A DUE PROCESS METAPHOR FOR PERFORMANCE APPRAISAL

Robert Folger, Mary A. Konovsky, and

Russell Cropanzano

ABSTRACT

The authors argue that performance appraisl (PA) effectiveness has suffered because of the overly rational nature of its conceptualization, which likens performance appraisal to test construction. We review and critique the underlying assumptions of the "test" metaphor for PA, which include: (a) work arrangements allow reliable and valid measurement, (b) raters will assess performance accurately, and (c) a rational, unitary criterion exists. The authors believe that a "political" approach to PA has arisen because of the deficiencies of the test approach to PA and because of the organizational realities within which PA is practiced. A political metaphor for PA also has problems, however, because it allows too easily the self-interested use of power for resolving the conflicts that arise in PA. The authors develop an alternative PA model, "a due process" metaphor that addresses the problems associated with the test and political metaphors. Elements of the due process model include adequate notice, fair hearing, and judgment based on evidence. The authors develop the implications of these due process elements for the practice of PA and for PA research. The due process metaphor addresses practical issues that have heretofore remained problematic in PA, such as resolving the conflict that inevitably arises in PA, the role of self-appraisal, and the nature of employee rights. The due process metaphor also suggests new areas of research and new emphases in existing research areas that we discuss in a concluding section.

129

Performance appraisal (PA) effectiveness has suffered because of its grounding in an overly rational conceptualization, namely a view that treats PA as analogous to the psychometric process of constructing a valid test. This PA "test metaphor" rests on three assumptions that we find problematic under a wide variety of circumstances commonly encountered within organizations. In general, the test metaphor conveys an impression of rationality that is inconsistent with behavioral evidence about managerial decision-making and judgmental capabilities (e.g., Bazerman, 1990; Kahneman, Slovic, & Tversky, 1982; Simon, 1957). We first review the test metaphor and its assumptions and then present an alternative model of PA, the "political metaphor." We discuss the shortcomings of this political model and present a "due process" remedy for those shortcomings. Finally, we discuss the theoretical and practical implications of a due process orientation toward performance appraisal.

THE TEST METAPHOR

The test metaphor represents the traditional academic way of thinking about PA, emphasizing the rating format and its psychometric properties rather than the PA process and its social context (cf. Landy & Farr, 1980). This metaphor's depiction of appraisers conjures up an image of research scientists. The scientist-appraiser image connotes a seeker of truth who records objective reality using reliable and valid measures. Banks and Roberson (1985) made the test metaphor explicit in arguing that "raters almost always function as test developers" (p. 128) and that "accurate appraisers may well be those who become expert at applying principles of test development" (p. 129). In other words, successful PA consists of using scientific tools with technical proficiency. Thus DeVries, Morrison, Shullman, and Gerlach (1981) have expressed the prevalent sentiment of test metaphor advocates: "Today's reality makes psychometric issues surrounding performance measurement more relevant than ever" (p. 28).

The test metaphor thus emphasizes rating formats as a means to improve PA accuracy (Schwab, Heneman, & Decottis, 1975). An adequate rating instrument should reliably measure true performance (Borman, 1975, 1978) and avoid biases such as halo, leniency, and differential dimensionality (Ronan & Prien, 1971). The PA instrument is presumed more objective when the rater scores only quantifiable behaviors, for example, leading to the use of Behaviorally Anchored Rating Scales (BARS) and the BES or BOS variants (e.g., Latham & Wexley, 1977; Smith & Kendall, 1963). Despite such efforts, it is unclear that increased instrument objectivity has actually improved rater accuracy (e.g., Borman & Dunnette, 1975; Burnaska & Hollmann, 1974; Kane & Lawler, 1979; Kavanagh, 1982; Saal & Landy, 1977). Landy and Farr (1980) concluded from an exhaustive review of the literature that despite years of

research, no rating instrument has yet proved to be demonstratively more valid than any other.

We think the problems of PA ratings stem not from deficient research effort, but rather from the directions those efforts have taken. Solutions to problems are constrained by how the problems get defined (Huber, 1980), and science follows directions guided by core assumptions. Application of the test metaphor relies implicitly on the following underlying assumptions (examined in the following sections): (a) that work arrangements allow for reliable and valid measurement; (b) that raters are capable of assessing performance accurately; and (c) that a rational, unitary criterion exists and has clear benefit to the organization. If any of these assumptions is dubious, the adequacy of the test metaphor's ideals of rationality and objectivity will be assailable—that is, the prospects for conducting PA guided solely by considerations of classic rationality will be brought into question.

Assumption #1: Work Arrangements Allow
Reliable and Valid Measurement

To get reliable and valid measures, the "test constructors" designing PA systems must have information about employees' task related behaviors and/or job outputs (cf. Kane & Lawler, 1979), and a rating format or other record-keeping device must incorporate the relevant information. Methods of recording performance, therefore, must take into account the characteristics of tasks performed. Lee (1985) adapted Thompson's (1967) and Ouchi's (1977) models to create a typology of task characteristics involving two dimensions. We examine the implications of this typology regarding constraints that are imposed on measurement capabilities.

Lee's first dimension, the *availability of reliable and valid outcome measures*, applies to jobs where tangible output can be monitored. Her second dimension is *knowledge of the transformation process*, or the means-end relationship. "Ends" are work performance outcomes that enhance the organization's well-being, such as by improving its efficiency or effectiveness; "means" are elements of the work process (e.g., the technology of the job itself) whereby job inputs are transformed into outputs. Thus the ends-means relationship indicates the extent to which specific and known behaviors produce job outcomes.

Lee noted that jobs differ in the extent to which they make information about means-ends relationship accessible and the availability of reliable, valid measures feasible. We argue that reliable, valid measurement is inherently problematic when task characteristics make information about means-ends relationships hard to acquire, and that such task characteristics are prevalent in many jobs. (Moreover, knowledge of the co-variation of means and ends may not be sufficient information when the critical determinants of task performance are *cognitive* and *emotional* processes associated with the transformation process.)

In terms of tasks for which reliable and valid measures are presumed feasible, Lee's examples included the jobs of assembly line workers, clerks, recruiters, and sales personnel. Knowledge of the appropriate ends is presumed non-controversial in the case of highly standardized assembly-line work, for example, because assumptions about the production outputs that constitute valued ends are ordinarily built into the production system itself (e.g., automobile manufacturers presumably value the number of cars the line can assemble within a given time period). Similarly, when a technology's transformation process requires the repetition of standardized behaviors, exhaustive information about the means of production is presumably accessible in a non-controversial fashion, thereby enhancing the prospects for reliable, valid measurement.

Certainly some jobs are more ideally suited than others to the requirements for reliable and valid measurement. But only simple, standardized tasks—or those that achieve ends with accepted value by means whose value is also accepted—fit easily into this category. Moreover, if many jobs instead entail at least some task characteristics involving incomplete knowledge about means-ends relationships, then the prospects for reliable and valid measurement are thereby diminished.

Fewer and fewer jobs, even at the lowest organizational levels, entail work where knowledge of means-ends relations is complete, and where readily available measures are reliable and valid. Three current trends corroborate this point. First is the continued and growing predominance of jobs in the service sector. The majority of the U.S. labor force (71%) already works in the service sector. In the past twenty years, 36 million net new jobs have been created in the United States (Quinn & Cagnon, 1986). On balance, 9 out of 10 of these jobs have been in the service sector (Heskitt, 1987). Jobs in the service sector yield products that are largely intangible, are produced and consumed simultaneously, and tend to involve the consumer in both production and delivery (Bowen & Schneider, 1988). The trait of intangibility makes the criteria for measuring the production outputs difficult to agree on because service output is largely perceptual and does not exist in time and space. Services are not possessed—they are experienced, created, or participated in (Shostack, 1981). A tight production/consumption link and the participation of the customer in the production of the service make the means-ends relationship subject to high levels of variability and unpredictability, thus imposing severe limits on reliable, valid measurement of job performance.

Jobs in the manufacturing sector also do not escape measurement difficulties. The characteristics of manufacturing jobs are changing in a way that makes measurement increasingly difficult. The advent of flexible manufacturing processes as one response to an increasingly competitive environment, for example, shifts the arena of competition from manufacturing to engineering (Jaikumar, 1986). Engineering performs the critical line function, and highly

variable specifications for plant production make standardization (and thus, measurement) of job performance more difficult. Finally (due to another economic trend—downsizing and re-organization), many companies are restructuring work itself so that employees are assigned job tasks formerly relegated to middle management levels. Autonomous work teams, for example, call for the development of employee self-management skills for planning and scheduling work and the development of interpersonal skills germane to conflict resolution and problem solving. Adequate performance of such behaviors is inherently more difficult to measure compared with the routine, standardized tasks formerly prevailing among rank and file employees (Cameron, 1980).

Information about means-ends relationships becomes less available as each of the above trends accelerates. Consequently, the prospects for reliable and valid measurement—even at the lowest organization levels—will become increasingly and more considerably doubtful with regard to successful accomplishment via traditional PA approaches.

We have been emphasizing how the characteristics of most jobs tend to preclude the attainment of sufficient knowledge about means-ends relationships in the first place. Turning next to implementation issues, we address in the following section some dilemmas created because raters—not just rating formats—also constitute part of the measurement process.

Assumption #2: Raters Will Assess Performance Accurately

To assess performance accurately, raters must have sufficient skill, knowledge, and motivation. We have already discussed the limits of available performance information due to current trends in job composition. We do not address rater motivation comprehensively, although our sections on the political and due process metaphors touch on some motivational issues. The skills necessary for well-conducted PA, however, represent a critical area for the test metaphor and hence draw attention to human information-processing and cognitive limitations—a topic with a considerable literature.

Suppose that a company had all the information necessary to construct rating scales possessing ideal psychometric properties. The cognitive literature suggests that even if such conditions existed, the cognitive limitations of raters would still impose constraints (and potential biases) on performance assessment. Many authors have noted these severe human limitations with respect to the encoding, retrieval, and evaluative use of performance relevant information (e.g., DeNisi, Cafferty, & Meglino, 1984; Feldman, 1981; Ilgen & Feldman, 1983). Furthermore, analyzing sources of cognitive deficiencies has not yet yielded proven strategies for more accurate performance appraisals. Without attempting a complete review of the literature, we will examine some of the major issues (cf. DeNisi & Williams, 1988; Lord & Foti, 1986).

According to social cognition research, raters have cognitive categories for performance dimensions (Lingle, Altom, & Medin, 1984). Categories are indexed by prototypes, with close matches placed in their respective categories. Although prototypes can aid efficient retrieval, in that specific behaviors need not be recalled (Ilgen & Feldman, 1983), they also lead to a variety of errors in rater judgment. Differences between behaviors are "smoothed out" in an effort to fit the behaviors into a category, for example, leading raters to disregard information that conflicts with their prototype (Nathan & Lord, 1983). Once categorized, behavior is invested with all of the attributes of the category (even some not observed), which can cause a rater to "recall" things that never took place (Phillips, 1984; Phillips & Lord, 1982).

Various aspects of the rating task also bias the outcome. These aspects include the time of delay between observation and rating (Murphy & Balzer, 1986), the way performance information is presented (Cafferty, DeNisi, & Williams, 1986), the conceptual similarity of the rating dimensions (Kozlowski, Kirsch, & Chao, 1986), the ratee's past performance (Murphy, Gannett, Herr, & Chen, 1986), the type of decision being made (Williams, DeNisi, Meglino, & Cafferty, 1986), and the extent to which the rater is a job expert (Smither, Barry, & Reilly, 1989).

Rater limitations in processing information about employees and task characteristics result inevitably in cognitive biases and partially distorted evaluations. Research on the upper limits of rater reliability provides evidence for this assertion. Borman (1978) trained raters, for example, then allowed them to view video taped performance vignettes as often as desired. He found that although convergent and discriminant validity were high, substantial disagreement occurred between raters. Weekley and Gier (1989), who examined the evaluations of expert judges at the 1984 Olympic figure skating competition, obtained similar results.

Research on the cognitive limitations of raters has helped to make salient that successful PA does not hinge on the psychometric properties of rating formats alone, and that a "technological fix" aimed only at the design of a rating format represents a logically incomplete approach to the potential difficulties inherent in all measurement efforts. When rating formats are completed by raters, the rater becomes an important component of the instrumentation package (cf. Campbell & Stanley, 1963), and the accuracy of measurement becomes a joint product of format and rater. Thus the measurement-related capacities of raters—as observers prior to performance ratings and as recorders when filling out rating scales—limit the accuracy of PA.

Assumption #3: A Rational, Unitary Criterion Exists

If, despite the preceding arguments, we assume that developments spawned by the test metaphor's use can yield objectively accurate ratings, an additional

question remains: *What* should be rated? Simon's (1983) discussion of conflicting values pinpoints a difficulty in achieving the rationality that test metaphor requires: the problem of selecting criteria for validating ratings. Specifically, Simon (1983) notes that "reason, taken by itself, is instrumental. It can't select our final goals, nor can it mediate for us in pure conflicts over what final goal to pursue—we have to solve these issues in some other way" (p. 106).

The ultimate criterion is a theoretical construct—a point made long ago by Thorndike (1949) but needing re-emphasis because it seems missing from the thrust of much psychometrically oriented research. Simon's (1983) contribution goes deeper, however, in noting that even at the conceptual level, choices among performance criteria entail resolving conflicts among competing *values*. There are potential disagreements between the rater and ratee and among other constituents about what constitutes the quality and type of performance deserving to be rewarded (i.e., what accomplishments benefit the firm) (Tsui & Milkovich, 1987). These differences of opinion cannot be resolved by refinements in measurement technology. Additionally, the different consequences of performance (e.g., effects on market share, profit margin, customer satisfaction, shareholder interests) attain their significance not solely on the basis of objective facts, but also as a function of the values held by key organizational decision makers.

One approach to resolving the ultimate criterion dilemma has been the use of multiple criteria. The PA literature often notes the need for a clear statement of the relative weights when multiple criteria are used, and measurement models (e.g., obtained via policy-capturing analyses) can reflect those weights. But these models do not *develop* the criteria, which must come from human decision makers exercising fallible judgment. Such decision makers might choose among criteria and assign weights by carefully considering the full range of possible organizational effects, but these criteria choices and weighting decisions are nonetheless value-dependent—because *all* decisions rest on value judgments as underlying assumptions or "givens" (Perelman, 1967; Simon, 1957, 1983). Furthermore, as Simon's (1983) analysis has intimated, limited attention and inherent uncertainties impose inevitable constraints precluding the ability of decision-makers to select criteria and assign weights on a neutral, purely objective basis (i.e., exhibiting classic rationality).

