

## **Uniformity and Affectivity: A Record of Failure**

Previous studies of the Social Security system, and of the physicians and adjudicators who are integral to it, support my unfolding thesis that uniformity and its companion, affectivity, have never been realized in an institution riddled with substantiated problems. Researchers have identified significant biases in both the procedural and substantive dimensions of disability adjudication, in the rules and in the adjudicators who apply them, and in the system's overall decision-making outcomes.

### **Biases Detected in Social Security's Decision-Making Process**

Chapter 2 explained that physicians provide the foundation for the disability decision-making process: they generate, interpret and explain the medical evidence on which the system relies. Despite the presumed objectivity of the evidence physicians provide, research has shown that both diagnostic and treatment decisions reflect gender and race bias. For example, studies indicate that women are not treated for heart disease as aggressively as are men. A 1991 study demonstrated that doctors referred fewer women than men for cardiac catheterization, despite the fact that women reported more debilitating symptoms resulting from their experience of angina than those reported by men who were referred for the procedure (Steingart et al. 1991). Another 1991 study revealed that women are less likely than men to undergo angiography, angioplasty, or coronary surgery when admitted to the hospital with a diagnosis of heart disease or chest pain (Ayanian and Epstein 1991). A study of black women reveals that regardless of income and region they receive less mammography than

white women, even when there is evidence that they visit their primary care physicians at similar rates (Burns et al. 1996). This finding may explain why black women with breast cancer have a decreased 5-year survival rate when compared to white women.

Other research confirms that African-American patients in general are disadvantaged. A 1993 study demonstrated that white men are more likely than black men to undergo invasive cardiac procedures even when financial incentives are absent (Whittle et al. 1993). These results were confirmed in a 1994 study at the Department of Veteran Affairs, which revealed that rates of cardiac catheterization were significantly higher for white patients when compared to black patients (Mirvis et al. 1994). In addition, white patients were more likely to receive surgery. This study also found that patients with vulvular disease, like coronary artery disease, experienced similar rates of racial bias in treatment. While racial differences varied in part based on region and facility, the researchers concluded that differences in resource utilization persisted even when economic factors were minimized. If, as these studies suggest, treating physicians are less likely to perform diagnostic tests on some groups, white women and African-American applicants are less likely to have corroborating evidence of disability. This problem is further exacerbated by the claimants' ability to pay for tests and related reports documenting their health care problems in disability claims. In a study by Blendon et al. (1995), 32 percent of African-American adults report they had problems paying health-related bills (including bills for physicians, hospitals, and prescription drugs) compared with 17 percent of whites.

In other instances, physicians may erroneously assume that female patients are disabled psychologically when in fact they are physically impaired. Several studies show that based solely on their gender, men and women are assessed differently with regard to psychological illness (Unger and Crawford 1992). A 1991 study of primary-care physicians found that "doctors classified as disturbed a larger proportion of nondisturbed women than nondisturbed men" (Redman et al. 1991, 527). In a 1983 study, physicians who were presented with identical men's and women's case-history vignettes on average judged the female patients as "more emotional" than the male patients (Colameco, Becker, and Simpson 1983).

In 1992, the Ninth Circuit Gender Bias Task Force surveyed claimant representatives, most of whom were attorneys, to determine whether they believed that the gender of the applicant influenced the testimony of med-

ical experts at Social Security disability hearings. Thirty-nine percent of the male representatives and 98 percent of the female representatives believed that the gender of the claimant did have an impact on medical experts (U.S. Court of Appeals [Ninth Circuit] 1992, 99). These findings suggest that as long as the Social Security disability system relies on exclusively medical evidence, the desired impartiality of the process may be corrupted by the biases of the physicians who generate and interpret that evidence.

The use of the Grid is similarly problematic. While the Grid is gender neutral on its face, its application can be gender biased. The Grid advantages individuals with no previous work history or with a strenuous work history (20 C.F.R. 404, subpt. P, app. 2, 201.00 (d)). The former classification favors women who work in the home, while the latter favors men, particularly those who previously held unskilled or semiskilled heavier occupations. The Grid also disadvantages people who, having worked in lighter occupations, are more likely to have transferable skills that would render them not disabled. For example, the Grid disadvantages women who have held traditionally female jobs in the service industry, such as nurses, clerical workers, and teachers, and who therefore would be more likely to have skills that would transfer to sedentary jobs.

The listings of impairments set a standard of severity that must be met for decision makers to presume disability. The impairments listed are described in terms of symptoms, signs, and laboratory findings (20 C.F.R. 404, subpt. P, app. 2, 201.00 (d)). Since the claimant's symptoms tend to be of little determinative value, adjudicators rely heavily on signs and findings, which are often based on results of biomedical research.

Women routinely have been excluded from biomedical research, and, as a consequence, their symptoms and diseases are less likely to be included in the listings than are diseases and symptoms that predominantly affect men. Such is the case although ample evidence indicates that "heart disease, AIDS, depression, and numerous other ostensibly 'gender neutral' conditions are expressed differently in men and women" (Dresser 1992).

In addition, Social Security's listings of impairments exclude two diseases believed by members of the medical community to be predominant among women: multiple chemical sensitivities and chronic fatigue syndrome (CFS). Both these maladies have posed significant problems for ALJs who have been reluctant, and even unwilling, to recognize them as disabling.<sup>1</sup> Only very recently has SSA engaged in a dialogue with the Chronic Fatigue and Immune Dysfunction Syndrome (CFIDS) Association to draft a ruling for the proper evaluation of CFS in disability pro-

ceedings, a step which recognizes the inadequacies of current methods for adjudicating CFS claims and others like them (NOSSCR 1998).

