

American Economic Association

Labor Economics and the Psychology of Organizations

Author(s): Edward P. Lazear

Source: *The Journal of Economic Perspectives*, Vol. 5, No. 2 (Spring, 1991), pp. 89-110

Published by: [American Economic Association](#)

Stable URL: <http://www.jstor.org/stable/1942687>

Accessed: 16/01/2014 11:38

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Economic Association is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Economic Perspectives*.

<http://www.jstor.org>

Labor Economics and the Psychology of Organizations

Edward P. Lazear

Industrial psychologists and sociologists have been exploring the institutions and practices that are internal to business organizations for decades. Attention by economists is more recent, and the economist's approach to analyzing internal labor markets is somewhat different from that of the psychologist or sociologist. The trademark of an economist is to focus on prices and income as central determinants of observed behavior. Other social scientists place a much lower weight on these variables, instead awarding the major role to social or psychological factors. Indeed, the word "institution" in economics generally connotes that the behavior is affected by constraints other than price.

The economic approach is more rigorous, more rational and probably better for prediction than that of the industrial psychologist. It is based on optimizing behavior in an environment where constraints are well-defined. The economic approach is also somewhat less accurate as a description of labor market phenomena. Thus, psychologists have concepts and data which would be useful to economists. A number of their ideas have already made their way into economics and more will follow.

This essay is a discussion of how economists are attempting to understand institutions within the organizations of the labor market. The institutions and issues discussed include mandatory retirement; discontinuous jumps in wages; pay compression; rights of tenure; up-or-out hierarchies; timing of raises, promotions and evaluations; the existence of partnerships; the use of bonuses vs. penalties; and pay as a motivator. Some of these questions seem primarily

■ *Edward P. Lazear is Isidore and Gladys Brown Professor of Urban and Labor Economics, University of Chicago, Chicago, Illinois, and Senior Fellow, The Hoover Institution, Stanford, California.*

economic, with a strong psychological component, while some are the reverse. But they are all questions that intrigue psychologists, sociologists, and economists alike, while remaining (at least so far) unanswered.

Mandatory Retirement

Mandatory retirement ages were used by most large organizations and many small ones, until it was made illegal (in most cases) by the Age Discrimination in Employment Act of 1974 and amendments to it in 1979. This presented a puzzle to economists. Why should a firm that was willing to pay a worker \$1000 per week be unwilling to employ that worker at any price in the next week, merely because the worker had turned 65? Surely productivity could not fall in such a discontinuous fashion.

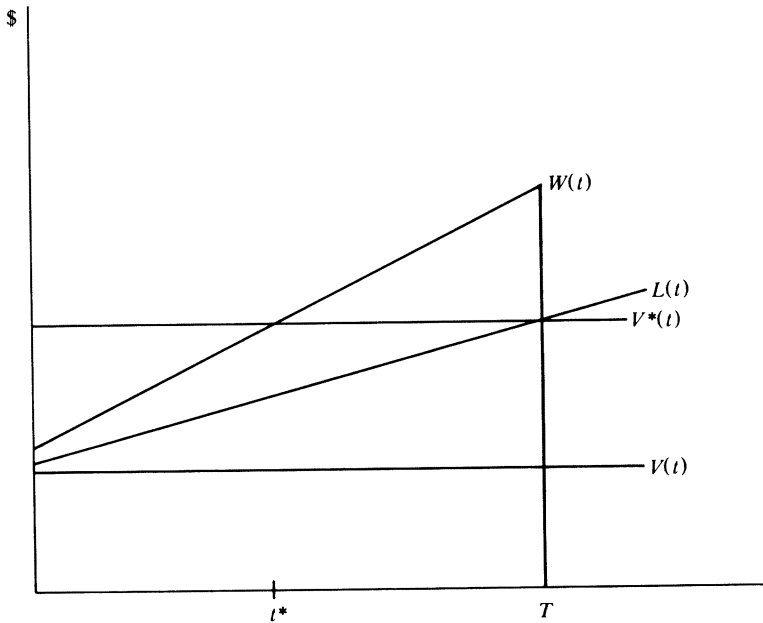
There were a number of attempts to explain the phenomenon in non-economic terms, but none was truly coherent. For example, some claimed that mandatory retirement was a way to ease out old workers without affecting company morale. But the knowledge that retirement is imminent may affect morale more adversely than a pay reduction with the option of staying on beyond the normal retirement age.

The explanation in Lazear (1979) takes a very different approach. There, it is argued that earnings should grow more rapidly than productivity as the worker acquires experience. In effect, young workers are paid less than they are worth and old workers are paid more. By doing this in a way that keeps the present value of lifetime wages equal to the present value of lifetime productivity, incentives are provided to workers that would be absent if they were to be paid a wage that followed life-cycle productivity more closely. The reason is that if workers reduce their effort, the risk of termination increases, and the workers would then forfeit the expected higher pay in the future. Thus, an upward sloping experience-earnings profile increases the costs of shirking, and acts as a motivator.

Figure 1 illustrates the point. Suppose a worker who works at the efficient level of effort has a flat productivity profile of $V^*(t)$ over his career. Suppose further that he has an alternative use of time (say, the value of leisure) shown by $L(t)$ which increases over the worker's lifetime. If wages were paid strictly according to productivity, then efficient retirement occurs at time T , when the value of leisure time exceeds the productivity of the worker.

The present value of the wage payment, $W(t)$, from 0 to T , must equal the present value of productivity, $V^*(t)$ over the life of the worker, or either the firm or the worker would seek out a different partner for this labor contract. If the worker were offered a constant wage $V^*(t)$ over his life, instead of a rising wage $W(t)$, he would be more likely to shirk as the retirement age T approaches.

Figure 1
Experience-Earnings Profiles and Mandatory Retirement



If a worker shirks, the worst thing that happens is that he is fired. But at T , the worker still picks up the value of his leisure, $L(T)$, which is equal to productivity $V^*(T)$, and thus to the wage. So shirking, say, one year before T costs nothing (since retirement will happen next year regardless) and yields the benefit of increased utility from reduced effort. This means that output will not actually be $V^*(t)$, but instead will be $V(t)$, reflecting the lower level of effort which occurs not only at the end, but throughout the entire lifetime because of backward induction.