We do not argue for abandoning the test metaphor with respect to its intended purpose, but instead argue that this purpose does not adequately address a larger issue: how to resolve conflicts of interest that arise when administrators, employees, shareholders, and other constituencies hold different values affecting the criteria chosen for PA purposes (a point we develop further in the context of the political metaphor). Decisions about those criteria should be guided by ideas about what makes the organization most effective, but a universal consensus about what it means for an organization

to be "effective" does not exist. Cameron (1980) argued that "researchers have not yet agreed on the most appropriate criteria for making evaluations of effectiveness, the characteristics that differentiate effective organizations from ineffective organizations, or even what constitutes organizational effectiveness in the first place" (p. 67). Tsui & Milkovich (1987) have shown that an organization's multiple constituencies further complicate the determination of a unified description of organizational effectiveness because of the unique opinions and values of each stakeholder group.

Cameron expanded upon the inherent difficulties faced when multiple constituencies define and assess organizational effectiveness. For example, the most frequently used approach—on which Lee's (1985) analysis implicitly relied—defines effectiveness as goal achievement and assesses outputs compared to goals. As Cameron (1980) pointed out, however, "an organization can be ineffective even though it reaches those goals" (p. 68), such as when the organization's goals are too low or are actually harmful to its well-being. Cameron further noted that many firms (e.g., large service organizations, business conglomerates, or research and development establishments) exhibit organized anarchy to a degree that makes *all* existing organizational effectiveness approaches inadequate.

Cameron identified characteristics of organized anarchies that include the following: (a) "Goals are generally ill-defined, complex, changing, and contradictory." (b) "Means-ends connections are not clear—that is, there is no obvious connection between the technology or the way work is done and the outcome." (c) "Widely differing criteria of success may be operating simultaneously in various parts of the organization. Pursuit of success in one part of the organization may even inhibit success in another part of the organization" (pp. 70-71). Moreover, Cameron argued that these features are becoming more and more widespread in organizations as environmental complexity and turbulence increases.

Finally, Cameron noted that choosing effectiveness criteria calls for answers to numerous questions: "What domain of activity should be the focus of the evaluation?" "Whose perspective, or which constituency's point of view should be considered?" "What level of analysis should be used?" "What time frame should be employed?" "What type of data should be used?" "What referent [e.g., compared to which competitors] should be employed?" (pp. 73-79). It is unclear what responses comprise the best answers to those questions, which parties should play a role in determining those responses, what those roles should be, or what processes best address such decisions. Obviously optimal responses vary with the situation, but it is also unclear how context affects the answers and which features function decisively to maximize organizational well-being.

The existence of multiple organizational actors and constituencies suggests that answers to Cameron's effectiveness-related questions are problematic. As

rchers have
lluations of
itions from
ffectiveness
vn that an
nination of
the unique

n multiple
ample, the
i implicitly
s outputs
rganization
h as when
well-being.
anizations,
nts) exhibit
anizational

include the
iging, and
is, there is
i done and
operating
cess in one
iart of the
eatures are
ronmental

or answers
ocus of the
iew should
time frame
at referent
73-79). It
ons, which
hose roles
ly optimal
ext affects
anizational

es suggests
ematic. As

Simon (1983), Pfeffer (1981), and others have noted, the inherent difficulty of resolving such issues on purely rational grounds (i.e., by demonstrably objective criteria) makes the resolution susceptible to organizational politics. Longenecker, Gioia, and Sims' (1987) examples of how PA decisions are influenced by organizational politics reinforces this argument that subjectivity in PA precludes rationality at the outset of decision making (when effectiveness criteria are chosen). Some managers that Longenecker et al. interviewed, for example, reported using the appraisal process to favor the goals of their own group rather than the whole organization; others reported softening performance appraisal standards in order not to have a detrimental effect on an employee's career.

An organizational politics view of PA has arisen partly because of such evidence (cf. Patz, 1975), but also because of the more general suspicion that PA practice does not always achieve the test metaphor's scientific ideals (Cascio, 1982). Frequently, appraisers act more like politicians than like test constructors or test administrators. This characterization suggests that although normatively appealing, the test metaphor is descriptively incomplete. Both of the remaining perspectives we explore, the political and due process metaphors, aim at providing descriptive accounts more closely attuned to actual or feasible practice.

THE POLITICAL METAPHOR

Pfeffer (1981) has characterized organizations as political entities. Pfeffer stated that "politics involves how differing preferences are resolved in conflicts over the allocation of scarce resources," and that political activities "attempt to influence decisions over critical issues that are not readily resolved through the introduction of new data and in which there are differing points of view" (p. 6). Thus PA is political because it is linked to the allocation of scarce resources (e.g., salary or promotion decisions) and because the resolution of different preferences or points of view cannot be completely accomplished by means of objective data. Pfeffer's views echo the thrust of the preceding section, in which we argued that rationality is a misleadingly simplistic metaphor for PA.

Longenecker et al. (1987) were among the first to present empirical evidence for a political model of PA decision making. Although finding it deplorable "that executives might deliberately distort and manipulate appraisals for political purposes," they nevertheless "found extensive evidence to indicate that, behind a mask of objectivity and rationality, executives engage in such manipulation in an intentional and systematic manner" (p. 183). They accounted for these findings by noting that "executives are political actors in an organization, and they often attempt to control their destinies and gain influence through internal political actions" (p. 184).

Patz (1975) reached similar conclusions from interviewing managers at various levels: "Most believe that superiors attempt to work the appraisal system to their own advantage by consciously or unconsciously reporting false evaluations to their subordinates" (p. 77). Examples of the reasons managers gave for distorting appraisals included that a subordinate receiving a very high evaluation was likely to be promoted out of a manager's work group, and that very low evaluations might reflect negatively on the supervisor. Cascio's (1982) review of PA corroborated this impression that "political considerations are organizational facts of life" (p. 31).

Other features of the political metaphor, as highlighted by Pfeffer (1981), have implications regarding the use of *power*. The political metaphor "presumes that parochial interests and preferences control choice" (p. 22) and hence "power would best predict changes and shifts in decisions and allocations" (p. 24). In addition:

> Political models of choice further presume that when preferences conflict, the power of the various social actors determines the outcome of the decision process. Power models hypothesize that those interests, subunits, or individuals within the organization who possess the greatest power, will receive the greatest rewards from the interplay of organizational politics . . . . Power is used to overcome the resistance of others and obtain one's way in the organization (p. 28).

Thus, rather than assuming that rational, objective criteria can be relied upon to dictate the decision outcomes, the political metaphor presumes "it is the relative power of the various social actors that provides both the sufficient and necessary way of resolving the decision" (p. 30).

The exercise of power pursuant to self-interest, however, is not the only political metaphor applicable to organizations. Another perspective is Tetlock's (1985) emphasis on the importance of *accountability*. Tetlock argued that a political actor's "primary goal is to maintain the positive regard of important constituencies to whom he or she feels accountable" (p. 299). Indeed, Tetlock indicated that accountability is not associated uniquely with the political role; rather, all human beings are intuitive politicians who often must be cognizant of "the interpersonal consequences of their conduct (How will others react if I do this? How effectively can I justify my views to others if challenged?)" (p. 306).

Tetlock's analysis rests on two assumptions: "the first posits that accountability of conduct is a universal problem of social life with which people must deal; the second posits that people are generally motivated to maintain the approval and respect of those to whom they are accountable" (p. 309). A further point equally key to his analysis and to PA, however, is that the effects of accountability will vary depending on the nature of the relationship among the parties associated with a decision-making situation. In particular, his

research identified conditions under which the nature of that relationship will be "conducive to complex and self-reflective information processing" (p. 310) and hence less vulnerable to decision-making based solely on self-interest. Tetlock noted that these conditions specifically included decision-making situations (such as PA) in which people know in advance that they will have to explain their stand on controversial issues to others. The allocation of scarce resources, such as that involved in PA, nearly always is controversial.

Research reviewed by Tetlock (e.g., Chaiken, 1980; Cvetkovich, 1978; Hagafors & Brehner, 1983; McAllister, Mitchell, & Beach 1979; Rozelle & Baxter, 1981; Tetlock, 1983) shows an additional circumstance especially conducive to thoughtful information-processing, namely accountability to others who views are not known in advance. The research shows that such conditions contribute to (a) use of cognitively complex decision and judgment strategies; (b) decisionmakers' enhanced awareness of their own cognitive processes; and (c) processing of evidence in forming impressions that is more data-driven than theory-driven (i.e., less susceptible to stereotypes or schemata). Tetlock (1985) has argued that these results stem from decisionmakers' enhanced motivation "to consider arguments and evidence on both sides of issues in order to prepare themselves for a wide variety of critical reactions to their views" (p. 315).

The conditions described by Tetlock to induce complex and reflective information processing are those often encountered when managers conducting PA's are accountable to a number of constituencies whose views do not coincide. They are accountable, for example, to the employees who are evaluated and who may have discrepant opinions about performance levels. These discrepant opinions are especially likely to arise when important and scarce organizational rewards are tied to PA evaluations. One important aspect of conducting PA, therefore, is managers' preparedness to anticipate and resolve such controversy. Managers are also accountable to organizational shareholders who demand that PA decisionmaking ultimately be linked to organizational effectiveness.

It is difficult to predict whether the self-interest or accountability motive will prevail in specific instances of PA. Conflicting interests, however, make one PA element quite salient: The decision-maker in charge of PA is like a judge in a civil proceedings case who must resolve competing claims as to which disputant is more entitled to a particular award or how a settlement will be allocated between two parties. This courtroom analogy to PA implies that the political metaphors discussed above neglect an important element relevant to PA: society's inclusion of legal procedures to resolve disputes when political considerations are otherwise likely to prevail. Society is a political arena in which groups and individuals pursue the maximization of their personal values. Nations limit the unbridled pursuit of self-interest by establishing systems of laws, alone with legal institutions, that resolve disputes within the framework

of those laws. The legal system is a vital substrate of a society's political system, therefore, and we believe that existing political approaches to PA have ignored the possibilities of due process for resolving the inevitable conflicts that arise in the PA context.

Recognizing a role for due process in PA, and more generally in all organizational decision making, helps close a fundamental gap between PA theory as described by traditional or rational approaches and PA practice as described by the political metaphor. With legal proceedings as an analog, a due process metaphor of PA helps locate a middle ground between the extremes of the test metaphor, which is descriptively deficient if based on unobtainable standards of rationality, and political metaphors that (with the exception of Tetlock's model) are normatively unacceptable if over-emphasizing opportunities for the unfettered exercise of power.

A due process metaphor provides a more feasible set of PA prescriptions compared with extreme versions of test and political metaphors. In the absence of conditions of perfect outcome rationality, for example, organizations may operate using standardized procedures rather than engaging in rational decision making on a continuous basis. In other words, procedural rationality is substituted for substantive rationality and choices are made according to rules and processes that have been adaptive and effective in the past (Pfeffer, 1981). A due process metaphor provides an alternative model of procedural rationality.

The objective of fair decision making—a pursuit emphasized by the due process metaphor—also has value in its own right. That objective represents a more normatively acceptable goal than the self-interested pursuit of political influence, as well as a goal whose pursuit reflects an apt descriptive characterization of actual management practices conducted responsibly. Finally, the rational PA test metaphor, both theoretically and practically, virtually ignores the conflict among different constituencies arising throughout the PA process. The due process metaphor legitimizes this inevitable conflict between competing groups and provides a mechanism for resolving conflict responsibly, according to the tenets of fair procedures.

We describe in the following section a due process perspective on PA. This due process metaphor can be seen as an extension of the political metaphor, in that it embodies key principles of government within the tradition of Western political institutions (particularly those growing out of the English common-law heritage). A principal feature of the due process metaphor is its derivation from legal practices designed to curb the effects of self-interested outcome maximization by attention to maintaining the rights of others. Indeed, this emphasis on *rights* rather than power is a distinguishing characteristic that separates the due process metaphor from existing descriptions of the political metaphor. Hence the advantage accrued by the due process conceptualization, when used as a supplement to the test and political metaphors, is that it can

capture those realities potentially attainable by responsible management, thereby achieving both descriptive validity and normative appeal.

THE DUE PROCESS METAPHOR

The roots of the due process metaphor lie in Thibaut and Walker's (1978) theory of dispute resolution, which described two types of disputes encountered in legal contexts. Conflicts about the most accurate view of reality comprise one type, where "the objective is to determine the truth according to a standard" (p. 541). These disputes involve precise statues such as speed limits. When the courts resolve this type of dispute, legal fact-finding can determine the outcome. The test metaphor highlights aspects of PA resembling this type of dispute. In contrast, "conflicts about the apportionment of outcomes, such as inconsistent claims to the division of assets or losses" (pp. 541-542), cannot be resolved by reference to the facts because these cases demand that legal decision-makers take into account complex and subtle particularities of the situation, especially if these have a bearing on alleged motives and other matters germane to the merits of disputants' claims. In other words, ruling on the merits of claims requires value judgments; and in such cases, a *flexible* common law standard makes it impossible to resolve disputes merely by determining facts. Our due process metaphor highlights aspects of PA resembling these latter types of disputes.

Performance appraisal ratings are frequently used for distributing rewards proportionately with ratings (Mohrman, Resnick-West, & Lawler, 1989). Because of limited reward pools, managers are faced with trade-offs in assigning employee rewards. Rewarding one employee at a high level will dictate reduced reward levels for other employees. When performance appraisal interviews allow discussion, such interviews can provide a forum for employees to "state their case" or bring evidence to bear on their performance levels relative to others' (in hopes of higher evaluations). Although employees do not contend directly against one another in the same manner as disputants in a courtroom, PA is analogous when used to distribute organizational rewards according to merit and employees are allowed such forms of input to the appraisal process.

When observed facts cannot be relied upon for a direct or conclusive determination, as is often the case in performance appraisals, Thibaut and Walker (1978) posit that conflicts are "best resolved with the aim of achieving distributive justice" (pp. 541-542). Furthermore, "distributive justice generally takes the form of evaluating the relative weight of each party's claims for a favorable distribution of the outcomes and then rendering an allocation decision that reflects these relative weights" (pp. 548-549). Such claims "maximize the party's perceived causal responsibility for, or contribution to, 'good' consequences . . . . or . . . minimize the party's attributed responsibility for a charge of 'bad' consequences" (p. 549).

Table 1. Due Process and its Relationship to
the Practice of Performance Appraisal

Element of Due Process	Descriptions Relevant to Performance Appraisal
A. Adequate Notice	Objectives and standards are established in advance, published, widely distributed, and explained.
	Employees have input into formulation of objectives and standards, or at least opportunities to question the content of the standards and objectives, the process whereby they were established, and the manner in which they will be implemented.
	Feedback is given on a regularly recurring and timely basis.
B. Fair Hearing	Standards for the admissability of PA "evidence" include the appraiser's familiarity with the appraiser's performance based on sufficiently frequent observation of behavior or work products.
	Employees have means to indicate their own viewpoint concerning their performance.
	Employees have opportunities to explain their own interpretation of PA "evidence" and to present arguments supporting that interpretation.
C. Judgment Based on Evidence	Steps are taken to have the appraiser apply standards consistently, without external pressure, corruption, or personal prejudice.
	Evaluations show efforts to use principles of honesty and fairness (employees have opportunities to question evaluations, and the explanations provided reflect such principles).
	Evaluations withstand scrutiny, including that which might be engendered by an appeal or other type of opportunity for recourse provided to employees.

