

Discretion and bias in performance evaluation*

Canice Prendergast

University of Chicago, Chicago IL, USA

Robert Topel

National Bureau of Economic Research, Cambridge MA, USA

1. Introduction

Most of the economics literature on compensation and organizations builds from the theory of agency.¹ For the most part, the literature analyzes situations in which agents' performance can be controlled by tying compensation to objective performance measures such as output or sales. It ignores the fact that most compensation arrangements involve superiors' *subjective*, and hence non-contractible, judgements about employee performance. In our view, much of what is interesting about actual employment relations follows from the observation that 'performance appraisal is a process by which humans judge other humans' [Milkovich and Wigdor (1991)].

This paper studies the implications of subjective performance evaluation for compensation policies and for the efficiency of employment relations. Our objectives are two fold. First, we propose the importance of subjectivity of evaluations to better understand organizational practices such as politicking, favoritism, and compression of wage scales. Second, we hope to orient the study of subjective performance evaluation in ways that are consistent with empirical evidence on what organizations actually do.

One interpretation of the agency literature is that it characterizes situations in which honest principals seek to control the behavior of agents who can't be trusted. Then subjectivity is not an issue; if a principal is known to honestly reveal his measure of an agent's performance, implicit contracts are effectively explicit. But this interpretation leaves aside at least two features of

Correspondence to Canice Prendergast, Chicago Business School, 1101 E 58th Street, Chicago, IL 60637, USA.

*This research was supported by the Centel Foundation/Robert P. Reuss Faculty Research Fund at the University of Chicago Graduate School of Business (Prendergast) and by the National Science Foundation (Topel). Opinions expressed are ours, as are the mistakes.

¹For a useful survey, see Hart and Holmstrom (1985)

real world employment relations. First, in most organizations agency relationships are multi-layered. Supervisors use subjective information to evaluate subordinates' performance and to allocate rewards, but supervisors are not themselves the residual claimants of subordinates' output. This leaves room for supervisors' preferences, and biases, to affect rewards by manipulating the appraisal system. The designs of compensation systems and organizations must account for this behavior.

Second, as we will argue later, in some cases it may serve the organization's interests to suppress information on agents' actual performance. This goes beyond the obvious financial incentive of firms to renege on payments when performance is non-verifiable. We argue that suppressing information on relative performance may enhance overall incentives and output.

2. Compensation with objective and subjective performance

Our principal focus is on the relationship between a worker's performance and his performance appraisal. The classic approach to employee compensation on the other hand, as in Holmstrom (1979), is to consider the optimal responsiveness of compensation to measures of a worker's performance. Among the insights from this approach are that (i) the provision of incentives to workers is traded off against the risk aversion of the worker and (ii) if a performance measure carries any information about the employee's effort, it should be included in the worker's compensation package.

These insights, though fundamental, appear inconsistent with empirical evidence on the compensation of many employees. Most workers face simple compensation schemes in which rewards are insensitive to performance, at least over short periods. One reason for this may be that certain aspects of a job are difficult to monitor and, so, must be determined by implicit rather than explicit contracting. Yet this does not explain why contractible and informative measures of performance are commonly excluded from employment contracts. For example, there are a plethora of statistics about baseball players' performance – hits, walks, batting average, and so on – yet incentive contracts that condition on these observables are not used.²

A possible explanation for the 'flatness' of compensation schemes has recently been made by Holmstrom and Milgrom (1991). They argue that workers often carry out many tasks and that there are substitution possibilities in the amount of effort devoted to various ones. A compensation scheme that makes pay sensitive to performance on one task may result in workers

²Third party awards for *overall* performance, such the Most Valuable Player award and the Cy Young award, do carry explicitly contracted bonuses. These exceptions are interesting for two reasons. First, they involve a mixture objective and subjective information on the overall value of a player's performance and, second, they are awarded by third parties who have no financial interest in a player's contract

shirking on another dimension. Thus baseball players who are paid for homeruns will avoid opportunities to bunt.³ In a sense, half a contract may be worse than no contract at all. One result of this inability to explicitly contract on all aspects of performance is that firms may forego explicit prices, even on contractible dimensions, in favor of subjective performance evaluation. This points out a well-known role of subjective performance evaluations: they can cater for dimensions not possible with objective measures. Our purpose in what follows is to illustrate potential pitfalls of subjective performance appraisal, noting throughout that subjective performance evaluation relies critically on the incentives of the evaluator in determining whether a person's performance is adequate.

