherishing multiculturalism. Professor Littleton attempts, not entirely suc-

observe real group differences of any type (for example, differential performance on a job screening test), we must act to ensure that those differences have no financial consequences for the group. If our only interest is securing equal rewards among groups, we can attain that objective through hiring quotas or race-norming of tests.⁹¹ If a group's basic claim of right is that its unique cultural style and set of contributions should be *legitimated*, however, only the abolition of such tests will suffice.

Conventional analysis rejects the idea that a firm must not only maximize profits but also subsidize social subgroups, even if the subsidy encourages socially useful behavior or provides more appropriate rewards to those groups.⁹² The notion that firms should compensate applicants for socially useful work that does not directly benefit the firm will probably meet with more hostility from the conventional analyst than the idea that non-profit-maximizing firms should broaden their conception of desirable output. Even if we assume, for the sake of argument, that subsidies are appropriate, the conventional analyst will ask why the private law firm should finance the subsidy.

cessfully, to develop an affirmative distributive principle that gives distinct social subgroups compelling claims to equal per capita income, independent of the claim that every individual is entitled to an equal share of her group's income. See Littleton, *supra* note 13.

Theorists such as Rawls arrive at a presumptive egalitarian position by rejecting merit-based distinctions based on natural talent. See J. RAWLS, *supra* note 52, at 100-04. Littleton, however, rejects such egalitarianism by making three rather broad assumptions: she credits the significance of merit as a moral category, she believes that distinctions in merit form a reasonable basis for distributive distinctions, and she seems to believe that the sort of merit that ought to ground legitimate distributive distinctions, although unequal among persons, must necessarily be equally distributed among groups. Why Littleton rejects a more thoroughgoing presumptive egalitarianism, why she believes merit, if it exists, is a good basis for the distribution of income, or why she believes that whatever merit she perceives is unevenly distributed among individuals but not among groups, remains unclear.

⁹¹ When tests are race-normed, a black whose score was in the 75th percentile of black test-takers is reported to have received the same score on the test as a white applicant in the 75th percentile of white test-takers, who have received, as a group, higher raw scores. For a fuller discussion of race-norming in the GATB, see FAIRNESS, *supra* note 1, at 20-21; and INTERIM REPORT: WITHIN-GROUP SCORING OF THE GENERAL APTITUDE TEST BATTERY 27-52 (A. Wigdor & J. Hartigan eds. 1988) [hereinafter INTERIM REPORT]. Given the widespread knowledge that blacks score roughly one standard deviation lower than whites on the GATB, race-norming in and of itself does not necessarily help blacks. Generally speaking, however, those who decide to use race-normed tests are committed to treating those with equal race-normed scores equally (sometimes, of course, by hiring in accord with normed test scores without regard to the applicants' racial identities). Race-norming, then, does not really deprive employers of information, because employers could recalculate raw scores from race-normed scores by using publicly available information. Rather, race-norming simply reflects a substantive commitment to interpreting raw test scores in a fashion that facilitates the hiring of workers from groups that do poorly on the test.

⁹² For a good restatement of the mainstream argument for private ordering, see Mnookin, *Divorce Bargaining: The Limits on Private Ordering*, 18 MICH. J.L. REFORM 1015 (1985).

The radical response to this question is difficult to anticipate, because the radical case for intervention does not explicitly justify subsidies. As a rhetorical matter, a reluctance to defend subsidies is understandable. The use of subsidy language implies that the distribution that would occur in the absence of the subsidy is in some sense legitimate; thus, the subsidy is a secondary, corrective transfer payment. Radical commentators, however, deny the propriety of a system that ignores the contributions of protected group members in its initial distribution of rewards. Their goal is to reformulate the background rules which frame free market choices.

I suspect, however, that radical commentators' reluctance to confront the "subsidy" question is ultimately more substantive than rhetorical. Substantively, the notion of a subsidy is unacceptable to radicals because the best case for subsidizing through antidiscrimination law depends on accepting the legitimacy of class division. Take our hypothetical in which women separate less readily from children. This attribute reduces the profitability of the law firm but promotes the welfare of children and society. If the state itself directly subsidizes child rearing, political considerations almost surely preclude a program that subsidizes women at grossly different rates depending on the opportunity costs of their time (what they could earn if they were to work outside the home).⁹³ It is difficult to imagine Congress enacting a program that gives greater direct credits to lawyers who stay at home than to their secretaries who do so, either on the perverse assumption that we should redistribute more to richer people or on the ugly assumption that it is more important to nurture the children of the rich than other children. But in a class-divided society, the levels of subsidy needed to ensure uniformly high levels of child care must vary with each woman's marketplace alternative. Women who could make more money if they were working outside the home will have to be offered more not to do so.

It is, of course, conceivable that a subsidizing collective would not think inducing a particular woman to nurture was socially desirable

⁹³ Proposals such as Littleton's, *see supra* note 13, which apparently seek to revalue women's particular cultural contributions through antidiscrimination norms without a thorough restructuring of the market system, face difficulties even in meeting their own goals. Antidiscrimination law can only instruct an employer not to penalize existing female employees for manifesting "feminine" traits; it does not create a new source of pay for women who have such traits. Thus, it is not necessarily the best exemplars of "feminine" traits who will be rewarded in Littleton's scheme. If a woman is not to be punished for her "feminine" traits, she must have the same access to power and money she would have had if she manifested "masculine" traits. The level of reward each woman receives will thus be *not* by the value of the female traits but by the opportunity cost of foregoing a woman's capacity as measured in "male" terms. A female lawyer gets paid as a male lawyer for mothering; a female secretary gets paid as a male secretary for mothering. The resultant inequality between female lawyer and secretary is surely not grounded in the evaluation of their relative mothering skills.

if her opportunity costs were especially high; but it is also conceivable that optimal subsidies would be set above the level needed to induce those women with the lowest potential market wage to choose nurturing. The only realistic way to enact such a subsidy system would be to force employers to hire those women whom they would have selected but for their incapacity to separate from children and demand that the employers not penalize these women for spending more time at home.⁹⁴ Each woman would then receive only that subsidy necessary to induce her to make the socially optimal set of decisions.⁹⁵ We have apparently targeted the spending effectively to minimize "excess transfers." We have done so, however, at the cost of duplicating and entrenching existing class stratification.

3. *Reparations.* — There is a distinct variant of the subsidy argument that is worth parsing out as a third strand of a theory of distributive invalidity. In this variant, the firm should be obliged to serve as a conduit for direct redistributive payments to members of the protected class. The purportedly more radical elaborations of such a theory hold that members of the protected class deserve these payments as reparations for past discrimination.⁹⁶ More conventional

⁹⁴ Obviously, *outside* the realm of explicit, conscious state intervention, such unequal subsidies occur all the time as a result of economic sharing in long-term relationships. Because women with high earning potential are often involved with female or male partners with similarly high earning potential, and because there is relatively little distinction in personal consumption levels within households, women in stable relationships who can take advantage of shared consumption are not usually penalized, *in consumption terms*, by their decision to choose socially devalued nurturing work. See Lazear, *Symposium on Women in the Labor Market*, 3 J. ECON. PERSP. 3, 5-6 (1989) (stating that women's income is typically lower than women's consumption, although measuring consumption, including the value of free time, is difficult). Naturally, even women in stable relationships who make few consumption sacrifices do sacrifice independence and other sorts of social power. Moreover, many women are not in stable relationships, and all run the risk that seemingly stable relationships will dissolve. Nonetheless, "nurturing" women who choose to do traditional "women's" work in traditional households in which the husbands have higher earning potentials than their wives will frequently consume as much as single women who participate in (male) market activities. Andrea Dworkin contends that she has found a feminist core in right-wing women's opposition to abortion and support for traditional marriage: each position minimizes the number of men with whom — and perhaps the absolute number of times that — intercourse will be demanded. See A. DWORKIN, *RIGHT-WING WOMEN* 71-105 (1983). Perhaps, too, right-wing traditionalists discern the only realistic path to "equality as acceptance" — pronounced differentiation without economic consequences — as submission to patriarchal dictatorship and social marginalization.

⁹⁵ It is unclear whether the law firm's partners, its clients, or its clients' customers would ultimately bear the cost of the subsidy. Thus, we cannot say whether the implicit tax to finance the subsidy would be proportional, regressive, or progressive when compared to, say, the federal income tax.

⁹⁶ Several authors defend the propriety of reparations. See, e.g., D. BELL, *AND WE ARE NOT SAVED* 123-39 (1987); Delgado, *The Imperial Scholar: Reflections on a Review of Civil Rights Literature*, 132 U. PA. L. REV. 561, 569 (1984); Matsuda, *Looking to the Bottom*, *supra* note 66, at 362-99. Reparations have been endorsed by decidedly "non-radical" writers as well. See, e.g., B. BITTKER, *THE CASE FOR BLACK REPARATIONS* (1973); R. NOZICK, *supra* note 8, at 152-53; Thomson, *Preferential Hiring*, 2 PHIL. & PUB. AFF. 364, 380-84 (1973).

liberal theories might argue instead that the redistributive payments serve forward-looking ends. By placing members of protected groups in positions for which they would not otherwise "qualify," the redistributive programs enhance opportunities for the next generation of protected group members significantly more than would programs focused solely on children.⁹⁷

Reparations theories take a number of forms.⁹⁸ Their most important common feature is the notion that someone receives reparations only if a special program offers her something to which she would not otherwise be legally entitled.⁹⁹ Thus, a party does not seek reparations until she has exhausted all her other claims. If, for instance, a statically invalid screening test excludes a protected group member from a job, a court order to employ her is not a grant of reparations. Rather, it is a recognition of her separate legal entitlement not to be the victim of illegitimate static discrimination.

Politically radical commentators usually defend reparations as a recognition of "group rights."¹⁰⁰ Despite this rhetoric linking reparations to group claims, however, I doubt that requiring shifts in private hiring decisions will in fact help the protected group as a group. On the other hand, if the fundamental arguments for reparations are actually individualistic, the case for firm-based reparations is stronger.¹⁰¹

Individualistic arguments for reparations take the following form. First, regardless of class, members of oppressed groups generally suffer the effects of animus directed toward the group and therefore deserve compensation as current victims.¹⁰² Even if group membership does not perfectly correlate with individual victimization, the fit is close enough to justify a conclusive presumption of some victimization. Some blacks may be slighted and insulted less often than others, or

⁹⁷ For a typical statement of this, as well as other forward-looking views, see Kennedy, *Persuasion and Distrust: A Comment on the Affirmative Action Debate*, 99 HARV. L. REV. 1327, 1329 (1986); Sullivan, *Sins of Discrimination: Last Term's Affirmative Action Cases*, 100 HARV. L. REV. 78, 80 (1986).

⁹⁸ Compare, e.g., R. NOZICK, *supra* note 8, at 152 (stating that reparations generally are designed to undo the impact of past rights violations) with Matsuda, *Looking to the Bottom*, *supra* note 66, at 389-91 (arguing that reparations predominantly bolster the power of a group that is still persecuted).

⁹⁹ See B. BITTKER, *supra* note 96, at 68-70, 135-37; Matsuda, *Looking to the Bottom*, *supra* note 66.

¹⁰⁰ See, e.g., Matsuda, *Looking to the Bottom*, *supra* note 66, at 376-77, 394.

¹⁰¹ The power of individualistic justifications poses another dilemma for radical commentators. Just as the justification for firm-based subsidization of protected group member traits requires an unarticulated tolerance of regressive transfer payments, the person seeking to funnel reparations through the firm may be forced to embrace individualistic premises that she would otherwise shun.

¹⁰² See, e.g., Boxill, *The Morality of Preferential Hiring*, 7 PHIL. & PUB. AFF. 246, 258-61 (1978); Thomson, *supra* note 96, at 381.

treated with less suspicion by state officials, or victimized by dominant group violence less frequently, but skin color is a reliable and easily administrable proxy for gauging subjection to these particular forms of discrimination.¹⁰³

Such current victimization demands appropriate compensation. Money, although frequently thought of as a universal balm, is not an especially apt form of compensation in this case. Since the bulk of the harms involve damage to one's capacity for self-respect and self-support — traditional dignitary offenses — remedies should be tailored to those injuries. Employment is an excellent solution. It arguably remedies some of the dignitary losses for which money compensates rather poorly,¹⁰⁴ and it may, prospectively, improve current victims' social status in ways that diminish the probability of continuing injury, thus obviating the need for ongoing compensation.¹⁰⁵

Even if a member of a protected group is not a victim of ongoing animus, he might still deserve individual compensation because past acts of bigotry directed at other members of the group have tended to limit his current capacity. Put another way, members of an oppressed group can claim an individualistic right to be put in the social position they would now occupy were it not for the historical oppression of their group. In general, the duty to put them in this position will fall on members of the class that has benefited from prior acts of bigotry, although the current members of that class may be less morally responsible than those who actually discriminated.¹⁰⁶ The firm-focused remedy deals with past violations more effectively than would a collectively-based money transfer payment. By giving individual members of protected groups a boost in their attempts to reach particular positions, the firm-focused plan approximates the distribution of jobs that would have obtained in the absence of historical oppression.

Finally, one may believe that an individualistic perspective justifies group-based reparations because oppression has helped create communities whose members do not define their welfare independently of other group members.¹⁰⁷ That is, to the extent that *X* and *Y* are each

¹⁰³ See Boxill, *supra* note 102, at 258–61.

¹⁰⁴ See *id.* at 261–62; Thomson, *supra* note 96, at 380–84.

¹⁰⁵ For an argument that reparations, if legitimately deserved, should be funded by the legislature, rather than tacitly funneled to victim group members by compelling higher wages, see Fischel & Lazear, *Comparable Worth and Discrimination in Labor Markets*, 53 U. CHI. L. REV. 891, 909–10 (1986) (arguing that the victimized group will experience higher unemployment if firms are ordered to pay group members more than would be warranted by their productivity).

¹⁰⁶ Those on whom the burden falls are analogous to the innocent beneficiaries of a third party's wrongdoing, who may well owe restitution despite their innocence. See Boxill, *supra* note 102, at 263–67.

¹⁰⁷ See Thurow, *A Theory of Groups and Economic Redistribution*, 9 PHIL. & PUB. AFF. 25, 32 (1979) (arguing that society often treats group and individual welfare as interdependent goals).

members of a group *G*, defined in part by the common experience of historical oppression, gains to *X* will have an unusually high impact on *Y*. Thus, in allocating goods *now*, we are better off allocating them to people whose success will benefit others. Once more, the firm focus seems especially apt. Although members of oppressed groups may gain utility from observing other members of the group consume more goods, as they would if group members received monetary reparations, the impact of welfare interdependency is likely to be more pronounced in the production sphere. Members of a protected group would probably gain more from seeing other members of the group occupying job positions from which the group has historically been excluded than from seeing group members consuming goods that they have traditionally lacked.

True group-based arguments for reparations, by contrast, focus not on the harms that each individual experiences, but on the direct harms that persecution has inflicted on the group as a cultural entity.¹⁰⁸ The appropriate response to such persecution is a revitalization of the group. Individuals will benefit derivatively, of course, if we assume that healthy group culture enhances individual lives.¹⁰⁹

¹⁰⁸ See Brest, *The Supreme Court, 1975 Term — Foreword: In Defense of the Antidiscrimination Principle*, 90 HARV. L. REV. 1, 51–52 (1976) (explaining but eventually rejecting the notion of group harm).

¹⁰⁹ Take the case of reparations for the internment of Japanese-Americans during World War II. There are several general discussions of the internment and proposals for reparations. See, e.g., UNITED STATES COMM'N ON WARTIME RELOCATION & INTERNMENT OF CIVILIANS, PERSONAL JUSTICE DENIED 117–133 (1982); Matsuda, *Looking to the Bottom*, *supra* note 66, at 363–68; Tang, *The Japanese Internment and Reparations*, 21 LOY. L.A.L. REV. 979 (1988). Civil claims for reparations were first made in *Hohri v. United States*, 586 F. Supp. 769 (D.D.C. 1984) *aff'd in part and rev'd and remanded in part*, 782 F.2d 227 (D.C. Cir. 1986), *vacated and remanded*, 482 U.S. 64 (1987). Congress first directly addressed the issue in 1948 in the American-Japanese Evacuation Claims Act, ch. 814, 62 Stat. 1231 (1948) (codified as amended at 50 U.S.C. §§ 1981–1987 (1988)), and in 1988 it passed a more expansive reparations statute. See *Restitution for World War II Internment of Japanese-Americans and Aleuts Act*, Pub. L. No. 100-383, 102 Stat. 903 (codified as amended at 50 U.S.C. § 1989 (1988)).

An individualist's case for reparations to *all* Japanese-Americans — whether interned or not — might be founded on the supposition that Japanese ancestry is still a good surrogate for predicting current victimization. It would also rest on the claim that Japanese-Americans generally would be in better positions today but for the internment, or on the idea that because Japanese-Americans form a socially meaningful group, and the welfare of group members is strongly interdependent, gains to some Japanese-Americans will have atypically strong positive impacts on others. An argument for reparations payments only to those who were actually interned, on the other hand, could be based on a simple duty to rectify tort-like offenses, or, again, on the recognition that failure to compensate individual members for their losses will have unusually strong negative effects when there is strong welfare interdependency within a group. This last claim seems to have dominated forward-looking individualistic discourse on the legal issue of compensating internees. This argument urges that the failure to compensate those who were directly wronged demoralizes all Japanese-Americans because the harm did not arise from a series of random, individual torts but from an assault that would not have occurred but for membership in a group that remains socially cohesive and meaningful today. Under

If employment tests screen out minority workers who historically have been (and still are) the victims of oppression, an individualist ought to tolerate use of the tests only as long as minority applicants get their proportionate share of jobs. Thus, race-norming of tests should be an adequate solution for individualists seeking reparations, because it enables minority workers to be hired in proportion to their application rates.¹¹⁰ Similarly, it should be an adequate remedy for those who believe that prospects for children in oppressed groups will be enhanced if their parents occupy "good" positions.

From a group rights perspective, however, the tests are themselves a significant source of injury. They play a role in the ongoing devaluation and destruction of the minority culture that will give rise to *future* claims for reparations.¹¹¹ Good test performance remains the legitimated cultural norm. Even if poor test performance has few direct economic consequences for the group as a whole, the use of tests as an *intra-group* sorting device surely penalizes cultural separatism, and the use of tests as a more general hiring norm delegitimizes those culturally distinct "ways of knowing," such as the use of "Black English," that might hamper test performance. Thus, the call to abolish general ability job screening is not part of a program to grant positions to people who would otherwise not deserve them, nor a way to compensate for other assaults on the individual fortunes of protected group members. Rather, it is a call to abolish an unacceptably racist practice, a practice that directly assaults the vitality of group life.

D. The Input Cost vs. Employee Output Distinction: Individual-Output-Centered Discrimination

The last conception of antidiscrimination law also requires that employers discard pure market norms. A profit-maximizing employer comparing two applicants' impact on firm profitability judges each in

this account of the harm, money transfers from the state were indeed an apt remedy for the individual losses.

In the group rights view, by contrast, the main cognizable injury resulting from internment must be the destruction of functioning Japanese-American communities *as communities*. Secondly, internment was a dignitary assault on the group, a libelous accusation of disloyalty that operated not so much to undermine the self-esteem of particular persons as to disparage the entire group. Reparations therefore should go primarily to institutions capable of reproducing a strong cultural identity or to programs that disseminate information about the group's history, such as ethnic studies programs. See Matsuda, *Looking to the Bottom*, *supra* note 66, at 385-88; Tang, *supra*, at 1023-24. Moreover, collective payment of reparations by the descendants of those who libeled the group would symbolize the withdrawal of the libel and a recognition of the urgency of renouncing the libelers' acts.

¹¹⁰ See *supra* note 91.