When female manifestations of diseases or entire maladies are excluded from the listings, applicants with these problems are likely to be denied disability benefits, particularly at the initial stages of determination. When a woman's disability is not included in the listings, her claim must then be heard by an ALJ, who considers not only medical evidence but also the applicant's credibility with regard to her pain and other symptoms. The judge is the first evaluator to consider whether the claimant's statements about her ability to function, her pain, and her work limitations are credible. Chapter 4 explores this issue in depth by analyzing how judges' credibility determinations, at least in my small sample of cases, may be affected by normative and otherwise repressed assumptions. In the next section I explore more broadly how Social Security adjudicators and judges perform when making disability decisions.

## Decision-Making Outcomes

### Initial and Reconsideration Stages

Efforts to study the question of disparity between SSA decision-making rules and practices began with government-supported examinations of the initial and reconsideration stages of the process. Funded by the U.S. Department of Health, Education and Welfare (HEW), Saad Nagi's *Disability and Rehabilitation: Legal, Clinical, and Self-Concepts and Measurements* (1969) was one of the earliest studies to examine the issues of objectivity and fairness in SSA disability decision making. Nagi was particularly interested to learn how the disability standard applied by DDS decision makers differed from the standard applied by teams of clinical experts who were not trained in the rules. Nagi explored this issue and HEW funded his project because disability decisions should be based on objective clinical evidence. Hence, Nagi sought to test whether decision makers were actually carrying out the purposes of the Social Security Act governing disability determinations (6).

Thus, focusing on the first two stages of the process, Nagi compared the results to two groups of decision makers who reviewed the identical claims of approximately 2,500 claimants. On the one hand, teams of clinicians—each consisting of a social worker, a physician, a psychologist, an occupational therapist, and a vocational counselor who were not trained

in the rules governing disability decisions—applied clinical expertise to determine eligibility for benefits. On the other hand, teams of adjudicators from state DDS offices who regularly made disability decisions applied Social Security criteria to the same sample of cases.

The results of Nagi's analysis revealed a significant rate of what he called "incongruencies." Reviewing the same medical evidence, clinical teams awarded benefits to more than 55 percent of the claims that DDS decision makers denied. The clinical teams found these applicants to be incapable of work activity outside the home or to be "totally unfit for [any] work," even work that could be done in the home (Nagi 1969, 94). (For purposes of Nagi's study, both these categories rendered people unable to do paid work and were thus equivalent to a DDS decision maker's approval of a claim.) While Social Security rules were supposedly founded on clinical evidence, Nagi's clinical teams thus drew conclusions that differed dramatically from those of DDS decision makers.

While the clinical teams would have awarded benefits to more than half of the claimants that the DDS teams denied, the incongruencies between the two teams were not uniform across the board. For example, the clinical teams would have awarded benefits in 76 percent of the cases involving psychological limitations, but DDS decision makers would have done so in just over 50 percent of such claims (Nagi 1969, 99). When it came to physical impairments, however, the two teams more closely agreed: DDS decision makers allowed benefits to almost 89 percent of the applicants that the clinical teams found to have severe physical limitations (Nagi 1969, 99).

Incongruencies were not Nagi's only important finding. When he looked closely to explain these discrepant findings, he noticed an interesting pattern with regard to gender. The DDS decision makers granted benefits to almost all the white male applicants that the clinical teams found to be disabled (65.2 percent and 68.6 percent, respectively) (Nagi 1969, 111). But the DDS teams granted benefits to far fewer white female applicants than the clinical teams did (46.7 percent versus 66 percent) (Nagi 1969, 111). Nagi attributes this discrepancy to DDS decision makers' preferences for physical impairments, previously described, and to the fact that white women were more likely than white men to allege mental limitations.

Interestingly, DDS decision makers and the clinical teams decided cases involving men and women of color similarly. DDS allowed benefits in 68.8 percent of the cases involving men of color; the clinical teams considered 70.5 percent of these same claimants unfit for work (Nagi 1969,

111). Similarly, DDS allowed benefits in 60.5 percent of cases involving female claimants of color, while the clinical teams granted benefits to 64.5 percent of these claimants (Nagi 1969, 111). Nagi attributes this congruence in clinical and DDS decisions involving people of color to certain factors: the lower educational levels of disability applicants of color, which led to a high allowance rate by DDS evaluators; and the tendency of applicants of color to allege physical rather than mental impairments, which almost always yielded higher award rates (Nagi 1969, 111–13).

Nagi's analysis of occupation-related findings, like his impairment-related findings, yielded some interesting implications related to gender, although he did not draw them out. While his investigation reveals no clear pattern of the effect of occupational history on disability decision making, it does suggest one very important point: unskilled laborers (mostly men) were allowed benefits more often than they were denied; however, clerical and sales workers (mostly women) were denied benefits more often than they were allowed (Nagi 1969, 122).<sup>2</sup> While these results might partially be explained by the vocational aspect of the disability equation (which tends to favor applicants who previously performed heavy as opposed to light work on the basis that their skills are less likely to transfer to lighter, more highly skilled jobs should they become too disabled to perform heavy work), Nagi neglected to analyze the obvious gender implications of these findings.