3. Reneging

Reneging occurs when contracted performance is not rewarded. Reneging is not a problem when performance is verifiable, because contracts can be made explicit and legally enforceable. The problem is more interesting when performance measures cannot be contracted upon because they are not verifiable by a third party. For example, suppose a firm promises to pay a worker well if a subjectively measured performance standard is met. Ex post, the firm has a clear incentive to claim that the standard has not been met in order to save on wages.

Reputation is the most obvious limit on dishonest behavior by firms. Firms that renege will face higher costs of recruiting in the future, but with imperfect information the costs of malfeasance may not be completely internalized. Given this, the literature in this area emphasizes ways of organizing employment that reduce or eliminate a firm's financial incentive to renege, thereby enhancing efficiency. There are several possibilities.

One organizational response is for firms to commit a fixed wage *bill*, with the division of wages among workers depending on some measure of relative performance [Carmichael (1983)]. This tournament structure can provide optimal incentives while eliminating the firm's incentive to claim poor performance. Even so, real world examples in which employers truly precommit to a wage bill are rare. The usual examples of alleged tournaments – partnerships in law and accounting firms or executive promotion contests – do not typically involve precommitment of the firm's total wage bill.

Two other insitutional mechanisms that reduce financial incentives to renege are up-or-out contracts and attaching wages to jobs or tasks. In up-

³The Saturn Division of General Motors offers its cars for a fixed price, and has dropped sales commissions from the compensation packages of their sales staff. This reduces the likelihood that shoppers will be pestered into buying options they don't want. If these harmful side effects are sufficiently costly, it is optimal not to offer commission.

or-out contracts the firm employs a worker for a fixed probationary period, during which it observes the worker's performance. At the end of the period the firm has the option of retaining the worker at wage x or terminating the relationship. The contract can enhance efficiency because it eliminates the firm's ability to save money by falsely claiming that the worker's performance was inadequate [Kahn and Huberman (1988)]. The firm must either pay x or terminate the worker.

Like fixed wage bill models, real world examples of up-or-out employment contracts are rare. Instead, most large organizations are characterized by long-term employment relationships and by promotion-based reward systems. Prendergast (1992a) shows that this system, in which wages are tied to job titles, can induce optimal skill collection by workers and obviate the firm's incentive to renege. The key idea is that skills must be task-specific, so the firm gains nothing by denying promotions (and raises) to qualified workers.

Our primary focus in the remainder of the paper is not on financial incentives to renege. Instead, we argue that most employees of large organizations are evaluated by supervisors, whose financial incentive to renege is limited or non-existent. Even so, it is common for subordinates to believe that their performance has been undervalued, or that the financial rewards for good performance have been given to other, less deserving candidates. Our emphasis in what follows is on *who* gets the resources that supervisors subjectively allocate.

4. Bias

The necessity of subjective performance evaluation raises issues of systematic bias in organizations. Evidence of potential bias in performance appraisals comes from a variety of sources. Bretz and Milkovitz (1989) find that supervisors often provide performance ratings higher than those warranted by employee performance. They attribute the difference to personal relations and the real and psychic costs of communicating poor evaluations to workers. Kraiger and Ford's (1985) survey of the effects of race on ratings reported that the race of *both* the rater and the ratee affected evaluations. Overall, supervisors give higher ratings to subordinates of their own race. Several studies by psychologists have found that the ultimate use to which appraisals are put affects appraisal outcomes. For example, ratings used to make administrative decisions such as merit pay or promotion are more lenient, and have less variance, than ratings used for employee feedback [Williams et al. (1985); Reilly and Balzar (1988)]. Other results illustrate political aspects of performance appraisals [Longnecker (1989)]. For example, Bjerke et al. (1987) find that navy supervisors evaluate favored subordinates so as to maximize the likelihood of promotion.