Now suppose that the worker is offered a steadily rising wage, $W(t)$. The cost of shirking is much higher. If the worker shirks just before T and is fired, the worker now loses the amount $W(T)$, which exceeds $L(T)$. The lower initial wage means that over the entire lifetime, the firm breaks even. Reputation works to keep the firm from terminating the worker when it needs to begin paying more than the worker's productivity. Thus, a profile with rising wages over a lifetime results in higher lifetime productivity, *and thus higher lifetime wages*, than would wages that followed lifetime productivity exactly. Since workers will prefer the higher wage profile, the market is dominated by firms that pay with an upward-sloping age-earnings profile.

But when wages rise over a lifetime of work, the labor supply decision is distorted. When wages are set apart from current productivity, workers will react to the wage level rather than their productivity. That is, the efficient wage arrangement requires an upward-sloping wage profile to provide motivation

against shirking, and it also requires mandatory retirement since workers facing a rising wage profile will not choose to leave at the efficient age, when the value of leisure exceeds their productivity. Remember, the efficient contract is designed such that the present value of wages paid equals the present value of output up through time T . Work beyond retirement age T would imply that the firm takes losses. Mandatory retirement is only “mandatory” in the sense that workers would prefer to keep on receiving wages that are higher than their actual level of productivity. But these same workers would prefer the rising wage profile with mandatory retirement to the flat one without, because the steep one results in higher lifetime earnings.

The point is more general. Workers make decisions on two margins: effort per hour and hours per lifetime. A wage profile that induces an efficient level of effort per year (or per hour) will not also guarantee an efficient choice of hours per lifetime. Another instrument is needed. Institutions that look like hours constraints, either in the form of mandatory retirement or daily restrictions on hours worked, can bring about efficiency on this dimension.

This explanation for mandatory retirement is based on optimizing behavior, and it is internally consistent. In addition, it is supported by the data. Lazear (1979) found that mandatory retirement is positively associated with both job tenure and wage growth; that is, long-term jobs tend to use mandatory retirement, and those with age-earnings profiles that are rising most quickly tend to have mandatory retirement. Mandatory retirement occurred primarily at age 65, because social security kicks in then, creating a discontinuous jump in the alternative wage function. These findings are consistent with the argument that mandatory retirement is reserved for those for whom wages exceed marginal products.

More evidence on how upward-sloping wage profiles are used for incentives is provided by Hutchens (1986; 1987). First, Hutchens finds that steep profiles are used for workers who are hired when young rather than old. He argues that the up-front implicit payment from worker to firm is like a fixed cost that workers are unlikely to pay unless the expected tenure is long. Workers hired early are more likely to have steeply rising wage profiles, mandatory retirement and pensions. Second, Hutchens (1987) matches information from the Dictionary of Occupational Titles with the National Longitudinal Survey. He finds that jobs where it is cheap to monitor effort frequently are not the jobs that have mandatory retirement, pensions or steep profiles. When monitoring is costly, jobs are more likely to use the upward-sloping profile as a substitute for incentive schemes like piece rates.

Discontinuous Jumps in Wages

The structure of promotions and raises within the modern corporation seems difficult to reconcile with standard economic theory, since raises often

seem too large to be consistent with encouraging greater supply. Consider an individual whose salary rises 50 percent upon promotion from vice-president to president. It is difficult to argue that the individual's skills have risen 50 percent on the day of the promotion, or that the raise is compensation for the disutility of being president. Examples like these have led some to argue that psychological factors are important in the determination of wages. For example, O'Reilly, Main and Crystal (1988) provide evidence that boards of directors set salaries to correspond to their own salaries (see also Anderson and Anthony, 1986; Bacon, 1982).

However, the compensation puzzle can be explained by viewing salaries as prizes, as in Lazear and Rosen (1981). The president's salary does not necessarily reflect personal productivity, but is chosen because the allure of the president's salary makes all workers more productive over their careers as they compete for the next promotion.¹

The model can be understood by using the metaphor of a tournament, like a tennis tournament. Winning the tournament and receiving the winner's prize is like being promoted to president of the company and earning the president's wage. Being second in the tournament is like making it as high as vice president and receiving the corresponding wage, and so forth. From this viewpoint, a tennis tournament has three important features. First, prizes are fixed in advance and based on a player's relative performance rather than absolute performance. If Edberg defeats Lendl, then Edberg wins the championship prize, which varies neither with the quality of play, the closeness of the match, or the effort exerted by either player. Analogously, the president's salary is fixed in advance and is not affected by the difficulty with which the promotion choice was made.

Second, the spread between the winner's and loser's prize affects effort. If the prize money is split evenly between winner and loser, there is little incentive to win. However, if the president earns significantly more than the vice president, VPs will work hard to win the president's job.

Third, there is an optimal spread. After all, why not increase the spread to extremely high levels? The reason is that additional effort has value in creating output, but it also imposes pain on the contestants. At some point, the cost of pain associated with incremental effort exceeds the value of output, and the firm will have to pay more for additional effort than the effort is worth. There is such a thing as too much effort. In fact, the relevant first-order condition says that the spread should be chosen so that the marginal cost of effort just equals the value to the firm of the output produced with it. Risk aversion also pushes toward smaller spreads. Thus, the size of the raise that workers receive in moving from VP to president is determined by the effect of the raise on the effort put forth by VPs.

¹Of course, discontinuities exist in product markets as well. For example, magazine prices change infrequently and by discrete jumps even though advertising space prices move more continuously.

One other point is relevant. Tennis matches involve luck, and a player may lose just because of bad luck. In fact, with identical players, each player will lose half the time. As luck becomes relatively more important, effort has a smaller effect on the chance of winning. Therefore, the returns to effort decline, and a larger spread is required to induce effort. In situations where the average deviation between pay and performance is large, a firm will need more inequality to induce effort. Put differently, in firms where extraneous luck is important, the salary structures should be more spread out.²

There is some evidence on this subject. Ehrenberg and Bognanno (1988) find that golf tournaments where prize money is more unequal (larger spreads), yield lower scores per hole. Bull, Schotter and Weigelt (1987) provide experimental evidence. They run classroom experiments where winning depends on "effort," which the players can choose, and on luck. Payoffs are made tournament-style, and there is a convex cost of choosing higher levels of effort to simulate the Lazear and Rosen model. They find that the prize structure induces behavior that converges to the predicted Nash equilibrium very quickly.

Although little data from real firms speaks to this issue, some does exist. Antle and Smith (1986) claim that managers are compensated in part on the deviation of their own firm's performance from industry performance, which is consistent with a relative compensation scheme of the tournament variety. Jensen and Murphy (1990) claim that output is much more variable than manager's compensation. If managers were risk-neutral, their argument is that managers should experience very large variations in compensation to reflect large variations in output. Instead, they believe that most of the motivation for managers comes through promotion, as in the tournament scenario.