The examples discussed here indicating disparity between administrative and clinical standards represent some of Nagi's most startling findings. The administrative and legislative response to Nagi's work was to impose more stringent regulatory control over the process and to fund several other studies that investigated similar issues. As for its relevance to my work, Nagi's analysis sets a trend not only in subject matter but also in manner of treatment: Particularly with respect to his findings related to gender but also, to a lesser extent, to kinds of impairment and occupational history, Nagi discovered something that potentially suggests that Social Security decision makers are biased, although he did not develop this thesis. While Nagi and subsequent researchers note that the disparities they detect may suggest inequities in the decision-making process, he and the others do little or nothing to elucidate these factors, their underlying causes, and ultimate effects.

The GAO (1976) also conducted a study, considerably smaller than Nagi's, concerned with the uniformity of the decisions of initial and reconsideration adjudicators. Personnel from 10 DDS state agencies were asked

to adjudicate a sample of 221 actual DI and SSI claims that had previously been decided at the initial and reconsideration levels. Each sample case was adjudicated by DDS examiners from at least two different states. The study showed a significant disagreement among the 10 states on the disposition of the 221 claims. The states completely agreed to approve or deny benefits on only 38 of the 221 claims (approximately 17 percent) (GAO 1976, 6). Even in this narrow margin of agreement on final decisions, the adjudicators often disagreed on the rationale behind their decisions.

In this study, as in others reviewed here, researchers were concerned with consistency in outcomes, not the presence and/or operation of bias or other factors of disparity in the process of decision making. As a result, the GAO interpreted its findings as evidence that the SSA had failed adequately to supervise state-level adjudicators and recommended, as a remedy, strengthening the federal-state administrative structure and enhancing training and information-disseminating activities.

In response to this GAO study, the Subcommittee on Social Security of the Senate Committee on Ways and Means studied the problem of inconsistencies at the initial and reconsideration stages (U.S. Senate 1978). Like the administrative response to Nagi's study, the committee determined that to remedy this problem of inconsistencies, more federal administrative control over DDS decision making was needed. Thus, still more legislation was proposed and eventually enacted to tackle the problem of inconsistency. This new legislation took the form of regulations describing the performance standards, administrative requirements, and procedures to be followed. Though the legislative hearing voiced much concern that decisions were not uniform, it entailed little or no discussion of possible patterns of bias against particular groups or kinds of disabilities. The committee rather focused almost entirely on restructuring the disability administration to strengthen Social Security's federal voice in the state-by-state decision-making process.

In a follow-up study to the 1976 GAO report that documented inconsistencies at the initial and reconsideration levels, HEW examined the consistency of initial disability decisions among and within the states (SSA 1981). As in the Nagi research and the 1976 GAO report, HEW's 1981 report was not as concerned with bias as with whether DDS decisions were inconsistent. The specific purpose of the study was to examine whether different states reviewing the same cases decided them similarly.

Trying to replicate the 1976 GAO study, the SSA's Office of Policy conceptualized the test in very much the same way as the GAO study. A

total of 504 cases were randomly assigned to states, and determinations of DDS adjudicators in different states were compared. This study also revealed disagreement within and between states to be significant. There was roughly a one in eight chance that any adjudicators from any two states would agree on a given claim. There was about a one in six chance that two examiners from one state would agree. That HEW found an even lower rate of consistency in 1981 than GAO found in 1976 was surprising, especially considering the effort that had been made since Nagi's 1969 study to assure that the disability program was based on objective and uniform federal standards.

To explain the enormous variation they had found, HEW statistically analyzed three characteristics of claimants—impairment, age, and education—to determine whether these factors led to the disparities in the state decisions. Were adjudicators in some states, for example, more sympathetic to claims involving physical impairments than mental impairments or to older applicants rather than younger applicants? The HEW did not consider gender and race variables in their analysis.

HEW researchers found that factors they considered did not seem to explain the variations, although they cautioned that their findings may be explained by the small size of their study sample. Ultimately, though, HEW explained the glaring inconsistencies in and among the states that they discovered as follows: (1) a certain amount of disagreement in decisions is to be expected; (2) the results are limited to eight states and cannot be generalized to all states; and (3) the states were in flux given the recent implementation of the Grid. As in all previous instances, HEW investigations concluded that efforts to improve the structure should focus on uniform SSA guidelines and federal instructions for individual adjudicators. They ignored the possible influence of bias in the decision-making process. They also neglected to consider how factors other than age, education, or impairment may have influenced the process.

One of the most significant studies of disparity in the initial decision-making process was published by the GAO in 1994. Researchers were charged with examining gender differences in the 1988 claims of DI applicants. They found that women in older age groups received DI benefits at lower rates than did men. The GAO, explaining two-thirds of the disparity, suggested that there were logical reasons why older women received benefits at lower rates. More specifically, the report opined that women apply for benefits with less severe impairments and that women are more likely to work in occupations with lower allowance rates, regardless of



gender. The GAO's explanations, however, should not go unquestioned. As I argued earlier in this chapter, what may appear to be less severe impairments may in fact be hidden forms of gender bias in medical diagnostic and treatment protocols. When women are offered fewer diagnostic tests or invasive procedures, the medical evidence is much more likely to render a claimant's impairment "less severe." Other forms of structural sexism may also explain why older women qualify less often for disability benefits based on their previous occupations. For example, the Social Security rules are more likely to grant benefits to men who have previously worked in strenuous occupations and to deny benefits to women who have skills that would transfer to sedentary jobs (Mills 1993).<sup>3</sup>

### The ALJ Hearing Process

A related set of inquiries has focused on the ALJ hearing process, the third stage of the Social Security disability adjudication process, and the stage with which my study is concerned. The first of these examinations was conducted in 1973 by Robert Dixon. Dixon examined ALJ reversal-rate statistics; these statistics summarize the percentages of ALJs who granted benefits to claims that had previously been denied. Dixon sought to determine whether fairness, equality of treatment, and correctness were being achieved at ALJ hearings. Dixon's study revealed a high degree of variation among ALJs' reversal rates, not only within a given region but also within individual hearing offices. Dixon's overall findings reveal that some ALJs granted benefits in as few as 8 to 12 percent of the cases they heard, while others granted benefits in as many as 84 to 88 percent.