4.1. *Bias and the structure of rewards*

Bias can cause inefficiencies on a number of dimensions. Employees who feel discriminated against may quit, with resulting turnover costs and lost human capital for the organization. In terms of workers' incentives, bias makes it difficult to distinguish genuinely good performance from favoritism. Other things the same, bias is a form of 'luck' that adds noise to the monitoring process. This will typically reduce incentives for effort, even perhaps among workers who are unduly favored. This is a formal sense in which bias reduces 'morale', leading to lower effort and output overall. Firms can compensate by increasing monetary rewards for effort, so one direct effect of bias may be greater wage inequality within the organization.

Other, indirect effects of bias can offset this tendency toward greater inequality. Specifically, bias is an important factor because superiors have rewards to dole out, which gives rise to inter-personal 'influence activities' as a form of rent-seeking in organizations [Milgrom and Roberts (1988, 1990)]. As compensation becomes more sensitive to subjectively measured performance, unproductive rent seeking will rise as workers seek to influence the evaluation process. For example, Argyris (1964) describes how managers covered by a bonus plan tied to budgets spent valuable work time bargaining with superiors to get a favorable budget standard. Firms may respond by designing compensation with more equity than would otherwise be optimal, which reduces rent seeking.

Even without influence activities, it may be optimal to reduce the inequality of rewards so as to offset the effects of supervisors' preferences. Prendergast and Topel (1992) argue that organizations typically use performance appraisals for (at least) two purposes: compensating individuals for their efforts *and* determining their true talents. The latter information is used to assign persons to different tasks and to identify training needs. Yet, as we noted above, supervisors are more likely to bias their evaluations when those evaluations have direct financial consequences for employees. This agency problem with supervisors means that greater inequality of rewards reduces the informativeness of supervisors' reports. The optimal response is to reduce wage inequality. Workers supply less effort and turnover among talented workers may rise, but these costs are offset by improved sorting of workers to tasks, raising overall productivity.

4.2. *Organizational responses to bias*

The existence of bias may affect the way firms organize their production and monitoring activities. For example, in many blue collar jobs supervisors have extremely limited discretion over rewards. Supervisors may have the authority to terminate workers, but less extreme decisions over pay and

promotion are based on observable factors such as seniority. We think of these arrangements as examples of 'rules rather than discretion', which limit the ability of supervisors or managers to exercise bias in personnel decisions.

This type of bureaucratic response has obvious costs: the wrong workers may be promoted and shirking is encouraged by the emphasis on equality. The alternative for management is to control bias by monitoring supervisors' decisions. This can be done directly; for example, managers may make their own reading on employee performance, which can be compared to the supervisor's. A formally similar solution is to rotate supervisors and workers, which achieves two results. First, rotation provides independent observations on an employee's ability (assuming supervisors' biases are imperfectly correlated). Second, to the extent that bias is the outcome of personal relationships and investments in influence, opportunities to build relationships are reduced. Of course this also points out the costs of rotation: productive matches of workers to either tasks or supervisors are less likely to develop. In rare cases, firms collect multiple readings by having each employee rank the contributions of his immediate colleagues. The opportunities for politicking in this arrangement are obvious.