My view is that firms often promote on the basis of inputs, rather than on the observed outcome. An example is a board of directors failing to penalize a chief executive officer for a decision that turned out badly, but was the right decision to make at the time. Conversely, a decision that yielded the company enormous profit will be reflected in increased compensation to the CEO, but will fall short of the increase in profits because much of it is attributable to luck rather than effort.

Other evidence on the reasons for discontinuous jumps in wages: O'Reilly, Main and Crystal (1988) claim that the incomes of the board of directors are important determinants of the chief executive officer's salary. They interpret this as evidence that peer group determines salary to a greater extent than competitive forces. However, Rosen (1982) and others have argued that the largest corporations are headed by the most able, and that these corporations offer the most prestigious and highest-paying directorships. It is not surprising, therefore, that directors' salaries (even in their non-director activity) and CEO's salary are correlated. This is likely to be true between any two job categories

²A general discussion of agency theory, of which the tournament model is a special case, is contained in Sappington's paper in this symposium.

across firms; CEOs' salaries are high in firms that pay their janitors, never mind their directors, more than average.

Pay Compression

Personnel departments (and deans, for that matter) often argue that wages must tend toward equality, because a salary structure that is too closely related to differences in individual productivity creates morale problems and destroys the team spirit of an organization. Industrial psychologists have rationalized this position by assuming that workers care about their relative positions as well as their absolute standards of living (Frank, 1985), but this explanation is somewhat unsatisfying. First, it is tantamount to assuming the answer: wages are equal because workers like equality. Second, a drive for sameness is not pervasive in this society. Individuals often seek to distinguish themselves in the clothes they wear, the cars they drive, and the houses in which they live. Why is differentiation sought in some areas and not others? Third, it is far from clear why negative effects on the morale of employees whose high productivity is ignored do not outweigh any positive effects on low quality employees. Indeed, it might well be argued that it is more important to keep the best workers happy than the worst ones.

An explanation that relies on economic efficiency is outlined in Lazear (1989). If jobs and salaries are awarded at least in part on a relative basis, as described above, then workers do well not only by performing well, but also by assuring that their rivals perform badly. This means that they are disinclined toward cooperation and, at the extreme, may engage in outright sabotage. In fact, the larger is the spread between winner's and loser's prize, the greater the incentive to be uncooperative.

One solution for the firm which desires to encourage cooperation is to compress the pay structure. Pay compression reduces effort, but it also makes cooperation less unattractive. In general, it is optimal on productivity grounds to compress the wage structure, at least to some extent, to further cooperation.

This reasoning has two clear implications. First, production technologies that require cooperation should be marked by a more compressed wage structure. The compensation of salesmen should not be compressed, because each salesman operates as an independent agent and there is little gain from encouraging cooperation among salesmen. By contrast, a team of individuals designing a new automobile need to cooperate with one another and there should be less variance in their salaries. Second, pay compression is an equilibrium response to a problem that exists only among some types of workers. Firms that have less cooperative workers optimally choose a more compressed wage structure. They also have lower output because uncooperative types are less productive. Thus, there should be positive correlation between wage mean and variance across firms.

Another possible argument for pay compression is that if firms cannot monitor the output of workers perfectly, then workers have an incentive to exaggerate their output and lobby for higher wages. Milgrom (1988) analyzes this situation in a principal-agent framework, and finds that ignoring some worker claims is an optimal response to this problem, since it reduces the time wasted on internal politics. But ignoring some of the measured output also tends to compress wages. In the extreme, if all variations in measured output were ignored, wages would be equal (Baker, 1990).

Another final possible explanation for pay compression is that workers have a taste for being higher ranked within their firm. They must be compensated for being of below-average ability and having to work with those who are better. Conversely, the more able take pleasure in knowing that they dominate others in their environment and therefore they are willing to accept lower wages. Frank (1984) makes this argument. While something in this story makes sense, my view is that other effects are probably more important. Frank's story means that an assistant professor at MIT must be compensated for having to associate with Samuelson, Solow, and Modigliani, whereas the same individual would accept lower wages at Kankakee Polytechnic because he may be big man on campus. However, young faculty actually seem to receive lower wages at MIT than they would elsewhere. This probably reflects their willingness to have lower pay in exchange for the greater opportunity to build their human capital. Additionally, being at MIT signals to other potential employers that a young professor has high ability. Since senior individuals have more human capital at MIT than peers elsewhere, they receive higher wages at MIT than they would elsewhere, even though they are above the institution's average. At least in this case, the information and human capital effects swamp any preference for being a big fish in a small pond, and offset the desire for high status within an organization.

Tenure

While academics face explicit tenure contracts, private and public sector organizations often have institutions that guarantee essentially the same thing (Pashigian, 1986). Tenure provides job security to workers, but carries a cost to firms. Why are workers awarded tenure?

The reason usually given among academics is protection of academic freedom; a scholar with tenure can express unpopular views without (as much) fear of retribution. Carmichael (1988) presents a related, but more general argument. In an organization where compensation is relative, like a rank-order tournament, current workers dislike hiring quality applicants because their rank is adversely affected by the new hire. (This ignores any complementarities between workers.) Thus, where workers make hiring (or promotion) decisions, incumbents must be insulated from negative consequences of employing able individuals. Carmichael argues that by insulating a worker's wage from his

competition, tenure improves the efficiency of the hiring and promotion process. This explains tenure in academic organizations and partnerships such as law firms, accounting firms, and physician practices. Explicit or implicit tenure also appears more prevalent in these types of organizations.

An alternative view of tenure is a statistical one. Even though explicit tenure is rarely found in hierarchical private firms, separation rates are so low for employees with more than a few years of experience that a sort of *de facto* tenure exists. Some have argued that tenure should be interpreted as very low layoff rates. An explanation of this phenomenon is inherent in Jovanovic's (1979) job matching model. Separations occur when workers are sorted inefficiently, so that their comparative advantage lies with another firm. As more and more time goes by, the sample of workers who remain with a firm is increasingly more likely to be appropriately matched to that firm. As a result, the separation rate falls and workers appear to have tenure. Harris and Weiss (1984) model tenure explicitly in this way. After a sufficient history of successes has been observed, tenure can be awarded, explicitly or implicitly.