To explain this disparity, Dixon hypothesized that ALJ decisions rested "significantly on the personality of the examiner as well as the actual record" (1973, 135), yet he did not consider what those personality influences might be. Rather than tackle these issues, Dixon, like researchers before him, instead concluded that more substantive regulations on disability determinations should be promulgated to insure uniform guidance. He went so far as to suggest that his findings were so disturbing that Congress should consider abolishing the ALJ hearing process altogether.

In 1978 Mashaw et al. presented the results of a major study they undertook of the Social Security disability hearing process. Their primary concerns were (1) whether the Social Security disability system had the capacity to produce accurate and consistent decisions efficiently; and (2)

whether the ALJ decision-making process could be redesigned to be both fair and cost-effective. Following an intensive investigation of four ALJ hearing offices that included interviews with judges and claimants; a statistical analysis of a random sample of disability hearing transcripts, evidence, and decisions; and observations of actual hearings, the authors made several important findings. Drawing on Dixon's finding that the variance among judges' awards of benefits ranged between 8 and 88 percent, the authors recognized a significant problem of inaccuracy in the ALJ decision-making process. They argued that this inaccuracy "is truly a product of subjective factors, probably relating primarily to the interpretive role of the ALJ rather than the investigative one" (Mashaw et al. 1978, 21). These conclusions were based on the premise that since cases are randomly assigned to judges and each judge should get a similar mix of cases, it is reasonable to expect that different judges, if impartial, should approve similar numbers of cases. While Mashaw and his colleagues followed previous researchers in recommending better specifications of standards and more training and conferences to improve accuracy and consistency among judges, the authors parted from earlier scholars in conceding that "these efforts will leave a substantial residuum of variance among the ALJ corps" (Mashaw et al. 1978, xxi-xxii).

From reviewing transcripts and observing hearings, Mashaw and his team gathered an important finding that helped to illuminate why they believed that ALJs' decisions may never be fully consistent. These researchers found, and the results from my small sample of cases confirm, that judges rarely gave adequate opening statements and often made insincere inquiries into the claimant's desire to be represented by counsel. Hence, judges frequently render such responsibilities an "empty formality" (Mashaw et al. 1978, 66). In other words, Mashaw's study, published nearly 20 years ago, detected evidence of the same pattern of disregard for the rules meant to ensure impartiality as I found in the hearing transcripts and decisions I reviewed. However, Mashaw et al. did not systematically analyze the hearing process for violations of rules, they did not systematically examine the violations they observed, and they did not attempt to explain the inaccuracies or inconsistencies they described as potential signs of bias. They seem rather to have regarded rule violations as an afterthought and did not consider their possible relation to the "subjective," "interpretive" factors that, as they saw it, rendered the ALJ hearing process variable and inaccurate. Thus, Mashaw and his colleagues con-

cluded that the high levels of inconsistency they found are unavoidable; inconsistency must be tolerated, or aspects of the process must be abandoned altogether.

### Overall Disability Decision-Making Process

Several studies have considered the issue of disparity in the overall decision-making process. In *Bureaucratic Justice* (1983), Jerry Mashaw's analysis further illuminates and attempts to resolve the myriad of dilemmas presented by the enormous Social Security bureaucratic decision-making process, which aspires both to be fair and impartial and to operate with dwindling resources and rising costs. Ultimately, Mashaw develops the conclusion he and his colleagues previously sketched (Mashaw et al. 1978) to recommend abolishing the current ALJ hearing process altogether. With this recommendation, he also suggests that the federal court appeals process and legislative oversight of Social Security rules and decisions likewise be curtailed. To examine a little more closely how he reaches that conclusion, it is useful to summarize his three models of justice, all of which he believes figure prominently in the disability process as it is currently organized.

*Bureaucratic rationality.* Bureaucratic rationality, the bureaucratic model SSA uses at the initial and reconsideration stages to make disability decisions, is concerned with distinguishing between true and false claims at the least possible cost: As a model for adjudicating claims that is both accurate and cost-effective, bureaucratic rationality seeks to "minimize the sum of error and other associated costs" (Mashaw 1983, 25). Mashaw regards bureaucratic rationality ideally as the sole or at least the predominant model to be used for disability decision making. Decisions made by this model, Mashaw argues, often differ with those generated by the other two models, professional treatment and moral judgment, because the basic values reflected in the models clash.

*Professional treatment.* The professional-treatment model brings professionals—physicians, vocational experts, counselors of all kinds—into the decision-making process to examine the needs of the claimant more or less holistically. Thus, according to Mashaw, the goal of the professional-treatment model is mostly to serve. When professionals become involved in the process, disability decision making becomes concerned with income support, medical care, vocational rehabilitation, and counseling—all needed to improve the claimants' well-being and perhaps self-sufficiency, but all at odds with the need to decide the greatest number of

claims as consistently as possible at the lowest possible cost. Thus, since professionals are influenced by facts, intuition, and judgment and not necessarily accuracy and efficiency, the professional-treatment model clashes with the bureaucratic-rationality model and introduces inefficiency into the system.