Favoritism is accentuated when the supervisor is not responsible for the performance of the subordinate. A means of aligning the supervisor's incentives with those of the organization is to tie rewards to promotion and to make the supervisor responsible for the output of the job to which his subordinates are promoted. For example, the supervisor could be given responsibility for promotions within his department, but where he is responsible for total output from his department. In this scenario, where wages are attached to jobs, the manager suffers by promoting on his whims if his favorite is not the most talented worker. As a result, the firm can reduce favoritism by requiring that supervisors maintain responsibility for their promoted subordinates. A related point is that the span of control afforded to a supervisor may be at least partly determined by a desire to reduce favoritism.

4.3. Monitoring bias

A third alternative for dealing with problems of bias is simply to measure it and punish its occurrence. There are a number of possible problems with using the observations of others to monitor favoritism. The first is that monitoring is difficult. Supervisors naturally have greatest contact with their subordinates, so they begin from an information advantage. A second problem concerns the fact that supervisors typically make many decisions that affect subordinates' productivities. For example, in addition to offering performance evaluations, they also assign subordinates to jobs, offer on-the-job training, take them to meet clients, and inform them of openings

elsewhere in the organization. This discretion affects the actual performance of employees, so monitoring raises the possibility that supervisors will 'sabotage' the performance of workers in order to justify their biased ratings. Ultimately, it may be more efficient for an organization to tolerate bias than to incur these costs.

A final problem with monitoring derives from the mechanics of the appeals process. In cases where evaluations are communicated to workers, workers typically have an option to appeal. This process is a method of monitoring supervisory bias. Two problems are typically encountered with appeals procedures. First, workers fear reprisals from supervisors if they report them for unfair treatment. So victimization can go unpunished. Second, empirical work shows that management is reluctant to reverse decisions made by supervisors, as supervisors 'lose face' as a result. For example, Freeman and Medoff (1984) find that 'employees fear reprisals from their supervisor ... [and] even more striking is the fact that when the procedures are used, "the percentage of decisions that upheld the original decision is very high", with about one half of the companies supporting the supervisor in every case' (p. 109).

This reluctance of managers to overturn supervisors' decisions suggests that managers may not enforce unpleasant decisions in the same way that supervisors show leniency towards their workers. An alternative possibility is that managers do not overturn supervisors, even when they appear wrong, to enhance trust in supervisors' decisions. A supervisor who is not trusted by management cannot induce the same effort levels from subordinates as a more able supervisor. A manager may side with the supervisor, even if he is biased (and believed to be such by management), to avoid the negative externality caused by decision reversal on future effort incentives.

For all these reasons we believe that the exercise of favoritism is extremely difficult to constrain, so that significant discretionary power is likely to remain in the hands of a supervisor.

5. Compression of ratings and rewards

Supervisors' preferences and their relationship with subordinates can affect appraisals and rewards even in the absence of personal bias. There is substantial evidence that supervisors have preferences about the distribution of rewards among subordinates. In many cases supervisors are reluctant to give poor ratings to subordinates, either because doing so is unpleasant or because supervisors prefer equity in outcomes. For example, when supervisors at Merck and Co. had discretion over the full distribution of appraisals there was a marked tendency toward uniformity in reported performance and, consequently, wages [Murphy (1991)]. This compression of rewards reduced incentives and aggregate productivity. In a similar way, the

U.S. Navy's 'zero tolerance' policy toward incidents of sexual harassment, which dictates expulsion from the service, reduced the likelihood that incidents are reported and that superior officers will act [see *The Economist* (1992)]. In both cases, preferences of those charged with administering rewards and punishment partially undermines the organization's goals.

These data suggest that organizations may have difficulty in implementing incentive schemes that involve discretion by supervisors. But this raises the question of why supervisors are given such wide latitude. Firms could require supervisors to report a fixed ranking of subordinates – say by categorizing workers into deciles – which would seem to overcome any incentive by supervisors to compress wages. Yet only about 20 percent of companies use forced rankings in their performance appraisal systems [Bretz and Milkovich (1989)].