¹¹¹ An individual might still believe that it is stigmatizing to be told by test givers that he is in some sense inadequate. His is not an individualistic claim for *reparations*, however, but a claim that the test works an ongoing injury.

terms of his "net value added." That is, each employee contributes to output, and each employee requires the employer to make expenditures on items such as wages and benefits and marginal capital costs associated with production. In some cases, no one would deny the legitimacy of the employer's focus on the marginal capital costs associated with employing a particular worker. It would appear ironic for an applicant for a typing pool to claim that he is as productive as another typist because he can type the same amount, with the same accuracy, in the same amount of time, as long as he is given a machine that is twice as expensive.

There are, however, title VII, disability discrimination, and age discrimination cases in which courts have permitted employers to screen out employees only if their output is lower, not if the input costs inevitably associated with their employment are higher.¹¹² These cases suggest that it may not be legitimate for employers to pay workers in accord with net value added; instead, this limited attack on market distributive norms demands that employers ignore differential inputs and treat applicants differently only insofar as their capacity to produce output is different. Thus, one might view these cases as examples of the courts accepting a mild version of what are

¹¹² See, e.g., *Duffy v. Wheeling Pittsburgh Steel Corp.*, 738 F.2d 1393, 1396 (3d Cir.), cert. denied, 469 U.S. 1087 (1984); *Wright v. Olin Corp.*, 697 F.2d 1172 (4th Cir. 1982). The issue may have first come into sharp focus when employers sought to fire pregnant women from jobs in which the employers claimed fetuses would be at risk, at least as long as the employers took no costly precautions to protect them. In *Wright*, the court held that such policies were justified if and only if a business necessity required their adoption. See *Wright*, 697 F.2d at 1186-90. The "necessity" that the employer would have to prove on remand, however, was *not* the need to save money, but the need to protect the fetus. See *id.* at 1188-90. For particularly important discussions of this issue, see Becker, *From Muller v. Oregon to Fetal Vulnerability Policies*, 53 U. CHI. L. REV. 1219 (1986); and Williams, *Firing the Woman to Protect the Fetus: The Reconciliation of Fetal Protection with Employment Opportunity Goals Under Title VII*, 69 GEO. L.J. 641 (1981).

The issue has also arisen in age discrimination cases, cases alleging discrimination against the physically disabled, and cases in which employers refuse to accommodate religiously observant employees. See, e.g., *Duffy*, 738 F.2d at 1396 (holding that a steel company violated the Age Discrimination in Employment Act, 29 U.S.C. §§ 621-634 (1988), when it fired four highly paid salespeople who also received higher pension, insurance, and social security benefits than their fellow workers); see also *Southeastern Community College v. Davis*, 442 U.S. 397 (1979) (holding that a community college need not accommodate its educational program to meet the needs of a hearing-impaired applicant); *Trans World Airlines, Inc. v. Hardison*, 432 U.S. 63 (1977) (holding that an airline is not required to accommodate an employee's desire to observe the sabbath on Saturday). See generally Earle & MacPherson, *Religious Discrimination in Employment*, 1987 DET. C.L. REV. 765 (arguing that title VII requires employers to accommodate workers' desire to celebrate the sabbath); Wegner, *supra* note 10 (noting that the cost of accommodation alone is not a defense to discrimination against the disabled, though the firm might reasonably argue that it ought not be forced to bear unreasonable costs or costs that will threaten its viability); Note, *The Cost Defense Under the Age Discrimination in Employment Act*, 1982 DUKE L.J. 580 (arguing that cost must be weighed against performance in decisions to discharge older workers).

typically considered more politically radical distributive invalidity theories.

It might be helpful in conceptualizing the difference between cases in which the employee's output is lower and those in which the employee's input costs are higher to think of the distinction between costly capital equipment that is universally helpful and costly equipment that aids only the disadvantaged person. A hearing aid, for instance, may enable a hearing-impaired employee to respond more readily to supervisor's commands, but it would not give a non-hearing-impaired worker any advantage over another non-hearing-impaired worker. On the other hand, one could imagine an employer providing aids to a protected group member that would help *any* worker's relative productivity. If, but perhaps only if, a worker requires only equipment that helps her in ways that would not help the bulk of workers, courts may, and perhaps should, prohibit employers from screening her out. Courts may treat the costly equipment as simply providing her an opportunity to demonstrate her capacity, not as specially boosting her capacity relative to other workers.

One might also want to differentiate between cases in which the employer seeks to avoid privately shouldering a social cost that will be borne in any event from cases in which the employer legitimately claims that screening out the employee will eliminate a needless social cost.¹¹³ Assume, for instance, that whether or not they take the job some applicants are predisposed to contract an illness that is especially expensive to treat.¹¹⁴ Assume, further, that an atypically high proportion of such applicants are members of protected groups. Finally, assume that the employer self-insures so that she will bear the costs of such an applicant's medical care. In such a case, the employer's selfish profit interests dictate screening out the applicant, even though doing so will not reduce the social costs of health care. The screening shifts costs; it does not eliminate them. If third-party insurance is more socially costly than self-insurance, forbidding screening may increase social costs by prompting the employer to switch away from

¹¹³ Many commentators do not believe compliance cost concerns are relevant to these "higher input cost" cases. See, e.g., Becker, *supra* note 112, at 1247, 1256; Logan, *Adapting Fetal Vulnerability Programs to Title VII*: Wright v. Olin, 9 EMPLOYEE REL. L.J. 605, 623-24 (1984).

¹¹⁴ See Stone, *AIDS and the Moral Economy of Insurance*, 1 AM. PROSPECT 62 (1990) (noting that when some sectors of a population are disproportionately likely to contract a disease, as is true in the case of AIDS, covering the disease through insurance redistributes resources to sufferers, because some know that they will not contract the disease); Vogel, *Containing Medical and Disability Costs by Cutting Unhealthy Employees: Does Section 510 of ERISA Provide a Remedy?*, 62 NOTRE DAME L. REV. 1024 (1987) (same); H. Greely, *The Death of Health Insurance* (Nov. 1, 1990) (unpublished manuscript on file at the Harvard Law School library) (noting that employers may use a variety of tools, ranging from genetic screening to compiling information on past health care costs, to screen out workers who will make atypically high demands on employer-funded health insurance programs).

self-insurance. But even this hypothesized increase in social costs is fundamentally a function of the employer's profit calculation (her decision not to self-insure unless it is profit-maximizing) and not an unavoidable social cost of hiring the worker.

Unless screening reduces social costs, an employer ought to be barred from screening out high risk applicants, especially if a disproportionate number of "costly" people are members of protected groups. The profit-maximization defense for a screening practice with disparate impact should ultimately be grounded in the assumption that rational actors in a market system achieve *social rationality* — that the firm's profit calculation will mirror a more general societal cost-benefit calculation.¹¹⁵ When private and social rationality diverge, however, social rationality should govern.

Contrast this scenario with a case in which the applicant is atypically vulnerable to the hazards of a particular job. Assume, for instance, that we can sort out legitimate from pretextual claims that women of child-bearing age will be exceptionally expensive to employ in a particular setting, either because a socially undesirable result (birth defects) will occur if they work there or because precautionary expenditures will be needed to avoid that result.¹¹⁶ Let us assume also that employers may not legally avoid bearing liability costs by making female employees waive claims based on potential birth defects¹¹⁷ and that, morally, we do not feel that employee choice is the appropriate way to deal with the issue.¹¹⁸

¹¹⁵ Cf. Engel, *An Approach to Corporate Responsibility*, 32 STAN. L. REV. 1, 34-37 (1979) (arguing that legislatures authorize corporations to maximize profits because profit-maximization presumptively serves social interests).

¹¹⁶ But see Becker, *supra* note 112, at 1232-37, 1269-73 (noting the minimal probability of any individual woman becoming pregnant and questioning the validity of empirical evidence on risks associated with women's exposure to hazards in the workplace); Williams, *supra* note 112, at 655-63 (questioning the scientific validity of claims that women of childbearing age are particularly vulnerable to environmental influences likely to cause birth defects). In addition to these empirical problems, I set aside for now the real difficulties inherent in screening out *all* women of childbearing age because a subset of such women may in fact become pregnant. See Becker, *supra* note 112, at 1232-33; Williams, *supra* note 112, at 696-703. In dealing with "general aptitude" tests in Part III, I return to the general problem of false negatives in the use of screening devices. See *infra* pp. 1224-34.

¹¹⁷ But see Becker, *supra* note 112, at 1244 (noting that under common law tort principles there would be no basis for recovery if title VII bans sex-specific fetal vulnerability policies and the employer fully informed the worker of possible fetal risks).

¹¹⁸ But see Becker, *supra* note 112, at 1241-43 (arguing that women employees should be free to subject themselves to risks and to balance the perils of joblessness or lower paid work against the perils of birth defects). Even ignoring the degree to which women are adequate egoless surrogate decisionmakers either for their unconceived potential children, or already-conceived fetuses, I am not persuaded that the anti-paternalist arguments here are any better than those made by opponents of *all* safety regulations. See, e.g., W. VISCUSI, *RISK BY CHOICE* 1-5 (1983) (arguing that employees should be able to waive knowingly all risks at the workplace). For the pro-paternalist case, see M. KELMAN, *supra* note 74, at 126-41; and Kennedy, *Distri-*

It is considerably less obvious here why the law should deny the employer the defense that hiring the employee will increase costs. Both social and private costs rise when these vulnerable workers are allowed onto a hazardous work site. Some courts and commentators, however, see no reason to allow even this defense.¹¹⁹ Commentators typically reach this result through a formalistic analysis that is ultimately of little help in working through the substantive quandary.¹²⁰ They urge that courts should classify cases in which employers protect against potential birth defects by barring women from certain jobs as disparate treatment rather than disparate impact cases.¹²¹ Unlike the employer in a disparate impact case, the employer in these cases does not use a formally neutral screening device (like an ability test), but rather excludes people on the basis of their membership in the protected group.¹²² Once a court determines that a case involves disparate treatment, the employer must show that his job requirement is a BFOQ,¹²³ rather than merely a "business necessity," as in a disparate impact case. Although the "business necessity" defense clearly anticipates that the employer will raise cost and profitability concerns,¹²⁴ such concerns are irrelevant to the BFOQ claim.¹²⁵ The applicant is qualified unless she simply cannot perform the job or threatens the safety of customers or other employees.¹²⁶

This formal distinction between business necessity and BFOQ, however, ultimately lacks normative force. The substantive dispute, however we classify it doctrinally, centers on the degree to which we should tolerate unequal job access when unequal access is cost efficient. Essentially, the normative argument against the efficiency de-

butive and Paternalist Motives in Contract and Tort Law, with Special Reference to Compulsory Terms and Unequal Bargaining Power, 41 MD. L. REV. 563, 624-49 (1982).

¹¹⁹ See, e.g., *Oil, Chem. & Atomic Workers Int'l Union v. American Cyanamid Co.*, 741 F.2d 444, 447-49 (D.C. Cir. 1984); *Hayes v. Shelby Memorial Hosp.*, 726 F.2d 1543, 1552 (11th Cir. 1984); *Wright v. Olin Corp.*, 697 F.2d 1172, 1191 n.29 (4th Cir. 1982); see also Becker, *supra* note 112, at 1253 n.152 (citing cases).

¹²⁰ See Becker, *supra* note 112, at 1228, 1251-63. The case law echoes to some extent Becker's concerns with formalistic analysis. See, e.g., *Burwell v. Eastern Air Lines, Inc.*, 633 F.2d 361, 369-70 (4th Cir. 1980) (noting that, although judges believe that the choice of theories may be outcome-determinative, courts have found it difficult to decide whether pregnancy-based restrictions require an employer to mount a BFOQ or a business necessity defense).

¹²¹ See, e.g., Becker, *supra* note 112, at 1248-51.

¹²² Obviously, the cases *could* be characterized at the formal level as disparate impact cases if we view the screening device as neutral, that is, if we view the possibility of pregnancy as an instance of vulnerability to job hazards, though one falling solely on women. See, e.g., *Wright*, 697 F.2d at 1185-86; Becker, *supra* note 112, at 1248-51.

¹²³ See *Dothard v. Rawlinson*, 433 U.S. 321, 332-33 (1977) (holding that an employer can justify overtly discriminatory practices when discrimination is reasonably necessary for operation of the business).

¹²⁴ See *supra* note 22.

¹²⁵ See *supra* note 38.

¹²⁶ See I. C. SULLIVAN, M. ZIMMER & R. RICHARDS, *supra* note 39, § 3.6.2, at 140-49.

fense must rest on two notions. The first, to which I return in Part III, is that employers are bound to hire and pay in accord with some notion of "merit," even if doing so neither maximizes profit nor social output. Even when the duty to engage in merit-based hiring is not absolute, but balanced prudentially against costs, it would remain a legally enforceable duty.

The second, related premise of the argument against the efficiency defense is that an employer should measure merit only on the basis of an employee's output capacity, not her-net-value added capacity. The basis for making this distinction is not immediately apparent: output capacities can be as much a function of genetic gifts or undeserved educational opportunities as is the capacity to add less to input costs. We seem comfortable, however, rewarding differential output. Take the example of a woman who has a high output capacity because she is dexterous but a low net-value-added capacity because her employer must make special expenditures to make the workplace safe for workers who may become pregnant. If we are comfortable rewarding a particular woman because she was born dexterous, even though she does not "deserve" the reward, why should we be bothered if we "penalize" her because she can become pregnant, and hence requires costlier protective capital, even though she also does not "deserve" that "handicap"? One explanation is that we remain drawn to the notion that productive capacity *is* deserved, that it is a function of both past efforts to educate oneself and current efforts to apply oneself. The capacity for pregnancy, by contrast, is clearly an immutable, biologically determined fact, even if actual pregnancy is not.¹²⁷

Another explanation for the distinction between output capacity and net-value-added capacity is that the employee's dexterity individuates her, while her capacity for childbearing does not. Merit, in this view, simply reflects those things that make us unique, rather than what we "deserve." To the extent, however, that productive virtues and handicaps are strongly correlated with group membership — if, for instance, superior upper-body strength makes nearly all men more qualified for a job than women — this distinction appears rather spurious. Likewise, to the extent that a "handicap" seems more individually chosen (unlike the capacity to bear children, actual pregnancy may be a matter of choice), the distinction weakens further.

¹²⁷ Women may frequently have very little choice about getting pregnant. How much choice one believes they have depends both on one's assumptions about the degree to which men control (through physical and social force) women's "decisions" to have intercourse and on one's assumptions about the availability of economically and psychically "free" forms of contraception. Moreover, even those women who claim to want to get pregnant, and do so, make this decision in a particular social and cultural context that can be considered to preclude the possibility of truly "free" choice. See generally C. MACKINNON, *FEMINISM UNMODIFIED* 94–96 (1987).

Thus, neither of these "principled" explanations seems particularly satisfactory. What I suspect actually supports intuitions that there is a difference between output capacity and net-value-added capacity is the ill-defined sense that the worker actively produces output, while her added input costs are passively caused by her peculiar status; in the absence of activity there can be no merit. Alternatively, there may be a sense that the job from the employee's perspective is to produce output; if the employee does that as well as the next person, she is as good as the next person at the task that has been set for her. Capital costs are somehow the concerns of the employer. If one sees merit as metaphorically measuring performance in a contest or race, employees with equal outputs seem, perhaps, to have "tied" in the race, as long as the added equipment that only one contestant requires does not bolster her capacity to run faster. Contestants in a race are almost surely considered equal, even if one is wearing more expensive orthopedic shoes, as long as the expensive shoes would not have benefited all other contestants as well.

Whether these explanations accurately capture the distinctions that people draw, or whether the distinctions *ought* to be drawn if they rest upon these ill-defined notions, is surely open to question. Part III of this Article deals with a partly parallel issue by discussing cases in which applicants who make meritocratic claims are ultimately distinguished not on the basis of their work output but on the basis of the relative costs that the employer will bear in identifying them as adequately productive workers.

III. GENERAL ABILITY JOB TESTS: ARE THEY CONSISTENT WITH THE DEMANDS OF MERITOCRACY AND ANTIDISCRIMINATION PRINCIPLES?

A. Are General Ability Job Tests Statically Discriminatory?

The question whether general ability job tests discriminate is of great current concern. The United States Employment Service (USES) is currently evaluating whether to modify, abolish, or retain the General Ability Test Battery (GATB).¹²⁸ In this Part, I argue that the tests do indeed impermissibly discriminate, even if we disclaim reliance on both dynamic and distributive visions of discrimination, the more controversial interpretations of the antidiscrimination principle that I detailed in Part II. I contend that, as an empirical matter, the tests have only modest statistical validity as predictors of productivity. I argue that the use of such low-validity screening devices, rather

¹²⁸ See sources cited *supra* note 7.

than what I will hypothesize to be more expensive ex post screening systems based on retaining only competent workers and discharging those who prove to be inadequately competent, may prove both privately rational for employers and socially cost minimizing. Nonetheless, such devices do not reward people in accord with their productivity and should thus be suspect if we adopt the individual output view of discrimination.

Although the GATB had been historically administered to a rather low proportion of those seeking jobs through the USES,¹²⁹ in 1986 the agency established pilot programs that were designed, when fully implemented, to administer the GATB to all of the roughly nineteen million job seekers it sees each year.¹³⁰ The National Academy of Sciences was asked to evaluate the extended use of the GATB, as well as the practice of race-norming the test, and its report on the GATB may well have led to the test's suspension.

The GATB is designed to measure a number of distinct skills, including numerical, verbal, and spatial relationship skills. The test also purportedly measures "general aptitude." The general aptitude (g) score allegedly measures not specific skills, but the capacity to learn, process new information, and solve problems.¹³¹ Employers who contacted a particular local USES branch were never required to inquire about applicants' GATB scores, nor were they required to make use of the GATB scores they discovered in the process of screening applicants.¹³² Prior to the suspension of the test, employers received race-normed GATB scores;¹³³ thus the GATB did not have an adverse impact on blacks as a group.

USES began to promote tests more aggressively in the late 1980s; most local offices, each of which operates with a certain degree of autonomy, have adopted a theory of universal validity that holds that g-loaded tests predict on-the-job performance on any job.¹³⁴ Those

¹²⁹ See INTERIM REPORT, *supra* note 91, at 9. In 1985, less than eight percent of USES registrants were given the GATB. However, nearly double the "traditional" number were given the GATB in a one-year period from 1986 to 1987 as part of a transitional experimental program aimed at increasing test use. See *id.*

¹³⁰ See FAIRNESS, *supra* note 1, at vii, 1-2.

¹³¹ For a fuller description of the GATB itself, see *id.* at 73-116.

¹³² See Wigdor & Hartigan, *The Case for Fairness*, 27 SOCIETY 12, 15 (1990) (indicating that many employers may make little use of the GATB scores they receive).

¹³³ For a description of race-norming, see *supra* note 91. In 1986, the Justice Department prepared to challenge race-norming as illicitly discriminatory against whites, but decided to delay the challenge. See FAIRNESS, *supra* note 1, at 21-22.

¹³⁴ See Levin, *Ability Testing for Job Selection: Are the Economic Claims Justified?*, in TEST POLICY AND THE POLITICS OF OPPORTUNITY ALLOCATION 211, 211-12 (B. Gifford ed. 1989) (noting that in 1987 public employment services in 37 states already used ability tests without local validity studies and that 90 offices used these tests as an integral part of their operational procedures for assessing and referring job applicants). There is, however, tremendous variation in the practices of the 1800 local employment service offices; it is difficult to assess the way in

who adopt this theory believe that even if no work has been done to demonstrate that the test predicts performance in a particular setting, the test is still worth using as a screening device.¹³⁵ Whether this decrease in skepticism about tests among test administrators has affected the actual practice of employers, however, is simply not known. It is generally difficult to estimate, even in a very rough fashion, how widespread the use of either the GATB or a variety of privately administered general ability tests has been. The most recent reasonably dependable survey, done in 1987 and 1988 by the Bureau of National Affairs Personnel Policies Forum,¹³⁶ indicates that although the use of "cognitive" screening tests by private employers is not standard practice, it is by no means rare. The BNA survey reveals that thirty-one percent of the respondents used mental ability tests in hiring.¹³⁷ Surveys done in the 1970s found similar levels of test usage.¹³⁸ These surveys do not seem especially dependable, however. Some are plagued by low return rates, while others are marred by non-randomness of the sample pool. Moreover, none is easy to interpret because one cannot tell how determinative the tests are for the employers who use them.¹³⁹

Given this shortage of information, test opponents should neither panic over the level of test use nor be sanguine that tests pose only a theoretical problem. I proceed on the fair assumption that tests have surely been used often enough to be worth discussing and that the resolution of the current controversy over whether to reinstate, reform, or abolish the GATB is significant.