Evaluating claims involves inferences about motives and intentions, but "lacking direct knowledge of the feelings or actions behind the participants' actions . . . . it seems almost inevitable that third parties [e.g., judges and juries] will bias contextual information . . . by asking questions and developing information that conforms to a model formed by their own experiences" (p. 549). In the PA context, for example, actor-observer biases imply that appraisers may neglect important contextual information, de-emphasizing environmental factors as influences on an employee's behavior, and instead emphasize the employee's personal responsibility for actions. Employees, on the other hand, would emphasize environmental and contextual information. Thibaut and Walker argue that disputants "ought to control the description of their respective inputs [because] . . . they can be relied upon to describe important contextual factors relating to the disputes that are likely to be

overlooked when information is developed from the narrower perspective of the decision-maker" (p. 550). Employee input to the appraisal process thus may be necessary for providing genuine balance to the performance information considered by an appraiser. Employee input to evaluation decision-making is just one example of the due process metaphor's implications for performance appraisal. Our discussion of procedural due process will identify further PA implications of the metaphor.

According to Forkosch (1958), due process's essential features[1] are adequate notice, a fair hearing, and judgment based on evidence; we examine each feature separately in the next section (see Table 1). Our discussion of due process also encompasses procedural justice and interactional justice (Bies & Moag, 1986). Procedural justice involves the formal, structural characteristics of a system (e.g., the existence of rules providing opportunities for participation or appeal of ratings), and interactional justice refers to the nature of the communication between parties (e.g., information conveyed in a manner communicating respect and honesty).

Adequate Notice in PA

Adequate notice is simple in principle yet profound in implication. It implies, for example, due notice given via the publication, distribution, and explanation of standards for performance (policies conforming to the legal practice of defendants' receiving written notice of charges against them). More importantly, the act of giving notice has implications for a variety of rights subsumed under the general notion of due process.

For example, the courts can hold citizens accountable to the dictum that "ignorance of law is no excuse" only if the laws were published. Similarly, employees should not be held accountable for their ignorance of performance standards unless sufficient efforts have been directed toward making these standards known and understood. Lawler, Mohrman, and Resnick (1984) presented descriptions of performance appraisals from 700 manager-subordinate pairs employed at GE. The subordinate participants reported often being surprised by appraisal interviews called on short notice. Furthermore, only one-third of the subordinates agreed that the appraisal interview was based on predetermined goals. Proper understanding also requires adequate explanation, which might entail being told not only what to do but also how and why. As a consequence, a degree of participation and two-way communication might be introduced into the process of PA.

Adequate notice in PA also provides employees information essential for questioning the application of standards, a necessary condition of our next due process principle: fair hearing. When adequate notice involves a chance for discussion of the standards, for example, employee input may help to formulate objectives and standards. Finally, the concept of adequate notice

extends to feedback given throughout the PA process, which should be given on a regular and timely basis. These and other issues are nascent to the concept of adequate notice in its legal context; appraisers and appraisees would both be better equipped to assure procedural fairness in PA if adherence to fair standards for adequate notice were observed.

Fair Hearing in PA

This aspect of due process guarantees consideration of all relevant evidence. Selznick (1969) has stated that the principles of evidentiary hearings include, first of all, informing the party of the tentative assessment of the facts (re-emphasizing the special importance of adequate notice) and allowing the affected party to present his or her own version of the facts. In PA, observing this principle would correspond to giving employees opportunities for hearing the appraiser's description of the performance observations that had been made, then allowing employees to provide their own commentary about those observations.

Selznick's second principle is that evidentiary hearings require the exclusion of unreliable material. Reliability is pursued in the courts not so much by technical requirements as by means of the adversarial process itself, as Selznick's third principle stipulates: Due process is said to involve "protection of the adversary principle as a high road to reliability in the assessment of legal fact" (p. 254). Application of the adversary principle to PA implies that the reliability of evidence concerning performance will be checked by allowing employees to challenge (or rebut arguments about) the sources and nature of the evidence.

By linking reliability to the adversary process, Selznick's third principle implies that two-way communication and self-appraisal (SA)—giving one's own version of the facts—should be a part of PA. In the psychometric tradition of the test metaphor, SAs are suspect because of alleged bias from a motivation to put one's best food forward; in the legal tradition, the same bias is encouraged because it is considered essential to the pursuit of justice. Due process requires disputants to present their case in the most favorable light, whereupon a judge or jury is responsible for weighing the merits of the respective presentations.

The contending parties in a civil suit, for example, both try to present the most favorable possible claim to the same benefits (e.g., contested property). When PA decisions govern the distribution of scarce resources such as pay and promotions, employees are placed in positions analogous to the contending parties in a civil suit. For a variety of reasons, appraisers cannot always directly observe performance related behavior; thus PA typically evaluates *evidence* of employee performance—not the performance itself. Employees want their performance to be evaluated favorably and in light of all relevant information.

Becau
vital
justifi
advoc
derive
stereo
An
have i
on idi
be ove
of a de
and a
Simile
admin
consti
Inadec
examp
emplo
In
subsur
counse
told at
Likew
of the
readily
inform
proces:
to prov
as wel
increas
objecti
manag
This pe
apprais
monito

The
judgme
PA assi
provisic
by a su

Because of appraisers' biases and cognitive limitations, employees provide a vital source of additional performance information. Thibaut and Walker's justification for adversary proceedings implies, in other words, that unless self-advocacy is permitted, decision makers are subject to the bias of mental models derived from their own unique perceptions (e.g., actor-observer biases, stereotypes, implicit personality theories).

Analyses based on social cognition research (e.g., Ilgen & Feldman, 1983) have identified precisely this propensity—namely appraisers' frequent reliance on idiosyncratic schemata and prototypes—as the predominant difficulty to be overcome in PA. Adversarial proceedings address this difficulty; the biases of a decision maker can potentially be offset when disputants present evidence and arguments the decision maker might not otherwise have recognized. Similarly, incorporating employee SA mechanisms into PA might help to make administrative appraisers more aware of such otherwise neglected factors as constraints on performance (cf. Peters, O'Connor, & Eulberg, 1985). Inadequate technology, machinery, and supplies often limit performance, for example, and the role these environmental factors play in constraining employee productivity is a critical element in a performance evaluation.

In legal proceedings, the right to counsel helps protect the principles subsumed under fair hearing. We do not claim employees should be given legal counsel for PA, but due process does imply that employees should be not only told about the appraisal rating itself but also told how the rating was derived. Likewise, they should be allowed to raise questions about an appraiser's version of the facts and justification for the process. Even if such opportunities are readily available, the competence to question an appraiser's judgment in an informed manner may require more intimate knowledge of the appraisal process than organizations normally provide. Thus, one way for organizations to provide employees with needed information would be to train employees, as well as appraisers, in the rating process. Precedents already exist for increased employee participation in appraisal processes. Management by objectives, for example, involves employees early in the appraisal process when managers and their employees develop mutually acceptable performance goals. This participatory process could be extended in both traditional performance appraisals and management by objectives to employee participation in monitoring performance and evaluating progress toward goals.

Judgment Based on Evidence

The third feature identified by Forkosch (1958) as essential to due process, judgment on the basis of evidence presented, has its parallel in the role that PA assigns an appraiser as decision maker. When employment contracts lack provision for formal grievance procedures or for the appeal of decisions made by a superior, the appraiser becomes the final arbiter of disputes about

outcomes. The appraiser's role thus requires special attention in light of Selznick's (1969) final principle, which stipulates that due process calls for "the maintenance of judicial integrity, including freedom from external pressures, personal corruption, and the more evident sources of bias" (p. 254).

If due process represents "the development of guiding ideas . . . for the limitation of arbitrary power" (Selznick, 1969, p. 250), what principles of due process should organizations follow to restrain autocratic tendencies? Selznick calls one of the "special ideals of due process" the principle that "the making and the application of law should affirm reason" (pp. 252-253). Similarly, Pound (1943) says that due process offers the following advice: "If one will is to be subjected to the will of another through the force of organized society, it is not to be done arbitrarily, but it is to be done upon some rational basis, which the person coerced, if reasonable, could appreciate" (p. 34).

Following this principle, PA should provide opportunities for explanation and discussion regarding the reasons for appraisal ratings, as well as the reasons for allocation decisions stemming from that appraisal. The perils of trying to generate appreciation merely through the size of raises or the rate of promotions alone parallel the recently identified difficulties in attempting to generate appreciation merely through PA ratings (Pearce & Porter, 1986). We believe appraisers should devote their efforts to making reasonable decisions rather than to attempts at placating employees with inflated ratings—but we add the corollary that reasonable decisions must be reasonably presented. In other words, appraisers must be willing to discuss not only the merits of a PA rating, but the nature and merits of the decision making process itself.

Another aspect of judgment based on evidence is that it entails "a reasoned weighing of the interests involved and a reasoned attempt to reconcile or adjust them" (Pound, 1943, p. 35). It is important to note that judgment based on evidence does not imply that the outcome will be more favorable to the employee. Indeed, if an individual performs poorly, fairness *requires* that corrective action be taken. It is well know that managers generally have trouble giving negative feedback (Larsen, 1984). When a worker is performing poorly, however, due process may actually be helpful to the supervisor. After gathering evidence and properly informing the subordinate of organizational expectations, the supervisor can be more comfortable about the fairness of negative feedback.

An additional means of providing a check on the integrity of a decision-maker is to institutionalize mechanisms for reviewing that person's decisions. The legal system's institutionalized mechanism is the structure of appellate courts. PA appeal mechanisms can be implemented in a number of ways. SAs provide one, informal appeal mechanism. Formal PA appeals can be implemented using procedures similar to progressive discipline systems. First employees may appeal directly to their immediate supervisor. Failure to get a satisfactory response at this point may then trigger appeals to management

ght of
ɔr "the
ssures,

or the
of due
:lznick
1aking
:ilarly,
ie will
ɔciety,
basis,

1ation
:asons
ing to
ɔtions
1erate
·elieve
rather
ld the
other
ating,

;oned
1djust
:d on
o the
· that
ouble
·ɔorly,
ering
ional
:ss of

sion-
,ions.
.ellate
SAs
n be
First
o get
ment

levels beyond the immediate supervisor. Ultimately an oversite committee comprised of individuals representing relevant constituencies may be used to determine the final appeal decision. Rather than assuming PA will properly embody due process by the mere creation of an appeal system, however, we believe close attention must be given to the underlying principles that appeal systems reflect.

Forkosch (1958) notes two features of appellate review. First, the courts have ruled that appeal is neither essential to due process nor a condition for impartial judgments. Second, because an appellate court ordinarily reviews only the procedural aspects of a lower court's decision, appeal is also not a sufficient condition for impartiality (i.e., an appellate court's lack of authority to review a case on its merits means that a lower court's substantive errors may go unchecked). Similarly, institutionalization of appeal mechanisms may be neither necessary nor sufficient to promote due process in PA decisions. We recommend that organizations consider implementing appeal procedures, but such procedures are not a panacea. Over-reliance on appeal as a corrective device may cause neglect of other ways to combat bias.

Additional bias-reducing efforts should include the best measurement technology available (e.g., job analysis, behavioral rating scales, systems of rater training). It is essential, however, that appraisers not regard a package of measurement technologies, or even a set of specific rules, as comprising the elements that ensure due process has been implemented. Measurement technologies may bear the *imprimatur* of science, but no scientifically devised system provides a fool-proof check on human frailties or completely foils the willful attempts of those determined to beat the system. Similarly, no set of rules is immune to attempts motivated by desires to circumvent them—and rules themselves can be mismanaged in application so as to foster bureaucratic paralysis by red tape and paperwork, rather than to institutionalize the essence of due process as a system of checks-and-balances protecting employee rights. Above all, therefore, due process mechanisms must be implemented in terms of guiding principles (i.e., designed with *process* goals in mind) rather than in a legalistic, mechanical, rote, or "cookbook" fashion.

The crux of our emphasis on the well-reasoned reconciliation and adjustment of conflicting interests is that there can be legitimate grounds for balancing one set of appraisal considerations against another (e.g., considerations of employees' merit vs. their market value and correlative prospects for employment elsewhere). Managers constrained to administer raises and promotions in ways that do not coincide with their perceptions of employees' performance, for example, are not necessarily acting unfairly. The fairness of such decisions depends on the nature of the principles used to justify those constraints. Legitimate principles can be used to justify "overpayment" relative to present productivity, for example, when an employee's valued expertise may be lost if that employee left the firm for higher pay elsewhere. Those constraints

that managers would be ashamed to admit are suspect; those they can explain and reasonably justify may well pass employee muster.

Addressing Possible Limits to the Legal/PA Analogy

One drawback in applying a legal metaphor of due process to PA is that the American legal system has features not present in PA systems (e.g., the neutrality of judges and juries as third-party decision makers). A due process model also contains other connotations that might be inappropriate if applied to PA in an unqualified fashion (e.g., the image of a trial as the final resort for resolving disputes; the adversarial and frequently zero-sum interactions between disputants). Nevertheless, some of these connotations contain kernels of truth that should not be denied just because the implications seem unpleasant. Administrative decisions about pay and promotions *do* require people to be judged and in essence put on trial, despite whatever discomfort appraisers might feel about being placed in the role of judge. Also, some allocations will be zero-sum (e.g., a fixed amount of money available for salary raises), even though appraisers in organizations should be more cognizant of possible integrative ("win/win") solutions.

Another nonanalogous element is the frequent use of PA for purposes other than making administrative decisions. One notable example is the use of PA for developmental purposes such as goal-setting and career counseling, as well as for coaching and motivating employees to improve sub-standard performance. These uses have sometimes been described as so diametrically opposed to the role of judge that they require a "split role" approach in which separate meetings are held for administrative versus developmental feedback (Meyer, Kay, & French, 1965).

We believe that this point has been overstated. Indeed, recent research conducted by Prince and Lawler (1986) at GE, where Meyer et al. tried the split-roles approach, indicates that there are many good reasons why developmental discussions should be linked directly to the discussion of administrative decisions such as salary increases. Prince and Lawler concluded that their results "refute the popular wisdom about the negative impact of salary discussion" (p. 371). They also point out, however, that "To date very little is known about the broader conceptual framework in which salary discussion is embedded" (p. 374). We believe the due process metaphor provides concepts useful for constructing such a framework.