There are several possible reasons for the rarity of forced ranking systems. First, supervisors can work around the system by rotating high evaluations over time. Then the losers in one year are the winners in the next, so average ratings and rewards are still compressed. Second, forced rankings may breed resentment and low morale when applied to small work groups. Smaller groups increase the variance of average talent across groups, so employees from different groups are judged by different standards. Forced rankings also increase competition for merit pay, which is counterproductive in environments where cooperation is important to production.⁴

Alternatively, it may be optimal to suppress information on relative performance. At issue is the amount of information that should be revealed to employees. Suppose that organizations base promotions on an assessment of workers' relative talents over a long period. Should the firm instigate a fast track, whereby a few workers are identified as stars? Or should workers be treated uniformly, on the theory that average effort will be higher when all are left in the dark about their prospects? The answer depends on how workers respond to good and bad evaluations. Available evidence suggests that those who receive good evaluations become encouraged to greater effort, while the less able become discouraged [see, for example, Tannenbaum et al. (1974)]. Discouraged effort among those who conclude that they are out of the race can offset any encouragement of the front-runners. Then compression of evaluations and rewards may be optimal, at least for some period of time. A related point is the well-known fact that workers tend to rate themselves higher, on average, than do their supervisors [Shore and Thornton (1986)]. Given this bias by workers, firms may be reluctant to reveal genuine promotion prospects.

These factors may explain why there is relatively late selection of 'high

⁴These considerations were recognized by Merck in reforming his evaluation scheme. They adopted forced rankings, but only in work groups that exceeded 100 employees.

fliers' in Japanese companies. Workers are typically not differentiated from others in their cohort for 12 to 15 years after joining a firm, after which there is segregation by assessed ability. According to Takeuchi (1985, p. 18), suppressing information in this way actually encourages effort: 'Japanese business organizations paradoxically use the principle of equality to motivate employees ... [because] the promotion of one or two persons will cause the remaining employees to lose their will to work'. In a similar vein, Hatvany and Pucik (1981, p. 13) quote a director of a major Japanese trading company who believes that 'the secret of Japanese management is to make everybody feel that he is slated for the top position in the firm' [see Prendergast (1992b) for details].

6. Fairness

The existence of opportunities for bias raises not only the issue of how should firms and workers respond to genuine bias but also how should genuinely unbiased supervisors act so as to avoid appearing biased. Suppose that a worker and a supervisor each get independent, error-ridden, observations on the worker's true performance. If the supervisor is known to be honest, then there is little reason for him not to report his observation honestly. However, this is not obviously the case when the worker believes that the supervisor could be dishonest. Then the worker may update the likelihood of an honest supervisor on the basis of his report, where beliefs that the supervisor is dishonest may result in lower future effort or a high propensity to quit.

Suppose that the supervisor observes the performance of the worker being less than he had anticipated. It could be that the supervisor's signal of performance underestimates the worker's true performance. If the supervisor reports his observation honestly, then the worker may believe that the supervisor is biased and supply less effort in the future. Given this, one possible (though unproved) conjecture is that the supervisor trims his report towards the mean if he observes poor performance by the worker. However, an issue then arises as to how the worker should respond. Given that the supervisor fails to report poor performance, should the worker work as hard as when the supervisor reports honestly? Again, we conjecture is that the workers exerts less effort for any contract offered.

7. Conclusion

Subjectivity is central to performance appraisal in most organizations. We have argued that important features of organizations and methods of compensation are meant to deal with subjectivity and its associated incen-

tives. Two related issues have been emphasized: a tendency towards uniformity of treatment and the potential for bias in performance appraisals. We argued that uniformity of treatment may play an efficiency role, especially when the less able become discouraged by knowing their relative position.

We have also argued that opportunities for supervisors to distort their opinions can give rise to inefficiencies on two margins. The first is rent seeking by workers, which is usually a waste of time. The second occurs because bias makes it difficult to determine the true talents of workers. Both were used to explain a tendency towards equity in organizations.

This paper has not provided definitive answers to well structured problems. Instead, our purpose was to outline unresolved issues and provide focus to the literature on incentive pay. We believe that subjective performance evaluation is a central, but understudied, factor in incentive and organizational design.