If one assumes, at least for argument's sake, that a screening device that produces a disparate impact on a protected group should not be used unless it saves the employer "significant" sums of money, then

which any particular office uses the test or to determine whether usage has generally increased. See FAIRNESS, *supra* note 1, at 66-69.

¹³⁵ See FAIRNESS, *supra* note 1, at 17-20. The claim that cognitive tests are universally valid should be distinguished from the claim that validity may be generalized, to some uncertain degree, from one setting to another, as long as a competent job analysis demonstrates that skills required in one setting are quite close to those required in another. The claim of universal validity asserts that validities can be generalized to any job of the same broad type *even in the absence of a particularized analysis* of the tasks required of the workers to be selected in the specific jobs in question. See *infra* pp. 1214-17.

¹³⁶ BUREAU OF NAT'L AFFAIRS, PERSONNEL POLICIES FORUM, SURVEY NO. 146, RECRUITING AND SELECTION PROCEDURES (1988).

¹³⁷ See *id.* at 17.

¹³⁸ The Prentice-Hall/American Society for Personnel Administrators survey, performed in 1975 and 1976, revealed that 49.1% of all respondents gave some form of hiring tests. See Friedman & Williams, *Current Use of Tests for Employment*, in 2 ABILITY TESTING: USES, CONSEQUENCES, AND CONTROVERSIES 99, 106 (1982). Considerably less than 10% of private employers, however gave *only* general ability job tests, though data does not reveal how many of the employers who gave two or more tests included a general ability test in their battery. See *id.* at 107.

¹³⁹ See *id.* at 103-04, 115.

the use of such general ability tests is certainly illegitimate unless it is statically valid. An employer obviously cannot save money by using a test if the test does not predict job performance; thus, tests cannot be justified unless they predict performance. But even if the predictor (test score) correlates with the criterion (performance), it remains a complex question whether use of the predictor will increase profitability. The answer to that question turns first on the relative productivity of different workers: a test that predicted how employees would perform relative to one another would be of little practical utility if the best and worst workers produced very similar levels of output. The answer turns next on the selection ratio — the proportion of applicants needed to fill available jobs. If an employer must hire 100% of the applicants, even a perfectly predictive test would be of no utility.

The claim that *g*-loaded tests like the GATB have been shown to be highly valid predictors of on-the-job performance in all settings has been made most forcefully by industrial psychologists John Hunter and Frank Schmidt. They claim that these tests are highly valid, that they are equally valid within job-types (for example, skilled versus manual labor) regardless of the particular job setting in which they are employed,¹⁴⁰ and that they predict job performance as accurately for minority as for nonminority applicants.¹⁴¹ In the following subsections, I argue that each of these claims is dubious.

¹⁴⁰ Until the mid-1970s, the mainstream position among personnel psychologists was that tests predict performance differently in different settings because each setting calls on people to use significantly different skills. Validities in reported studies do in fact vary widely. Because industrial psychologists were not sensitive to the possibility that the reported validities could be different for artifactual reasons (for instance, sample size, transcription error, etc.), the only plausible explanation they could offer was that true validities indeed differed from one setting to another. The most significant mainstream texts reflected this position. See, e.g., E. GHISELLI, *THE VALIDITY OF OCCUPATIONAL APTITUDE TESTS* 125–27 (1966); Ghiselli, *The Validity of Aptitude Tests in Personnel Selection*, 26 *PERSONNEL PSYCHOLOGY* 461, 475–77 (1973). The EEOC and the older American Psychological Association principles also endorsed this perspective. See *Uniform Guidelines on Employee Selection Procedures*, 29 C.F.R. §§ 1607.7–8 (1990); AM. PSYCHOLOGICAL ASS'N, *STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTS* 31–33, 36 (rev. ed. 1974) (demanding local validation of studies given assumption of situation specificity). The 1985 APA principles are ambiguous as to whether they accept or reject either the “universal validity” hypothesis or the “situational specificity” hypothesis. See AM. EDUC. RESEARCH ASS'N, AM. PSYCHOLOGICAL ASS'N, NAT'L COUNCIL ON MEASUREMENT IN EDUC., *STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTING* 12, 59–62 (1985).

In contrast, proponents of the universal validity hypothesis have argued that psychologists observed different validity coefficients largely as a result of artifactual elements, particularly sampling error. See J. HUNTER, *VALIDITY GENERALIZATION FOR 12,000 JOBS: AN APPLICATION OF SYNTHETIC VALIDITY AND VALIDITY GENERALIZATION TO THE GENERAL APTITUDE TEST BATTERY (GATB)* 31–34 (1983); Schmidt, Hunter & Pearlman, *Task Differences as Moderators of Aptitude Test Validity in Selection: A Red Herring*, 66 *J. APPLIED PSYCHOLOGY* 166, 166–67 (1981).

¹⁴¹ For a good summary of these views, see Schmidt, Hunter, Pearlman & Hirsch, *Forty Questions About Validity Generalization and Meta-Analysis*, 38 *PERSONNEL PSYCHOLOGY* 697,

1. *The Validity of General Ability Tests.* — The claim that tests are highly valid predictors of productivity is problematic for three distinct reasons. First, when test validators correlate a predictor (the screening test) with a criterion (performance on the job), the information they produce is useful only if the criterion they measure is a reasonable surrogate for actual productivity.¹⁴² Industrial psychologists, however, do not correlate test scores with pay, the best presumptive measure of worker productivity in a market system. Economists have generally found quite low correlations between the “ability” that the tests purportedly measure and earnings.¹⁴³ Unless

712-13, 723-24, 756-60, 768-70, 772-74 (1985) [hereinafter *Forty Questions*]. Hunter and Schmidt make use of a statistical technique called meta-analysis, which is a tool for summarizing large numbers of research studies. The meta-analyst essentially assumes that one can average the correlations found in a number of distinct studies to locate the true universal correlation, at least as long as the deviations from the average observed in particular studies are explicable on the basis of chance factors (particularly sampling error). For general descriptions of meta-analysis, see J. HUNTER, F. SCHMIDT & G. JACKSON, *META-ANALYSIS: CUMULATING RESEARCH FINDINGS ACROSS STUDIES* (1982); and G. GLASS, B. MCGAW & M. SMITH, *META-ANALYSIS IN SOCIAL RESEARCH* (1981). In their meta-analyses of local validation studies, Hunter and Schmidt conclude that the mean validities found in the local validation studies are positive (that is, test or predictor scores correlate with productivity or criterion “scores”). See *Forty Questions, supra*, at 723, 762, 775-76. Second, they conclude that the observed variance among the local studies is a product of artifactual elements (predominantly sampling error). See *id.* at 705, 757-58, 765-66. There are, for instance, no more local studies finding zero or negative correlations between predictor and criterion than one would expect to occur simply as a result of chance. Third, they conclude that tests are equally valid for blacks and whites and do not underpredict the performance of black test takers as a group. See Hunter, Hunter & Schmidt, *Differential Validity of Employment Tests by Race: A Comprehensive Review and Analysis*, 86 *PSYCHOLOGICAL BULL.* 721, 732-34 (1979).

¹⁴² There is little reason to believe that any local validation studies measure “productivity” in a meaningful sense. See Levin, *supra* note 134, at 215-16. For a general discussion of the difficulties in measuring productivity, see Jorgensen & Griliches, *The Explanation of Productivity Change*, 34 *REV. OF ECON. STUD.* 249 (1967); and Searle & Waite, *Current Efforts To Measure Productivity in the Public Sector*, in *NEW DEVELOPMENTS IN PRODUCTIVITY MEASUREMENT AND ANALYSIS* 333, 340-49 (J. Kendrick & B. Vaccara eds. 1980). In a sense, the basic practical and theoretical problem that test validators face is the measurement of either the absolute or relative productivity of particular workers. If we are not satisfied that we can reasonably measure the contribution an employee makes to her firm, we cannot simply observe that tests do correlate with other measurable factors and, for convenience, dub an aggregation of those factors “productivity.”

¹⁴³ Test score differentials for similarly educated workers do not seem to have much impact on earnings. See, e.g., Willis, *Wage Determinants*, in 1 *HANDBOOK OF LABOR ECONOMICS* 525, 572-90 (1986). Willis notes:

Given the complexity of the issues and the non-representative character of the data sets that have been employed in the literature on ability bias, it is difficult to reach any firm conclusions about the magnitude or even the direction of the bias in U.S. data My impression is that the simple Mincer-type earnings function [which assumes people with equal educations will earn equal amounts, regardless of differences in ability] does a surprisingly good job of estimating the returns to education

Id. at 589-90; see also Griliches & Mason, *Education, Income and Ability*, 80 *J. POL. ECON.* S74, S88, S99 (1972) (noting that, with other factors held constant, increased scores on the

employers cannot observe productivity¹⁴⁴ or observe productivity but do not pay in accord with it,¹⁴⁵ these findings cast grave doubt on

GATB-like Armed Forces Qualifying Test (AFQT) have an extremely limited impact on earnings: while an additional year of schooling adds about 4.6% to income, a 10% improvement in AFQT scores typically adds only 1%); Hause, *Earnings Profile: Ability and Schooling*, 80 J. POL. ECON. S108, S130-31 (1972) (finding that, within a sample schooling class, one standard deviation in measured ability is associated with earnings differentials ranging from 10 to 13% for those who are most educated, but that there are "[n]egligible differences . . . for those with less than a high school education"); Willis & Rosen, *Education and Self-Selection*, 87 J. POL. ECON. S7, S26 (1979) (finding that an increase of one standard deviation in a high school graduate's math or verbal achievement score leads to a slight wage increase 25 years out of school and leads to a slight decrease in the graduate's wages at his first job). *But see* Taubman, *Comment*, 80 J. POL. ECON. S104, S106 (1972) (noting that, for a sample of 4000 high school graduates, math ability, though not IQ generally, has as much impact on earnings as does education).

I do not mean to imply that this evidence has utterly foreclosed the possibility that people tend to earn more when they are more capable in the sense the tests measure. *See, e.g.*, Griliches, *Estimating the Returns to Schooling: Some Econometric Problems*, 45 ECONOMETRICA 1, 6-8, 11 (1977) (reasoning that if people continue to educate themselves only if they are "capable," we may misidentify returns that actually reflect ability as returns to education); Lillard, *Inequality: Earnings vs. Human Wealth*, 67 AM. ECON. REV. 42, 45-46 (1979) (noting that returns to ability are often underestimated since they are more pronounced in middle age than early life, and most subjects in studies that follow people through their lives to measure lifetime economic performance are young men under the age of 35).

Nonetheless, it is safe to say that large (one standard deviation) differences in "ability" do not correspond to differences in earnings that even approximate a standard deviation.

¹⁴⁴ *See, e.g.*, L. THURLOW, DANGEROUS CURRENTS 201-02 (1983) (arguing that employers cannot ascertain the productivity of each employee); J. Bishop, *Employment Testing and Incentives to Learn* (1988) (Cornell Univ. Working Paper No. 88-12) (claiming that students have inadequate incentives to develop skills in school, rather than to acquire degrees simply as credentials, because productivity is unobservable). Hunter and Schmidt, however, cannot rely on this reasonably plausible explanation because they (wrongly) believe that productivity is both observable and observed (for example, by supervisors or by firms that give work sample tests).

¹⁴⁵ *See, e.g.*, G. BECKER, HUMAN CAPITAL 10 (1961) (arguing that pay may exceed current marginal product for senior workers and be lower for junior workers to give proper incentives for workers to invest in plant-specific training); Akerlof, *Labor Contracts as Partial Gift Exchanges*, 97 Q.J. ECON. 543, 555 (1982) (explaining that firms may pay more than the market-clearing wage in order to affect norms determining a fair day's work); Bowles, *The Production Process in a Competitive Economy*, 75 AM. ECON. REV. 16, 25-26, 32 (1985) (reasoning that pay may exceed marginal product in order to increase the implicit penalty for termination when workers "shirk"); Parsons, *The Employment Relationship: Job Attachment, Work Effort and the Nature of Contracts*, in 2 HANDBOOK OF LABOR ECONOMICS 784, 808-10 (O. Ashenfelter & R. Lavard eds. 1986) (making same argument as in G. BECKER, *supra*); Rosen, *Implicit Contracts: A Survey*, 23 J. ECON. LITERATURE 1144, 1145, 1147-48 (1985) (noting that pay may exceed marginal product in some periods and be less in others because workers implicitly purchase "insurance" contracts against wage fluctuation); Yellen, *Efficiency Wage Models of Unemployment*, 74 AM. ECON. REV. 200, 204 (1984) (following Akerlof in noting that pay may be greater than marginal product as a "gift" in return for extra effort from workers). While the question is complex, I believe it is doubtful whether any of these theories in fact suggests that lifetime earnings differentials among workers will not correspond to productivity differences, although they suggest that at any point in time, a particular worker may be paid more or less than the value of her marginal product.

Pay would not correspond to productivity, even over a lifetime, if a worker is subject to

testing advocates' claims that ability test scores predict output. Nevertheless, test validators typically correlate the predictor with performance on job sample tests or with supervisor ratings.¹⁴⁶

Using job sample tests as an indicator of productivity is conceptually problematic, because the capacity to perform well on the sorts of highly individualized tasks that job sample tests can measure may correlate only weakly with the capacity to contribute to the product of the group in which production occurs.¹⁴⁷ The existing job sample tests on which Hunter and Schmidt rely in extolling ability tests do not even do a particularly good job of measuring individualized skills.¹⁴⁸ Positive correlations between ability tests and job sample tests may simply reflect shared methods variance (people proficient at

discrimination. See, e.g., R. CHERRY, *supra* note 28, at 75-76. But those who have studied the connection between ability and pay, see sources cited *supra* note 143, do not seem to find that the ability-pay nexus is stronger for majority-group male workers.

Whether pay ultimately corresponds to productivity or not, the failure of Hunter and Schmidt to account for what appear to be the rather trivial correlations between test scores and earnings is a symptom of economic naiveté: nowhere do they even acknowledge that these low correlations would be the central problem to explain if test scores actually predicted productivity.

¹⁴⁶ See Levin, *supra* note 134, at 213-15; Pearlman, Schmidt, & Hunter, *Validity Generalization Results for Tests Used to Predict Job Proficiency and Training Success in Clerical Occupations*, 65 J. APPLIED PSYCHOLOGY 373, 377-79 (1980) (noting that criterion measures include supervisor ratings and rankings, work samples, and measures of output quantity and quality).

¹⁴⁷ The local validation studies on which advocates of universal test validity rely make no real effort to solve the problem of measuring the value added or marginal shifts in plant product for which a particular employee is responsible. This is not surprising because workers contribute to team or plant production in a wide variety of ways, few of which are readily observed. See Alchian & Demsetz, *Production, Information Costs, and Economic Organization*, 62 AM. ECON. REV. 777, 779 (1972); see also L. THURLOW, *supra* note 144, at 205-14 (discussing the implications of both the unobservability of the individual product and the team nature of production). A particular worker may or may not educate other workers about plant-specific features; she may or may not raise co-worker morale; she may or may not perform individualized tasks quickly and in a way that minimizes unacceptable production outcomes. If she interacts with outsiders (customers, clients, or the general public), she may or may not make a positive impression and increase firm goodwill.

¹⁴⁸ Hunter and Schmidt rely on only three major job sample tests in reaching their conclusion that testing could dramatically improve the productivity of the federal government work force. The three job samples involved IRS workers, customs inspectors, and social insurance claims examiners. See Schmidt, Hunter, Outerbridge & Trattner, *The Economic Impact of Job Selection Methods on Size, Productivity, and Payroll Costs of the Federal Work Force*, 39 PERSONNEL PSYCHOLOGY 1, 5-8 (1986). Levin has reviewed these studies and concluded that none is a genuine work sample. Rather, each retests to a significant degree the same type of purely cognitive skills that the predictor was designed to measure. Thus, observed correlations may be attributed to a considerable extent to the fact that the criterion is not adequately conceptually distinct from the predictor. See Levin, *supra* note 134, at 216-27. For example, the job sample test involving IRS workers measured only the capacity to make decisions about legally appropriate responses to taxpayer delinquent accounts. It made no effort to measure other job-related skills, such as locating and contacting taxpayers, negotiating settlements, and gathering information about financial assets. The supposed job sample of customs inspectors was just a job information test in which workers watched videotapes of an inspector at work and noted any mistakes they spotted. See *id.* at 218, 220-21.

one type of test will do well on other, similar tests) rather than a correlation between predictor and output.

Supervisor ratings also present difficulties as reliable measures of output. Although they ostensibly measure all of a worker's contributions to production, rather than just her capacity to perform isolated tasks, supervisor ratings may be biased against certain groups who have performed poorly on tests, such as blacks.¹⁴⁹ In addition, supervisors may exaggerate the productivity-enhancing impact of the sort of cognitive knowledge that good test takers will have in relative abundance.¹⁵⁰ Thus, once again, we may well see deceptively positive correlations:¹⁵¹ those who do poorly on the test will receive low ratings

¹⁴⁹ See Kraiger & Ford, *A Meta-Analysis of Rater Race Effects in Performance Ratings*, 70 J. APPLIED PSYCHOLOGY 56, 59-60 (1985). Kraiger and Ford found that both black and white raters give systematically higher ratings to those of their own race, regardless of the type of rating, rating purpose, or rater training. See *id.* at 60. White raters rated the average white ratee higher than 64% of black ratees; black raters rated the average black ratee higher than 67% of white ratees. See *id.* at 59-60. Because supervisors are usually white, blacks will more often be systematically underrated.

Kraiger and Ford's study is hardly perfect, however. Standing on its own, it would leave open the possibility that white raters rate black workers negatively because they are in fact poorer workers, while black raters, feeling strong political and racial solidarity with other blacks in a racially hostile environment, overrate black supervisees. In another study, for instance, Ford found no important differences between black performance on subjective criteria (ratings) and on objective criteria (supposedly insensitive to the judgments of raters) in a set of 53 tests using both types of criteria. See Ford, Kraiger & Schechtman, *Study of Race Effects in Objective Indices and Subjective Evaluations of Performance: A Meta-Analysis of Performance Criteria*, 99 PSYCHOLOGICAL BULL. 330, 333 (1986). No matter what kind of performance was measured or what kind of measurements were used, blacks did between .112 and .336 standard deviations worse than whites. See *id.* at 333 (table 1). The performance of blacks is relatively better, however, on direct job performance measures than on cognitive performance or job knowledge tests. See *id.* at 334. This fact may well reflect the biased understanding of job knowledge used in the tests. However, this second study may be misleading by *underestimating* black performance. The supposedly "objective" measures, see *id.* at 332, may be distorted by racial bias. Surely, the level of customer complaints may well be related to the race of the worker. Moreover, what an evaluator identifies as an employee "accident" may depend on the race of the participant. It is also possible that race affects the supervisor's recording of absenteeism, tardiness, or "units produced" by workers.

¹⁵⁰ See Hunter, *A Causal Analysis of Cognitive Ability, Job Knowledge, Job Performance, and Supervisor Ratings*, in PERFORMANCE MEASUREMENT AND THEORY 257, 265 (F. Lane, S. Zedeck & J. Cleveland eds. 1983) (reporting that scores on "job-knowledge" tests correlate more highly with supervisor ratings than does work-sample performance). This overevaluation of cognitive skills may reflect unconscious racism, in that a supervisor may simply assume that the particular skills that whites more typically display are the most important.