Much has been made of differences in turnover rates and lifetime employment between Japan and the United States. Hall (1982) points out that U.S. jobs are longer-lived than previously thought. But a more recent study by Hashimoto and Raisian (1985) shows that longer tenure in Japan is a fact, not a statistical artifact. Neither the Carmichael explanation nor the statistical argument seems likely to explain these international differences.

Part of the difference in separation patterns between countries is purely institutional or legal. Workers cannot be dismissed in most European countries, even for genuine business reasons, without receiving some severance pay. I have found elsewhere (Lazear, 1990) that the effect of these laws on employment are quite significant. Unless economists can build an economic theory of why these laws differ by country, we are left with these sorts of institutional explanations providing much of the mileage.

Up-or-Out Hierarchies

Some organizations use this puzzling promotion rule: either a worker is promoted to the next level up, or that worker is terminated. Nothing in the middle is tolerated. Universities generally use this rule; either promotion to a tenured position is granted or the professor is told to leave. In the military, soldiers must make it to a particular rank after a specified number of years, or they are denied the option to stay on at their current rank. In law firms, an associate either makes partner after a few years or is encouraged to seek employment elsewhere.

This discontinuous pattern of employment behavior is difficult to explain. Why was the firm happy to have the worker one day and unwilling to tolerate that worker under any circumstances, at any wage, on the next day? This

resembles the puzzle of mandatory retirement, except that the decision is made mid-career rather than at retirement. One argument is that workers who are denied promotion harbor ill-will and have adverse effects on morale. But then why is this tolerated in some organizations, but not in others?

Kahn and Huberman (1988) argue that up-or-out promotion is visible to all workers, and thereby easily verified. If an employer is forced either to promote workers or fire them, the employer's statement that a worker's performance does not measure up implies loss of that worker. If the employer had the option of merely announcing a lower wage, then employers might behave opportunistically and claim that workers were worse than they actually are. If the consequence of doing that is losing the worker, then the firm has no incentive to lie about a worker's productivity. Thus, up-or-out promotion helps ensure truth-telling by firms, and workers should prefer it.

Timing of Raises, Promotions, and Evaluations

The frequency of evaluations and raises varies from occupation to occupation. Serious evaluations are rare for academics, being restricted to promotion times. In investment banking, where feedback on performance is more readily obtained, workers receive raises and bonuses on a quarterly basis. Where output is very easily measured—for example, among farm workers—payment is made by piece, so that feedback is immediate. Firms usually give raises and performance evaluations to their workers with some regularity. But why don't more firms try alternatives like paying workers a fixed amount per year and then rewarding them with a bonus on separation that is contingent on their performance with the firm?

Psychologists are well-acquainted with the theory and evidence on reinforcement and its effects on behavior (Thorndike, 1913; Ferster and Skinner, 1957). Their ideas can be incorporated into an economic framework (Lazear, 1990a). In the context of studying the timing of raises, promotions, and evaluations, the economic approach has two advantages over psychological analysis; first, the action that is predicted to follow a particular stimulus is derived directly from optimizing behavior by the individual; second, the payoff structure that evolves must be compatible with a market economy. Pigeons pecking at a lever are different from workers in many senses, but most important in a labor market context is that the cages of pigeons prevent them from working for another researcher even if they are unhappy with the reinforcement schedule. Owners do not have the same freedom, since they must attract workers in an economy where work is voluntary.

Psychologists understand that intermittent reinforcement is required to keep a subject interested in the task. But why? How can economists rationalize this idea with standard utility theory? The metaphor of a slot machine is

helpful. Slots that offer a \$1000 jackpot pay off smaller amounts too, like rewards of \$3. Why provide these small rewards? The entire act of gambling suggests that risk aversion is not a helpful answer in this context. Nor does time preference help, because the receipt of a small \$3 prize is unlikely to alter the consumption trajectory.

One explanation for small intermittent rewards is based on information. Even if most machines pay off, any particular machine may be broken, or the casino may be dishonest, or the player may believe that some machines are unlucky. A few dollars reassure the player that this machine has a working money mechanism. This idea also provides an endogenous meaning to time preference, without merely assuming that individuals would like to have something now rather than later. The premium that is paid for having information early is a measure of time preference.

Receiving early evaluation allows workers to move to firms for which they are better suited. Workers are willing to pay for this information since it will help them abandon jobs that do not pay off, and this will mean that expected lifetime income will be higher.

In a competitive economy, the frequency of evaluation and reinforcement boils down to a comparison between the value of the information to the worker and the costs to the firm of providing it. There are some straightforward implications of this analysis (Lazear, 1990a). First, the larger the evaluation cost, the less frequent the evaluation. In jobs like academics, where output and measurement are somewhat subjective, serious evaluation is rare. Secretaries whose output is more easily measured should be evaluated and reinforced more frequently than their bosses, whose output is more difficult to characterize. This explanation of patterns of market reinforcement is surely more appealing than a psychological one that relies, say, on different “now orientations” among the various education, income, or ethnic groups.

Second, the value of frequent reinforcement rises with the value of the alternative. This implies that workers with a great deal of firm-specific capital need be reinforced only infrequently; they are unlikely to switch jobs, since specific capital means that their value at the current firm exceeds the value elsewhere. Lawyers in large firms, who have significant client- or case-specific capital, are evaluated infrequently; and Gilson and Mnookin (1985) argue that there is not much variation in compensation among partners. (Note that the job switch need not occur between firms. It is just as reasonable that the early information will be used by the current employer to sort workers to their most productive uses.)

A corollary is that evaluation and reinforcement should occur more frequently before much is invested in specific capital, but infrequently after the investment has been made. Thus, graduate students take more exams during the first year of the program than during the last. By the time of dissertation writing, the student’s skills in a particular field sufficiently exceed those in other fields to make a switch less probable.

Partnerships

The existence of free-rider effects implies that effort in a partnership is likely to be minimal, since compensation is based on firm profits rather than on an individual's effort. Despite these apparent difficulties, partnerships are widespread. In fact, profit sharing, which poses similar analytical problems, has become an important part of compensation for workers in many countries, most notably Japan. Why do firms use profit-sharing and partnerships if the free-rider effects are so pronounced? It would seem that direct supervision would dominate, at least in large firms. Personnel analysts have argued that group incentive plans create team spirit and encourage worker discipline, but are not much more explicit (Armstrong and Lorentzen, 1982). While this conclusion may well be true, it is important to understand how these forces work to change the nature of interaction within the firm.³

Peer pressure is analyzed in Kandel and Lazear (1989). A competitive firm can use peer pressure to motivate workers under certain circumstances. The key is that peer pressure is not free, and the value of disutility associated with using a compensation scheme that encourages peer pressure must be lower than the cost of measuring individual output directly.