In our view, a due process metaphor can integrate the roles of judge and coach by emphasizing the importance of procedures designed to resolve disputes, such as procedures analogous to those used in the legal system. Whether used for administrative or developmental purposes, PA feedback should involve a process for uncovering and addressing disagreements. Too often managers assume that they know what is best for the employee and the

organization's well-being, rather than recognizing the value of feedback from the employee—in the form of evidence (and even arguments) about what the employee has actually accomplished, as well as in the form of advice about what the organization should be seeking to accomplish. This need for two-way communication follows directly from the observations of Simon (1983), Pfeffer (1981), and others about the multiple constituencies and divergent interests prevalent in organizations.[2]

In the case of motivational discussions about goal-setting, *implicit* disagreements might more often be uncovered by following the model of due process. Managers and employees are potential disputants, for example, when managers consider setting new goals for employees' performance. If goals are assigned rather than derived through employee participation in a decision-making process, the potential disputes involve not only whether the goal is *attainable* but also whether the goal is even *appropriate* for the organization's well-being (e.g., the employee might believe increased quotas are detrimental to product quality). To the extent that attention to due process provides additional mechanisms for employee input, enhances existing mechanisms, or simply encourages appraiser and appraisee to take one another's viewpoints more seriously, increased opportunities will be created for the exposure of disagreements about the appropriateness of various organizational goals. Why encourage such arguments? As Staw and Boettger (1990) have noted in the course of reporting empirical evidence to be described below, the most efficient methods of production can actually cause damage when misguided goals are pursued. The same principle is expressed colloquially, albeit in jest, when someone says "I have no idea where we're going, but we're sure making good time."

Unfortunately, some discussions of participatory goal-setting's potential benefits for enhanced employee commitment emphasize only how such benefits can be derived because of increased information that flows *to* employees; the due process metaphor stresses the additional benefit of information gained *from* employees when they have opportunities to describe their own reasons for believing that some goals are better than others (cf. Nemeth & Staw, 1989; Near & Miceli, 1987).

Our discussion in this section has addressed possible limits to the due process metaphor by comparison with the legal context from which it was derived. Another way to address the metaphor's potential limitations, however, further applies its own essential insight (the value of competing views) by calling for direct comparisons between the due process metaphor and the other two considered earlier. Comparing the due process metaphor with the other metaphors allows us to address possible misunderstandings that might otherwise arise. Advocates of the test and political metaphors might challenge the utility of the due process metaphor in a variety of ways. As a way of elaborating further implications of the due process metaphor and clarifying its use, we consider separate critiques from the perspective of the test and political metaphors below.

TEST VERSUS DUE PROCESS

Comparing the test and due process metaphors can clarify whether more than one metaphor is needed in the first place. If those metaphors do not differ substantially in important ways, then why create conceptual litter with neologisms—or with "fresh paths" that turn out to be dead-end streets? Rather than engaging in a simplistic compare-and-contrast exercise, we want to identify some of the sharpest possible attacks that test metaphor advocates might direct against the due process metaphor.

Test metaphor advocates might first suggest that there is no need to consider possible implications or applications regarding the due process metaphor, because any additional advice related to "being fair" represents no more than mere elaborations of principles and practices already implicit in the test metaphor itself and long championed by test-metaphor advocates. Does the field require, for example, a due process metaphor as grounds for suggesting that employees ought to be informed of what their ratings are and how they were derived? Or that employees should be allowed to question the results? These things are already recognized as sound practice, so the argument runs, and hence there is no need to make these practices "laws" or to treat them so seriously as to threaten an overly "legalistic" orientation. If the test and due process metaphors do indeed overlap to such a considerable extent, then the due process metaphor is trivial and not worth the trouble of pursuing any further.

How can the due process metaphor be both trivially redundant with the test metaphor, and at the same time be the subject of attack by test-metaphor advocates who consider the due process metaphor to be controversial? Indeed, some of our test-metaphor colleagues find the due process metaphor to be offensive and dangerous. One of the most controversial aspects of the due process metaphor involves the extent to which it characterizes performance appraisals as inherently adversarial and lauds that feature as a worthwhile, necessary, and even morally obligatory part of appraisal as a process.

A related test-metaphor advocate criticism might suggest that the adversarial implications of a due process model not only give an undesirable image to PA, but also introduce concerns that are irrelevant to the PA process. Test metaphor advocates assume that PA can and should rely on valid observation, and that this type of observation should be used as the basis for resolving disputes.

In rebuttal we point to our earlier review of the test metaphor, where we raised serious questions about the availability of valid measures. We find it naive to assume that contentiousness can be abolished by recommending as an alternative reliance on results from purportedly objective indicators, observations, reports, and recorded data. PA decisions inevitably entail some subjectivity and the exercise of human judgment. Test metaphor advocates

assume the problem of valid measures can be dealt with through discussions of issues such as criterion contamination, deficiency, and relevance. The due process metaphor sees these issues as the tip of an iceberg that ultimately reveals the impossibility of achieving test metaphor goals.

A further "danger" from the adversarial potential of due process—one making this metaphor definitely controversial rather than trivial—is an implied threat to the maintenance of power and the traditional organizational authority structure. The test metaphor emphasizes the importance of rater training and implicitly assumes that the supervisor-as-assessor is the ultimate source of truth. We argue that this hidden assumption of the test metaphor may unwittingly, or even on at least some occasions perhaps intentionally, connote a misleading image of managers as the source of *all* truth, thereby effectively diminishing the viability of the employee's role as an additional source of potentially important observations, insights, and commentaries regarding their own performance. In other words, we believe that there are dangers in over-emphasizing the importance of maintaining power and authority in managerial roles.

The types of dangers that might be encountered, of course, vary with the nature of the situation. Our earlier critique of the test metaphor had already noted the difficulties it faces because of increasing uncertainties about means-ends links and rapid changes related to the technology of the production process. Such dangers recall Katz and Kahn's (1966) warnings about the need to encourage innovation and spontaneity in the workplace—an unlikely occurrence when strict adherence to authority is the norm. Over-emphasis on the importance of maintaining power and authority in managerial roles can sound the death-knell to creativity, can thwart efforts to keep pace by means of flexibility, and can— in the extreme—foster the other forms of organizational pathology so amply documented by work such as Milgram's (1974) on the harm from blind obedience to authority and Janis' on the symptoms of Groupthink (1982).

More recently, Staw and Boettger (1990) have shown that even something so mundane as simple actions to correct faulty tasks or misdirected work roles can vary greatly in its probability of occurrence as a function of how authority relations are designed and implemented. An ironic finding of theirs, for example, concerned a goal setting effect that they had predicted and yet that seems surprising in light of goal-setting's general effectiveness. Staw and Boettger noted that although goal-setting techniques may indeed be highly reliable in maximizing performance, nothing about such techniques in and of themselves guarantees that the *direction* set by the goal will necessarily be one of true benefit for the organization (our discussion of the test metaphor's *Assumption #3* is also relevant). As they put it, goal setting "may simply increase behavior in the prescribed direction"—the difficulty being that "an increase in the targeted direction may come at the expense of more spontaneous or innovative actions" (p. 7).

It is one thing to assert that superiors ought to, and in fact do, have the responsibility of making final judgments; it is quite another matter—and, we would argue, a serious mistake—to assert or imply that this legitimate authority entails a unique and direct path to truth. (Just as emphatically, however, we are not asserting that employees have a monopoly on truth, or that the employees' perspective is necessarily or even more frequently any better than a given manager's perspective). We do question, however, whether the power accompanying such responsibility ought to be unchecked. Indeed, the power associated with the manager's ability to use PA ratings for compensation and advancement decisions is so substantial, and has such potential impact on the lives of employees, that any system of PA ought to incorporate mechanisms designed to abate the use of unfettered power. Managers are in positions of great power already. Efforts to sanction and protect their power even further, whether under the guise of access to truth or possession of legitimate authority, recall the dangers indicated by Lord Acton's dictum: Power corrupts, and absolute power corrupts absolutely. Our position extends the familiar notion that authority must entail responsibility as a concomitant; again, the point is to recognize an inevitable tension between power and accountability, and to design systems of checks and balances so that this essential tension is managed in a healthy and constructive manner.

The due process metaphor directly addresses issues concerning the balance of power between employee and manager. As indicated by the Thibaut and Walker (1975) model of procedural justice, power is differentiated at two stages of the decision-making process: (a) presentation and interpretation of evidence, and (b) the final decision. This model asserts that a procedurally fair method of resolving disputes consists of allowing the decision-maker to exercise total control of the second stage, while allowing disputants some indirect control over the first stage via their participation in the process.

In light of questions about management power and prerogatives, note that an important distinction exists between two types of arguments for allowing employees some form of control over, or participation in, the stage of the PA process at which performance evidence is collected and interpreted. The first is based on the principle of triangulating on truth from different directions, namely that employees may have access to performance information not available to the manager and may have arguments about the interpretation of evidence that would not otherwise have occurred to the manager. This argument emphasizes the same aims as those pursued by the test metaphor— the search for truth—but notes that such aims can be pursued in more than one way. In short, this type of argument suggests that even if the due process metaphor and the test metaphor were identical with respect to desired ends (i.e., if due process were conceived as a method for improving the efficiency by which truth and accuracy could be obtained, and thus shared with the test

have the
er—and,
egitimate
hatically,
on truth,
ently any
however,
checked.
ratings
has such
ought to
power.
tion and
to truth
by Lord
solutely.
t entail
evitable
f checks
thy and

balance
aut and
o stages
vidence,
method
ise total
control

ote that
allowing
the PA
The first
rections,
ion not
retation
er. This
aphor—
re than
process
ed ends
ficiency
the test

metaphor the desired end of scientific objectivity), the two metaphors would differ in promoting different *means* for the same ends.

There is also, however, a second type of argument for allowing some form of employee input to the PA process—an argument that differentiates the two metaphors in terms of ends rather than means. In particular, an alternative or additional option for a desired end, rather than pursuing accuracy or objective truth exclusively, is based on the presumed rights of employees to have some say in matters that affect their destiny. This latter form of argument bears on issues regarding management philosophy and ultimately involves principles of morality, which may well be controversial (in a fashion similar to controversy about the employment-at-will doctrine; see Youngblood & Bierman, 1985). The second type of argument deserves serious consideration; active debate about the scope and nature of employee rights should be encouraged. Unfortunately, perhaps one reason why such debate has not been vigorously pursued is that the issue of employee rights does not get raised in the first place, or is quickly dismissed as meaningless or irrelevant if and when it is raised.

Finally, due process's adversarial nature might suggest that allowing a give-and-take discussion between rater and ratee might cause some ratees' scores to be determined by litigatory skill rather than by performance-related contributions. Thus the test-metaphor advocate might claim that the due process metaphor encourages an overly litigious and unduly adversarial atmosphere, whereas objective ratings of past achievement make contention and argument unncessary. Once again the difficulty of attaining objective truth in PA ratings deserves further comment.

Contemporary philosophers of science have often attacked the "received view" promoted by logical positivism. Among the components of that view most repeatedly attacked, to the point where its successful defense now seems completely unlikely, is the original empiricist claim that observations as sense-data can be attained in a direct and uninterpreted manner. Rather, the modern view is that all observations are to some degree "theory laden" (a term originally coined by Hanson, 1972). The theory-laden nature of PA ratings is vastly underplayed by the psychometric tradition on which the test metaphor is based. Instead of being based on the assumption that "past achievements" are merely data to be observed, the due process metaphor fits the assumption that PA ratings are inevitably imbedded within implicit *theories* about what constitutes good performance.[3]

Performance ratings contain implicit assumptions of a theory-laden nature because such assumptions enter the performance appraisal process at two points: (a) as part of the act of establishing performance criteria (i.e., in the creation of "implicit theories of performance"), and (b) as part of the act of observation itself (viz., because observation is "laden" or "driven" by theory in the sense of background knowledge, assumptions, and experience). We address each point in turn.

1.   *Establishing performance criteria involves having a theory of performance.*   To become a believer that different "theories of performance" exist, observe the variety of diverging orientations and strategies pursued by different individuals in one firm. Consider the example of a customer in a department store returning a previously purchased item without a sales receipt. Both direct sales personnel and managers will have opinions about this event, but these opinions may be quite different. The divergence in their approaches to the business of retailing carries with it a divergence in theories of performance.

Quite simply, sales personnel may be in favor of accepting returned merchandise without a sales receipt because of the good will it engenders on the part of the customer, the likelihood of encouraging return business, and the implications this has for future sales. Management, on the other hand, may be under directives from higher organizational levels to tighten fiscal controls and to establish constraints on salespersons' behavior that will accomplish this fiscal directive. Not accepting returned purchases unless there is an accompanying sales slip may be one such constraint. From the managers' perspective, this constraint accomplishes a budgetary goal; from the employees' perspective, this constraint is counter to their goal of increasing sales through customer service. Not only may employee and management perspectives differ, but managers themselves may be at odds with each other. Some managers' sales philosophies may include recognition of the impact of quality customer service directly on the bottom line and other sales managers may believe that tight controls are the route to profit. Each of these multiple perspectives are indicative of different theories of performance, and each will have implications for appraising performance.

A comprehensive treatment explicating performance theories as a construct, however, involves a lengthier and more complicated set of issues than the above example has suggested so far. Put another way, full-blown theories of performance involve multiple elements and multiple levels of analysis. Part of the basis for divergence between various implicit theories of performance, for example, begins with the prospect for differences in the central goal or "mission" of organizations. In our example, each individual may provide a very different answer to the question "What are we trying to accomplish?"—even though they are in the same organization. One individual may define the primary strategic orientation in terms of budgetary controls; another may define the primary goal as providing a special quality of service to customers.

Responding to problems created by multiple theories of performance, the test model essentially recommends methods designed to ensure that everyone uses the same theory. At the very least, solutions such as behavioral anchors and particular types of training approaches are supposed to make the divergences among theories-in-practice as minimal as possible. In this sense of what it means for there to be different theories of performance, therefore,

the test metaphor actually admits that potential problems exist (e.g., discussions about the problems of criterion contamination, deficiency, and relevance). The metaphor's advocates would probably add that routes toward increasingly more sufficient solutions to those problems are continually being charted. The design of rating-scale instruments and formats, and the development of new approaches to training raters, as well as other "technological fix" approaches, are all intended so that everyone involved will—to the greatest possible degree—think about performance the same way. Perhaps such attempts may someday succeed, but the evidence so far is not very encouraging.