¹⁵¹ Although I believe that, on balance, we *overestimate* correlations by finding criteria that correlate well with the test without measuring true productivity, I would find plausible the claim that we *underestimate* the true correlation between the sort of cognitive ability the GATB is designed to measure and performance, because we do not measure "cognitive ability" accurately due to *predictor* unreliability. If one believed that there is some abstract cognitive talent that GATB-like tests measure, one would recognize that GATB scores would be somewhat random measures of it, because the test is poorly administered. Because the test is both highly "speeded" (difficult to finish) and does not penalize wrong answers, a person who is either coached to guess blindly on questions that she has not had a chance to think through or who simply

and those who do well will receive high ratings for reasons that have little to do with actual productivity. It should also be noted that when the same worker's productivity is simultaneously measured by these two commonplace measures — supervisor ratings and job samples — the correlation between the two measures is quite low, a fact that casts further doubt on either measure's ability to test productivity accurately.¹⁵²

Second, even if the chosen criteria are in fact accurate measures of productivity, the critical question is whether the use of cognitive tests can be justified on the basis that such tests have a significantly higher validity than alternative screening devices such as education, job experience, peer evaluations, and interviews.¹⁵³ In fact, observed mean validities (correlations between test and criterion) for general ability job tests, while positive, are not demonstrably higher than those for many of these other screening devices. Although Hunter and Schmidt state that mean validities are typically higher than the observed validities for alternative screening devices,¹⁵⁴ other analysts have found mean validities for cognitive tests that are either barely distinguishable from, or actually lower than, validities for alternative screening devices.¹⁵⁵ The differences among the authors' estimates

happens to do so will get a better score than someone unaware of the test's idiosyncrasy. See FAIRNESS, *supra* note 1, at 110–15. Such predictor unreliability obviously does not affect the validity of the predictor, but if the deeper question is whether cognitive “ability” correlates with performance, such unreliability mutes the actual relationship.

¹⁵² See Heneman, *The Relationship Between Supervisory Ratings and Results-Oriented Measures of Performance: A Meta-Analysis*, 39 PERSONNEL PSYCHOLOGY 811, 816 (1986) (finding that in 23 validation studies for which both direct output measures and supervisory ratings were available, the mean correlation between the two measures was only .27); Hunter, *Cognitive Ability, Cognitive Aptitudes, Job Knowledge, and Job Performance*, 29 J. VOCATIONAL BEHAV. 340, 351–52 (1986) (finding mean correlations of only .19 in military studies and only .27 in seven civilian studies).

¹⁵³ The question is particularly important if such tests have greater adverse impact on racial minorities or women. Studies about the relative adverse impact of alternative employment screening devices seem quite inconclusive. See, e.g., Reilly & Chao, *Validity and Fairness of Some Alternative Employee Selection Procedures*, 35 PERSONNEL PSYCHOLOGY 1, 55 (1982). But see Schmidt, Greenthal, Hunter, Berner & Seaton, *Job Sample vs. Paper-and-Pencil Trades and Technical Tests: Adverse Impact and Examinee Attitudes*, 30 PERSONNEL PSYCHOLOGY 187, 193–94 (1977) (noting that job sample tests for metal trade workers have less adverse impact on minorities than content-valid written achievement tests).

¹⁵⁴ In Hunter and Schmidt's validity generalization studies, corrected validities for cognitive tests range from about .56 (for high complexity jobs) to about .23 (for unskilled jobs). See Schmidt, *The Problem of Group Differences in Ability Test Scores in Employment Selection*, 33 J. VOCATIONAL BEHAV. 272, 278–79 (1988); see also Hunter, *Cognitive Ability, Cognitive Aptitudes, Job Knowledge, and Job Performance*, 29 J. VOCATIONAL BEHAV. 340, 344–45 (1986) (noting similar estimates of validity for GATB clone). At the same time, according to Hunter and Schmidt the validity of general psychomotor ability tests rises from .3 to .48 as one descends the typical occupational ladder. See Hunter & Schmidt, *Ability Tests: Economic Benefits Versus the Issue of Fairness*, 21 INDUS. REL. L.J. 293, 296 (1982).

¹⁵⁵ See, e.g., Reilly & Chao, *supra* note 153, at 6, 15, 24 (concluding that biodata — resumé credentials such as educational achievement and employment history — have a mean validity

result primarily from the degree to which each "corrects" observed mean validities for unreliability (or "attenuation")¹⁵⁶ and range restriction.¹⁵⁷ Hunter and Schmidt's estimates of *uncorrected* cognitive validities are in precisely the same range as those of other analysts; but their corrections for range restriction and unreliability increase validities and are psychometrically inappropriate.¹⁵⁸

of between .32 and .46, peer ratings have a mean validity of .31, and interviews have a mean validity of .19 in predicting job proficiency); Schmitt, Gooding, Noe & Kirsch, *Meta-Analyses of Validity Studies Published Between 1964 and 1982 and the Investigation of Study Characteristics*, 37 PERSONNEL PSYCHOLOGY 407, 415 (table 4) (1984) (finding mean validities of .248 for cognitive tests, .243 for biodata, and .427 for supervisor and peer evaluations)

¹⁵⁶ One corrects for the effect of criterion unreliability by dividing the measured mean validity coefficient by the square root of the mean reliability coefficient. Corrections for unreliability (or attenuation) are psychometrically appropriate under some circumstances. See FAIRNESS, *supra* note 1, at 165-66. If one assumes productivity measurements are randomly unreliable, true validities will surely be understated. To illustrate, imagine an attempt to measure the Law School Admissions Test's (LSAT) predictive validity, using first year grades as the criterion. If student transcripts were frequently illegible, the measured correlation between LSAT scores and grades would be suppressed, because there would be no reason for the predictor to correlate at all with a set of misread grades.

¹⁵⁷ To correct for range restriction, one compares the standard deviation on the predictive test in the concurrent study sample to the standard deviation on the test for the applicant pool. The validity coefficient for the whole applicant population is then assumed to be the measured validity coefficient in the concurrent sample multiplied by the inverse of the ratio of study sample standard deviation to population standard deviation. See FAIRNESS, *supra* note 1, at 124-27; see also Linn, Harnisch & Dunbar, *Corrections for Range Restriction: An Empirical Investigation of Conditions Resulting in Conservative Corrections*, 66 J. APPLIED PSYCHOLOGY 655 (1981) (examining the general effectiveness of range restriction corrections). The purpose of range restriction corrections is illustrated by a simple analogy. If a law school admits only those applicants scoring in a very narrow, high range on the LSAT, we might not be surprised to find that the small differences in LSAT scores among that quite similar group did a poor job predicting differences in grades. It would still be true, however, that if the school admitted some students with much lower LSAT scores, those students would get lower grades. Thus, to get a more accurate measure of validity here one would "correct" measured validity.

¹⁵⁸ In a sense, there is no abstract or context-independent answer to the question of how one should correct for range restriction. If the only alternative to job screening through tests were lottery hiring from the population as a whole, then the linear range restriction corrections I described would likely *understate* the correlation between predictor and criterion for highly complex jobs, because some of those who are neither tested nor hired would be so incapable of doing a particular job that their test scores would *overpredict* performance. But if applicants gravitate to jobs for which they are more suited, even in the absence of test-based screening, the ratio between the standard deviation of applicants and accepted workers would be very close to one. Hunter, however, assumes that the ratio is .8, a figure that implies that workers do rather little self-selection. See J. HUNTER, *supra* note 140, at 59; see also FAIRNESS, *supra* note 1, at 128 (discussing possible corrections for range restriction). To continue the law school analogy, it would be inappropriate for institutions to correct measured validities for the LSAT as a grade-predictive test if, first, few persons with low test scores applied to schools that accepted only high-test-score applicants or, second, few applicants were admitted irrespective of test scores. A strong case can be made that it would be prudent not to correct GATB validities for range restriction at all. This position was the consensus of the National Academy of Sciences Advisory Committee. See *id.* at 128.

Hunter and Schmidt's corrections for unreliability are also problematic because they simply assume, without justification, that all errors in criterion measurement are random. To the extent

Third, and finally, because the vast bulk of validation studies are produced either by defendants in job discrimination suits or academics seeking to publish studies, one would expect the studies available for meta-analysis to show higher validities than would the universe of actual studies. Both defendants in suits and academics have a stake in finding significant positive correlations, either to avoid liability or to produce a study that will be considered worthy of publication.¹⁵⁹

2. *The Universal Validity of General Ability Tests.* — Hunter and Schmidt's claim that tests are not just highly valid, but *universally* valid without regard to distinctions in work setting, has important procedural implications for title VII law. If Hunter and Schmidt's claim is correct, it should not be necessary for any particular employer to validate the use of a test for her particular place of employment.

that criterion measurement and test measurement are biased in the same way, however (for example, each is racially biased), the unreliability correction is not only inappropriate, but results in an error of sign. Imagine using a radar gun to measure the correlation between the speed of a car at point one and point two. If the radar gun randomly mismeasured speeds on occasion, the true correlation would be understated. But if it happened to register 80 miles per hour every time a black car passed, whatever the speed of the car, correlations between speed at the two points would be *overstated*. In addition, Hunter and Schmidt's corrections for random attenuation assume lower levels of reliability than the studies they meta-analyze assert. See *id.* at 165-66. Hunter estimates that reliability is only .6, based on what appears to be his subjective reaction to his survey of the criterion measurement literature. See J. HUNTER, *supra* note 140, at 41. The effect of this peculiarly low estimate of reliability on what he takes to be true validity is quite stunning. For example, using a .6 reliability estimate increases observed validities by 29%; using a .8 reliability estimate would lead one to "correct" observed validities by only 12%. See FAIRNESS, *supra* note 1, at 166.

Thus, if one makes no correction for range restriction and assumes reliabilities of .8, one corrects measured validities by only 12%, whereas Hunter's more extreme suppositions, assuming .6 reliability and a .8 ratio between study sample and population standard deviations on the test, will yield a 61% increase in observed correlations. See *id.* at 167. Moreover, if one meta-analyzes only the most recent 264 validation studies of GATB that are available rather than an additional 500 older studies arguably less carefully performed, and if one uses the more conservative corrections, one sees that GATB validities are *extremely* modest. See *id.* at 160-63, 167-68. Overall, GVN (general intelligence, verbal, and numerical aptitude) validities are only .22, not .47 as Hunter asserted in his enthusiastic endorsement of the GATB. See *id.* at 168. Thus, only four percent of the variance in job performance is explained by variance in test score performance.

¹⁵⁹ Hunter and Schmidt argue that this "file drawer" problem is not serious. First, they claim that unavailable studies would have to exhibit an implausibly large string of negative findings to lower validities significantly. See *Forty Questions*, *supra* note 141, at 722; Schmidt & Hunter, *Development of a General Solution to the Problem of Validity Generalization*, 62 J. APPLIED PSYCHOLOGY 529, 535-37 (1977). Hunter and Schmidt seem to believe that in evaluating the possibility that the validities are lower, we should assume that unavailable studies would be just like available ones, and that what we should try to ascertain is how unlikely it would be if unavailable studies uniformly happened to have far lower validities by chance. But those who worry about the file drawer problem start with the more reasonable supposition that *negative correlations typically go unreported* (because of publication and reporting bias), or even undiscovered (because researchers may abandon experiments when preliminary findings indicate that they will find no positive results). See G. GLASS, B. MCGAW & M. SMITH, *META-ANALYSIS IN SOCIAL RESEARCH* 63-68 (1981).

Instead, if courts accept the levels of universal validity posited by Hunter and Schmidt as justification for the use of tests with racially disparate impact, aptitude tests will become automatically permissible screening devices.¹⁶⁰

Although claims of universal validity are not as insupportable as the claim that validities are *high*, courts should still reject testing proponents' claims that tests need not be validated in each particular work setting. The claim that validities vary only slightly within a broad job type inevitably rests on the implausible assumption that the relative productivity levels of different workers is essentially insensitive to the mode of job organization.¹⁶¹ In fact, organizational changes can vastly increase worker productivity.¹⁶² Moreover, it seems unlikely as a practical matter that there might be better production modes that would increase absolute productivity levels without altering relative worker productivity,¹⁶³ thereby reducing test validity.¹⁶⁴

¹⁶⁰ A small but not insignificant number of courts already accept the validity generalization hypothesis; defendants in such courts need not perform a local validation study. See, e.g., *Friend v. Leidinger*, 588 F.2d 61, 65 (4th Cir. 1978); *Cormier v. P.P.G. Indus.*, 519 F. Supp. 211, 261-62 (W.D. La. 1981), *aff'd*, 702 F.2d 567 (5th Cir. 1983); *Pegues v. Mississippi State Employment Serv.*, 488 F. Supp. 239, 252-54 (N.D. Miss. 1980), *aff'd in part on other grounds*, 699 F.2d 760 (5th Cir.), *cert. denied*, 464 U.S. 991 (1983).

¹⁶¹ See *Forty Questions*, *supra* note 141, at 768-69. Hunter and Schmidt do not deny the sensitivity of output to organizational intervention. See *id.* Instead, they argue that although absolute output levels might shift, the relative outputs of different workers will remain the same. See *id.* This claim also seems implausible, however, in that different organizations will clearly demand different levels of autonomous decisionmaking from workers or a different mix of mechanical and cognitive skills. See M. PIORE & C. SABEL, *THE SECOND INDUSTRIAL DIVIDE* 205-20 (1984) (arguing that industry is currently shifting away from a mass production mode in which the employees make relatively few autonomous decisions toward flexible production, shorter production runs, and greater product variety, a mode in which employee creativity is more valuable).

¹⁶² When Toyota and General Motors entered a joint venture to produce cars at a Fremont, California plant, they retrained a notoriously unproductive workforce. The companies radically reorganized the plant, emphasizing the use of autonomous production groups rather than an assembly line mode of production. Productivity levels increased by 50%. See Brown & Reich, *When Does Union-Management Cooperation Work? A Look at NUMMI and GM-Van Nuys*, 31 CAL. MANAGEMENT REV. 26, 27-34 (1989); Levin, *Issues of Agreement and Contention in Employment Testing*, 33 J. VOCATIONAL BEHAV. 398, 400-01 (1988).

¹⁶³ If it were possible to increase absolute productivity for a group of workers, and if one were to assign only low-test-score workers to a better-organized subsection of a plant, the differentiation between low- and high-test-score workers would at least decrease and perhaps disappear altogether. Obviously, such a redesign of job assignments is implausible, but it is no more implausible than the opposite assumption made by Hunter and Schmidt: that workers at each test level are randomly assigned to jobs with more and less output-enhancing capital and labor organizational structures. Moreover, if tests are supposed to predict relative productivity across jobs, low-test-score workers who are employed by a highly productive plant will exceed the performance level of high-test-score workers at a less productive plant. Although validation studies have tended to be localized by plant, there is no inherent reason that they should be, because the underlying claim seems to be that test-biased differentiation in output is an insoluble, immutable pre-social fact about individuals.

¹⁶⁴ See *Forty Questions*, *supra* note 141, at 744, 770.

In any event, claims of universal validity are statistically grounded in an unsatisfactory and conclusory procedure. Hunter and Schmidt assume that the mean validities measured in a variety of job settings represent the actual validity for what is simply presumed to be a uniform group of workplaces. As long as variance around this mean can be explained in terms of artifactual elements (such as sampling error), they assume that the source of variance is the artifactual elements rather than the existence of some situationally specific mediators. This error in statistical inference, known as "affirming the consequent,"¹⁶⁵ leads Hunter and Schmidt to understate the possibility of genuine situational specificity.¹⁶⁶ In other words, they underesti-

¹⁶⁵ James, Demaree & Mulaik, *A Note on Validity Generalization Procedures*, 71 J. APPLIED PSYCHOLOGY 440, 442 (1986); James, Demaree, Mulaik & Mumford, *Validity Generalization: Rejoinder to Schmidt, Hunter, and Raju*, 73 J. APPLIED PSYCHOLOGY 673, 675 (1988).

¹⁶⁶ To offer a simple example, assume that one selected 10 groups of 100 persons, each with a different mean and median height. If one sampled two persons from each group, the observed variance in average height for each group would almost surely be explicable in terms of sampling error, given the tiny sample size, but it would *not* be reasonable to infer that the groups in fact had the same mean height. One could (as Hunter and Schmidt do) first assume no variation between groups; then assume that the mean for the 20 person sample was the mean for the whole group of 1000; and finally assume this mean for the whole group was in turn the mean for each subgroup, at least as long as the observed means for the sample of each subgroup were not more dispersed around the presumed common mean than one would expect given the massive sampling error. But one would have simply rejected, by fiat, a perfectly plausible alternative hypothesis — mean height could in fact be *different* from group to group. One cannot know which hypothesis is true because of the inadequate sample size. Typical employment validity studies have small sample sizes. One study has found that the median sample size of studies that are typically subject to meta-analysis is 68. See Lent, Aurbach & Levin, *Research Design and Validity Assessment*, 24 PERSONNEL PSYCHOLOGY 247, 258 (table 5) (1971). Given these small sample sizes, we cannot simply infer that variations in reported validities are a function of artifacts from the possibility that they *could* be. A number of statisticians have done "Monte Carlo" studies, in which samples of artificially differentiated groups are selected randomly. These studies conclude that when the sample sizes are less than 200, small and moderate distinctions in actual underlying validities will typically go undetected. See, e.g., Sackett, Harris & Orr, *On Seeking Moderator Variables in the Meta-Analysis of Correlational Data: A Monte Carlo Investigation of Statistical Power and Resistance to Type I Error*, 71 J. APPLIED PSYCHOLOGY 302, 305-06 (1986).

It is perhaps worth reviewing the Hunter-Schmidt procedure briefly to see why these "Monte Carlo" tests are important. Hunter and Schmidt look at the observed variance in a large number of validity studies, which themselves have relatively low sample sizes. No "situational moderators" are directly measured or identified; if variance is no greater than that which would be observed given the effect of artifactual elements, such moderators are presumed to be absent. But as a result of "second order sampling error," meta-analysis will not recognize the presence of situational moderators; that is, it will not detect real distinctions between situations, but will instead attribute different validities to sampling error. See *id.*

Hunter and Schmidt also use an inappropriate "75% rule of thumb": if more than 75% of variance is explained by artifacts, they reject the possibility of situational specificity, arguing that the remaining variance results from otherwise unmeasured artifacts. See Schmidt, Hunter & Raju, *Validity Generalization and Situational Specificity: A Second Look at the 75% Rule and Fisher's z Transformation*, 73 J. APPLIED PSYCHOLOGY 665, 666-67 (1988). Commentators have justly criticized this rule not only because of the incapacity of small sample size studies to

mate the possibility that different job settings have unique characteristics that will affect the validity of the general ability test. The flaws in their own approach cast doubts on their claim of universal validity for general ability tests.¹⁶⁷

3. *The Claim of Lack of Bias.* — Finally, Hunter and Schmidt are wrong to argue that tests are not biased against blacks. In section B, I discuss at length how the use of a test can be biased even if it predicts job performance equally well for members of minority and majority groups. For now, I address a narrower point. Hunter and Schmidt have consistently denied that tests are differentially valid — that is, they deny that test scores predict intragroup variations in the performance of minority group members less well than they predict intragroup variations among majority group members.¹⁶⁸ But they provide insufficient evidence to support this strong claim. Hunter and Schmidt rely primarily upon studies of black-white validity pairs — that is, studies in which test validators separately correlated test scores with job performance for the black and white workers. According to Hunter and Schmidt, these studies show that significant differences in measured validity are found at random rates.¹⁶⁹

detect real distinctions, but also because of the cavalier dismissal of substantial unexplained variance (up to 25% of such variance). After all, all artifactual sources of error other than sampling error (such as range restriction, criterion unreliability, and test unreliability) explain only 7.5% of variance. See James, Demaree & Mulaik, *supra* note 165, at 444-45.