Is it necessary for a firm to have profit-sharing to create peer pressure? There are two questions here. First, can workers feel peer pressure in the absence of profit-sharing? Second, will a worker monitor co-workers without profit sharing?

The answer to the first question is directly related to the definition of the relevant group. A worker who shirks necessarily hurts someone. If workers do not share in the profits, then shareholders are hurt. If workers share in the profits, then, additionally, co-workers are hurt. If shareholders were in the workers' group of "peers," then even in the absence of monitoring, guilt might prevent workers from shirking.⁴ However, if workers empathize only with co-workers, profit-sharing will be necessary to make peer pressure a force toward more effort. Without profit-sharing, shirking has no adverse consequences for others in the reference group.

But profit-sharing may not be sufficient. Absent guilt, peer pressure can only result if one worker is willing to monitor another. But when the enterprise is very large, it does not pay for any individual to monitor others, especially when reporting on a co-worker imposes personal costs on the informer. This suggests that the motivation behind profit-sharing in large Japanese firms has little to do with mutual monitoring, but instead probably relates to risk-sharing.

³Farrell and Scotchmer (1988) consider some of these issues with an application to sharing of information.

⁴Guilt is defined as disutility received when others are hurt, even if the identity of those responsible for the harm remains unknown. Shame is defined as disutility received because others observe that an individual has shirked. Thus, the distinction between shame and guilt hinges on the necessity for observability.

In relatively small firms, where empathy is strong because co-workers know each other personally, peer pressure can be an effective motivator. This way of thinking about motivation helps explain why partnerships tend to be among individuals of similar types. Partnerships among lawyers and among doctors are common, but partnerships between a lawyer and a doctor are rare. If mutual monitoring is to work at all, peer pressure must (almost by definition) be among peers, which implies an ability to evaluate one another. Lawyers are better able than doctors to evaluate other lawyers. Also, secretaries within law firms are rarely equity holders, although nothing in theory prevents them from being paid in equity even if their shares are not equal to those of the lawyers. But in the case of secretaries and lawyers, the smaller share held by secretaries would imply greater free rider effects. Thus, equity ownership should start among the highest paid workers and move down the hierarchy. The individuals with the largest proportion of their income in the form of profit-sharing should be those with the largest incomes. This is clearly borne out in the modern corporation.

If peer pressure is important, it can explain the popularity of orientation meetings, quality circles and company picnics. Firms sometimes spend considerable amounts on apparently meaningless indoctrination. For example, the U.S. military puts all new personnel through the same basic training, irrespective of subsequent job assignment.⁵ It is implausible that a soldier destined to be a Pentagon-based computer programmer is made more productive directly by learning to crawl beneath barbed wire with live ammunition screaming overhead. But practices that resemble hazing more than training can be explained as development of team spirit, a sense of loyalty to peers. This translates into guilt or shame associated with letting friends down on the job and can raise the equilibrium level of effort, and thus the utility of all workers.

One implication of this view is that firms where workers receive a large part of compensation in the form of profit-sharing should also spend relatively more on orientation and other indoctrinating activities. The argument is consistent with the observation that quality circles and company songs are less common in the United States than in Japan, where workers receive a larger part of compensation as profit-sharing. But it is not fully satisfying because even in Japan, workers at one plant can gain by free riding on workers at other plants, to whom they may feel little connection.

Bonuses vs. Penalties

Psychologists have long argued about the choice of carrot or stick as motivator, and the idea of positive and negative reinforcement finds its way

⁵Arguably, academics do the same to graduate students, but there is a difference. Graduate students are shopping for a field at the beginning of the program and learning a bit of everything is useful. But Ph.D. statisticians are never going to be infantrymen, even if they turn out to be superb marksmen.

into the earliest behavioral literature. But the argument has been somewhat religious, or at best descriptive, rather than analytic.

Consider two compensation schemes, both based on output. In the first, a worker is told that he will receive a monthly salary of \$10,000 plus a bonus of \$1 for each unit of output sold. In the second case, he is told that he will receive a monthly salary of \$15,000, but will be penalized \$1 for every unit short of 5000 that he sells. (It is convenient, but unnecessary, to assume that sales cannot exceed 5000 so that the penalty is never a negative number.) These schemes may sound different, but they are identical. The bonus scheme is written as $\text{Salary} = 10,000 + Q$, where Q is the number of units sold. The second is written as $\text{Salary} = 15,000 - (5000 - Q)$. But the second equation can be rewritten easily as $10,000 + Q$ so the penalty scheme is algebraically equivalent to the bonus scheme.

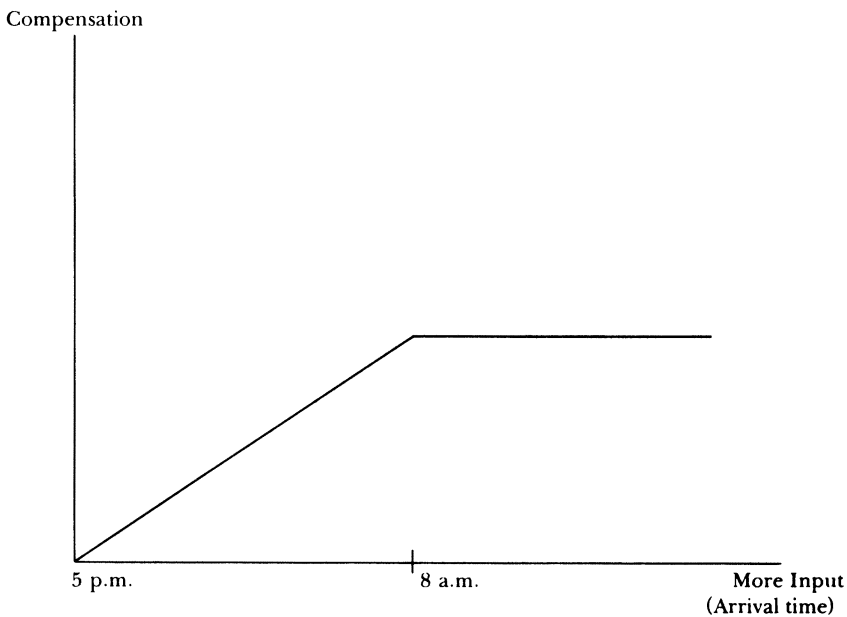
It is difficult to believe that workers faced with these schedules cannot see quickly that they are the same. One need not understand how to use a spread sheet to see that obtaining no sales results in an income of \$10,000, that sales of 5000 units results in \$15,000 and that sales of 2000 results in \$12,000 under either scheme.