References

- Argyris, C., 1964, *Integrating the individual and the organization* (Wiley, New York).
- Bjerke, D., J. Cleveland, R. Morrison and W. Wilson, 1987, *Officer fitness report evaluation study*, Report, TR-88-4 (Navy Personnel Research and Development Center, Washington, DC).
- Bretz, R. and G. Milkovich, 1989, *Performance appraisal in large organizations Practice and research implications*, Working paper no. 89-17 (Center for Advanced Human Resource Studies, Cornell University, Ithaca, NY).
- Carmichael, H. L., 1983, Firm-specific capital and promotion ladders, *Bell Journal of Economics* 14, 251-258.
- The Economist, 1992, *Naval operations*, July 4-10, p. 51.
- Freeman, R. and J. Medoff, 1984, *What do unions do?* (Basic Books, New York).
- Hart, O. and B. Holmstrom, 1985, *The theory of contracts*, in: T. Bewley, ed., *Advances in economic theory* (Cambridge University Press, Cambridge).
- Hatvany, N. and V. Pucik, 1981, *Japanese managerial practices and productivity*, *Organizational Dynamics*, Spring, 5-21.
- Holmstrom, B., 1979, *Moral hazard and observability*, *Bell Journal of Economics* 10, 74-91.
- Holmstrom, B. and P. Milgrom, 1991, *Multi-task principal-agent analyses: Incentive contracts, Asset ownership and job design*, *Journal of LEO* 7, 24-52.
- Kahn, C. and G. Huberman, 1988, *Two sided uncertainty and 'up-to-out' contracts*, *Journal of Labor Economics* 6, 423-444.
- Kraiger, K. and J. Ford, 1985, *A meta-analysis of ratee race effects in performance ratings*, *Journal of Applied Psychology* 70, no. 1, 56-65.
- Longenecker, C., 1989, *Truth or consequences: Politics and performance appraisals*, *Business Horizons*, Nov., 1-7.
- Milgrom, P. and J. Roberts, 1988, *An economic approach to influence activities and organizational responses*, *American Journal of Sociology* 94, July, S154-179.
- Milgrom, P. and J. Roberts, 1990, *The efficiency of equity in organizational decision processes*, *American Economic Review* 80, no. 2, 154-159.
- Milkovich, G. and A. Wigdor, 1991, *Pay for performance* (National Academy Press, Washington, DC).
- Murphy, K. J., 1991, *Performance measurement and appraisal*, in: W. Bruns, ed., *Performance measurement, evaluation and incentives* (Harvard Business School, Cambridge, MA).
- Prendergast, C., 1992a, *The role of promotion in inducing specific human capital acquisition*, *Quarterly Journal of Economics*, forthcoming.
- Prendergast, C., 1992b, *Career development and specific human capital acquisition*, *Journal of the Japanese and International Economies*, forthcoming.

- Prendergast, C. and R Topel, 1992, The costs of favoritism. Mimeo. (University of Chicago, Chicago, IL).
- Reilly, C. and W. Balzer, 1988, Effects of purpose on observation and evaluation of teaching performance, Mimeo (Bowling Green University, Bowling Green, OH).
- Shore, L. and G. Thornton, 1986, Effects of gender on self- and supervisory ratings, *Academy of Management Journal* 29, 115-129.
- Takeuchi, H., 1985, Motivation and productivity, in: L. Thurow, ed., *Management challenge* (MIT Press, Cambridge, MA).
- Tannenbaum, A., B. Kravik, M. Posner and G. Weiser, 1974, *Hierarchy in organization: An international comparison* (Jossey-Bass, San Francisco, CA).
- Williams, K., P. Wickert and R. Peters, 1985, Appraisal salience: Effects of instructions to subjectively organize information, in: *Proceedings of the Southern Management Association Meetings*, Orlando, FL.