¹⁶⁷ Despite the flaws in their analysis, Hunter and Schmidt's meta-analytic procedure reminds us of certain statistical truths that should give us pause in accepting the sort of local validation studies that have generally been the mainstay of title VII litigation. Hunter and Schmidt's work should make us recognize that the use of local validation studies will frequently be unduly *permissive* to discriminating employers and will treat both employers and applicants differently for unacceptable, random reasons. Given the small sample size in the typical validation study, we would expect to find substantial validities in many studies simply as a result of chance even though the underlying population validities were actually near zero. Hunter and Schmidt focus largely on the problem of employers being unable to demonstrate the validity of their screening device, even when the test is valid. Studies with sample sizes around 65 have only a 50% chance of detecting true validities as high as .5. See Schmidt, Hunter & Urry, *Statistical Power in Criterion-Related Validation Studies*, 61 J. APPLIED PSYCHOLOGY 473, 475-76 (1976). But Hunter and Schmidt's data indicates the opposite problem is significant as well. The 95% confidence level where $N=68$ is 48 correlation points wide; thus, when a correlation is observed to be .48 (which would be atypically substantial), there is still a five percent chance that the correlation is actually zero. It might be necessary for courts or administrative agencies to adopt a per se rule of universal test *invalidity* to cope with this problem, even if one believed that underlying validities *differed* according to setting in a way that made tests reasonably predictive in some instances. See *id.* at 481.

¹⁶⁸ See, e.g., Hunter, Hunter, & Schmidt, *supra* note 141, 733; Schmidt, Berner & Hunter, *Racial Differences in Validity of Employment Tests: Reality or Illusion?*, 58 J. APPLIED PSYCHOLOGY 5, 6-8 (1973). Hunter and Schmidt's position echoes that taken by Virginia Boehm in Boehm, *Differential Prediction: A Methodological Artifact?*, 62 J. APPLIED PSYCHOLOGY 146, 153 (1977).

¹⁶⁹ See Boehm, *supra* note 168, at 153; Hunter, Hunter & Schmidt, *supra* note 141, at 733. In other words, if we selected from a white-only group a group of samples the size of the actual

Some of Hunter and Schmidt's evidence, however, suggests the opposite view. One can infer from their own figures that a meta-analysis of black-only and white-only validity studies in fact shows that black validities *are* lower. Hunter and Schmidt misinterpret their own data, however, to claim that "[t]he overall difference between the mean validities for whites and blacks was zero."¹⁷⁰ They obtain that result only by eliminating eighty-five pairs of studies where validities for blacks were negative.¹⁷¹ In doing so, they eliminate nearly ten percent of the total sample of validity pairs. Although their article does not contain adequate data to recalculate the actual mean validities for the full universe of black-only studies, it is obvious that black validities would be markedly lower than white validities if these studies had not been removed.¹⁷²

It may be that tests are racially biased, not because they are differentially valid, but because they systematically underpredict the performance of blacks as a group.¹⁷³ Most industrial psychologists,

black samples and labeled these samples the "greens," significant green-white disparities would exist as frequently as black-white disparities. Given the existing small sample sizes in black-only validity studies, we cannot prove that black validities are significantly lower than white validities. Because the black sample sizes are so low, however, this finding is hardly startling. It certainly does not settle the question whether real disparities exist. Just as we cannot reject the possibility of situational specificity due to second-order sampling error, *see supra* note 166, so we cannot reject the possibility of differential validities because sample sizes are too small to detect such differentials.

¹⁷⁰ Hunter, Hunter & Schmidt, *supra* note 141, at 731.

¹⁷¹ *See id.* at 727-30.

¹⁷² It may be the case, as Hunter and Schmidt argue, that the findings of negative validity were unwarranted. *See id.* at 727-28. The general justification for using a meta-analytic procedure, however, is that when large numbers of studies are summarized, the individual errors in each study will cancel one another out. Even if actual negative correlations were rare, one should expect negative correlations in samples when actual population correlations are zero or near zero. Of course, many *positive* correlations observed in black-only studies would presumably occur even if the underlying population correlation were zero.

Analysis of black-white pair studies validating the GATB weakly confirms the proposition of differential validity, although given the low validities for majority group tests there is not much room for massive differential validities. White-only tests have a mean validity of .19, while black-only tests have a mean validity of .12. *See* FAIRNESS, *supra* note 1, at 176, 188. In one quarter of black-only validity studies of the GATB, correlations are .03 or lower. *See id.* The findings of differential validity are rather striking, given the strong possibility that black-only validities are overstated to the extent that more "assimilationist" blacks do better on culturally biased tests and are overrated by supervisors.

¹⁷³ *See* FAIRNESS, *supra* note 1, at 178; Linn, *Single-Group Validity, Differential Validity, and Differential Prediction*, 63 J. APPLIED PSYCHOLOGY 507, 510-11 (1978). Validities could hypothetically be equal for a test that grossly under- or overpredicts black performance. Assume one could predict the rank order of black workers' on-the-job performance solely by knowing their test scores, just as one could for whites. The correlation between test and criterion would be perfect; intragroup validities for each group would be 1. It would still be theoretically plausible, however, that a black whose score is equal to a white's would be expected to do better on the job (if the test underpredicts black performance) or worse (if the test overpredicts). The regression equation predicting performance [where *P* (performance) is a function of the test

however, believe that tests *overpredict* black performance.¹⁷⁴ This claim suggests forms of bias in the workplace that commentators do not generally recognize, such as the possibility that existing work organization demands even greater suppression of culturally specific black traits than do the tests.¹⁷⁵ It is important to realize that we must inevitably assess a factually controversial record when dealing with claims of racial bias. The ideology of the civil rights laws, broadly understood, directs us to react with vigilant skepticism to claims that blacks are less capable than whites. We must reject such claims whenever evidence is ambiguous,¹⁷⁶ even when the claim is

score, given as some constant C (the intercept) plus the test score (T) multiplied by a constant coefficient, S (the slope)], would then be different for whites and blacks. Whether this hypothetical possibility is plausible is discussed in note 175, *infra*.

¹⁷⁴ See FAIRNESS, *supra* note 1, at 180-82, 185, 188; Bartlett, Bobko, Mosier & Hannan, *Testing for Fairness with a Moderated Multiple Regression Strategy: An Alternative to Differential Analysis*, 31 PERSONNEL PSYCHOLOGY 233, 235-36 (1978); Gael, Grant & Ritchie, *Employment Test Validation for Minority and Nonminority Clerks with Work Sample Criteria*, 60 J. APPLIED PSYCHOLOGY 420, 424-25 (1975); Gael, Grant & Ritchie, *Employment Test Validation for Minority and Nonminority Telephone Operators*, 60 J. APPLIED PSYCHOLOGY 411, 416 (1975); Linn, *supra* note 173, at 511.

¹⁷⁵ Most industrial psychologists believe that tests would properly predict each group's aggregate performance if validities were equal. See Hunter, Hunter & Schmidt, *supra* note 141, at 733-34. The basis of this intuition is that the claim of bias is based on the idea that tests are unduly culturally specific. Were that the case, intra-black differentiation on the tests would be largely a function of "assimilation" rather than either "abstract cognitive skills" (assuming, for the sake of argument, such skills exist) or "assimilation" skills that are as useful in the workplace as they are on the tests. But, if validities are equal, intra-black test performance varies in precisely the same way as intra-majority group performance. See *id.*

This plausible hypothesis, however, is incomplete. To the extent that superior black performance on the job (again, assuming for the sake of argument that it can be properly measured) were in significant part a function of "assimilation" rather than skills, which might be the case when there is racist supervision or culturally insensitive plant organization, we would still expect the black-only validity coefficients to be positive. In such a scenario, we would also expect that black test scores would *overpredict* on-the-job performance as long as existing job organizations typically suppress black performance even more than tests. Alternatively, they would *underpredict* performance if the tests demand even more "assimilative" skills than employers do. Empirical findings of overprediction, see sources cited *supra* note 174, might well be in part an artifact of measuring the predictive powers of tests solely in plants that are most likely to be organized in a racist fashion. It is certainly possible that, as a group, employers who use tests in hiring would be unlikely to remedy organizational features that suppress black performance at the production stage given that they know that such tests have adverse racial impact. Although the tests might appear unbiased or even "too fair" to blacks (at least statically) given the population of firms that are studied, they might not be unbiased if the tests were used more extensively. Thus, even if one does not adopt a dynamic validity perspective and demand test validation in an "acceptable" work setting, the claim of across-the-board static validity must remain suspect if hostile work settings are atypically prevalent among test givers.

¹⁷⁶ The impetus behind the dismantling of the American system of enforced legal segregation, from *Brown v. Board of Education*, 347 U.S. 483 (1954), through the Civil Rights Acts, was not just the normative ideal of legal equality but the supposition of *factual* equality — the rejection of the idea of a natural hierarchy of groups. The stigma of enforced separation that *Brown* rightly decied was the assertion of white superiority implicit in the need of the politically

not that blacks are inherently less capable but simply less capable because of social disadvantage. It is surely debatable whether there is any meaning to the concept of general intelligence or aptitude, or if, by contrast, all intelligence consists merely of situation-specific problem-solving skills.¹⁷⁷ It is even more doubtful whether one can measure aptitude by testing an individual's capacity to solve a particular set of problems that must necessarily be situation-specific.¹⁷⁸ Finally, it is evident that the effort to measure this "capacity" has all too frequently been consciously motivated by the desire to prove the superiority of politically dominant ethnic groups or classes or to justify overtly racist eugenics programs.¹⁷⁹

4. *Utility.* — Even if tests were more predictive than I have argued here, their utility would remain debatable. Employers should not be able to use even valid tests unless their use saves money for the firm.¹⁸⁰ Yet arguments that particular employers would save large sums of money by using tests have not been persuasive.¹⁸¹ Arguably,

dominant whites to keep blacks distant. See Black, *The Lawfulness of the Segregation Decisions*, 69 YALE L.J. 421, 424–25 (1960). The pervasiveness of private discrimination was a function of white supremacist beliefs, not an evenhanded, non-hierarchical separatist ideal. Even the pro-segregationist sentiment that "blacks don't want to mix with whites either" typically rested on the assumption that white culture was too advanced or complex for inferior people, rather than on the belief that blacks might want to maintain an inherently valuable cultural integrity. See, e.g., N. BARTLEY, *THE RISE OF MASSIVE RESISTANCE* 12–13, 184–85, 237–39 (1969); Akins, *Social Equality and The Fourteenth Amendment: The Original Understanding*, 4 HOUS. L. REV. 639, 646–50 (1967); Brown, *White South is a Minority Group*, 17 ALA. LAW. 438, 442 (1956).

To the extent that desegregation decisions and civil rights statutes were deliberately intended to oppose the white supremacist ideology that sustained imposed separation, the statutes should be read as hostile to that ideology, as well as to the specific forbidden practices that ordinarily followed from it. Anyone who desires to exclude blacks faces an extremely heavy burden in showing that some aspect of traditional supremacist thinking (for example, that black job applicants lack cognitive ability) is actually correct.

¹⁷⁷ See Glasser, *Education and Thinking: The Role of Knowledge*, 39 AM. PSYCHOLOGIST 93, 98 (1984).

¹⁷⁸ See Flynn, *Massive IQ Gains in Fourteen Nations: What IQ Tests Really Measure*, 101 PSYCHOLOGICAL BULL. 171, 184 (1987).

¹⁷⁹ See J. BLUM, *PSEUDOSCIENCE AND MENTAL ABILITY* 99–100 (1978); S. GOULD, *THE MISMEASURE OF MAN* 30–72 (1981); Delgado, Bradley, Burkenroad, Chavez, Doering, Lardiere, Reeves, Smith & Windhausen, *Can Science Be Inopportune? Constitutional Validity of Governmental Restrictions on Race-IQ Research*, 31 UCLA L. REV. 128, 131–44 (1983). The fact that scientific racists of the past have made what we see now as egregious technical errors, as well as more grievous moral ones, does not necessarily prove that today's testing advocates are either technically incompetent or motivated by racial animus. Nevertheless, we must read all social scientific data, past and present, in the context of a broader cultural history.

¹⁸⁰ See *supra* pp. 1165–66.

¹⁸¹ Whether a particular employer would increase plant-wide productivity by using tests to screen depends on three factors: the predictive validity of tests relative to alternative screening devices, the employer's selection ratio (the capacity of the employer to choose those employees she most desires to fill openings), and the degree to which different employees are differentially productive. The standard formula, generally employed by psychometricians, is that $G = (r)(s)(A)$,

the law should allow employers to defend practices with disparate impact only when the practices reduce both the firm's costs and social costs.¹⁸² Claims that widespread use of tests would dramatically improve aggregate economic performance, however, have even less empirical support than claims that individual firms are helped by tests.¹⁸³

where G is the dollar gain per worker per year due to using a test in hiring, r is the correlation between test score and productivity (validity), s is the standard deviation of yearly productivity per worker in dollars, and A is the average test score of applicants selected, standardized so that the mean for applicants is zero and the standard deviation is 1. This formula was originally suggested in Brogden, *On the Interpretation of the Correlation Coefficient as a Measure of Predictive Efficiency*, 37 J. EDUC. PSYCHOLOGY 65 (1946). For further discussions of the equation, see FAIRNESS, *supra* note 1, at 235–36; and Hunter & Schmidt, *Fitting People to Jobs: The Impact of Personnel Selection on National Productivity*, in 1 HUMAN PERFORMANCE AND PRODUCTIVITY 233, 236–40 (M. Dunnette & E. Fleishman eds. 1982). Hunter and Schmidt estimate that individual employers can save \$80 billion annually by testing. They assume validities of .5, a standard deviation of per-worker productivity (SDy) of 40% of annual wages, and a 10% selection ratio. Yet a *high* estimate of the increase in validity of tests over other screening devices would be more like .2 than .5, the standard deviation of productivity is probably closer to 20% than 40% of annual wages, and a selection ratio of six to one appears a quite high estimate. For a fuller discussion of the problem in Hunter and Schmidt's attempts to measure the gains from testing, see M. Kelman, Concepts of Discrimination, with Special Reference to "General Ability" Job Testing 123–28 (Working Paper No. 65, John M. Olin Program in Law and Economics, Stanford University (1990)). Using these numbers, the utility of tests drops from \$80 billion to between \$10 and \$13 billion, depending on assumptions about worker tenure. These were the estimates ultimately adopted by the National Academy of Science's Committee on the GATB. See FAIRNESS, *supra* note 1, at 238–41. I am not certain there are any plausible estimates of SDy . Although I have no particular objection to the use of the 20% figure, I view it as essentially fanciful. All the other estimates strike me as plausible upper-bound estimates.

¹⁸² See *supra* p. 1201.

¹⁸³ It would be quite difficult to estimate the impact of universal testing on aggregate productivity even assuming, counterfactually, that we knew the extent to which a particular employer with a reasonably low selection ratio could be aided by testing. Workers rejected by one employer almost invariably find work elsewhere. At the economy-wide level, the selection ratio very nearly approaches one. Assuming, for the sake of argument, that the failure to test is costly to a particular employer, it is still plausible that perfectly valid tests that increase a particular employer's productivity *hurt* economy-wide productivity. At the general equilibrium level, the relevant inquiry is not how to maximize A 's product alone — to note how much more high-test-score Applicant A would produce at the job he most desires, Job X , than low-test-score Applicant B would — but how to maximize the *sum* of the two workers' products at the two jobs they will fill, X and Y . See FAIRNESS, *supra* note 1, at 245; Levin, *supra* note 134, at 226, 228.

Given the difficulty of determining whether private optimizing decisions are harmful or helpful to aggregate productivity, Hunter and Schmidt's claim that they have "conservatively" estimated GNP losses from the failure to test at roughly \$87.5 billion per year is dubious. See Hunter & Schmidt, *supra* note 181, at 268. First, Hunter and Schmidt's decision to compare the gains from testing to random assignment is puzzling. See *id.* at 263, 266. Even if there were no other pre-employment screening devices (such as educational requirements), *no* employment system would ultimately be a random assignment system as long as employers have discretion to promote, retain, and fire different workers. Furthermore, in most mainstream economists' models of labor markets, workers match themselves to jobs quite well in the first instance. See, e.g., Heckman & Sedlacek, *Heterogeneity, Aggregation, and Market Wage Func-*

B. Do Low-Validity Tests Comport with the Demands of Meritocracy and Antidiscrimination Principles?

Although I believe that I have demonstrated that tests are biased in the sense that they mismeasure the probability that minority applicants will perform a job adequately, my primary interest in this section is in exploring whether the use of tests should be considered biased even if the tests do *not* mispredict minority applicants' performance. To focus attention on this issue, I make a number of simplifying factual and normative assumptions.

I assume, first, that when validating a test, we do not require that the employer alter prevailing workplace organization to increase the output of those who perform poorly on both the test and the job as currently constituted. Similarly, we do not require the employer to distribute jobs in accord with norms not based on output, although whether the employer is entitled simply to maximize profits remains an issue. Second, I assume that test validities are positive, but low — that is, tests predict performance, but not very well.¹⁸⁴ Third, I assume that general aptitude tests are equally valid across all jobs, although I believe that the discussion in Part IIIA showed that assumption to be incorrect.¹⁸⁵ Fourth, I assume that the only alternative available to employers other than screening applicants through tests is lottery hiring; there are no other valid *ex ante* screening devices. Employers could, however, fire unsuccessful employees, and applicants might self-select to a substantial degree to avoid the psychological and financial costs of failing on a job. Fifth, I assume that tests are unbiased in the sense that tests are equally valid for blacks and whites *and* that tests properly predict the performance of blacks as a group, although, again, I believe this assumption to be incorrect.¹⁸⁶ Finally, I assume that blacks do substantially worse than whites on general aptitude screening tests.¹⁸⁷

There is a strong argument that, although the screening tests themselves are unbiased in such a case, the *use* of the tests is nevertheless illegitimately discriminatory. Even if people who did worse

tions: An Empirical Model of Self-Selection in the Labor Market, 93 J. POL. ECON. 1077, 1078-80 (1985); Willis & Rosen, *supra* note 143, at S28. Hunter and Schmidt present no data to demonstrate that labor is actually misallocated when "qualified" employees are displaced by less qualified ones. Alternatively, one would expect Hunter and Schmidt, or other testing advocates, to attempt to demonstrate that companies or industries that abandoned general ability tests in favor of alternative employee selection procedures experienced otherwise inexplicable productivity declines, or that the economy generally experiences such declines when the use of tests diminishes. No such econometric evidence exists.

¹⁸⁴ See *supra* pp. 1208-14.

¹⁸⁵ See *supra* pp. 1214-17.

¹⁸⁶ See *supra* pp. 1217-20.

¹⁸⁷ See *supra* note 1.

on the test did not constitute a socially cohesive group (such as a racial or ethnic group), but simply represented a random group of bad test takers, reliance on the test would be problematic for persons who believe in individualistic meritocracy. Although the use of the test would be privately rational for employers, it would arguably breach what might be termed a duty to hire people in accord with "merit." Merit, in this sense, would be the capacity to do work as well as another, rather than the capacity either to do well on tests *or* to maximize one's employer's profits.

The case for imposing an affirmative obligation on employers to hire in accord with this sort of merit is considerably stronger if the "victims" of the employer's privately rational use of the test are an historically oppressed racial group. Thus, those who advocate individual output theories of discrimination should oppose testing under these circumstances. Finally, title VII could arguably be read to preclude even the use of higher validity tests with a racially disparate impact if broad utilitarian concerns can trump meritocratic principles.

1. *Concepts of Racial Bias.* — The Cleary definition of bias that psychometricians generally use,¹⁸⁸ and which both the American Psychological Association¹⁸⁹ and the EEOC¹⁹⁰ endorse, declares a test unbiased so long as it predicts minority performance on the job as well as it predicts nonminority performance.¹⁹¹ This is not, however, the only definition of bias that psychometricians might employ. One might argue, for instance, that a selection device is biased unless the proportion of potentially successful applicants to the number selected is the same for each group.¹⁹² Alternatively, one might describe a

¹⁸⁸ See Cleary, *Test Bias: Prediction of Grades for Negro and White Students in Integrated Colleges*, 5 J. EDUC. MEASUREMENT 115 (1968). Hunter and Schmidt also support the Cleary standard. See Hunter & Schmidt, *Critical Analysis of the Statistical and Ethical Implications of Various Definitions of Test Bias*, 83 PSYCHOLOGICAL BULL. 1053 (1976); Hunter, Schmidt & Rauschenberger, *Methodological, Statistical, and Ethical Issues in the Study of Bias in Psychological Tests*, in PERSPECTIVES ON BIAS IN MENTAL TESTING 41, 58-62 (C. Reynolds & R. Brown eds. 1984).