Still, as an empirical proposition, psychologists tell us that the framing of the salary matters. The work by Kahneman and Tversky (1979) and Tversky and Kahneman (1981) is the best known in this area, although the interested reader might also check Loewenstein (1987; 1988) for an interesting extension of their model. As an empirical proposition, it seems clear that the way a proposition is phrased affects the outcome. For example, Sears refers to its second in a line of four models of lawn mowers as Sears' "second-best mower." It would be laughable to encounter a description of the same machine as "Sears' third-worst," although these statements have the same meaning.

The main objection to the framing approach is that it is not very good at predicting when the market will choose a penalty specification and when it will choose a bonus specification. For example, workers are usually given large Christmas bonuses for good performance rather than small Christmas penalties. But those same workers are usually penalized by being docked pay for arriving late, rather than rewarded with an on-time bonus. Framing matters, but here are two work situations where opposite terminology is used. We need to be able to predict when one is used over the other, as well as to predict the differences in behavior associated with each. Vocabulary may be important. A bonus for arriving on time may have a different meaning from being docked for lateness. Does an economic framework shed some light on this distinction? While this remains one of the toughest issues to model, economics helps at least to clarify what is at issue.

Perhaps the choice between using a bonus or a penalty is a non-price way of conveying information about the expectations of the firm. It is not obvious that bonuses and penalties are needed here. After all, a non-linear wage schedule can convey the expectations of firms without any mention of bonuses or penalties, and such a varying schedule can motivate workers to gravitate to

Figure 2
A Penalty Schedule



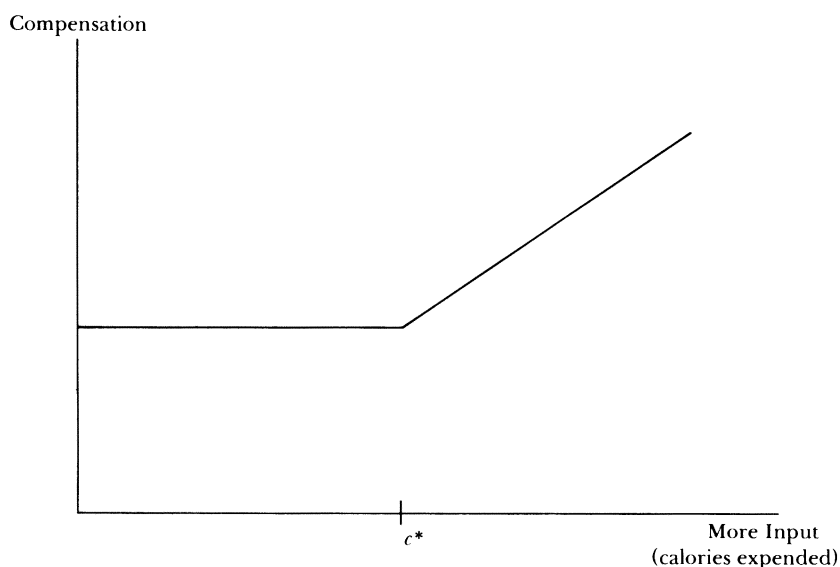
the most efficient behavior without relying on non-price inducements. If the firm receives extra value from having the worker appear at 8 a.m. sharp, then the value could be reflected directly in compensation. Workers could be paid more than the normal hourly wage for the half hour between 8:00 and 8:30 A.M., to provide them with the right incentives to arrive on time.

Standard English usage of “bonus” connotes extra credit; the idea that taking the desired action brings reward, but failing to take action does not reduce compensation. At one level, this is nonsensical. If an action brings reward, then the opportunity cost of no action is the value of that reward. But there is a sense in which extra credit has meaning. It may imply a kinked payoff schedule, where the worker is rewarded for exceeding a standard, but not punished for falling short of it. Conversely, a penalty implies a kink in the reverse direction; the worker is punished for falling short of a standard, but not rewarded for exceeding it.

To understand this point, consider the example of being docked for arriving late. If the desired arrival time is 8 a.m., then the worker’s reward schedule might be as in Figure 2. A worker is docked for arriving after 8 a.m., but is not rewarded for coming early. This is a kink which punishes falling short of a standard, but does not reward exceeding it.

Figure 3 illustrates an example of a bonus. It pertains to the following story. An individual leaves his automobile’s headlights on and returns to find the battery dead. To get the car started, it is necessary to push it. He hires some

Figure 3
A Bonus Schedule



workers to push the car, and pays them on the basis of calories burned. Until c^* calories are burned, the car does not move at all. Beyond c^* , the speed of the car and the probability that it will start increase with additional calories burned by the worker. Calories from 0 to c^* have no value, whereas those beyond c^* have positive value, related to the increasing probability that the car will start and the value of getting it running.

These definitions suggest that penalty schemes should be used when output or input above some critical level has no value. In the case of the worker docked pay, arriving earlier than 8 A.M. is not rewarded because a machine is unavailable for him to work on before starting time. Instead, he is penalized for input less than 8 hours, but not rewarded for input greater than 8 hours. Conversely, bonus schemes should be used when output or input below some critical level has no effect on value received. In the case of car pushing, all input levels between 0 and c^* yield the same value so the worker is given a bonus for effort above c^* .

In a more global sense, it is fair to argue that there is no such thing as extra credit. For example, offering extra credit on an exam only has meaning if the standard does not adjust to reflect average performance. If the instructor grades on a curve so that 10 percent receive A's and 10 percent receive F's, there is no sense in which any problem can be for extra credit. If all students obtained full points on the extra credit problem, the final distribution of grades

would not be altered. Unless the instructor raises the number of A's and B's and decreases the number of D's and F's, the extra credit problem becomes a required problem.

The more general point is that if compensation is relative, then bonuses and penalties have little meaning. It is meaningful to talk about bonuses and penalties, even in the sense of a non-linearity, only if the scale is truly an absolute one. Whether there are any true absolutes is too deep a question to be considered here.