*Moral judgment.* Mashaw's moral-judgment model refers to the decision-making processes Social Security ALJs use at the hearing level rather than the processes DDS adjudicators use at the initial and reconsideration levels. Moral judgment, according to Mashaw, approaches decision making as a clarification of values. At issue is whether the applicant "deserves" benefits. He further characterizes this model by "its promise of a full and equal opportunity to obtain one's entitlement" (Mashaw 1983, 31). While Mashaw recognizes that all three models are ultimately concerned with attaining the same goal—that is, factually correct decisions—he finds the moral-judgment model to be the least efficient means to this end, because its regulators, ALJs, by the nature of their task, have the greatest "error proneness and inefficiency" (Mashaw 1983, 41).

Mashaw is convinced that a bureaucratic model that allows an organization to govern itself with as little interference as possible from outside regulators is capable of achieving the goal of uniform decision making and even the public perception of fairness. According to Mashaw, outside regulators such as ALJs, federal judges, and legislators place constraints on the Social Security disability decision-making system that undermine the bureaucracy's ultimate goal, which is to distinguish efficiently between true and false claims. Thus, Mashaw believes that Social Security and bureaucracies like it should be permitted to operate on their own through what he calls internal law (that is, written instructions and interpretations, standard bureaucratic routines, and developmental and decisional practices). Toward this end, he expands Dixon's (1973) and his own previous (Mashaw et al. 1978) suggestion and explicitly recommends abolishing the current ALJ and federal court appeals process and curtailing legislative oversight. He would replace these agencies with what he calls a superbureau, an institution that would combine a "judicial chamber" with the functions of other federal agencies such as the GAO and the Justice Department's Office of Legal Counsel. Toward this end, such a superbureau would supervise the entire functioning of the bureaucracy and oversee instances of maladministration (Mashaw 1983, 226).

Mashaw's contention that Social Security's bureaucracy should be trusted to ensure fairness and equity without ALJ, federal court, or leg-

islative oversight is incomprehensible. In my view, and based on all the previous research, more, not less, sensitivity and individualized attention needs to be incorporated into the disability decision-making process. As Justice William Brennan has suggested, such a bureaucracy should not be distant and detached but instead should embrace a passion that puts bureaucrats and line workers in touch with “the dreams and disappointments of those with whom they deal” (1988, 19). Mashaw’s superbureau seems to embody none of these important concerns. Indeed his solution ignores the commitment made by Congress and the courts to provide a scrupulous and conscientious inquiry into each individual claim. Checks and balances provided by the system’s current formulation are required to preserve and protect the juridical process from the stereotyping and indifference Mashaw and other researchers have detected.<sup>4</sup>

In *The Disabled State* (1984), Deborah Stone reveals the presence of an inherent contradiction within the Social Security disability program. Her analysis surpassed most previous efforts to explain disparity in the system and clearly established the need for studies of the possible influence of bias in the decision-making process. Concerned about the expansion of disability programs in the United States and abroad, Stone’s analysis of the U.S. Social Security system exposes the political mechanisms the United States has used to resolve the difficult issues concerning the distribution of benefits to people with disabilities. In the process, she discovers that the goal of objectivity in disability decision making is inherently contradictory with the flexibility of the clinical standards, which rely on physicians’ impressions, findings, and interpretations.

Stone argues persuasively that because clinical criteria cannot draw adequate boundaries between disability and ability to work, they cannot provide the objectivity the U.S. Social Security disability system seeks. She explains this shortcoming as follows:

[C]linical tests can provide a measure of some phenomenon, but no single test can tell whether a person is “impaired” or “disabled” or still able to work. Someone has to decide what level of any measurement is indicative of an inability to function and how the information from different tests ought to be combined. (1984, 128)

Stone contends that even clinical “tests” that are supposedly objective are open to interpretation: Laboratory studies and X-rays show “alarming discrepancies,” patients can manipulate some test results by making only

partial effort, and physicians' interpretations of laboratory findings show "wide discrepancies" (1984, 129, 131). Given that the evidence underlying a disability claim is so easily manipulated and that the rules are not conducive to being objectively applied, Stone is convinced that the accuracy and consistency of decision makers is unlikely to improve with more rules or tougher SSA oversight.

Stone's interpretation of the contradiction embedded in a system that on the one hand strives for objectivity and on the other relies on highly subjective evidence does not directly address the issue of potential bias on the part of Social Security disability decision makers. Rather, like that of Dixon and Mashaw, Stone's analysis suggests that the problem of inaccuracy and imprecision in Social Security disability decision making is irremediable by stricter regulation and oversight and, as such, is probably a foregone conclusion. Stone, however, raises a concern relevant to the current study. Taking her idea just one or two steps further, it is arguable that because the clinical evidence on which the system relies is highly flexible and standards for applying that evidence are imprecisely defined, the system may, in this way, implicitly encourage decision makers to inject their personal views (and hence their biases) into the disability decision-making process. Institutional pressures on ALJs to decide cases quickly, to deny as many as possible, and to act as defense, prosecutor, and judge are three other means by which the system may encourage decision makers to inject personal views into the process. Although I refer to these institutional pressures in passing when analyzing my results, it has not been my primary focus.