¹⁸⁹ See AM. EDUC. RESEARCH ASS'N, AM. PSYCHOLOGICAL ASS'N & NAT'L COUNCIL ON MEASUREMENT IN EDUC., STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTING 12-13 (1985).

¹⁹⁰ See EEOC, Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607.16V (1990).

¹⁹¹ See Cleary, *supra* note 188, at 115. In technical terms, both the slope and intercept of minority regression lines predicting on-the-job performance must be the same as the slope and intercept of nonminority regression lines if the test is to be deemed unbiased. Thus, assume a predictor score, P , is designed to predict a criterion score, C , according to a formula $K + xP = C$; K is the intercept and x is the slope of the regression line. Cf. *id.* at 122 (showing results of similar regression analysis).

¹⁹² See Thorndike, *Concepts of Culture-Fairness*, 8 J. EDUC. MEASUREMENT 63, 63 (1971) (arguing that a test is unbiased only to the extent that it selects minority and majority group applicants as they would be selected by a perfectly valid test).

selection mechanism as biased unless the probability of selection for an individual who could in fact succeed on the job is the same regardless of whether the individual is a member of the majority or minority group.¹⁹³ Finally, one could pronounce a selection device biased unless the probability of success for a selected individual is equal for both the majority and minority groups.¹⁹⁴

It may not be immediately apparent why these definitional controversies matter. To explain, I begin by positing a minimally predictive test that is fair in the Cleary sense: it is equally, but not highly, valid for blacks and whites. The test in such a case might be considered unbiased in the limited sense that the test instrument cannot be improved at all by simply removing "unfair" questions. The test *is* biased, however, in the sense that it is particularly incapable of recognizing the ways in which poor test takers, who are disproportionately minority group members, could nevertheless succeed on the job.¹⁹⁵ In other words, the test may not contain a pattern of questions that is irrelevant to job performance and yet distinguishes minority and nonminority test takers; instead, the problem may be that the test does not contain enough questions. The test's failure to measure a broader range of skills would hurt minority group members most often, but minorities would be no more hurt as individuals than would nonminority bad test takers.¹⁹⁶

Assume that an applicant pool consists of 100 blacks and 200 whites. Half (fifty) of the blacks pass this "fair" test, while three-fourths (150) of the whites do. Employers hire all people who pass

¹⁹³ See Darlington, *Another Look at "Cultural Fairness,"* 8 J. EDUC. MEASUREMENT 71, 75 (1971). In describing four distinct views of what constitutes a fair test, Darlington suggests that a test could be considered biased if the correlation between group membership and test performance is greater than the correlation between group membership and job performance. See *id.* at 72-77. A test would be unbiased in this sense if "conditional probabilities" were equal — that is, if the probability of selection given the relevant condition (that the applicants are equally capable) is the same for members of all groups. See *id.*

¹⁹⁴ See Einhorn & Bass, *Methodological Considerations Relevant to Discrimination in Employment Testing,* 75 PSYCHOLOGICAL BULL. 261, 261-63 (1971).

¹⁹⁵ Workers may well have coping and learning techniques that are not measured by the tests. See, e.g., Scribner, *Studying "Working" Intelligence,* in EVERYDAY COGNITION: ITS DEVELOPMENT IN SOCIAL CONTEXT 9, 10-13 (B. Rogoff & J. Lave eds. 1984); Scribner, *Thinking in Action: Some Characteristics of Practical Thought,* in PRACTICAL INTELLIGENCE: THE NATURE AND ORIGINS OF COMPETENCE IN THE EVERYDAY WORLD 13, 26-28 (R. Sternberg & R. Wagner eds. 1986).

¹⁹⁶ An analogy may be helpful. Assume that we are trying to assess "qualifications" for a sport — perhaps archery — for which upper-body strength is useful, but only to a very limited degree. Upper-body strength is equally useful for men and women; it will correlate equally well with archery success for each gender, but not very well for either. Assume women generally have less upper-body strength. If we selected the archery team solely on the basis of upper-body strength, we would have used a measure unbiased in the Cleary sense. However, the test would remain particularly inapt in capturing the ways in which those with less upper-body strength, who are typically women, may succeed as archers.

The ratio of whites who can potentially succeed to those who are hired $[(140 + 20)/(140 + 20) = 1]$ is lower than the analogous ratio for blacks $[(50 + 20)/(50 + 10) = 1.17]$.

On the other hand, the test is biased in favor of blacks if we focus on individuals who in fact are hired and determine whether among those hired, blacks are as likely as whites to succeed. Recall that in this view a test is biased unless the following equation is met:

$$\frac{I_W}{I_W + IV_W} = \frac{I_B}{I_B + IV_B}$$

Here, however, whites who are hired have a higher probability of success $(140/160 = .87)$ than do blacks $(50/60 = .83)$.

Finally, the test is biased *against* blacks if we focus on applicants who will succeed on the job and try to ensure that an equal proportion of those persons is picked from each group. A selection procedure is unbiased in this view only if the probability of selection for persons with an equal potential for success is the same regardless of race — that is, if the following equation is met:

$$\frac{I_W}{I_W + II_W} = \frac{I_B}{I_B + II_B}$$

Here, 140/160 or 87% of potentially successful whites are selected while only 50/70 or 71% of potentially successful blacks are. Twenty of the forty blacks who could have succeeded were not hired (20% of the total black applicants), compared to only 10% of the white applicants (20 of 200). Moreover, whites who will in fact turn out to be poor workers get hired considerably more frequently than do blacks who will turn out to be poor workers. Using our numbers, forty of the 200 whites would be poor workers if all were hired, and twenty of these (50%) in fact get hired (fifteen who pass the test and five who fail). Thirty of the 100 blacks would be poor workers if all were hired, and only ten of these (33%) in fact get hired (five who passed the test and five who failed.)

I believe this last definition best expresses the vision of “meritocracy” I explore in this section, and is thus the most morally attractive of the views. This view alone asks whether parties equally capable of succeeding on the job have an equal chance of being hired. If we adopt this view, then for almost all values of pass and accuracy rates, a test that predicts equally well for blacks and whites will become less fair as the relative pass rates on the test diverge and as validities decline. To return to the example above, if only twenty of 100 blacks passed the test, while every other assumption remained the same, 87% of whites who could succeed would now be hired, compared to 45% of potentially successful blacks. (The employer would have hired all

twenty blacks who passed, eighteen of whom would succeed. He also would have hired sixteen of those who failed, eight of whom would succeed. Of the sixty-four whom he did not hire, thirty-two would have succeeded. Thus he hired 26/58 or 45% of those who could have succeeded.)

Now imagine the pass and hiring rates we initially posited, but assume that seven out of ten of those who fail the test succeed on the job and only eight out of ten test passers do — that is, that the test predicts success very weakly. In such a case, only 63% of good black workers are hired. (Of the fifty blacks who pass the test and are hired, forty will succeed; of the ten blacks who are hired despite failing the test, seven will succeed. Of the forty blacks who are rejected, twenty-eight would succeed. Thus 47/75 or 63% of good black workers are hired.) On the other hand, 82% of the whites who would succeed are hired. (All 150 whites who pass the test are hired and 120 of them succeed; ten whites who fail are hired and seven of those will succeed; of the forty whites who are rejected, twenty-eight would succeed. Thus 127/155 or 82% of good white workers are hired.) Moreover, the percentage of poor white workers who would be hired is now 73% (33/45), while the percentage of poor black workers who would be hired is only 52% (13/25).

Finally, change both conditions at once. Assume twenty blacks pass and eighty fail *and* that 70% of those who fail tests and 80% of those who pass tests succeed on the job. In such a case, of those blacks who would succeed if all blacks were hired, only 38% (27/72) will be hired, because so many blacks will have failed the test. The employer ends up hiring 73% of the whites who will be failures on the job and only 38% of the blacks who will succeed. *A far higher proportion of whites who will fail are hired than the proportion of blacks who would succeed.*¹⁹⁷

¹⁹⁷ This is not a merely hypothetical problem. Even if we assume that the GATB has a reasonably high validity — say, a validity of .3 — the relationship between test success and job success is modest. A white test taker who scored in the 16th percentile on the test (one standard deviation below the white mean) would still have a 38% chance of being an above-average worker — the same chance that an average black test taker would have. A black test taker who scored in the 16th percentile of black test takers would still have roughly a 27% chance of being an above-average worker, while a white in the 84th percentile of white test takers (one standard deviation above the mean) would have a 42% chance of being a below-average worker; that is, roughly one in four black test takers who did especially poorly on the test within the black group would still perform better on the job than four in ten of the whites who did especially well within the white group. See FAIRNESS, *supra* note 1, at 276–77.

The conceptual question here is the probability that a person whose predictor score is X standard deviations away from the mean is likely to finish above the criterion mean, given a particular correlation between predictor and criterion. If the correlation (C) is .3, as we have assumed, we must figure out the probability that someone whose score predicts he will perform .3 standard deviations (C times X) below mean job performance will actually have a job rating of average or better. If we assume that job performance is normally distributed, there is a 38%

2. *Individualist Claims Against Weakly Valid Screening Tests.* — Ultimately, whether the use of tests under the conditions just described should be considered biased depends on one's definition of individual entitlements (if one approaches this question as one of individual rights) and on whether one believes that group results are ethically relevant. One way to explore the basis of an objection to the use of such tests is to imagine, first, that those who do badly on the test constitute no socially recognizable group — say, for example, there is a random group of people with “test anxiety.” To the extent that an individual possessing no culturally meaningful group identity has a colorable fairness claim against the use of the test, his appeal would be to liberal individualist principles. His claim could be stated as a right of an abstract universal citizen, without regard to his more particular situation of group membership.¹⁹⁸

chance that someone expected to be .3 standard deviations below a mean will have a score at that mean or better: one simply assumes the true mean is .3 standard deviations below the population mean and asks how many observations would be at OSD or above. Thirty-eight percent of observations of a normal probability distribution are more than .3 standard deviations away from the mean in each direction. See R. JOHNSON, *ELEMENTARY STATISTICS* 210–16, 431 (2d ed. 1985). If validities drop to .1, a person with a score one standard deviation below the mean would be predicted to perform one-tenth of a standard deviation below the mean on the job, and the probability that she would finish with a criterion score at or above the actual mean would be 46%, while someone finishing a standard deviation above the mean would have only a 54% chance of being in the top half. A black applicant one standard deviation below the black test mean — hence, two standard deviations below the white mean — would still have a 42% chance of being an above-average worker. Finally, if GATB validities were most typically about .2 (which, if one had to choose a single “universal” number, is probably the most plausible upper-bound estimate of the validity gains beyond standard biodata and self-selection screening), 42% of those with scores in the 16th percentile of whites (one standard deviation below the white mean, roughly at the black mean) and 34% of those two standard deviations from the white mean would be expected to be in the top half of job performers. See *id.*

¹⁹⁸ Assume a medical school admissions officer has narrowed the applicant pool for a class of 500 to 1000 applicants, 500 women and 500 men, who seem to him equally capable. He believes, quite reasonably, that it is bad to waste school resources on those who will not practice medicine, and he knows that 96% of admitted women become practitioners while 98% of men do. Moreover, he knows there are no more particularized predictors that work better than gender. Admitting only men will indeed meet a valid goal, maximizing the production of doctors, but it should drive meritocrats crazy. Each woman is in fact someone who will either become a doctor or not, but the chances that those women who would in fact practice (who are, as it turns out, 96% of the female applicant population) get admitted is zero, while the chances that those men who will not actually practice get admitted is 100%.

From the purely individualistic meritocrat's perspective, nothing at all turns on the fact that the discriminating criterion in this case is group membership itself. A psychological test that measured “propensity to finish school” by distinguishing those who fail the test (and in fact become practitioners 96% of the time) from those who pass (who become practitioners 98% of the time) would be similarly troublesome if we focused our attention on rewarding merit. The fact that false negatives are concentrated on a particular group (whether women in the medical school hypothetical or minority applicants in the real world of GATB screening) provides an independent cause for concern.

Even an individual applicant in a protected group could not found his claim on the standard title VII assertion that he must be hired unless an employer has a business reason for not hiring him,¹⁹⁹ because the employer might well have such a reason. Assume the employer uses the last test I described, in which 20% of blacks (whom we will now relabel the "test anxious") pass compared to 75% of whites (for now, the "test lovers"), and that 70% of persons who fail and 80% of those who pass the test succeed on the job. Of the employees hired, 78.6% will succeed and 21.4% will fail. If the employer instead hires 196 employees randomly, a slightly lower percentage (75.6%) will succeed and, of course, a slightly higher proportion (24.4%) will fail. It is privately rational to screen if the test costs little to administer, because it is more expensive for the employer to have 3% more of her work force fail than to use the tests.²⁰⁰

If ex post screening is costlier to utilize than the ex ante screening test, the applicant cannot claim that the employer has rejected him without any justification. Instead, the applicant must claim that he has some sort of affirmative right to meritocratic treatment and that the employer has a duty to fulfill this right, even at some cost. There are two problems with the applicant's claim. First, how might one define "merit" in such a way that the employer's refusal to hire in this case would breach the employee's right to be treated in accord with merit? Second, why should this particular employer bear a potentially costly affirmative duty to treat the applicant in accord with "merit" as we define it?

(a) *Equal Opportunity or Meritocratic Claims.* — It is relatively easy to dispose of the assertion that an applicant receives an equal opportunity simply by virtue of having the chance to compete on a test that is equally predictive for himself and others. Tests that bear *no* relationship to performance would still be equally valid for everyone: they would have a universal validity of zero. Nonetheless, it would make little sense to say that people had an equal opportunity to get particular jobs just because they had been given a test that did not *underestimate* their capacity to succeed on the job. Although this scenario is extreme, it illustrates a valuable point. Because tests "objectively" differentiate those who take them, there is a strong tendency to believe, wrongly, that doing better on tests *itself* gives rise to a claim of merit.²⁰¹ Richard Fallon may be right when he

¹⁹⁹ See *supra* note 23.

²⁰⁰ If the "test anxious" were a protected group, like blacks or women, the employer could not use the test if she could eliminate the bad workers nearly as cheaply by using an alternative system that had less disparate impact. See *supra* pp. 1174-75.

²⁰¹ See generally Crouse & Trusheim, *The Case Against the SAT*, 93 PUB. INTEREST 97, 102 (1988) (noting that the Educational Testing Service has attempted at times to justify the Scholastic Aptitude Test without regard to its predictive validity, presumably relying on the notion

argues that the use of tests, no matter how irrelevant to success, precludes the use of subjective, discretionary criteria and thus eliminates both the possibility for and the appearance of a process beset by personal animus or bigotry.²⁰² But the personal animus of decisionmakers is just one barrier to merit-based hiring; its absence is only a necessary, not a sufficient, condition for ensuring merit-based hiring.

It is more difficult to argue that a person whose ex ante probability of success is in fact lower than another's is treated unfairly if she is not hired. The low-scoring applicant must claim the following: each applicant is, ex post, either a good worker or a bad worker. If the applicant is in fact a good worker, indistinguishable in the relevant sort of merit (the ability to do the job, as opposed to the ability to take tests) from high-test-score good workers, it seems unfair that her chances of being hired are relatively low, compared to the members of the group of high-test-score workers. If she is in fact a poor worker, indistinguishable in actual merit from a high-test-score poor worker, her chances of being hired are also dramatically poorer. Thus, due to her performance on the test, no matter what her level of merit, she will be treated worse than some of her equals. It is not clear how one could consider a system meritocratic when people of equal factual (ex post) merit systematically end up in different places.

The problem with this argument is that in *any* ex ante probabilistic screening system, the existence of false positives and negatives ensures that factual equals will not be treated equally. Thus, if one conceives the above argument as the basis for an absolute meritocratic claim, only ex post retention or firing would be justified.

Alternatively, the above observation about ex ante screening could be seen as providing an additional, vital perspective on the meaning of equal opportunity, a perspective which should be balanced prudentially against the cost of reliance on ex post screening. The refusal to allow ex ante screening is less justifiable if failure rates on the job differ dramatically for those who pass and fail exams. Although it would inevitably still be true that someone who tended to fail exams but who was in fact a good worker would have a lower chance of being hired, the cost of ignoring the test and using an ex post screening system (such as firing) would rise as the number of those hired who would fail on the job increases.

Good performance on the test itself, however, could "balance" against the ex post meritocratic claim only if good test performance were itself meritorious. Good performance on tests that screen more

that test achievement was independently meritorious). The more sophisticated position assumes that valued achievement will result from possession of the trait ostensibly measured by the exam. *See id.* at 98, 104.

²⁰² *See* Fallon, *supra* note 19, at 838-39.

cheaply than ex post firing could be deemed "meritorious" only if we believe that being able to demonstrate one's skill to others cheaply is a virtue. If one defines a worker's "merit" as his capacity to increase the firm's profitability, those who test poorly but work well are not in fact as meritorious as those who test and work well, because it is impossible to hire those who "predict" poorly yet work well without simultaneously instituting what we hypothesize to be a profit-reducing system of ex post retention or firing. The situation bears some similarity to the cases I discussed in Part II in which one applicant produced the same output as another, but his net value added to the firm was lower because his hiring forced the firm to bear increased real input costs.²⁰³ As in those cases, whatever intuition we may have that the increased costs associated with hiring those who "predict" poorly but perform well do not reflect adversely on the worker's merit is best explained by the imagery of a contest or race. The use of the ex post screening system does not boost the bad test taker's productive capacity; it simply gives her a chance to compete.

Less imagistic arguments also support this intuition. When we discuss equal opportunity and meritocracy generally, we may take it for granted that someone must bear different levels of background social costs (for example, compensatory education expenses) in order to achieve equal opportunity.²⁰⁴ We seldom consider it adequate to answer claims that a particular person has not had a fair opportunity to reach his potential by noting society will have to incur extra costs to allow this individual to reach his potential. We generally think of such costs as raising distinct utilitarian concerns; the pursuit of equal opportunity is not the sole social good.²⁰⁵ Nevertheless, if two people would be equally productive workers if each were trained, but training

²⁰³ See *supra* pp. 1198-1204. Note, too, that although they do not always seem to, courts should explicitly recognize that to the extent that race or gender itself is an inexpensive proxy for discerning desirable employment traits, any requirement that employers use individualized tests or validate screening devices will in fact increase costs. Mainstream economists seem to take it for granted that "statistical discrimination" may be privately rational and that the use of more exacting screening is costly. See, e.g., Phelps, *supra* note 28, at 659. Still, it is quite clear that title VII bars such statistical discrimination. See, e.g., *Weeks v. Southern Bell Tel. & Tel. Co.*, 408 F.2d 228, 235 n.5 (5th Cir. 1969) (considering the possibility of allowing employers to use statistical rather than individualized screening devices only when "it is impossible or highly impractical to deal with women on an individualized basis"). This underscores the fact that title VII is *more* than an employer-protective piece of paternalist legislation that attempts to help non-profit-maximizers see profit opportunities that they miss due to racism or sexism.

²⁰⁴ See generally Jencks, *supra* note 51 (arguing that the meaning of equal educational opportunity depends on one's view of the ideal society, which in turn affects the distribution of resources required to achieve "equality").

²⁰⁵ See generally Fallon, *supra* note 19, at 817-19, 823-25, 831-34 (describing why a purely merit-based system might be compromised to account for the unequal distribution of certain attributes among racial or ethnic groups).

costs more for one person, *P*, and *P* is morally blameless in creating the need for higher training costs, it would be troublesome to argue that *P* had an equal chance despite the absence of training because his potential product, net of training costs, was lower than the person with whom he was competing. Although we ought not dismiss concerns about training costs, they are fundamentally utilitarian concerns, *not* deontological, rights-based concerns that bear on *P*'s claim to equal treatment.