Pay as a Motivator

One school of industrial psychologists has argued that motivation through pay is ineffective and even counterproductive because it causes workers to lose interest in their intrinsically interesting job (Lawler, 1973; Deci, 1972). Recently, Frey (1991) has argued that there is an economic logic to the notion that extrinsic rewards can actually have a negative effect on effort. He suggests that individuals rationally respond to the amount of "work morale" attributed to them. If work morale is thought to be lower than it is, then workers rationally reduce their morale, which results in less effort. Using extrinsic rewards signals the worker that the principal believes work morale is low. The worker reduces morale as a result so that when the extrinsic reward is withdrawn, effort is lower than it was before.

It is clear that an inappropriately designed compensation scheme can be counterproductive. For example, a piece rate that is tied to quantity and ignores quality will induce the worker to produce lower quality items (Lazear, 1986). Centrally planned economies often suffer from this problem. But it need not be this way. A worker could be penalized for producing poor quality goods. Much has to do with observability. If some attributes are easily observed, while others are more difficult, pay on the basis of the easily observed attributes may get the worker to focus on those aspects, to the exclusion of others (Baker, Jensen and Murphy, 1988).

These questions of observability and its impact on compensation are beginning to be modeled by economists. In addition to the papers cited above, another early step is by Milgrom (1988). He argues that since managers cannot observe the effort of their workers perfectly, workers spend resources convincing managers of their value; for example, workers devote time to that part of perceived output that is most easily produced. This activity is unproductive, so managers must screen out part of the signal.

A related story appears in Baker (1990). He argues that principals cannot always be precise in the instructions that they give to the agent because agents often have better information than the principals. If a contract is too tightly constrained to measurable output, agents will not behave optimally. He shows

that there is an optimal piece rate that varies directly with the signal-to-noise ratio.

Applying Psychology to Economics

To this point, this article has argued that economic analysis is often very helpful in examining problems within labor markets which might, at first glance, appear to be caused by institutional inertia or non-economic psychological behavior on the part of workers. But the usefulness of links between economics and psychology run in both directions. This concluding section will explore a couple of the psychological concepts that are proving especially useful to economists: cognitive dissonance and bounded rationality.

The psychological notion of cognitive dissonance involves a situation where people are confronted with a phenomenon that conflicts with their previously held beliefs, thus creating internal pressure for an after-the-fact rationalization of the unexpected phenomenon. In Akerlof and Dickens (1982), individuals choose their beliefs and then process information to reinforce those beliefs. In this way, for example, workers may rationalize working with dangerous substances. One implication of this model is that safety legislation can be Pareto improving because it allows workers (as voters) to analyze safety hazards before being exposed to them.⁶

Cognitive dissonance raises interesting intertemporal problems, since individuals may behave one way in one period, but differently after they choose their beliefs. Recent work on addiction is akin to this idea. In Becker and Murphy (1988), individuals take actions in one period that affect their ability to enjoy a good in subsequent periods. Thus, addiction is explained, not as a mistake, but rather as a decision that is optimal even during the earliest step in the process. The adverse consequences are merely the (anticipated) costs that must be borne so that the enjoyment can be taken earlier. This is a particular form of non-separability over time. Consumption in one period affects directly the marginal utility of consumption in another period. It may be rational to begin smoking even though the addiction to it is painful. The unpleasantness of addiction is the price paid for the enjoyment received in becoming addicted. As a result, individuals can express regret after the fact.

The theory is ingenious, although it does not square with all the facts. For example, it has difficulty explaining the life cycle pattern of addiction. Other things constant, old persons would be more likely to become addicted because

⁶Frank (1987) makes use of similar concepts when he asks whether, in a game theoretic context, it pays to be able to choose beliefs. This idea goes back at least to Schelling (1960), who argued that sometimes there is strength in weakness. Being able to precommit to an irrational (non sub-game perfect) strategy can make a player better off.

the period over which costs must be borne is shorter. A counterargument might be that ability to enjoy addictive substances is age-related, but this is not a completely satisfying answer.

Of course, researchers have understood that the assumption of intertemporal separability of utility is a convenience, but not necessarily a fact. In the labor context, it may help explain why workers become attached to particular organizations and are willing to turn down higher wage offers that they would have accepted early in their careers. Firm-specific human capital only explains why outside offers are lower than those at the current firm, not why a worker would refuse an offer that actually is higher. Workers become attached to their organizations in a way that is not unlike addiction.

Group norms can be another expression of cognitive dissonance. Once it is recognized that individuals can choose beliefs, it is not difficult to imagine that organizations can affect those choices as well. The earlier discussion on orientation and indoctrination is related to cognitive dissonance on a group level. Kandel and Lazear (1989) model this by assuming that deviations from the equilibrium value of effort are disliked by other workers, thereby bringing disutility to the deviator. The extent of this disutility affects the equilibrium level of effort. Thus, a (Nash) equilibrium effort level becomes a self-enforcing norm.

One can go somewhat further. Because different firm “personalities” imply different amounts of disutility associated with deviating from the norm, an infinite number of equilibrium effort levels can result. But only one level is Pareto optimal. It may be that a competitive labor market will result in survival only of firms that have a personality consistent with a norm set to the efficient level of effort.

Psychologists (and industrial psychologists in particular) are fond of claiming that economic models are uninteresting because workers do not always behave rationally. Herbert Simon (1957) and his theories of satisficing are among the best-known attempts to give some economic content to these ideas. More recently, theories of bounded rationality have proceeded down two main lines. The first, with a primarily macroeconomic thrust (Akerlof and Yellen, 1985a; 1985b) relies mainly on transactions or information costs to prevent full optimization. It is argued that near the optimum, mistakes produce only small deviations in profit; but in an economy that has some market imperfections, the same mistakes produce first-order deviations from Pareto optimality.⁷

The second view of bounded rationality relies on limited memory and is primarily game-theoretic (Kreps, Milgrom, Roberts, Wilson, 1982; Aumann and

⁷The intuition can be understood this way: A competitive firm sets marginal cost equal to price, which is the marginal social benefit of a good. A small reduction in output implies a reduction in costs and benefits that are nearly equal, because they were exactly equal at the optimum. But a monopolist creates a wedge between social value and cost so a reduction in output brings about a significant discrete change in net value, but not in profit.

Sorin, 1986). This has a relation to computer science. Computer scientists discuss bounds on computation and rationality in terms of storage space, computation speed, and communication speed. Most models of bounded rationality in economics rely on the first constraint of memory limitation. For example, a union may only be able to keep track of a certain number of moves made by an employer, so its negotiation strategy is not fully rational in the sense of using all historical data.