Probably the single most important study undertaken so far on the issue of impartiality of Social Security decision makers is the 1992 GAO study of the impact of claimants' race on disability decision making. The principal objective of this research was to analyze the circumstances surrounding the lower proportion of allowances among black applicants relative to white applicants. Using quantitative methods, GAO researchers analyzed issues and decisions at the initial, reconsideration, and ALJ-hearing levels. The researchers reviewed claims filed in 1983, including approximately 700,000 cases involving white applicants and 245,000 cases involving black applicants.

The GAO found that demographic characteristics and impairments could largely explain the difference in allowance rates to black and white DI applicants (claimants with a work history) at the initial and reconsideration levels. Thus, the black allowance rate was lower than the white rate "primarily because Black applicants were concentrated in age groups and

had impairments, such as hypertension, that had low allowance rates regardless of race” (GAO 1992, 4).

However, for SSI applicants (those without a work history) between the ages of 18 and 24, the racial difference was “largely *unexplained* by differences in education, sex, geographic location, percent urban population or impairment type” (5). Thirty-four percent of African Americans who applied were allowed benefits compared with 47 percent of Whites (GAO 1992, 5; emphasis added). The only explanation for the difference was race.

At the ALJ-appeals level, ALJs granted benefits to 55 percent of black DI applicants compared to 66 percent of white DI applicants. They granted benefits to 51 percent of black SSI applicants compared to 60 percent of white SSI applicants. “For the most part, GAO could not explain the racial difference [in ALJ decisions] by other factors, such as demographics or impairment type” (GAO 1992, 5). The researchers found that blacks living in the Chicago region who applied for DI benefits were 17 percent less likely than whites to receive benefits at the hearing level. Blacks living in the Chicago region who applied for SSI benefits were 10 percent less likely than whites to receive benefits when their cases were considered by ALJs. The Boston region revealed similar results. Blacks who applied for DI benefits were 14 percent less likely than whites to win their claims before ALJs. Black SSI applicants were 10 percent less likely than whites to receive benefits from ALJs. The San Francisco region results revealed a 12 percent difference for black DI applicants and a 7 percent difference for black SSI applicants when their cases were adjudicated by ALJs (43). Although the 1992 GAO study does quantitatively analyze the disability decision-making process for bias against black applicants, it does not consider why, how, and under what circumstances race plays a role or how other biases may potentially influence decision makers.

One other study examined the effects of gender in disability decision making. The preliminary and final reports of the Ninth Circuit Gender Bias Task Force (U.S. Court of Appeals [Ninth Circuit] 1992, 1993) further suggest that ALJ decision making may be biased. Using surveys of claimant representatives and focus groups of attorneys who specialize in representing applicants before ALJs, the task force documented instances in which women respondents believed that gender influenced judges. For example, the statistics generated by the task force study indicate that all of the female representatives who responded to the survey believed that the gender of their female claimants had some effect on how ALJs evaluated credibility: 32 percent of female representatives believed that ALJs in gen-



eral attributed greater credibility to the testimony of men than women, and 80 percent of female respondents believed that male ALJs attributed less credibility to the testimony of women than men (U.S. Court of Appeals [Ninth Circuit] 1992, 99). The task force found that pain is one of the symptoms ALJs most commonly consider in assessing credibility. Since women are far more likely than men to claim disability because of chronic pain syndrome and since 75 percent of applicants with chronic pain syndrome are denied benefits, the task force concluded that these judges' decisions may reflect gender bias (U.S. Court of Appeals [Ninth Circuit] 1992, 93–104).

The male judges who responded to the study survey did not agree. All 83 male ALJ respondents reported that gender plays no role in disability decision making (U.S. Court of Appeals [Ninth Circuit] 1992, 99). In contrast, four of the seven female ALJs who responded reported that disparaging remarks about women occur “frequently” or “somewhat frequently” (99).

Although the Ninth Circuit Gender Bias Task Force explored the effects of gender in both the hearing process and its outcome, the investigators did not examine ALJ hearing transcripts and decisions to learn systematically how that bias may enter into the system and how it operates once there. Using these findings as a springboard, my study looks directly at ALJ hearing transcripts and decisions precisely to devise a means to investigate these issues.

### **Recent Studies of Decision Making**

Recent efforts to study and improve decision making in disability claims have followed in the wake of the 1992 GAO race bias study. In March 1993, the Social Security Administration's Disability Hearings Quality Review Process (DHQRP) became fully operational. This unit developed and implemented a peer review method by which ALJs, acting as Reviewing Judges (RJs), analyze the decisions of their peers to “promote fair and accurate hearing decisions” (SSA 1997, 13). While the two DHQRP reports issued to date covering the periods 1992–93 and 1995–96 (SSA 1995, 1997) make no mention of the 1992 GAO race bias study, there is no doubt that the quality review process was prompted, in large measure, by the GAO's suggestion that there were racial disparities in the system (GAO 1997; Skoler 1994). The history is critical as the DHQRP's findings should be analyzed in light of SSA's motivation to improve their badly tarnished image.



Consistent with the SSA's history of systematically denying allegations of prejudice, neither of the two reports issued by the Office of Program and Integrity Reviews have examined the extent to which race or gender bias may be operating in the decision-making process. Indeed, the SSA's DHQRP has peer reviewed over 9,000 cases using a systematic method for evaluating ALJ decision making; not one question on the 18-page evaluation form studies the extent to which race or gender bias might be operating in the process. Nevertheless, SSA's findings merit attention, as these are the first studies to examine ALJ compliance with key rules.