Assuming that the added costs of using an ex post firing system rather than an ex ante test are exactly the same as the added costs of training *P* compared to a more readily-trained worker, a hypothetical poor test taker but good worker *X* would seem to have a substantially *stronger* claim than the costlier-to-train *P*. This is because we adhere to the view that meritocrats may take people as they are, not as they might have been.²⁰⁶ Arguably, *P* might have been as good a worker as others, but he simply is not. *X*, however, *is*, by hypothesis, as good a worker as others. Society simply must bear higher transaction costs to *discover* that fact.

It might well be the case that at the societal level, although not the firm level, there are utilitarian, efficiency arguments for ensuring that capable people who otherwise would be screened out get a chance to "show their stuff."²⁰⁷ First, one can blur utilitarian and deontological arguments through a focus on the demoralization costs of unfair practices, as Frank Michelman did in his celebrated defense of the compensation requirement in takings law.²⁰⁸ It is obviously a speculative question whether people experience either great disutility when they or others are deprived of opportunity, or tremendous gains in utility when someone not expected to be capable turns out to do the job. It is clear, however, that private firms cannot capture all such utility gains or losses, although they may capture some if employee morale improves when "outsiders" succeed.²⁰⁹

Second, although a particular employer might find it irrational to screen ex post, such screening might encourage socially rational human capital formation among low-test-score workers who otherwise would

²⁰⁶ See *id.* at 835-37.

²⁰⁷ These "efficiency" arguments are merely suggestive. Whether the abolition of testing is actually efficient in the ways described is probably an unanswerable empirical question.

²⁰⁸ See Michelman, *Property, Utility, and Fairness*, 80 HARV. L. REV. 1165 (1967).

²⁰⁹ Popular culture reflects our attraction to the individual without credentials who gets the chance to do the job and wow the crowds. We seem to revel in seeing movies in which understudies get their break when the star goes down, see, e.g., *42nd Street* (Warner 1933); *Funny Girl* (Columbia 1968), people without MBA's start making the big business decisions, see, e.g., *Trading Places* (Paramount 1983); *Working Girl* (Twentieth Century Fox 1988), or low-rent, seemingly second-rate athletes get to compete and win in the big time, see, e.g., *Breaking Away* (Twentieth Century Fox 1979); *Major League* (Paramount 1989); *Rocky* (United Artists 1976).

have little incentive to acquire skills.²¹⁰ Applicants will engage in the optimal amount of training only if they will receive rewards for self-development. At the social level, the added output from self-development could exceed the sum of added screening costs and self-development training costs.

Particular employers, however, would not rationally bear the additional costs of ex post screening unless they could overcome free rider and coordination problems by ensuring that nearly all employers discontinued ex ante testing.²¹¹ Let us assume, to simplify, that self-development training costs five and increases a low-test-score worker's output by fifty, equalizing it with the output of high-test-score workers. We should further assume that it costs the firm twenty to observe this output rise for low-test-score workers. The firm will not abandon testing unless it can be assured that low-test-score workers will have done the self-development training, but low-test-score workers will not do the training unless employers generally guarantee that they will not test. Thus, if firms generally screen ex ante, the human capital investments, although socially rational, will not be made, and the individual firm's decision to continue ex ante screening will remain privately rational.

The main problem for the meritocrat in abolishing screening tests may well be more pragmatic than principled. If such tests were abolished, there would be pressure to loosen controls on discharge, not as a *logical* corollary of screening test abolition, but as a practical response to the needs of the employer. Because the employer must hire more randomly, the argument goes, she must be allowed to fire, after short probationary periods, as costlessly as possible. To the degree that the meritocrat distrusts the discretion inherent in ex post screening systems,²¹² he may despair that actual merit will not be rewarded if we expand the domain of such systems.

²¹⁰ See, e.g., Lundberg & Startz, *Private Discrimination and Social Intervention in Competitive Labor Markets*, 73 AM. ECON. REV. 340, 342-46 (1983); see also Schwab, *Is Statistical Discrimination Efficient?*, 76 AM. ECON. REV. 228 (1986) (presenting a different model of the inefficiency of statistical discrimination that focuses on the inefficient allocation of workers between jobs in which they get to manifest their individual skills, such as self-employment, and the jobs in discriminatory markets in which workers are compensated according to the average productivity of people exhibiting their discernable traits).

²¹¹ Any one employer's decision to abolish testing will have little impact on would-be low-test-score applicants' decisions to improve performance, so each will hope the signal comes from other employers instead. See generally M. OLSON, *THE LOGIC OF COLLECTIVE ACTION* 9-16 (1965) (discussing the competitive norms that inhibit individuals from assuming the cost of a public good).

²¹² A significant theme in the minority critique of the supposed critical legal studies hostility to rights is the belief that in any informal system, outsiders (non-whites, gays and lesbians, etc.) will be victimized. See, e.g., Delgado, Dunn, Brown, Lee & Hubbert, *Fairness and Formality: Minimizing the Risk of Prejudice in Alternative Dispute Resolution*, 1985 WIS. L. REV. 1359, 1387-89; Williams, *Alchemical Notes: Reconstructing Ideals from Deconstructed Rights*, 22

Although this dilemma should be troubling to conservative individualist meritocrats, it will be even worse for political leftists who have long been committed both to abolishing tests and to regulating more strictly the employer's decision in firing.²¹³ Some on the left may be unwilling to sacrifice just cause requirements for discharge as the price of eliminating testing, particularly if they fear that undisciplined firing decisions may produce the same racially unbalanced work force that would have resulted from testing in the first place.²¹⁴

(b) *The Scope of Employer Obligations.* — Assume, for the sake of argument, that I have demonstrated that the sorts of "fair" but minimally predictive tests described at the beginning of section B deprive the poor test taker of the same opportunities as others similarly situated. It is still a problematic ethical (and ultimately legal) question whether anyone has violated that applicant's rights, and whether anyone has (or should have) a duty to remedy the situation.

The hypothetical duty of the employer can be described in negative terms as a duty not to discriminate, or in affirmative terms as a duty to establish an ex post screening system that we have assumed is more costly than ex ante testing. Whichever description is more apt, one should note that, in this legal culture, legal duties may be correlative to meritocratic claims, although these claims have been most powerfully made by members of protected groups rather than by individuals facing no systematic pattern of oppression. In *Jones v. Alfred Mayer Co.*,²¹⁵ for example, Justice Stewart, arguing that the thirteenth amendment empowered Congress to bar private discrimination in housing sales, wrote:

So long as a Negro citizen who wants to buy or rent a home can be turned away simply because he is not white, he cannot be said to enjoy "the same right . . . as is enjoyed by white citizens . . ." ²¹⁶

HARV. C.R.-C.L. L. REV. 401, 406-08 (1987). Of course, this position has long been taken by some white male critical legal studies scholars as well. Compare, e.g., Abel, *The Contradictions of Informal Justice*, in 1 THE POLITICS OF INFORMAL JUSTICE 267, 295-301 (R. Abel ed. 1982) (expressing fear of the exploitation of the disempowered in informal settings) with Simon, *Legal Informality and Redistributive Politics*, 19 CLEARINGHOUSE REV. 384, 384-88 (1985) (responding that anti-informalist arguments may be flawed in failing to account for the breakdown of solidarity and care in formalist regimes).

²¹³ See, e.g., Cornell, *Dialogic Reciprocity and the Critique of Employment at Will*, 10 CARDOZO L. REV. 1575 (1989); Minda & Raab, *Time for an Unjust Dismissal Statute in New York*, 54 BROOKLYN L. REV. 1137, 1161-70, 1188-1214 (1989); Note, *Protecting At Will Employees Against Wrongful Discharge*, 93 HARV. L. REV. 1816, 1836-44 (1980).

²¹⁴ Given that supervisors rate workers in a racially biased fashion in performing job evaluations, see *supra* note 149, it is likely that their recommendations as to which workers should be discharged will be biased as well.

²¹⁵ 392 U.S. 409 (1968).

²¹⁶ *Id.* at 421 (quoting 42 U.S.C. § 1982 (1988) (emphasis added by the Court)). The Court continued:

The claim that a particular social practice denies equal rights can take a number of forms in the liberal legal tradition. In the more libertarian vein of Lockean liberalism, a person has "equal rights" to purchase housing as long as the state does not erect barriers for that person that others do not face. In other words, as long as everyone has an equal legal *privilege* to purchase a house, everyone has equal rights. In contrast, in the substantively egalitarian tradition, persons have equal rights only when outcomes are equal: blacks must end up with the same level of housing amenities or their "rights" to housing have been violated.²¹⁷

Justice Stewart's position is different from both of these. He does not demand that the state distribute housing equally, yet he is not willing to tolerate blacks' lack of access to housing whenever the state does not mandate segregation or increase burdens on the minority population. Rather, in Justice Stewart's view, substantive results may differ among individuals, but only if the differences reflect the purchaser's buying power. The particular seller's freedom of disposition is limited by the buyer's claim that he should be allowed to purchase as much as a white buyer willing to pay the same amount. In other words, the seller, a particular private person, owes a duty to the buyer. Whether that duty is categorized negatively (a duty not to discriminate) or affirmatively (a duty to sell to the black buyer when, by hypothesis, the seller is sacrificing utility that he might derive by selling to his most preferred buyer) is, for now, less central than the fact that the duty cannot be adequately explained without regard to the black purchaser's meritocratic claims.

The ramifications of Justice Stewart's position for our study are best illustrated by an examination of the effects of discrimination in the product market in tandem with the labor market. The devaluation of dollars held by blacks will effectively erode the real wages that black employees receive in the labor market. For example, imagine two workers, one black, one white, with identical productive skills. As long as the black earner's income is devalued by his inability to purchase identical goods for the same price as the white worker, merely eliminating discrimination in labor markets (so that the two workers receive the same pay) will not adequately ensure that each is treated in accordance with productivity, the relevant "merit" that the labor market purportedly measures. Although each purchaser is

Negro citizens . . . who saw in the Thirteenth Amendment a promise of freedom . . . would be left with a "mere paper guarantee" if Congress were powerless to assure that a dollar in the hands of a Negro will purchase the same thing as a dollar in the hands of a white man.

Id. at 443 (quoting CONG. GLOBE, 39th Cong., 1st Sess. 1151 (1866) (statement of Rep. Thayer)).

²¹⁷ For the classic discussions of "negative" and "positive" liberty, see I. BERLIN, *Two Concepts of Liberty*, in *FOUR ESSAYS ON LIBERTY* 120, 131-34 (1969).

equally privileged to negotiate, if the state fails to establish a duty to sell to buyers offering an equal price, and some sellers would exercise their legal privilege to choose white buyers, a black buyer does not have "equal opportunity" to get the same goods as his equally skilled white counterpart. Productivity is the critical factor that ought to determine social position; private discrimination prevents that result.

It is surely easier to justify imposing a duty to ensure that buying power corresponds to merit in the *Jones* case, however, than in our job screening cases. When sellers turn down equal dollar bids from black offerors we can initially explain their refusal to accept those bids only in terms of either their own racial animus or their desire to pander to the racist sentiments of neighbors.²¹⁸ Even if one disregards the buyers' meritocratic claims, one can readily imagine prohibiting the sellers' behavior based on its connection with the history of racist mistreatment. Whether one focuses on the seller's actual motivation²¹⁹ or on the unambiguously bigoted cultural meaning of the refusal to deal,²²⁰ the seller imposes a stigmatic injury distinct from the ultimate material devaluation of the black buyer's skills. Unlike the house seller, the employer could be motivated simply by a desire to cut costs. Because such ends are not unambiguously racist, it is unclear whether the employer's practice would be generally perceived as biased.

From the individual seller's perspective, however, the distinctions may be slimmer than I have implied. The seller who refuses to accept equal dollar offers from blacks may not be acting out of personal racial animus, but could be exploiting the general existence of discrimination by holding out for a higher economic return. Nonetheless, even in that case, the individual seller's economic motive is grounded in *someone's* explicit racism. In contrast, employers *as a group* may by hypothesis simply want to cut screening costs. In this view, Justice Stewart's ban on housing discrimination does not so much vindicate a black buyer's right to equal opportunity as it protects her against a particular "tort-like" offense²²¹ — exposure to undeniable racism. The fact that it is difficult to characterize the putative "tortfeasor's" duty as either negative (refraining from acts of racism) or affirmative (for-

²¹⁸ Thus, in Professor Brest's terms, the decisions are impermissibly race-dependent; decisions would have been different but for the race of the offeror. See Brest, *supra* note 108, at 6.

²¹⁹ See *id.*; Ely, *Legislative and Administrative Motivation in Constitutional Law*, 79 YALE L.J. 1205 (1970); Simon, *Racially Prejudiced Governmental Actions: A Motivation Theory of the Constitutional Ban Against Racial Discrimination*, 15 SAN DIEGO L. REV. 1041 (1978).

²²⁰ See Lawrence, *supra* note 26, at 355-78.

²²¹ Cf. Delgado, *Words that Wound: A Tort Action for Racial Insults, Epithets, and Name-Calling*, 17 HARV. C.R.-C.L. L. REV. 133 (1982) (developing a tort action for racial insults). For a trenchant development of the notion that expressing racism is fundamentally an assault similar to actionable inflictions of emotional distress, see Austin, *Employer Abuse, Worker Resistance, and the Tort of Intentional Infliction of Emotional Distress*, 41 STAN. L. REV. 1, 12-15, 53-56 (1988).

going the perceived gains from racism) is of little moment. The distinction between the seller and employer cases I have posited is that someone, whether the "tortfeasor" or not, is necessarily racist only in the seller case.

Although these distinctions are powerful, they may not ultimately outweigh the similarities. The primary harm to the rejected black buyer may be dominantly dignitary — the harm lies in the feeling of racial rejection, not in the breach of an affirmative right to equal opportunity. But consider the following: a seller in a post-*Jones* world tells a prospective black purchaser, "I sell to your kind only because the law makes me. Everyone around here will hate me for it." The seller's comments may well be deemed non-actionable only because they are constitutionally protected speech.²²² The important point, however, is that the buyer would *still* have a meritocratic interest to vindicate by demanding the seller accept his equal bids, even if he could not be protected against harm from racist speech.

Some might argue that even the buyer's remaining interest is essentially dignitary, not meritocratic: if the state affirms his right to buy, the unseen others who enacted the protective legislation are thereby refusing to endorse the racist statement of the particular seller. Thus, the state may view itself as compensating for the racist outrage, which nonetheless remains the fundamental interest to be vindicated, by a rule forbidding *acting upon* racist impulses. I find this account inadequate. Were the state interested only in demonstrating official disapproval of particular acts of racism, rather than in protecting meritocratic claims, it would likely evidence its objections through antiracist public education programs, proclamations, and the like.

Assume now what is concededly a problematic proposition — that the buyer has a right not simply to be free from the consequences of racial animus but to be ensured the sort of equal opportunity I believe Justice Stewart ultimately sought to protect. One might still question why, in the housing seller's case, the correlative affirmative duty is lodged in the individual seller rather than in the community more generally, and whether, in the employment context, that duty can legitimately be lodged in the particular employer in the form of a prohibition of what may be privately rational job screening methods.

In the housing context, it is quite clear that reluctant individual sellers must bear the costs of providing more equal housing opportunity. In fact, to the extent that unequal opportunity is a function of

²²² Several authors have considered the degree to which overtly racist speech should be protected. See, e.g., Au, *Freedom from Fear*, 15 LINCOLN L. REV. 45 (1984); Jones, *Article 4 of the International Convention on the Elimination of All Forms of Racial Discrimination and the First Amendment*, 23 HOW. L.J. 429 (1980); Kretzmer, *Freedom of Speech and Racism*, 8 CARDOZO L. REV. 445 (1987); Matsuda, *Public Response to Racist Speech: Considering the Victim's Story*, 87 MICH. L. REV. 2320 (1989).

decisions not readily attributed to particular individuals, it will *not* be collectively corrected at all. For instance, assume that the market value of housing in residentially segregated white areas rises above the market value in both integrated and all-black areas. Black property holders in those areas will see the value of their assets decrease relative to whites who initially purchased housing for the same price. This seems to breach equal opportunity claims: equally skilled workers with parallel consumption patterns end up holding assets of different value solely on the basis of race. However, the equal opportunity claim cannot be vindicated because we can identify no particular dutyholder who might have prevented the unequal result (as our earlier hypothetical seller could remedy the problem of devaluing black income by accepting equal dollar offers from black purchasers). We do not undertake the collective alternative of levying a general, community-wide tax to compensate the black homeowner for the loss of wealth that may result either from his integrating a neighborhood or his remaining in one that becomes segregated by white flight.

Even if it is sensible in the housing context to place the duty to vindicate meritocratic claims on particular individuals rather than on the community, though, it is not as clearly defensible in the job-selection context. The cost that the racist seller bears due to state regulation of his selling practices is arguably distinguishable from the cost that would be forced upon the employer by a prohibition of job testing. The cost to the seller is the "psychic cost" of being forced to refrain from acting on his racist desires, not the unavoidable "resource cost" imposed on an employer by a more costly *ex post* screening system.²²³

More significantly, the social legitimacy of the seller's "cost" may be low compared to the employer's. Obviously, a strictly economic, preference-satisfying utilitarian cannot distinguish between the displeasure experienced by a person forced to act against his racist tastes and the displeasure experienced by a consumer who loses access to the greater quantity of gadgets that might be produced if the gadget factory used the cheaper screening method (or the displeasure experienced by an employer who loses access to the higher profits she would earn with the cheaper screening method). Less strictly utilitarian regimes, however, usually discount "illegitimate" pains and pleasures.²²⁴ In a sense, it may seem more reasonable to ask that an

²²³ Some might differentiate the costs because one is psychic and the other is "real," although whether much can rest on this distinction, given the theoretical possibility of monetizing any subjective experiences, is problematic in ways I need not detail here. Several authors discuss whether it is desirable to put a money value on all experiences. See, e.g., S. KELMAN, *WHAT PRICE INCENTIVES?* 54-77 (1981); Abel, *Torts*, in *THE POLITICS OF LAW* 185, 195-96 (D. Kairys ed. 1982); Tribe, *Policy Science: Analysis or Ideology?*, 2 *PHIL. & PUB. AFF.* 66, 84-94 (1972).

²²⁴ The legal process school's founding fathers, Henry Hart and Albert Sacks, may have

individual bear a cost to exercise preferences that seem quirky, particular, *and* immoral. It may seem more fair to expect the community generally to share a burden that is both morally unproblematic and broad-based, a category that might seem to describe the ex post sorting schemes.

This argument is somewhat appealing, but there are two reasonably satisfactory responses to it. First, it is not clear how the collective could reasonably ascertain and finance the added costs of ex post screening. Nor is it clear that the cost of eliminating ex ante screening actually falls disproportionately on the employer (or the collective owners of the employer company) rather than on the community of consumers more generally. Second, and more importantly, one cannot establish a system that identifies and compensates particular people who are in fact victimized in the hiring process by the use of a less legitimate, ex ante screening system. By hypothesis, such people simply cannot be identified by any ex ante screening device. Any system that compensated *all* those who failed the tests would not remedy the relevant breach — the failure to ensure that all those with equal talent wind up in the same place — because it would either overcompensate those who would have in fact failed on the job or undercompensate those who would have succeeded. On the other hand, it is possible to administer, through the employer, a simple negative rule that meets our concerns: “Don’t use a test with the properties we describe.”