2. *Observation is inherently and hence inevitably "theoretical."* The impact of this viewpoint on the meaning of observation, and the correspondingly changed perspective about what is entailed by measurement, has been chronicled, for example, by Hanson (1971). Hanson describes earlier measurement views by analogy, suggesting that scientists had at one time naively assumed that observation by means of measuring instruments was akin to capturing the essence of objects by taking pictures with a camera. Some test metaphor advocates seem to act as if they hold similar views. If so, those advocates might find at least some cause for alarm in Hanson's report of a series of developments in physics:

> Without a doubt . . . 'dust-bowl empiricism' derives some of its appeal from such an uncriticized view of the nature of measurement. Only during the scientific revolutions of this century—Relativity Theory and in Quantum Mechanics—have modifications of such a pervasive, powerful, and perennial view been lastingly effected. When practical operationalists (like Mach and Einstein and Bridgman) began searching for the 'cash value' of terms like *mass* and simultaneity and *time*, a certain 'involvement' of subject matter and observer became clear. 'Interaction' is now the watchword. (p. 16)

In particular, Hanson noted "a pervasive interaction between . . . events and our theories of measuring technique"—a comment followed by the additional remark that "numbers emerging from measuring encounters may be the result not of simple, objective data-registration, but of a most intricate enravelment of subject matter, probe, and theory" (p. 17).

In science, different theories are subject to continual debate regarding not only their *validity* in some sense of absolute truth (faithful representation of underlying reality), but also with respect to their *usefulness* for various purposes. Our opinion is that when PA ratings are viewed as theory-laden, they should be similarly subjected to a managed discussion regarding differences of interpretation—and that employees' viewpoints about the validity and utility of PA ratings are a legitimate source for encouraging a constructive exchange about the relevant assumptions and values engendering judgments about performance.

The potential influence of "litigatory skill" also raises some further issues even more complex than those just addressed. We mentioned, for example, that PA ratings might be contaminated by employee differences in skillful argumentation—a dilemma inherent to the due process metaphor, and a serious matter indeed. Note, however, that the same problem occurs within the American adversary system of legal proceedings. Despite the inevitable presence of differing effectiveness among disputants (e.g., the effect of one's financial means on the type of professional counsel that can be employed on one's behalf), the system continues to use the principle of maintaining each disputant's entitlement to self-advocacy.

The charge that the due process metaphor would encourage an overly litigious atmosphere has particularly negative connotations. We believe, however, that organizations are better served by mechanisms deliberately surfacing sources of potential conflict within their own boundaries—lest, in the absence of such provisions, they become increasingly susceptible to legal challenges from outside (e.g., grievance actions). It is better to expose conflict early and deal with it head-on, before it gets out of hand, than to wait until too late and pay a more severe penalty.

Others, including management scholars as far back as Mary Parker Follett (1941), have taken a less defensive view and have instead argued forcefully for the *constructive* functions of organizational conflict (e.g., Baron, 1990; Coser, 1956; Katz & Kahn, 1966). Indeed, there is a striking resemblance between some forms of employee activity labeled *counter-role behavior* and the appropriate challenge to authority that we have in mind under the purview of due process: "Of the many possible counter-role behaviors, taking action to correct a faulty task or misdirected work role is probably one of its most clearly functional forms" (Staw & Boettger, 1990, p. 537). Another form of counter-role behavior, whistle-blowing, might actually save a company from consequences of greater magnitude incurred when problems go unchecked for long periods of time (e.g., with regard to the Challenger space shuttle, engineering a solution to the O-ring problem versus the loss of life incurred when no early engineering solution was developed). Counter-role behaviors such as whistle-blowing can work toward a firm's long-term benefit even during the short-term period of confrontation and conflict when those actions seem more like a mere nuisance (cf. Organ, 1988, on citizenship behavior).

The question then becomes how to *manage* appropriately the debates that may arise from allowing the expression of opposing viewpoints. The due process metaphor addresses precisely this issue, whereas the test metaphor implies that (a) it should never come up in the first place, and (b) if it does, it should be resisted and ultimately swept away by psychometric technology— or if necessary, by management fiat. The test metaphor implicitly assumes that differences of opinion are irrelevant, because objective truth can be discovered and used to resolve disputes. From the perspective of the due process metaphor,

rther issues
ir example,
; in skillful
nd a serious
within the
inevitable
ct of one's
nployed on
uining each

an overly
ve believe,
deliberately
es—lest, in
ole to legal
ose conflict
wait until

iker Follett
rcefully for
990; Coser,
ce between
r and the
ne purview
cing action
of its most
er form of
pany from
hecked for
ce shuttle,
e incurred
behaviors
ven during
tions seem
).

ebates that
. The due
metaphor
if it does,
hnology—
sumes that
discovered
metaphor,

on the other hand, differences of opinion are inevitable because (a) the self-interest of various parties will not coincide, and (b) various parties have different views of the organization's best interest. The due process metaphor grants each party the right to articulate the best possible case for his or her own perspective.

Managing the adversarial quality of conflicting self-interests becomes, from the perspective of the due process metaphor, a duty of the person placed in charge of conducting the PA. Proper management thereby requires recognizing the sources of opposed interests. When PA is used to determine compensation and advancement, the self-interests of different employees are necessarily opposed to one another because of the "fixed pie" or zero-sum nature that characterizes the organization's limited and scarce resources. This fact alone, which is under-emphasized by the test metaphor, makes management of adversarial relations essential.

Additional elaborations concerning the management of adversarial relations among employees lie outside the scope of the due process metaphor per se. This metaphor does draw attention to the related literatures of procedural justice, negotiation, and conflict, however, which supply other models for the resolution of conflict. In particular, useful leads might be obtained by examining recent work on the process of mediation (e.g., Sheppard, 1984) and tactics for seeking integrative or collaborative ("win/win") solutions to problems initially posed in purely distributive (zero-sum) terms (e.g., Pruitt, 1981). Creative opportunities for minimizing contentious behaviors should be explored when resolving PA differences of opinion that surface when applying the due process metaphor. Many such opportunities exist in a management context (e.g., offering job enrichment to employees disappointed with their PA ratings). The range of outcomes that managers can supply thus extends beyond what a judge can do in a courtroom trial, and hence there are enhanced possibilities in organizations for resolving disputes. Brett, Goldberg, and Ury (1990) offer a useful framework that might provide the basis for a more extensive elaboration of these notions.

Another possible difference between managers and judges also requires discussion. A judge may be said to act as a neutral and distinterested third party, whereas a manager's position might be more apt to engender a degree of special interest, personal or parochial (as illustrated by our discussion of politics-and-PA research by Longenecker et al., 1987, and Patz, 1985). Judges, however, do *not* act as totally *disinterested* parties to disputes; rather, the judge's role requires a viewpoint reflecting the good of the state, or society as a whole. Similarly, the conscientious manager's own self-interest is supposed to be suppressed and subjugated to the interests of the organization as a whole. Although perhaps difficult to achieve in practice, this characterization of the manager's role is at least consistent with the ideals of the due process metaphor.

A judge also has the duty of recognizing and taking into account the biases of disputing parties. This role is crucial in the American system of adversarial proceedings, because the disputing parties are expected and even encouraged to present their own cases in the best possible light. Test metaphor proponents instead emphasize using the rating process to *remove* as many forms of bias as possible (because PA is presumably merely an objective "reading" of evidence that allows the PA administrator to retain impartial neutrality)—which has led to the exclusion of self-appraisals (SAs) as a legitimate PA input. Campbell and Lee (1988), for example, concluded that SAs can serve *only* developmental purposes and should *never* be used for evaluative purposes. In contrast, the due process metaphor gives SAs a value analogous with the legal right to present one's case in the best possible light. These implications of the test versus due process metaphors diverge because of differing fundamental assumptions.

Campbell and Lee (1988) explicitly identified certain key assumptions of the test metaphor as related to classic psychometric theory:

> If we view performance ratings as some mix of "true score" and error (Cronbach, 1955), we might conclude that supervisor ratings will reflect more true score and less error than self-evaluations, because supervisory ratings are not self-generated. But it is logically possible that SAs will reflect true performance better than supervisory ratings, or that *neither* set of ratings will adequately indicate true performance. Nonetheless, because most organizations using supervisory ratings assume that such evaluations adequately capture true performance, that convention is maintained here. (p. 303)

The assumed validity in supervisory ratings, and hence use of the supervisor's judgment as a proxy criterion, becomes merely a "convention . . . maintained here," adopted merely out of deference to routine management practice! Yet notably, Campbell and Lee themselves nonetheless cite several references that challenge the validity of supervisory ratings. Sources of invalidity include informational constraints (Mitchell, 1983), the difficulty of separating the duties and boundaries of highly interdependent jobs (Kiggundu, 1981), and inadequate observation opportunities (cf. Kane & Lawler, 1979; Wexley & Klimoski, 1984) because of extensive manager responsibilities (Borman, 1978; Mintzberg, 1973).

For Campbell and Lee, any disagreement in supervisor and subordinate ratings of subordinate performance instead constitutes *prima facie* evidence for abandoning the administrative/evaluative use of SAs (even as an adjunct to the supervisor's ratings). They argued that disagreement between supervisor and subordinate ratings should be treated as a matter of inter-rater reliability rather than as an issue regarding validity—because if validity were an issue, the supervisor's ratings would have to be the criterion (re the passage cited above). This approach, however, begs the question and installs validity by fiat rather than by evidence.

unt the biases
of adversarial
:n encouraged
or proponents
forms of bias
g" of evidence
which has led
out. Campbell
levelopmental
ntrast, the due
ght to present
est versus due
imptions.
mptions of the

nbach, 1955),
ess error than
it is logically
tings, or that
because most
lately capture

e supervisor's
. . maintained
practice! Yet
eferences that
llidity include
eparating the
u, 1981), and
'9; Wexley &
lorman, 1978;

d subordinate
facie evidence
as an adjunct
ten supervisor
iter reliability
were an issue,
passage cited
talidity by fiat

From a due process perspective, bias is not exclusively a reliability issue but instead involves precisely the questions about validity that Campbell and Lee would prefer to assume out of existence. They discounted *any* potential for validity in SAs because of extant evidence that such ratings can contribute to leniency bias (e.g., Meyer, 1980; Shrauger & Osberg, 1981; Thornton, 1980; Tsui & Barry, 1986). Even inflated self-ratings might yield useful information, however, when that leniency from multiple sources of one type (subordinates) offsets the prevailing form of bias from a single, normally all-powerful source of another type (supervisors). For example, subordinates might emphasize contextual constraints on performance (e.g., sub-standard machinery or raw materials) in order to convince a supervisor to inflate PA ratings. This additional contextual information, however, also may provide important performance-relevant information the supervisor would not normally consider, thereby rendering PA ratings more accurate. A supervisor as evaluator can reach an independent judgment after being presented with the subordinates' SAs. (Analogously, the judge in a civil case hears competing self-inflated presentations of merit and must form an independent conclusion, taking the presumed bias of each party's claims about merit into account.)

We do not presume that leniency problems can be swept under the rug; instead, we call for investigations of new methods aimed at solving such problems. For example, leniency might be addressed in an aggregate fashion, such as by giving a higher level of management the oversight capacity to ensure that one lower unit's aggregate scores are not inflated relative to another's. In recognizing potential problems of leniency in SAs, we simultaneously call for others to join in taking such matters seriously—rather than foreclosing investigation of the issues, as occurs when management's judgment gets declared valid as the default option.

Differences between SAs and supervisory ratings are to be expected; the need to understand why and how employees may interpret their contributions from a different perspective than management's is a reason for encouraging supervisors to collect SA data and to discuss discrepant opinions. This need is especially critical because such discrepancies may reflect something other than inter-rater unreliability based on multiple sources' observations about the same object of judgment. The test metaphor's emphasis on reliability implicitly assumes that the object of judgment is the employee's *behavior* (re the extensive literature on behaviorally based rating systems). For the purpose of compensation, promotion, and other decisions regarding the allocation of scarce resources among employees on the basis of merit, however, the relevant criterion is the employee's *performance contribution* to the organization's effective functioning.

The performance criterion comprises the heart of the validity issue that Campbell and Lee pass over so lightly in deferring to a supervisor's judgment. Differences between supervisor and subordinate ratings may reflect not just

different opinions about an employee's behavior, but different evaluative *weights* to various performance components—weights that reflect differing views about what constitutes effective performance (cf. Schmitt, Noe, & Gottschalk, 1986). Implementing supervisory views, of course, reflects existing differences in power—the topic to which we turn now because of its relevance to the political metaphor.

POWER VERSUS DUE PROCESS

Contrasted with the test metaphor emphasis on rationality and objective truth, Pfeffer's (1981) version of the political metaphor tends to stress the role of power directed toward self-interest or the preferences of individuals and coalitions. Pfeffer also noted that rationality-based approaches assume "a unified purpose or set of preferences" (p. 18), whereas the political metaphor implies that "consistency and unity in the goals, information, and decision processes is problematic" (p. 20). The political and due process metaphors thus share a common element missing from the test metaphor—they imply that realistic descriptions of organizations should reflect differences and possible conflicts regarding the preferences, values, interests, and goals of various organizational actors. The due process and political metaphors diverge primarily in their implications regarding the resolution of differences of opinion and competing interests. The political metaphor posits that power inevitably dictates the outcome of decisions, whereas the due process metaphor highlights a mechanism for imposing constraints on the exercise of power.

Pfeffer (1981) considered a variant of the due process metaphor and denied its viability in organizations. The variant, called a *multiple advocacy system*, was proposed by a political scientist, Alexander George (1972), for foreign policymaking. Closely paralleling the due process metaphor, George's (1972) system defined "the role of the chief executive as a 'magistrate' who listens in a structured setting to different, well-prepared advocates making the best case for alternative options" (p. 751). George argued that discourse in such a system "is more likely to secure a critical examination and weighting" of differing viewpoints (p. 752). Thus he proposed the multiple advocacy system as a way of managing conflict in order to promote high quality decisions. Pfeffer (1981) cited Bower's (1965) research as indicating that conflict indeed motivates careful analysis and constructive thinking (cf. Tetlock, 1985)—but Pfeffer nonetheless disavowed the multiple advocacy system's organizational applicability because of its presumed naive assumptions concerning power differences. Citing "the assumption of what is required to fulfill the magistrate role and the assumptions of equal power, resources, and access," he labeled the multiple advocacy system "an interesting model but probably not . . .

effective in practice" (p. 344). Because the multiple advocacy system closely parallels the due process metaphor, we examine these challenges in some detail.

Pfeffer's (1981) chief objection to the magistrate role was that it "expects the chief executive to have great wisdom and insight" because it "makes a presumption of the chief executive as being above the fray, of being able to get beyond parochial interests and personal concerns to make a wise and intelligent choice" (p. 344). This objection, however, misconstrues the expectations of advocacy-based proposals such as those shaped by the due process metaphor. No one should expect magistrates to be all-knowing and infinitely wise, nor should anyone expect supervisors to be above parochial interests and personal concerns. The very likelihood that such interests and concerns will exist, however, makes employee advocacy one of the due process metaphor's vital features. Such advocacy serves a check-and-balance function by constraining a supervisor's parochialism and self-interest.