The SSA DHQRP solicits ALJs to serve as RJs for a four-month period. The judges selected as RJs represent a cross section of experience and are selected from different regions. Since its inception, 108 ALJs have served as RJs, representing approximately 10 percent of the ALJ corps.<sup>5</sup>

SSA's findings focus on several issues that are relevant to determining whether judges comply with procedural or substantive safeguards. For example, the data collection form requires RJs to assess such matters as whether ALJs adequately inform claimants regarding their right to representation and whether ALJs obtain appropriate waivers of that right from claimants; whether ALJs meet their obligation to protect the claimant's rights; and whether ALJs provide an opening statement, identify witnesses, and obtain specific testimony (the specific rules governing these procedures are described in more detail in chapters 4, 5, and 6). In addition, the RJs are asked to determine whether the ultimate decision is supported by substantial evidence, the standard the Appeals Council uses when deciding whether the ALJ decision should be affirmed. If the RJs conclude that the decision was not supported by substantial evidence, RJs elaborate why based on a number of factors. Finally RJs are asked whether ALJs abused their discretion or made errors of law, both of which are standards for Appeals Council review and remand or reversal. The RJs are then asked to perform an "independent" review of the entire record to determine if they would have reached the same conclusion as ALJs making the respective decisions.

None of their findings are particularly startling. Overall, the RJs concluded that most hearing denials were supported by substantial evidence (82 percent for 1992–93 data and 92 percent for 1995–96 data) (SSA 1997). In 1992–93, RJs concluded that they would have made the opposite decision in only 4 percent of denied cases and in 1995–96, RJs concluded that they would have made the opposite decision in only 2 percent of denied cases. Similar data were generated for approved claims. RJs would have

made the opposite decision in 5 and 6 percent of cases they reviewed in 1992–93 and 1995–96, respectively. In 1992–93, RJs believed that 14 percent of denied cases needed more documentation and by 1995–96, only 6 percent of denied cases needed more documentation.

During 1992–93, in 9 percent of the denied cases reviewed by RJs, there was an “abuse of discretion” and by 1995–96, only 5 percent of the denied cases involved an “abuse of discretion.” In general, when abuse of discretion was detected, there were three primary explanations: expert medical or vocational testimony was either inadequate or misrepresented in the decision; appropriate procedures were not followed, such as ALJs did not send posthearing evidence to claimants or their representatives; or claimants were pressured to proceed without representatives.

Errors of law were more common. In 1992–93 cases, 19 percent of hearing denials contained errors of law and in 1995–96, 16 percent of hearing denials contained errors of law. Of those cases containing errors of law, RJs identified several areas in which ALJs made such errors, including but not limited to: credibility (20 percent for 1992–93 and 3 percent for 1995–96); procedural issues (12 percent for 1992–93 and 9 percent for 1995–96) and evidentiary issues (33 percent for 1992–93 and 34 percent for 1995–96).

Overall, RJs concluded that in 1995–96, ALJs protected claimants rights in 96 percent of the cases; in 1992–93, they found this protection evident in 95 percent of the cases. For 1995–96, 95 percent of denied claimants rights were protected whereas 89 percent of denied claimants rights were protected during the 1992–93 review.

In those cases in which RJs concluded that the denial of benefits was not supported by the evidence, RJs believed that ALJs failed to obtain such information as the impact of other symptoms or pain on the claimant, activities of daily living, the treating physician’s opinion, and past relevant work (SSA 1997). In addition, the RJs found that in those denials where there was insufficient evidence to support the decision, claimant’s evidence or testimony was lacking in 63 percent of the cases. As was detected in previous studies, it seems that the evaluation of mental impairments remained problematic for some ALJs. In the 1992–93 cases reviewed by RJs, only 78 percent of mental impairment denials were supported by substantial evidence; in 1995–96 cases, 86 percent of mental impairment denials were supported by substantial evidence.

The findings with regard to procedural issues are particularly interesting insofar as many of the SSA’s inquiries mimic the areas of inquiry in

this study. The RJs found that most judges administer the oath to both allowed and denied cases (97 percent in 1992–93 and 1995–96) and in 90 percent of denied cases, witnesses were introduced (1995–96 data only) (SSA 1997). RJs also found that in 1992–93, 26 percent of denied and awarded claimants did not receive an opening statement. By 1995–96, 18 percent of all claimants did not receive an opening statement. Data from 1995–96 revealed that ALJs are more likely to give opening statements in denied cases (15 percent of denied claimants did not receive opening statements while 21 percent of claimants who received favorable decisions were given an opening statement).

Questions relating to the interrogation of claimants revealed that RJs found some ALJs hearing and decision making practices to be lacking. Taken from the 1995–96 data, ALJs had the most difficulty with gathering evidence related to claimants medication and their adherence to medical treatment prescribed by a physician (78 percent), daily activities (81 percent), sources of current income (72 percent), and physical capabilities (72 percent) (SSA 1997). Similarly, these problems surfaced when questioning the experts present at the hearing. Again relying on 1995–96 data, the ALJs did not adequately address the ME's evaluation of pain and/or symptoms (adequately addressed in only 73 percent of denials), the ME's professional qualifications (adequately addressed in only 53 percent of denials), assessment of mental capacity (adequately addressed in only 60 percent of denials), and asking the ME if more claimant information is needed (adequately addressed in only 51 percent of denials). Similar problems were detected in the questioning of VEs.