In summary, the Court’s discussion in *Jones* can best be interpreted as protecting meritocracy by preventing the earnings of blacks from being devalued by the racist practices of sellers in the product market. Although it is plausible to interpret *Jones* simply as a bar to the dignitary affront inherent in sellers’ failure to deal on equal terms, I think that account cannot explain why *Jones* regulates sellers’ behavior, rather than their speech. To the extent that the *Jones* result simply protects blacks against expressions of direct or indirect animus, it implies little for the job screening cases on which we have focused, because our employers, while breaching meritocratic demands, show no animus when they are simply minimizing costs. If, however, one believes that the *Jones* Court sought to guarantee that blacks who are equally “meritorious” as whites would live as “well” as whites (in the sense of enjoying equal access to goods), then *Jones* implies that capable black workers must be given a chance to show their merit. Finally, to allow both the black buyer and the low-scoring but high-

differentiated between the satisfaction of legitimate wants, which they dubbed “valid human wants,” and illegitimate ones. See H. HART & A. SACKS, *THE LEGAL PROCESS* 113 (tent. ed. 1958). But their account of this distinction is elusive, at best. For more sophisticated discussions of whether consequentialist utilitarians should account for arguably illegitimate pleasures, see Adams, *Motive Utilitarianism*, 73 *J. PHIL.* 467 (1976); and Sen, *Rights and Agency*, 11 *PHIL. & PUB. AFF.* 3, 7–12 (1982).

capacity black job applicant to vindicate their meritocratic claims, the state must establish correlative duties for individual sellers and employers, respectively. In neither case can the breach of right be readily remedied in any more collectively funded manner.

3. *Individualistic Arguments Against Using Low-Validity Tests That Have Disparate Impact on Identifiable Social Groups.* — Because in reality those who fail to get the same economic rewards as others with equal job performance capacity often constitute a socially identifiable group (women or blacks) rather than a “group” that one can identify only by reference to members’ specific disadvantaging factor (poor test taking), the use of the sort of facially “fair” but inaccurate test we have been considering may seem problematic. Of course, the *Jones* decision arguably protects the meritocratic claims of black citizens specifically, and it is worth exploring why that might be the case.

People frequently refer to legal claims deriving from one’s status position as group rights claims.²²⁵ Whether they use the term in a Hohfeldian “legal” sense²²⁶ is often ambiguous. The wholly legal quality that a group right would presumably possess is that it could be vindicated only in a class action suit in which the chosen remedy would bind the entire plaintiff class. The individual would have no particular vindicable claim simply as a member of the group. Presumably, too, his interests would be subordinate to collective interests.²²⁷ Thus, to apply this principle to the employment case, no individual black applicant could be guaranteed meritocratic fair treatment if we decided that blacks as a group would be better served by a program that happened to deny the individual meritocratic fairness.

Often, however, people use the term “group rights” to refer to individually vindicable claims emanating from group membership.²²⁸ The claim to better treatment of blacks need not be based on the idea

²²⁵ Both the most virulent conservative opponents of accounting for the racial identity of a test taker and its strongest progressive proponents make this characterization. Compare, e.g., Blits & Gottfredson, *Equality at Last, or Lasting Inequality? Race-Norming in Employment Testing*, 27 SOCIETY 4, 9–11 (1990) (opposing group-based remedies) with Lynd, *Communal Rights*, 62 TEX. L. REV. 1417, 1423–30 (1984) (advocating group-based rights) and Matsuda, *Looking to the Bottom*, *supra* note 66, at 373–88 (same).

²²⁶ See Hohfeld, *Some Fundamental Legal Conceptions as Applied in Judicial Reasoning*, 23 YALE L.J. 16, 20–25 (1913) (noting that all rights claims should be understood in terms of the correlative duties they imply).

²²⁷ This fact, of course, leads to the frequently discussed problem of intra-class conflicts in situations in which individualized remedies are difficult. See, e.g., Rhode, *Class Conflicts in Class Actions*, 34 STAN. L. REV. 1183 (1982).

²²⁸ In the title VII context, an applicant cannot challenge an invalid employment selection criterion unless it discriminates on the basis of a protected trait. Similarly, even if it is irrational to do so, landlords may “discriminate” by refusing to rent to lawyers, because lawyers are not a protected group. See, e.g., *Kramarsky v. Stahl Management*, N.Y.L.J., Nov. 9, 1977, at 11, col. 2 (N.Y. Sup. Ct. Nov. 8, 1977).

that blacks constitute a moral entity, but rather can arise from a belief that concern for group status will ultimately redound to the benefit of particular individuals. There is little doubt that, when people identify themselves as members of discrete groups, the success of other members of the group affects them.²²⁹ Whether one tends to emphasize the more immediate self-esteem effects on individuals or the more long-term benefits of successful role models is not pertinent to the ultimate argument. As long as people are more affected by the achievements of people in their group than by the achievement of people in general, the relative status of each group, as a whole, has significant consequences in the lives of particular people.²³⁰

Racial subgrouping may be particularly appropriate in the American context. Given the enormous historical disparities in the life circumstances of blacks and whites both during and after slavery,²³¹ social practices that disadvantage blacks as a group will inexorably both confirm blacks' sense of exclusion from the larger community and create or reinforce self-doubt.²³² A message of inferiority is inevitably reinforced because one aspect of blacks' cultural subjugation by the white majority has been the attempt by whites to suggest that black failure is both natural and inevitable.

The reinforcement of self-doubt may be particularly powerful when tests of the form we are describing are used alongside explicit affirmative action programs. For instance, employers might systematically and openly add points to black applicants' test scores or hire a higher proportion of black applicants with low test scores. This will clearly create the appearance that blacks are receiving preferential treatment. However, it is still readily conceivable (again, depending on the size of intergroup test score disparities and the predictive validity of the test) that both qualified and unqualified blacks as a group would still be employed less frequently than similarly skilled whites.²³³ It becomes difficult to maintain a strong sense of self-worth in the face of relative group failure, especially in light of the seemingly compensatory advantages already employed ("objective" color-blind tests conceded to have no straightforward cultural bias, combined with test-score weighting or preferential hiring). To recognize that in such a system one is *still* less likely to be employed than a similarly capable

²²⁹ See generally Thurow, *supra* note 107, at 31-32 (arguing that an individual's welfare function can legitimately embody preferences about the relative performance of different economic groups).

²³⁰ See *supra* pp. 1196-97.

²³¹ See J. FRANKLIN, *supra* note 24; L. HIGGENBOTHAM, IN THE MATTER OF COLOR: RACE AND THE AMERICAN LEGAL PROCESS (1978).

²³² See, e.g., K. CLARK, DARK GHETTO 63-67, 127-29 (1965); A. KARDINER & L. OVESEY, THE MARK OF OPPRESSION 386-87 (1951).

²³³ See *supra* note 197.

white is again not impossible, but seems to be unlikely. That this account of the illegitimate inequality of tests has not even been noted and rejected, but rather ignored, by all the reviewing courts in title VII and fourteenth amendment cases, suggests that it is not an intuitively obvious response to the social practices of using screening tests with or without "preference" for minority applicants.

The case for the sort of group focus that is hostile to tests seems even more compelling if we believe that test scores are relatively immutable in the short run when compared to job performance. Black children may believe that they have relatively little control over their scores on standardized tests, and hence over their employability. If they are discouraged from acquiring actual skills because they see that their employment prospects will be relatively bleak regardless of their actual skill, an objective disparity in job competence between the races may increase over time.²³⁴

The failure to treat black applicants as a group in accordance with their actual productive capacity, then, is ultimately a far more troublesome practice than the failure to treat the randomly test-anxious individual meritocratically. As individuals, blacks have been and are subject to systematic abuse and hostile mistreatment. There is no reason to believe test-anxious people have experienced such general discrimination. For the group, the economic advancement of significant numbers of black job applicants may help break the cycles of poverty and despair that are the legacy of slavery,²³⁵ physical terrorism,²³⁶ cultural assaults,²³⁷ and enforced ghetto-

²³⁴ Gaps in achievement test scores between blacks and whites do seem to change rather substantially in short periods of time. See, e.g., Jones, *White-Black Achievement Differences: The Narrowing Gap*, 39 AM. PSYCHOLOGIST 1207 (1984). Moreover, the black-white economic gap has diminished over the post-war period, due largely to educational improvements for blacks as well as improved relative returns on education (presumably caused in some part reduced discrimination in employment). See Smith & Welch, *Black Economic Progress After Myrdal*, 27 J. ECON. LITERATURE 519, 521-28, 547-57 (1989).

There appears to be no such shift, however, in relative performance on *g*-loaded tests, of which the GATB is typical. Thus, to the extent that such *g*-loaded tests come to dominate employment decisions, there is simply less reason for blacks to study. As Jones notes:

Career opportunities for black citizens today are free from many of the limitations that were typical for earlier generations. . . . [T]he recognition by black students of many career options for which educational achievement is required may be contributing to achievement gains

Jones, *supra*, at 1212.

²³⁵ See generally E. GENOVESE, ROLL, JORDAN, ROLL: *THE WORLD THE SLAVES MADE* (1974) (discussing the impact of slavery on slaves). For a trenchant general discussion of the ongoing impact of slavery on black life, see D. BELL, *AND WE ARE NOT SAVED* (1987).

²³⁶ See generally J. FRANKLIN, *supra* note 24, at 431-37 (discussing the history of anti-black racial violence in America); H. SHAPIRO, *WHITE VIOLENCE AND BLACK RESPONSE* (1988) (same); Matsuda, *supra* note 222, at 2330 nn. 59-60 (discussing ongoing violence against blacks).

²³⁷ Indigenous culture may at times be expropriated, see, e.g., Matsuda, *Looking to the Bottom*, *supra* note 66, at 336-37, or suppressed, see, e.g., E. CRONON, *BLACK MOSES: THE*

zation.²³⁸ The economic advancement of particular test-anxious persons may not visibly alter the fate of any community, for the test anxious do not form a community. Furthermore, it is highly unlikely that significant numbers of people will derive joy or sorrow from the increased success or failure of the invisibly handicapped anxious test takers. By contrast, blacks are visible both to one another and to non-blacks. As a result, their acceptance and rejection can have social meaning. Finally, incurring costs to ensure that blacks are treated in accord with their static productive capacity may partly compensate for the ways in which society both underspends on developing blacks' latent capacity and continues to breach legal duties to treat blacks in accord with static merit, even when it would be costless to do so.

4. *Deontological vs. Utilitarian Claims: A Last Hurdle for Those Who Would Justify Test Use.* — Let us finally assume, solely for argument's sake, that we have a test that is not only racially unbiased in the Cleary sense, but perfectly valid, so that it could not be deemed to violate anyone's claims to ex post meritocratic treatment. It still is not obvious that those with higher test scores are *entitled* to the jobs that by hypothesis they are better able to perform.²³⁹ We once more confront the conflict between utilitarian and deontological views of entitlements, but this time in two different contexts: worlds with and without racial or gender differentiation.

First, assume total social homogeneity. We may face a choice between maximizing the sum of two applicants' outputs or simply giving to the person best able to perform a task the job she prefers.²⁴⁰ A conflict between these decision rules may arise if the more talented person does the alternative job quite well, while the less talented one does considerably worse on the less desirable job than on the more desirable one, thus making the social utility of the unconventional arrangement higher.

In claiming the job she prefers, the talented person could appeal either to libertarian and contractual principles or to meritocratic principles. Her claim is libertarian to the extent that it relies on the privilege of all parties to make the employment contracts they choose and meritocratic to the extent that it relies on some notion that the more talented party "deserves" her pick of prizes. The libertarian

STORY OF MARCUS GARVEY AND THE UNIVERSAL NEGRO IMPROVEMENT ASSOCIATION 73-77, 100-02, 110-18, 119-31, 142 (1955); D. GARROW, *THE FBI AND MARTIN LUTHER KING, JR.* 208-13 (1981), or treated with disrespect, see, e.g., G. FREDRICKSON, *THE BLACK IMAGE IN THE WHITE MIND* 520-32 (1971).

²³⁸ See, e.g., J. FRANKLIN, *supra* note 24, at 429-31 (discussing enforced racial segregation); Mahoney, *Law and Racial Geography: Public Housing and the Economy in New Orleans*, 42 STAN. L. REV. 1251, 1268-84 (1990) (same).

²³⁹ See Daniels, *Merit and Meritocracy*, 7 PHIL. & PUB. AFF. 206, 218-19 (1978).

²⁴⁰ See *id.* at 210.

argument, however, is incomplete, unless the libertarian absolutely rejects utilitarian state intervention to overcome the transaction costs that prevent employers from negotiating to divide the surplus they would obtain if they assigned employees to achieve the socially efficient, output-maximizing result.²⁴¹ If libertarians reject a particular "interference" simply by defining as "natural" the distribution that would obtain in the absence of that interference, their argument is vacuous and circular, for a distinct "natural" distribution would obtain in the revised legal regime as well.²⁴²

Because of the arbitrariness of the prevailing set of contracts, the libertarian argument will generally be backed up by either a meritocratic moral claim or a utilitarian one. At the theoretical level, however, meritocratic and libertarian views may clash. Assume, for instance, that transaction costs preclude employers from identifying talented applicants from off the beaten track. The libertarian simply treats undiscovered talents as unlucky but claimless. The meritocrat, on the other hand, sees a breach of societal obligation to ensure the identification and reward of merit.

The applicant's meritocratic claim is, of course, subject to the family of disputes concerning whether the applicant deserves (or need deserve) his merit, given that his talents depend on heredity and social advantages which are not themselves merited.²⁴³ But even if one assumes that there is some bedrock element of genuine, fully deserved merit, the basis of the non-utilitarian meritocratic claim is not wholly obvious. At times, meritocratic claims have devolved into horizontal equity injunctions: treat likes alike. By hypothesis, however, the utilitarian might distinguish two equally talented persons (one assigned to her preferred, one to her less preferred job) on the basis of their relative skills at the less-preferred job.²⁴⁴ As usual, horizontal equity will not take us far.²⁴⁵

Alternatively, meritocratic claims might be based on Kantian notions that the failure to reward virtue (or "punish" vice) violates basic

²⁴¹ Of course, it is unimaginable that a state rule would or could explicitly either lower transaction cost barriers, duplicate a market without transaction costs, or attempt to compensate those harmed by the shift from the "free contract" to the regulated equilibrium. Nonetheless, given the prevailing pattern of transaction cost barriers that preclude a different set of contracts from being made, the notion that the particular market distribution should be entitled to any normatively privileged position is problematic.

²⁴² See M. KELMAN, *supra* note 74, at 164-67; Michelman & Kennedy, *Are Property and Contract Efficient?*, 8 HOFSTRA L. REV. 711, 729-35 (1980).

²⁴³ See *supra* p. 1186.

²⁴⁴ See generally Westen, *The Empty Idea of Equality*, 95 HARV. L. REV. 537, 543-50 (1982) (arguing that egalitarian principles are contentless until we recognize independent moral bases of similarity and difference).

²⁴⁵ See *id.*; see also Kaplow, *Horizontal Equity: Measures in Search of a Principle*, 42 NAT'L TAX J. 139, 139-42 (1989) (criticizing the use of horizontal equity-based arguments).

moral precepts, by making us either complicit in vice, inattentive to the good, or violative of rectification or balancing notions.²⁴⁶ Even if one takes these arguments to heart, the nature of the reward that is due the virtuous is hardly obvious. Recognition of an applicant's "superiority" may be quite sufficient from a moral perspective. Again, the point is not to resolve longstanding debates over meritocracy or libertarianism, but merely to remind those who write about the "fair" use of tests that the concepts of fairness they employ are not self-evident and require further examination.

If we decide that utilitarian interests can trump either libertarian or meritocratic claims, we may be particularly moved to regulate tests when we eliminate the assumption of a socially homogeneous world and note that tests adversely affect a particular minority group's achievement. Although it is hypothetically conceivable that we could increase aggregate output by assigning two workers to maximize their joint product rather than by allowing the better worker to choose the job at which she is more productive, it is implausible to assume that any collective body will have the sort of information needed to make such allocative decisions. It is *not* implausible, however, to assume that we can make some utilitarian judgments about the impact of long-term job segregation or discouragement of minority workers. My claim is *not* that utility judgments will be obvious (assuming that the judgments will require trading off some increase in social product for the welfare gains from a particular distribution), but that once made, and if made in such a way that the distributive gains outweigh the product loss, such judgments provide a legitimate basis for making social decisions that affect individual claims. The utility calculations are then the basis of policy choices, and individual entitlements simply reflect particular social policies.²⁴⁷

IV. CONCLUSION

Broad injunctions against discrimination in hiring leave courts and administrative agencies great interpretive latitude, both conceptually and pragmatically. Obviously, employers cannot use screening devices that do not predict on-the-job performance and disproportionately exclude blacks from the workplace. We should be loath to accept, however, even the use of weakly predictive screening devices that may disproportionately exclude capable blacks. I have argued that the use of such weak predictors is almost surely illegitimate if the employer could establish an equally cost-effective *ex post* screening system with

²⁴⁶ See, e.g., M. MOORE, *LAW AND PSYCHIATRY* 233-43 (1984).

²⁴⁷ For a fuller exposition of this classically legal realist view of entitlement, see Kelman, *Taking Takings Seriously: An Essay for Centrists*, 74 *CALIF. L. REV.* 1829, 1833-46 (1986).

less adverse impact. Moreover, even if alternative screening devices were costlier, the pressing social commitment to treat protected group members in accord with their relevant "merit" — their capacity to do the job — suggests that we ought to be extremely wary of low-validity tests.

An employer similarly cannot cite lower productivity as a basis for refusing to hire a member of a protected group if lower productivity results from the employer's failure to protect the applicant from independently illegal working conditions that limit the applicant's effectiveness. It is unclear, however, whether the employer is obliged to make other simple workplace reforms that mute intergroup inequalities.

Employers should not invariably be allowed simply to exclude members of the protected group for not meeting a set of "organizational goals" if those goals are neither obvious nor uncontroversial. The use of unduly narrow goals is especially troublesome if the goals are set without the input of protected group members. Moreover, even employers who claim to be straightforward profit maximizers may be barred from screening out perfectly productive employees who increase firm input costs. Antidiscrimination law may thus embrace a partial theory of distributive justice separate from the simple demand that employers refrain from non-profit-maximizing race or gender consciousness. The theory may require private employers to hire and pay in accord with a worker's gross output capacity and may forbid firms from considering the impact of hiring on input costs, particularly if the firm seeks only to avoid bearing a social cost that will be borne in any event. Whether otherwise profit-maximizing employers must account in their decisions for other distributive justice concerns, such as the failure of markets to recognize adequately the genuine social contributions of protected group members, is an open ethical question that has not yet engaged those with legal authority. I have found no reason to believe that Congress intended title VII to embrace the more radical theories of distributive discrimination nor to suspect that there is anything resembling widespread current political support for such a reinterpretation. More politically radical theorists must not only persuade other political actors that certain distributive policies (such as subsidies or reparations) are appropriate, but must also better articulate reasons for making private firms the medium for the implementation of these public policies.

In the short run, of course, it might seem unlikely that the current Supreme Court — which is apparently committed to narrowing the reach of civil rights law²⁴⁸ — will attempt to invalidate the use of

²⁴⁸ See, e.g., *Patterson v. McLean Credit Union*, 491 U.S. 164 (1989); *Wards Cove Packing Co. v. Atonio*, 490 U.S. 642 (1989); *City of Richmond v. J.A. Croson Co.*, 488 U.S. 469 (1989).

disparate impact hiring devices by expanding the concepts of dynamic or distributive invalidity beyond their rather limited current judicial interpretations. Nevertheless, the use of general aptitude tests should be sharply circumscribed even by a Court firmly wedded to the idea that title VII does no more than protect the capacity of minority-group citizens to get existing jobs that they are already capable of doing. This is true first because, as an empirical matter, the tests themselves generally have low validities, and have even lower validity for minority group members. Second, as a practical matter, ascertaining when such a test really predicts performance on a particular job is very difficult given the small sample sizes in most validity studies. We would expect to see some high validities, simply as a matter of chance, even when a job setting was factually indistinct from others in which validities have proven low. Third, at a more theoretical level, the use of low-validity tests in a setting in which blacks do much worse on tests generates unacceptable levels of false negatives among minority group applicants, preventing far too many capable applicants from getting jobs at which they would actually succeed. Neoconservatives who seem to be seeking a radical *extension* of testing should take note. It is actually more consistent with neo-conservative *principles*, which purport to place a high premium on rewarding the deserving, to ban the use of such general aptitude tests.

Copyright © 1991 by The Harvard Law Review Association. Copyright of Harvard Law Review is the property of Harvard Law Review Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.