Employee self-advocacy opportunities can expose differences of opinion in the first place, for example, and then promote discussions that resolve them. Encouraging the employee-advocate role can thereby offer potential benefit to the organization, especially when PA discussions uncover more appropriate performance criteria than had been used previously. An organization further stands to benefit when employees more clearly understand what the existing criteria entail and why those criteria are used. Hence advocacy-based discussions can foster an employee's acceptance of performance criteria and enhance commitment to maximizing them. By means of such discussions, PA sessions become a learning opportunity for both appraiser and appraisee. The due process metaphor does not expect a supervisor as "magistrate" to be inherently wise, therefore, but presumes that he or she can be made *wiser* (or at least better informed). Similarly, as we noted, advocacy procedures also can beget useful occasions during which employees become better informed.

Pfeffer's objection to the role of magistrate also lacks force precisely because when a supervisor is given the authority to make allocations among employees based on PAs, the organization has already placed that supervisor in the role of magistrate. Regardless of how difficult the role may be, a supervisor who conducts PAs is in effect a "judge" making determinations of employees' relative merits. American citizens find it unthinkable to be brought before a judge without the opportunity for presenting their own case. Comparable rights ought not to disappear once a citizen becomes a member of an organization. The due process metaphor thus stipulates that when resource allocations are based on PA, employees have a right to present arguments and interpretations regarding their contributions to the organization's well-being—and, if need be, to challenge a supervisor's opposing views.

We have claimed that advocacy procedures offer potential benefits to the organization and thus can function as much more than a "mere" employee right or an organization's moral obligation. Moral obligations should also

matter, however, and should be given consideration alongside cost-benefit calculations. Pfeffer's version of the political metaphor gives short shrift to moral and ethical considerations, whereas the due process metaphor brings them to a center-stage position. In this regard, we side with Tetlock's (1985) version of the political metaphor, which places greater emphasis on accountability than on power. Accountability, however, varies in its implications depending on the constituency to whom one is held accountable (cf. Cummings & Anton, 1990). Presently, in the typical organization, most PA systems hold supervisory appraisers accountable chiefly to those at *higher* levels in the organizational chart. In many political institutions of Western society (and to a growing degree elsewhere), on the other hand, the typical politician faces accountability from *below*—that is, from the citizens whom the politician represents—because such officials are often elected rather than appointed. Clearly, if our suggestions regarding ways to incorporate self-advocacy into PA systems were adopted, organizations would become more like democracies than they are now (and less like dictatorships).

The advocacy implications for appraisal methods and techniques remaned to be worked out, hence we do not take an a priori stand regarding how much of a challenge to an organization's goal-and-authority structure is healthy. Rather, in some respects we would be content with the less ambitious goal of having an additional group (employees) be more often considered a bona fide constituency in the first place. As Simon (1957), Barnard (1938), and others have noted, an economic exchange that involves hiring personnel as full-time employees (as opposed to the type of exchange wherein a person contracts with someone else on a one-time basis, such as for purposes of having one's driveway paved or shutters repaired) typically also involves a certain volitional *surrendering* of authority: Upon accepting the terms of employment, the employee voluntarily agrees to accept management's authority within a "zone of indifference." Management practices in some firms at times seem to reflect partial forgetfulness that (a) the exercise of organizational authority is inherently circumscribed by what was granted voluntarily at the outset, and (b) it remains continually circumscribed by the degree of good will that can expand or contract the zone of indifference as time goes by.

Pfeffer's second objection to advocacy-based systems contends that "it is almost impossible to equalize power, much less analytical resources, bargaining skills, and information" (p. 344). Here also Pfeffer has applied too severe a standard, however, as half a loaf is better than none; without some form of employee advocacy, *all* power, resources, and information would be held by supervisors. Allowing employee advocacy shifts the balance of power somewhat, but need not attempt to equalize it. Pure equality might entail decision-making by consensus—an improbability dictating instead that someone be given authority to make final decisions. Such authority necessarily imposes a power differential whose very existence, however, demands that it

cost-benefit
hort shrift to
aphor brings
tlock's (1985)
emphasis on
varies in its
l accountable
ization, most
hose at *higher*
s of Western
d, the typical
itizens whom
d rather than
orporate self-
become more

ques remaned
ing how much
re is healthy.
mbitious goal
dered a bona
8), and others
el as full-time
contracts with
ine's driveway
in volitional
ployment, the
vithin a "zone
eem to reflect
authority is
le outset, and
will that can

nds that "it is
es, bargaining
too severe a
some form of
ild be held by
ce of power
might entail
instead that
ity necessarily
mands that it

not be misused. Employee advocacy diminishes that power (but not ultimate authority) by helping to prevent power from being used in an arbitrary or capricious manner.

Pfeffer's reference to the impossibility of equalizing "bargaining skills" raises a slightly different set of issues. There is, on the one hand, a discrepancy in the strength of the bargaining positions of supervisors vis-à-vis subordinates and, on the other hand, the possibility that one employee is a more skillful bargainer than another. In the case of the former discrepancy, supervisors have an advantage because their arguments are made from a position of greater power and authority; as Pfeffer has commented, "Those social actors with relatively more power are better able to get their perception of the problem and the environment accepted" (p. 141). Even if discrepancies between SA and supervisor ratings become completely negotiable, however, two employees with identical performance might not receive the same final rating if one were more persuasive than the other. Previously we noted that the same dilemma arises in the legal system, which nevertheless tolerates differential argumentative skills as a lesser evil than the abrogation of citizens' right to present their own points of view.

Note that the dilemma of differential argumentative skill is not resolved by prohibiting discussion and argument concerning SA and supervisor rating discrepancies (as in a system where the supervisor obtains SA ratings but does not consult with employees regarding their interpretation), because employees may also differ in their willingness to make exaggerated and self-aggrandizing claims in the first place (i.e., inflated SAs). Rather, it is an advantage of discussing SA ratings with employees that those discussions provide additional opportunities for such inflationary proclivities to be revealed, examined, and addressed.

There may be no structural or systematic solution, however, that fully resolves this dilemma. Instead, the prospects for surmounting this difficulty depend on the supervisor's ability to make independent judgments, which recalls Pfeffer's objection that the requirements of the magistrate role are too severe to be humanly possible. Two points should be made regarding that ability. In the first place, these independent judgments are the responsibility of a supervisor under *any* PA system. A supervisor must always act as an impartial judge, and incorporating SA ratings into the judgment process expands the range of information on which the supervisor can draw in making decisions.

Second, the supervisor's ability to draw independent conclusions also benefits from his or her access to information other than the SA ratings and employees' interpretation of those ratings. For precisely this reason, the due process and test metaphors are supplementary rather than either/or perspectives. The more accurate, reliable, and valid the data that supervisors can obtain from their own observations (via a good job analysis, behaviorally based rating scales, and other tools guided by sound psychometric considerations), the better their ability not to be so easily swayed by differences among employees' persuasive ability.

We think the value of discussions with employees about discrepancies between SA and supervisor ratings outweighs the possible drawbacks. The research by Schmitt et al. (1986), which revealed that supervisors and subordinates differentially weighted performance dimensions, indicates particularly strong grounds for this conclusion. A key difference between employees' and supervisors' perspectives concerns the interpretation of what constitutes a worthwhile contribution to organizational effectiveness. Where these differences exist, they ought to be exposed and discussed. The discussion process is essential to clarifying the nature of job responsibilities. By the means of such discussion, there are increased prospects for both employees and supervisors to be further educated about how the organization best functions.

IMPLICATIONS FOR RESEARCH
AND THEORY DEVELOPMENT

The preceding exploration of PA metaphors has implications for empirical investigation and theoretical development. In this section we first examine unfilled research gaps related to our discussion of the test metaphor assumptions, followed by a review of research and theory development related directly to the due process metaphor.

Further Exploration of the "Test Metaphor"

The first test metaphor assumption we identified asserted that the nature of most jobs allows development of reliable and valid measures. We centered our discussion around Lee's (1985) typology of jobs. Her classification scheme, based on the availability of knowledge concerning (a) job outputs and (b) the means-ends relationship, provides the kernel of a theory of job classification for PA purposes. Further development of this job typology might allow appraisers to match job type to appropriate PA goals (i.e., is accurate rating even a possibility?), rater format, rater training, and PA processes. In other words, a well-developed and accurate typology of tasks might provide the basis for a contingency model of PA. No single PA approach will be consistently the best across all situations, and different PA approaches will have strengths and weaknesses in relation to the type of job being assessed and the nature of the performance questions being asked.

Landy and Farr (1980) noted that no performance appraisal technique has demonstrated a clear advantage over any other, and they called for a moratorium on future format research. Perhaps, however, different formats are most accurate for different types of tasks. Consequently, when the validity coefficients for a variety of instruments are averaged across multiple tasks, overall each shows a small-to-moderate relationship, and none show any clear

advantage. Rather than asking which PA format is the most accurate, the contingency approach implied by Lee suggests that we should consider which format works best for which task. There may be, for example, some types of tasks that cannot be appraised accurately. There has yet been little empirical exploration addressing the descriptive or prescriptive accuracy of Lee's model of PA.

The second test metaphor assumption we identified addressed the limits of raters' abilities to assess performance accurately. Although such limitations have been widely recognized, we noted that current understanding of these limitations has not yet yielded many strategies for more accurate performance appraisal. Further research linking what we currently know about rater's cognitive limitations to the effectiveness of various feasible strategies for avoiding these biases during PA would be a logical development and application of cognitive theory in the PA context.

Existing research provides an indication of the possibilities for this approach. Recently, for example, DeNisi, Robbins and Cafferty (1989) used cognitive theories to devise a system of diary keeping. They noted that the accuracy of performance ratings depends on how the diaries are organized. This diary-keeping approach also has advantages from a justice perspective. In a laboratory study, Greenberg (1987a) demonstrated that the keeping of observational diaries enhanced the perceived fairness of appraisals relative to systems not using such diaries.

A second possible method for enhancing accuracy, involving the use of *subordinate* participation in gathering information, also deserves further investigation. Ilgen and Feldman (1983) and Feldman (1981) have noted that raters' cognitive schemas influence their evaluations by directing attention, affecting recall, and guiding interpretation. Accuracy may thus be enhanced to the extent that varying perspectives are presented. Indeed, this is one reason why research on Behaviorally Anchored Rating Scales has called for employee participation in instrument construction (DeVries, Morrison, Shullman, & Gerlach, 1981; Smith & Kendall, 1963).

As with diary-keeping, participation may also increase fairness. For example, Silverman and Wexley (1984) found that employees reacted more favorably to a performance appraisal instrument when they participated in its design. More research is needed on the role of participation, however, because there seem to be situations where it is less important. In one study, for example, Hillary and Wexley (1974) found that subordinates rated their participation as less important when they perceived themselves as lacking adequate job knowledge, but the conditions most likely to contribute to such perceptions have not been identified.

The final test metaphor assumption was based on the premise that a unified performance criterion can be defined for job performance. We argued that the existence of multiple constituencies in organizations make it highly unlikely

that all parties affected directly or indirectly by the PA process could agree on performance criteria. On a practical level, this dilemma is usually handled by fiat, giving management exclusive authority over performance criteria. Nevertheless, it is critical to explore the theoretical implications of multiple constituency approaches.

Tsui (1990) provides some direction in this regard. Three questions surfaced in her application of the multiple constituencies approach to human resources departments, and these questions are directly applicable to PA:

- Which constituencies are relevant?
- What are the effectiveness criteria associated with each constituency?
- What factors influence these criteria?

Answers to these three questions would provide the conceptual core of a multiple constituencies framework for PA. Research and scholarship that provides the answers to each of these questions would yield the information necessary for both prescriptive and descriptive multiple constituency theories of performance appraisals.

Research and Theory Based on the "Due Process Metaphor"

The due process metaphor elevates the goal of fairness in PA to the same level as accuracy and brings dispute resolution issues to front and center stage. The due process metaphor also speaks to the issue of employee rights in the face of managerial power. Finally, the due process metaphor highlights the critical role of process in resolving managerial dilemmas such as those posed by PA. Implications of each of these points for empirical research and theory development are discussed below.

Fairness versus Accuracy

The due process metaphor distinguishes between the goals of accuracy and justice in PA and highlights the importance of fairness as a legitimate PA goal. Previous PA research has not always recognized this important distinction. Note, for example, that the criterion measure in Landy, Barnes, and Murphy (1978) was the following item: "Has performance been fairly and accurately evaluated?" (p. 752), which confounds the goals of accuracy and fairness.

Vroom and Yetton's (1973) normative decision-making model illustrates the importance of decision-making fairness. The effectiveness of organizational decisions is influenced both by quality (i.e., rationality) and by the extent to which a decision is accepted by subordinates. A decision can be ineffective because it did not utilize all the available information or because it was resisted and opposed by those who had to implement and abide by it. Vroom and

Yetton provide ample evidence that decision-making fairness affects decision acceptance.

The determinants of fairness and accuracy perceptions are sometimes different as are their outcomes, and confounding the goals of accuracy and fairness in PA research limits theoretical development of explanatory models. At the very least, separate measures of fairness and accuracy should be developed to determine the similarities and differences in antecedents and outcomes for employee fairness and accuracy perceptions. These separate explanatory models for employee and manager accuracy and fairness perceptions will have some overlap, but it is likely that there will be important differences. Specifically, the role of PA procedures and dispute resolution factors, which we discuss below, are likely to be highly visible in models of PA fairness where they have not previously played a salient role in models of PA accuracy. We may also learn that the determinants of manager PA perceptions are different from the determinants of employee PA perceptions.

Full understanding of the implications of fairness regarding PA, however, calls for further exploration of the multiple meanings of fairness in the PA context. Distinctions need to be made, for example, between accuracy and fairness of the rating scales and the fairness of the overall PA process. In stressing only the psychometric qualities of PA scales, existing research neglects format properties that might enhance perceived fairness. As Latham (1986, p. 125) noted, "the objective characteristics of the appraisal system such as the scales themselves have not been studied in this regard" (cf. Mount, 1983).

Moreover, the due process metaphor implies that objective characteristics of the scales themselves will not be all that affects their acceptability. The procedures used in *developing* PA rating scales may be important determinants of employee reactions as well. Silverman and Wesley (1984), for example, found that workers allowed to participate in designing a behaviorally anchored rating scale manifested more favorable perceptions of their performance appraisals.

A full understanding of fairness in PA would also require that distinctions be made between procedural and distributive fairness. New avenues of PA research and theory along these lines are suggested by emerging lines of investigation within the procedural justice literature.

Procedural versus Distributive Justice

Just as PA research from the perspective of the test metaphor has drawn attention to the accuracy of the outcomes received as a result of PA decision-making, the due process metaphor and related psycho-legal research draw attention to procedural justice, or the fairness of the processes associated with PA decision-making. Not only is fairness an important criterion for judging the success of PA, but fairness may also subsume accuracy because accuracy is an important indicator of, among other things, fairness. A key theoretical

contribution in PA research would be identifying the relative contribution of procedural fairness of PA systems against the contribution of distributive or outcome-related components, which have been emphasized in research and theory inspired by the test metaphor.