In a review of the RJ data relating to unrepresented claimants' cases, it was found that in 1992–93, 77 percent of unrepresented claimants whose claims were denied were adequately informed of their right to representation. In 1995–96, 80 percent of unrepresented denied claimants were adequately informed of their right to representation. In addition, RJs found that 80 percent of those claimants whose cases were denied, using the 1992–93 data, were supported by substantial evidence, compared to 89 percent in 1995–96. The SSA found that the difference between represented and unrepresented claimants for 1995–96 was statistically significant (94 percent for represented claimants who were denied). This finding suggests that ALJs may be more inclined to provide stronger evidentiary decisions in cases involving claimants who were represented. This also suggests that claimants who are unrepresented may not only be deprived of adequate instructions regarding their right to be represented,

but might also receive decisions that are not supported by substantial evidence. Statistically significant differences which disadvantaged unrepresented claimants were also detected with regard to errors of law.

Overall, these findings suggest that RJs believe that their fellow ALJs do a fair to excellent job when hearing cases and evaluating claims, especially with regard to denied applicants. In 1992–93, RJs concluded that they would make the same decision in 89 percent of the 1992–93 denied cases and in 94 percent of the 1995–96 denied cases. As for allowances, RJs would have made a different decision in 83 percent of the 1992–93 allowed cases and 72 percent of the 1995–96 allowed cases. Hence, at least according to RJs, ALJs are more often correct when they deny claims, than when they grant them.

The variations that RJs detected in these efforts motivated SSA to improve both the training and practices of ALJs. Indeed, SSA's most recent report (1997) finds "dramatic" improvement between the two study periods (1992–93 and 1995–96) with "ALJs improving significantly in many areas of case adjudication" (69). What is most stunning about these results however, is the absence of any attempt to explore whether prejudice persists in the ALJs outcomes.

In 1994, Dan Skoler, Associate Commissioner for OHA, described the development of the DHQRP peer review program and its relationship to detections of bias. This effort, he claimed, would "identify and explore factors in hearing-level claims that might contribute to the disparity dilemma" (19). Yet nothing in the SSA reports suggests that this peer review process was concerned with or ever considered the impact of bias on decision making. Indeed, even though race and gender data are available for the 9,000 cases reviewed for these two periods, no analysis of this data has been undertaken or published. While SSA analyzed the relevance of age and impairment on ALJ and RJ reviews, it neglected to analyze the data in light of other factors such as race or gender. As with previous studies, the SSA remains unwilling, in any meaningful way, to scrutinize ALJ decision making practices for bias.

One final recent study merits mention. In 1997, the GAO issued a report in which they compared initial and reconsideration decision making, to the awards and denials of ALJs. Twenty years on, the federal government is still trying to determine why disparities between these two sets of decision makers persist. Essentially, the GAO concludes that DDS examiners and ALJs differ over claimants ability to function: ALJs are much more likely than DDS examiners to conclude that claimants have

severe limitations which prevent them from working. They attribute those differences to a number of policy and procedural factors, including but not limited to ALJ reliance on claimant testimony and the opinions of treating physicians. The GAO seems heartened by the downward shift in award rates by ALJs and the SSA's support for programs like DHQRP that monitor the decision-making practices of ALJs to ensure that their decisions comport with agency policy. This GAO study however, like previous studies, is not concerned with race or gender disparities in decision making, but rather with making decision making consistent between the two sets of decision makers. Implicit in the GAO report are two guiding assumptions. First, ALJs grant too many disability claims, especially when their decisions are compared to DDS examiners. Second, ALJs should be the decision maker of last resort; the GAO assumes that most claims should be decided by DDS examiners at the initial or reconsideration stages of the process. The GAO recognizes the long-standing problem in inconsistent decision making between DDS and OHA and concludes that confidence in the decision-making process and the costs associated with expensive ALJ appeals demands system reform. They recommend unification initiatives, similar to the redesign plan, to rectify the consistencies they detect.

Despite the disability system's procedural and substantive safeguards, the evidence presented here raises serious questions about the fairness of and thus the potential influence of bias in disability decision making. The mechanisms designed to ensure impartiality, including the listings, the Grid, and the five-step sequential evaluation process, may operate to standardize and mask bias rather than to eliminate it. Moreover, the findings of Nagi, the GAO, the SSA, Dixon, and Mashaw, and more recently of the DHQRP, which document incongruencies, inconsistencies, and inaccuracies, suggest that factors other than the stated guidelines influence decision makers, including ALJs, and that errors in determinations are detectable and prevalent. But no researcher examined how and at what point subjectivity or personal judgment may enter the hearing or decision-making process; none of these previous studies, that is, took the issue of impartiality head on. Nor did they propose any real options for addressing the disparities detected.

Deborah Stone's political analysis raises the question of how clinical standards permit bias to occur—a question that, to some extent, presupposes that bias is an inevitable feature of disability decision making. The 1992 GAO race study indicates that personal bias probably does influence

the process, but it does not explain how that prejudice operates in practice. The reports of the Ninth Circuit Gender Bias Task Force provide a starting point for an analysis by suggesting that the gender of a claimant may affect such matters as how judges elicit evidence from women. The DHQRP reports (SSA 1995, 1997) tentatively suggest that even the peer review process has revealed some procedural and substantive irregularities in the system, despite its silence on the role of race or gender in the dynamic between judge and claimant.

In the next chapter I explore in greater detail the cases in my sample. My findings further suggest and illuminate exactly how judges fail to apply the rules uniformly and with the affectivity the system requires.