The existing PA research suggests that procedural justice is at least as important as distributive justice (Greenberg, 1987b). Additional research, however, also suggests that procedural justice may be more important that distributive justice in the PA context, at least with regard to certain types of dependent variables. Folger and Konovsky (1989) followed Greenberg's (1986) lead in examining employee concerns about fairness in PA and found that procedural fairness played a more important role than distributive fairness (equity considerations) in accounting for variance associated with organizational commitment and trust in supervisor. Such data illustrate the importance of research that compares the effects of procedural and distributive justice on different outcome variables. Continued development of this line of research is important for sorting out the mechanisms underlying the influence of due process and distributive factors on employee reactions to PA.

One central theoretical issue is identifying the factors that determine when due process is merely a means of achieving distributive justice and when due process has effects in its own right above and beyond those of achieving distributive justice. Research by Tyler (1987) and Tyler and Folger (1980) in the legal context (summarized by Lind & Tyler, 1988), and by Folger and Konovsky (1989) in the context of PA, suggests that procedural justice is more closely related to the evaluation of system or institutional characteristics and distributive justice more closely related to evaluation of specific outcomes. More comparative research of this nature will contribute to our theoretical understanding of the role of different justice aspects in PA decision-making.

Procedural Justice Components

In order to conduct the type of comparative research and theory building described above, the components of due process or procedural justice in the appraisal context need to be established. A number of theoretical frameworks for understanding procedural justice are available (e.g., Forkosch, 1958; Leventhal, 1980; Thibaut & Walker, 1978; Tyler, 1987), but the applicability of these frameworks to the PA context has not been fully investigated.

Work by Tyler and his colleagues (e.g., Tyler, 1984; Tyler & Caine, 1981; Tyler & Folger, 1980; Tyler, Rasinski, & Spodick, 1985), however, has included numerous field studies and has extended procedural justice research beyond the legal domain. Tyler (1987) found, for example, that acceptance of a judgment was affected more by *consideration* of one's opinions than by voice itself. Similarly, Bies (1987) has noted that perceptions of procedural justice are affected not only by the structure of rules and regulations used to make

decisions, but also by the way those decisions are communicated. He identified several different types of "social accounts" used when one party harms another—presentations designed to minimize the likelihood of retaliation by the latter against the former. Similarly, Folger, Rosenfield, and Robinson (1983) have shown that resentment of unfair outcomes is not solely a function of the procedure used to obtain those outcomes; rather, resentment is exacerbated when a questionable procedure is used without sufficient justification. As yet there is no counterpart in PA research, where it would seem exceedingly useful to investigate reactions to performance ratings as a function of different types of justification systems.

Although procedural justice components are being empirically explored in the legal context, the extent of procedural fairness research in the PA context is limited. Both Landy et al. (1978) and Dipboye and de Pontbriand (1981), for example, investigated a small set of procedural components relevant to the PA context (e.g., frequency of evaluation and existence of a formal program). Greenberg (1987a) extended our understanding of PA procedures by identifying five aspects of procedural justice including employee input, two-way communication, right to challenge, rater familiarity with ratee's work, and application of consistent standards. Folger and Konovsky (1989) provided a four-factor model of procedural fairness in PA. The due process metaphor suggests that findings in the area of psychology and law *may* be applicable to the context of PA; the generalizability of these findings has not been empirically established, however, nor has the full range of procedural components relevant to the PA process been identified.

Dispute Resolution and Employee Rights

We pointed out earlier that one novel implication of the due process metaphor is its acknowledgement of the underlying adversarial nature of PA. In contrast to the test metaphor, the due process metaphor does not deny self-interested motives on the part of ratees but instead embraces these as essential to the process of seeing that justice is done. The due process metaphor, therefore, implies that research and theory on dispute resolution is absolutely relevant to the PA context. In the broadest context, the due process metaphor draws attention to all procedural mechanisms for resolving disputes. There are a wide variety of such mechanisms, but their applicability in the PA context is unexplored. The field needs a theory of dispute resolution in the PA context.

The inherently adversarial nature of the PA process also focuses attention on employee rights throughout that process. The due process metaphor's emphasis on rights contrasts sharply with the neglect of rights in mainstream organizational theory but is consistent with the importance of rights in daily affairs (Keeley, 1983). Virtually all organizational participants have rights and act accordingly. In the area of employee rights, descriptive systems need to

be developed and the validity of these systems assessed among various organizational constituencies. Our discussion has indicated some potential avenues for observing employee rights throughout the PA process. We have, for example, already stressed the importance SAs would acquire if PA adopted the adversary principle of employee voice in decision-making, and we have noted the desirability of linking investigations of advocacy to research and theory on dispute resolution. What remains is the need for a comprehensive framework addressing the full spectrum of employee rights and related dispute resolution mechanisms specifically in the context of PA.

CONCLUSION

We close by clarifying some aspects of metaphors, adding a few words of caution, and then briefly summarizing the thrust of our position. At times we have used *metaphor* and *model* interchangeably (cf. Black, 1962; Hesse, 1966), but primarily for variety of expression. None of the three perspectives constitutes a formal model. Nevertheless, metaphors play an important part in science, and that role deserves comment.

Hesse has offered a "view of theoretical explanation as metaphoric redescription" (1966, p. 157). Pointing to the ubiquity (if not the necessity) of devices such as metaphor, simile, and analogy, she noted that "both scientific arguments and ordinary language employ analogy as the normal and not the exceptional case" (p. 131). Moreover, because metaphors connote "associated ideas and beliefs that come to mind" (p. 159), they can contribute to "a *program* for exploration or a *framework* through which . . . [phenomena are] seen" (p. 161, emphasis in original).

Hesse also supplied a caveat: "For . . . terms to constitute a metaphor it is necessary that there should be patent falsehood or even absurdity in taking the [metaphor] literally" (p. 160). Recall the fable of blind men encountering an elephant and later reporting their discoveries (e.g., the assertion, from the man who grabbed the tail, that "an elephant is very like a rope"). To say "PA is very like due process" would be just as absurd— but so would applying either other metaphor too literally. On the one hand, distinctions without a difference generate more heat than light (although obviously we think that our distinctions among PA metaphors make a difference and that they shed new light on an old topic, we can only hope that the light will be greater than the heat). On the other hand, distinctions can also lead to distortions (e.g., the blind men and the elephant). Alternative metaphors of PA, therefore, should be treated as an opportunity to apply three supplementary viewpoints—what might be called the "three faces of performance appraisal."[4]

Thinking about PA in different ways has several potential benefits. Explicitly identifying the core analogies that underlie each perspective on PA, for example, has the benefit of making more explicit an analogy whose further exploration might reveal remaining heuristically fruitful implications. The analogy within a given metaphor implies a comparison (between a system being studied, such as PA, and an image brought to mind), and as Hesse has pointed out, "we do not know how far the comparison extends—it is precisely in its extension that the fruitfulness of the model may lie" (p. 162). A related benefit stems from recognizing that these supplementary perspectives reflect alternative meta-theoretical positions about the fundamental nature of the appraisal process. In light of that prospect, we now offer some final summary comments on our position vis-à-vis the most traditional meta-theory of PA, the test metaphor.

Emphatically, we are not saying that performance appraisal accuracy is unimportant; indeed, an inaccurate evaluation system probably could not exist without its fairness being questioned eventually. We are, however, making three points about accuracy. First, perfect accuracy is extremely unlikely, despite present (and even foreseeable) advances in technology. Second, key decisions about criteria, which precede the time when accuracy becomes an issue, would still be based on human values and biases. Third, even if both accuracy and criteria were perfect, employees could still feel unfairly treated if they object to some aspect of the decision-making or implementation process on grounds of procedural injustice. This perceived injustice can result in problems for the organization, especially if the power aspect of the political metaphor (i.e., "might makes right") represents the only dispute-resolution mechanism available.

As differentiated from test and political approaches to PA, the due process metaphor stresses that appraising people is a matter of judging them, not simply measuring them as if they were to be fitted for new clothes. The due process metaphor argues for procedures appropriate to the management of disagreement rather than for procedures suited only for testing employees. Disagreements and disputes, inevitable within organizations that have multiple constituencies, competing preferences, and differing viewpoints, are best handled by methods modeled after the legal system's governing principle of procedural due process.

The need to gather evidence for PA decisions as accurately as possible always exists. But accuracy has inherent limits, and we have argued that no technological fix can guarantee the correctness of decisions such as those involved in making appraisal judgments. The due process metaphor suggests that if we cannot be sure of being correct, we can at least try to be fair—and to press for much more research concerning how, when, and why PA procedures come to be regarded as fair.

NOTES

1.   See Forkosch (1958) for a discussion of the distinction between substantive and procedural due process. Also note that we use a much broader interpretation of due process than do Aram and Salipante (1981), whose discussion of this principle's application to organizational settings focused primarily on the use of formal grievance mechanisms.

2.   Moreover, Staw (1983, p. 38) has noted that sources of the need for such two-way communication go beyond the existence of multiple criteria for effectiveness. As he put it, even when "organizations seem to be running efficiently . . . [they should] engage in some . . . stretching exercises"; most notably, his first such recommendation was the following: "Question the assumptions of the implicit theories on which performance is judged." He also echoed the thrust of our remarks by indicating that it is "probably healthy for there to be a periodic review of assumptions."

3.   As seen in the prior passage that Footnote 2 cited, Staw (1983) has also pointed out the problems created because "we seem to have . . . [too] much confidence in our implicit theories of performance" (p. 33).

4.   We thank Arthur P. Brief for this suggestion.

REFERENCES

Aram, J. D., & Salipante, P.F., Jr. (1981). An evaluation of organizational due process in the resolution of employee-employer conflict. *Academy of Management Review, 6,* 197-204.

Banks, C.G., & Roberson, L. (1985). Performance appraisers as test developers. *Academy of Management Review, 10,* 128-142.

Baron, R.A. (1990). Conflict in organizations. In K. R. Murphy & S.E. Saal (Eds.), *Psychology in organizations: Integrating science and practice,* (pp. 197-216). Hillsdale, NJ: Erlbaum.

Bazerman, M.H. (1990). *Judgment in managerial decision making.* New York: John Wiley & Sons.

Barnard, C. (1938). *The functions of the executive.* Cambridge, MA: Harvard University Press.

Bies, R.J. (1987). The predicament of injustice; The management of moral outrage. In L.L. Cummings & B.M. Staw (Eds.), *Research in organizational behavior* (Vol. 9, pp. 289-319). Greenwich, CT: JAI Press.

Bies, R.J., & Moag, J.S. (1986). Interactional Justice: communication criteria for fairness. In B.H. Sheppard (Ed.), *Research on negotaiton in organizations* (Vol. 1, pp. 43-44). Greenwich, CT: JAI Press.

Black, M. (1962). *Models and metaphors.* Ithaca, NY: Cornell University Press.

Borman, W.C. (1975). Effects of instructions to void halo error on reliabiilty and validity of performance evaluation ratings. *Journal of Applied Psychology, 60,* 556-560.

Borman, W.C. (1978). Exploring upper limits of reliability and validity in job performance ratings. *Journal of Applied Psychology, 63,* 135-144.

Borman, W.C., & Dunnette, M.D. (1975). Behavior-based versus trait-oriented performance ratings: An empirical study. *Journal of Applied Psychology, 60,* 561-565.

Bowen, D.E., & Schneider, B. (1988). Service marketing and management: Implications for organizational behavior. In B.M. Staw & L.L. Cummings (Eds.), *Research in organizational behavior* (Vol. 10, pp. 43-80). Greenwich, CT: JAI Press.

Bower, J. L. (1965). The role of conflict in economic decision-making groups: Some empirical results. *Quarterly Journal of Economics, 79,* 263-277.

Brett, J.M., Goldberg, S.B., & Ury, W.L. (1990). Designing systems for resolving disputes in organizations. *American Psychologist, 45,* 162-170.

Burnaska, R.F., & Hollmann, T.D. (1974). An empirical comparison of the relative effects of rater response biases on three rating scale formats. *Journal of Applied Psychology, 59,* 307-312.

Cafferty, T.P., DeNisi, A.S., & Williams, K.J. (1986). Search and retrieval patterns for performance information: Effects on evaluations on multiple targets. *Journal of Personality and Social Psychology, 50,* 676-683.

Cameron, K. (1980, Autumn). Critical questions in assessing organizational effectiveness. *Organizational Dynamics,* pp. 66-80.

Campbell, D.J., & Lee, C. (1988). Self-appraisal in performance evaluation: Development versus evaluation. *Academy of Management Review, 13,* 302-314.

Campbell, D.T., & Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research.* Chicago, IL: Rand McNally College Publishing Company.

Cascio, W.F. (1982). *Applied psychology in personnel management* (2nd ed.). Reston, VA: Reston Publishing Company.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39,* 752-766.

Coser, L. (1956). *The functions of social conflict.* New York: Free Press.

Cronbach, L.J. (1955). Processes affecting scores on understanding of others and assumed "similarity." *Psychological Bulletin, 52,* 177-193.

Cummings, L.L. & Anton, R.J. (1990). The logical and appreciative dimensions of accountability. In S. Srivastava & D.L. Cooperider and Associates (Eds.), *Appreciative management and leadership* (p. 257-286). San Francisco, CA: Jossey-Bass Publishers, Inc.

Cvetkovich, G. (1978). Cognitive accommodation, language, and social responsibility. *Social Psychology, 41,* 149-155.

DeNisi, A.S., Cafferty, T.P., & Meglino, B.M. (1984). A cognitive viefw of the performance appraisal process: A model and research propositions. *Organizational behavior and human performance, 33,* 360-396.

DeNisi, A.S., & Williams, K.J. (1988). Cognitive research in performance appraisal. In K. Rowland & G.S. Ferris (Eds.), *Research in personnel and human resources management* (Vol. 6, pp. 109-156). Greenwich, CT: JAI Press.

DeVries, D.L., Morrison, A.M., Shullman, S.L., & Gerlach, M.L. (1981). *Performance appraisal on the line.* New York: Wiley.

Dipboye, R.L., & de Pontbriand, R. (1981). Correlates of employee reactions to performance appraisals and appraisal systems. *Journal of Applied Psychology, 66,* 248-251.

Feldman, J.M. (1981). Beyond attribution theory: Cognitive approaches to performance appraisal. *Journal of Applied Psychology, 60,* 736-741.

Folger, R., & Konovsky, M.A. (1989). Effects of procedural and distributive justice on reactions to pay raise decisions. *Academy of Management Journal, 32,* 115-130.

Folger, R., Rosenfield, D., & Robinson, T. (1983). Relative deprivation and procedural justifications. *Journal of Personality and Social Psychology, 45,* 268-273.

Follett, M.P. (1941). *Dynamic administration.* London: Pitman.

Forkosch, M.D. (1958). American democracy and procedural due process. *Brooklyn Law Review, 24,* 173-253.

George, A. (1972). The case of multiple advocacy in making foreign policy. *American Political Science Review, 66,* 751-785.

Greenberg, J. (1987a). Using diaries to promote procedural justice in performance appraisal. *Social Justice Research, 1,* 219-234.

Greenberg, J. (1987b). Reactions to procedural justice in payment distributions: Do the means justify the ends? *Journal of Applied Psychology, 72,* 55-61.

Hagafors, R., & Brehner, B. (1983). Does having to justify one's decisions change the nature of the judgment process? *Organizational Behavior and Human Performance, 31,* 223-232.

Hanson, N.R. (1971). *Observation and explanation.* New York: Harper & Row.

Hanson, N.R. (1972). *Patterns of discovery.* Cambridge: Cambridge University Press.

Heskett, J.L. (1987, January 26). Thank heaven for the service sector. *Business Week,* p. 22.

Hesse, M.B. (1966). *Models and analogies in science.* Notre Dame, IN: University of Notre Dame Press.

Hillary, J.M., & Wexley, K.N. (1974). Participation in appraisal interviews conducted in a training session. *Journal of Applied Psychology, 59,* 168-171.

Huber, G. (1980). *Management decisionmaking.* Glenview, IL: Scott, Foresman, and Company.

Ilgen, D.R. & Feldman, J.M. (1983). Performance appraisal: A process focus. In B.M. Staw & L.L. Cummings (Eds.), *Research in organizational behavior* (Vol. 5, pp. 141-197). Greenwich, CT: JAI Press.

Jaikumar, R. (1986, November-December). Post-industrial manufacturing. *Harvard Business Review,* 69-79.

Janis, I.L. (1982). *Groupthink.* Boston: Houghton Miffling Company.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

Kane, J., & Lawler, E. (1979). Performance appraisal effectiveness: Its assessment and determinants. In B.M. Staw (Ed.), *Research in organizational behavior* (Vol. 1, pp. 425-478). Greenwich, CT: JAI Press.

Katz, D. & Kahn, R.L. (1966). *The social psychology of organizations.* New York: Wiley.

Kavanagh, M.J. (1982). Evaluating performance. In K.M. Rowland & G.S. Ferris (Eds.), *Personnel management* (pp. 187-226). Boston, MA: Allyn & Bacon.

Keeley, M. (1983). Values in organizational theory and management education. *Academy of Management Review, 8,* 376-386.

Kiggundu, M.N. (1981). Task interdependence and the theory of job design. *Academy of Management Review, 6,* 499-508.

Kozlowski, S.W.J., Kirsch, M.P., & Chao, G.T. (1986). Job knowledge rateé familiarity, conceptual similarity, and halo error: An exploration. *Journal of Applied Psychology, 71,* 45-49.

Landy, F.J., Barnes, J.L., & Murphy, K.R. (1978). Correlates of perceived fairness and accuracy of performance evaluation. *Journal of Applied Psychology, 63,* 751-754.

Landy, F.J., & Farr, J.L. (1980). Performance rating. *Psychological Bulletin, 82,* 72-107.

Larsen, J.R., Jr. (1984). The performance feedback process: A preliminary model. *Organizational Behavior and Human Performance, 33,* 42-76.

Latham, G.P. (1986). Job performance and appraisal. In C.L. Cooper & I. Robertson (Eds.), *Review of industrial and organizational psychology.* Chichester, UK: Wiley.

Latham, G., & Wexley, K. (1977). Behavioral observation scales. *Personnel Psychology, 30,* 255-268.

Lawler, E.E., Mohrman, A.M., & Resnick, S. (1984, Summer). Performance appraisal revisited. *Organizational Dynamics,* pp. 20-35.

Lee, C. (1985). Increasing performance appraisal effectiveness: Matching task types, appraisal process, and rater training. *Academy of Management Review, 10,* 322-331.

Leventhal, G.S. (1980). What should be done with equity theory? In K. J. Gergen, M.S. Greenberg, & R.H. Willis (Eds.), *Social exchange: Advances in theory and research* (pp. 27-55). New York: Plenum.

Lind, E.A., & Tyler, T.R. (1988). *The social psychology of procedural justice.* New York: Plenum.

Lingle, J.H., Altom, M.W., & Medin, D.L. (1984). Of cabbages and kings: Assessing extendibility of natural object concept models to social things. In R.S. Wyer & T.K. Srull (Eds.), *Handbook of Social Cognition* (Vol. 1, pp. 71-117). Hillsdale, NJ: Erlbaum.

Longenecker, C.O., Gioia, D.A., Sims, H.P. (1987). Behind the mask: The politics of employee appraisal. *Academy of Management executive, 1,* 183-193.

Lord, R.G., & Foti, R.J. (1986). Schema theories, information processing, and organizational behavior. In H.P. Sims, Jr., & D.A. Gioia (Eds.), *The thinking organization* (pp. 20-48). San Francisco, CA: Jossey-Bass.

McAllister, P.W., Mitchell, T.R., & Beach, L.R. (1979). The contingency model for the selection of decision strategies: An empirical test of the effects of significance, accountability, and reversibility. *Organizational Behavior and Human Performance, 24*, 228-244.

Meyer, H.H. (1980). Self-appraisal of job performance. *Personnel Psychology, 33*, 291-295.

Meyer, H.H., Kay, E., & French, J.R.P.,Jr. (1965). Split roles in performance appraisal. *Harvard Business Review, 43*, 123-129.

Milgram, S. (1974). *Obedience to authority*. New York: Harper & Row.

Mintzberg, H. (1973). *The nature of managerial work*. New York: Harper & Row.

Mitchell, T.R. (1983). The effects of social, task, and situational factors on motivation, performance, and appraisal. In F. Lundy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement and theory*. Hillsdale, NJ: Lawrence Erlbaum.

Mohrman, A.M., Jr., Resnick-West, S.M., & Lawler, E.E., III. (1989). *Designing performance appraisal systems: Aligning appraisal and organizational realities*. San Francisco, CA: Jossey-Bass.

Mount, M.K. (1983). Comparisons of managerial and employee satisfaction with a performamce appraisal system. *Personnel Psychology, 36*, 99-110.

Murphy, K.R., & Balzer, W.K. (1986). Systematic distortions in memory-based performance evaluations: Consequences for rating accuracy. *Journal of Applied Psychology, 71*, 39-44.

Murphy, K.R., Balzer, W.K., Lockhart, M.C., & Eisenman, E.J. (1985). Effects of previous performance on evaluations of present performance. *Journal of Applied Psychology, 70*, 72-84.

Murphy, K.R., Gannett, B.A., Herr, B.A., & Chen, J.A. (1986). Effects of subsequent performance on evaluations of previous performance. *Journal of Applied Psychology, 71*, 427-431.

Nathan, B.R., & Lord, R.G. (1983). Cognitive categorizational and dimensional schemata: A process approach to the study of halo in performance rating. *Journal of Applied Psychology, 68*, 102-114.

Near, J.P., & Miceli, M.P. (1987). Whistle-blowers in organizations: Dissidents or reformers? In L.L. Cummings & B.M. Staw (Eds.), *Research in organizational behavior* (Vol. 7, pp. 321-368). Greenwich, CT: JAI Press.

Nemeth, C.J., & Staw, B.M. (1989). The tradeoffs of social control and innovation in groups and organizations. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 22, pp. 175-210). New York: Academic Press.

Organ, D.W. (1988). *Organizational citizenship behavior: The good solider syndrome*. Lexington, MA: Lexington Books.

Ouchi,W.G. (1977). The relationship between organizational structure and organizational control. *Administrative Science Quarterly, 22*, 95-113.

Patz, A.L. (1975). Performance appraisal: Useful but still resisted. *Harvard Business Review, 53*, 74-80.

Pearce, J.L., & Porter, L.W. (1986). Employee responses to formal performance appraisal feedback. *Journal of Applied Psychology, 71*, 211-218.

Perelman, C. (1967). *Justice*. New York: Random House.

Peters, L.H., O'Connor, E.J., & Eulberg, J.R. (1985). Situational constraints: Sources, consequences, and considerations. In K.M. Rowland & G.R. Ferris (Eds.), *Research in personnel and human resources management* (pp. 79-113). Greenwich, CT: JAI Press.

Pfeffer, J. (1981). *Power in organizations*. Marshfield, MA: Pitman.

Phillips, J.S. (1984). The accuracy of leadership ratings: A cognitive categorization perspective. *Organizational Behavior and Human Performance, 28*, 143-163.

Pound, R. (1943). A survey of social interests. *Harvard Law Review, 57*, 1-39.

Prince, J.B., & Lawler, E.E. (1986). Does salary discussion hurt the developmental performance appraisal? *Organizational Behavioral and Human Decision Processes, 37,* 357-375.

Pruitt, D.G. (1981). *Negotiation behavior.* New York: Academic Press.

Quinn, J.B., & Cagnon, C.E. (1986, November-December). Will services follow manufacturing into decline? *Harvard Business Review,* p. 95.

Ronan, W.W., & Prien, E.P. (1971). *Perspectives on the measurement of human performance.* New York: Appleton-Century-Croft.

Rozelle, R.M., & Baxter, J.C. (1981). Influence of role pressures on the perceiver: Judgments of videotaped interviews varying judge accountability and responsibility. *Journal of Applied Psychology, 66,* 437-441.

Saal, F.E., & Landy, F.J. (1977). The mixed standard rating scale: An evaluation. *Organizational Behavior and Human Performance, 18,* 19-35.

Schmitt, M., Noe, R.A., & Gottschalk, R. (1986). Using the lens model to magnify raters' consistency, matching, and shared bias. *Academy of Management Journal, 29,* 130-139.

Schwab, D.P., Heneman, H.G., & DeCottis, T.A.S. (1975). Behaviorally anchored rating scales: A review of the literature. *Personnel Psychology, 28,* 549-562.

Selznick, P. (1969). *Law, society, and industrial justice.* New York: Russell Sage Foundation.

Sheppard, B.H. (1984). Third party conflict intervention: A procedural framework. In Staw, B.M., & Cummings, L.L. (Eds.), *Research in organizational behavior* (Vol. 6, pp. 141-190). Greenwich, CT: JAI Press.

Shostack, G.L. (1981). How to design a service, In J.H. Donnelly, & W. R. George (Eds.), *Marketing of services.* Chicago: American Marketing Association.

Shrauger, J.S., & Osberg, T.M. (1981). The relative accuracy of self-predictions and judgments by others in psychological assessment. *Psychological Bulletin, 90,* 322-351.

Silverman, S.B., & Wexley, K.N. (1984). Reactions of employees to performance appraisal interviews as a function of their participation in rating scale development. *Personnel Psychology, 37,* 703-710.

Simon, H.A. (1957). Causal ordering and identifiability. In H.A. Simon (Ed.), *Models of man.* New York: Wiley.

Simon, H.A. (1983). *Reason in human affairs.* Stanford, CA: Stanford University Press.

Smith, P.C., & Kendall, L.M. (1963). Retranslation of expectations: An approach to the construction of unamiguous anchors of rating scales. *Journal of Applied Psychology, 47,* 149-155.

Smither, J.B., Barry, S.R., & Reilly, R.R. (1989). An investigation of the validity of expert true score estimates in appraisal research. *Journal of Applied Psychology, 74,* 143-151.

Staw, B.M., & Boettger, R.D. (1990). Task revision: A neglected form of work performance. *Academy of Management Journal, 33,* 534-559.

Tetlock, P.E. (1983). Accountability and the complexity of thought. *Journal of Personality and Social Psychology, 45,* 74-83.

Tetlock, P.E. (1985). Accountability: The neglected social context of judgment and choice. In L.L. Cummings & B.M. Staw (Eds.), *Research in organizational behavior* (Vol. 7, pp. 297-332). Greenwich, CT: JAI Press.

Thibaut, J.W. & Walker, L. (1975). *Procedural justice: A psychological analysis.* Hillsdale, NJ: Erlbaum.

Thibaut, J.W. & Walker, L. (1978). A theory of procedure. *California Law Review, 66,* 541-566.

Thompson, J.D. (1967). *Organizaions in action.* New York: McGraw-Hill.

Thorndike, R.L. (1949). *Personnel selection: Test and measurement techniques.* New York: Wiley.

Thornton, G.L. III. (1980). Psychometric properties of self-appraisals of job performance. *Personnel Psychology, 33,* 263-271.

Tsui, A.S. (1991). Multiple constituency model of effectiveness: An empirical examination at the human resource subunit level. *Administrative Science Quarterly, 35,* 458-484.

Tsui, A.S., & Barry, B. (1986). Interpersonal affect and rating errors. *Academy of Management Journal, 29,* 586-599.

Tsui, A.S., & Milkovich, G.R. (1987). Personnel department activities: Constituency perspectives and preferences. *Personnel Psychology, 3,* 519-537.

Tyler, T.R. (1984). The role of perceived injustice in defendants evaluation of their courtroom experience. *Law and Society Review, 18,* 51-74.

Tyler, T.R.. (1987). Conditions leading to value expressive effects in judgments of procedural justice: A test of four models. *Journal of Personality and Social Psychology, 52,* 333-344.

Tyler, T.R., & Caine, A. (1981). The influence of outcomes and procedures on satisfaction with formal leaders. *Journal of Personality and Social Psychology, 41,* 642-655.

Tyler, T.R., & Folger, R. (1980). Distributional and procedural aspects of satisfaction with citizen-police encounters. *Basic and Applied Social Psychology, 1,* 281-292.

Tyler, T.R., Rasinski, K., & Spodick, N. (1985). The influence of voice on satisfaction with leaders: Exploring the meaning of process control. *Journal of Personality and Social Psychology, 48,* 72-81.

Vroom, V.H., & Gier, J.A. (1989). Ceilings in the reliability and validity of performance ratings: The case of expert raters. *Academy of Management Journal, 32,* 213-222.

Wexley, K.M., & Klimoski, R. (1984). Performance appraisal: An update. In K. Rowland & G.R. Ferris (Eds.), *Research in personnel and human resources management* (Vol. 2, pp. 35-79). Greenwich, CT: JAI Press.

Williams, K.J., DeNisi, A.S., Meglino, B.M., & Cafferty, T.P. (1986). Initial decisions and subsequent performance ratings. *Journal of Applied Psychology, 71,* 189-195.

Youngblood, S.A., & Bierman, L. (1985). Due process and employment-at-will: A legal and behavioral analysis. In K.M. Rowland & G.R. Ferris (Eds.), *Research in personnel and human resources management* (Vol. 3, pp. 185-230). Greenwich, CT: JAI Press.