More recently, researchers have begun to think in terms of computation speed constraints, which are qualitatively different. For example, a salesman must make a decision on whether to push item A or item B on a particular customer. This decision may be a complicated one, involving the customer's horizon and the firm's plans for future products. While the information may be available and known to the salesman, processing it may require time. The salesman may have an advantage in gathering information about the customer, but the salesman may not be the best data processor in the firm. Other managers are likely to be better at computing an optimal strategy, given the data. This suggests a methodology for predicting the structure of the firm. Top-down organizations are likely to prevail when data gathering and communication are relatively easy, but computation of strategies is difficult. Bottom-up organizations are likely to prevail when data gathering and communications are difficult, but when given the data, optimal strategies are obvious. Sah and Stiglitz (1986) have come closest to modeling the organization in this fashion, where the choice between hierarchical or parallel command depends on minimizing such statistical error.

Conclusion

Economics is sometimes accused of being sterile, unrealistic, and inhumane. We are often charged with ignoring psychological and institutional issues that may have most of the explanatory power. This stereotype has had some truth, but it is becoming much less accurate. In labor economics and other areas, previously non-economic issues are being systematically incorporated into economic analyses. Issues like mandatory retirement, discontinuous wage jumps, up-or-out promotion rules, tenure, the pattern of evaluations, and peer pressure are now incorporated into standard economic models.

The fact that some issues in economics are fuzzy does not preclude their being analyzed with economic models. The fact that we have only scratched the surface leaves opportunity for economists to help clarify the discussion of organizations and labor institutions.

■ *Financial support from the National Science Foundation is gratefully acknowledged.*

References

- Akerlof, George A., and William Dickens,** "Economic Consequences of Cognitive Dissonance," *American Economic Review*, 1982, 72, 307-319.
- Akerlof, George A., and Janet L. Yellen,** "Can Small Deviations from Rationality Make Significant Differences to Economic Equilibria?" *American Economic Review*, September 1985a, 75, 708-720.
- Akerlof, George A., and Janet L. Yellen,** "A Near Rational Model of the Business Cycle with Wage and Price Inertia," *Quarterly Journal of Economics*, 1985b, 100, 823-838.
- Anderson, Charles A., and Robert N. Anthony,** *The New Corporate Directors: Insights for Board Members and Executives*. New York: Wiley, 1986.
- Antle, Rick, and Abbie Smith,** "An Empirical Investigation into the Relative Performance Evaluation of Corporate Executives," *Journal of Accounting Research*, Spring 1986, 24, 1-39.
- Armstrong, Michael, and John F. Lorentzen,** *Handbook of Personnel Management Practice*. Englewood Cliffs, N.J.: Prentice-Hall, 1982.
- Aumann, Robert, and Sylvan Sorin,** "Cooperation and Bounded Rationality," IMSSS mimeo 1986; forthcoming, *Games and Economic Decisions*.
- Bacon, Jeremy,** *Corporate Directorship Practices: The Compensation Committee*, Conference Board Report #829. New York: Conference Board, 1982.
- Baker, George P.,** "Piece Rate Contracts and Performance Measurement Error," Harvard Business School Working Paper 1990; forthcoming, *Journal of Political Economy*.
- Baker, George P., Michael C. Jensen, and Kevin J. Murphy,** "Compensation and Incentives: Practice vs. Theory," *Journal of Finance*, July 1988, 43, 593-616.
- Becker, Gary S., and Kevin M. Murphy,** "A Theory of Rational Addiction," *Journal of Political Economy*, August 1988, 96, 675-700.
- Bull, Clive, Andrew Schotter, and Keith Weigelt,** "Tournaments and Piece Rates: An Experimental Study," *Journal of Political Economy*, February 1987, 95, 1-33.
- Carmichael, H. Lorne,** "Incentives in Academics: Why Is There Tenure?" *Journal of Political Economy*, June 1988, 96, 453-472.
- Coase, Ronald,** "The Problem of Social Cost," *Journal of Law and Economics*, October 1960, 1-44.
- Deci, Edward L.,** "The Effects of Contingent and Noncontingent Rewards and Controls on Intrinsic Motivation," *Organization Behavior and Human Performance*, October 1972, 8, 217-229.
- Ehrenberg, Ronald G., and Michael L. Bognanno,** "Do Tournaments Have Incentive Effects?" NBER Working Papers Series, No. 2638, June 1988.
- Farrell, Joseph, and Suzanne Scotchmer,** "Partnerships," *Quarterly Journal of Economics*, May 1988, 103, 279-297.
- Ferster, Charles B., and B.F. Skinner,** *Schedules of Reinforcement*. New York: Appleton-Century-Crofts, 1957.
- Frank, Robert H.,** "Are Workers Paid Their Marginal Products?" *American Economic Review*, September 1984, 74, 549-71.
- Frank, Robert H.,** *Choosing the Right Pond: The Economics and Politics of the Quest for Status*. New York: Oxford University Press, 1985.
- Frank, Robert H.,** "If *Homo Economicus* Could Choose His Own Utility Function, Would He Want One with a Conscience?" *American Economic Review*, September 1987, 77, 593-604.
- Frey, Bruno,** "Monitoring and Crowding Out Work Morale," University of Chicago, 1991.
- Gilson, Ronald J., and Ronald H. Mnookin,** "Sharing among the Human Capitalists: An Economic Inquiry into the Corporate Law Firm and How Partners Split Profits," *Stanford Law Review*, January 1985, 37.
- Hall, Robert E.,** "The Importance of Lifetime Jobs in the U.S. Economy," *American Economic Review*, 1982, 72, 716-24.
- Harris, Milton, and Yoram Weiss,** "Job Matching with Finite Horizon and Risk Aversion," *Journal of Political Economy*, August 1984, 92, 758-79.
- Hashimoto, Masanori, and John Raisian,** "Employment Tenure and Earnings Profiles in Japan and the United States," *American Economic Review*, September 1985, 75, 721-735.
- Hutchens, Robert,** "Delayed Payment Contracts and a Firm's Propensity to Hire Older Workers," *Journal of Labor Economics*, October 1986, 4, 439-457.
- Hutchens, Robert,** "A Test of Lazear's Theory of Delayed Payment Contracts," *Journal of Labor Economics*, October 1987, 5, S153-S170.
- Jensen, Michael C., and Kevin J. Murphy,** "Performance Pay and Top Management In-

centives," *Journal of Political Economy*, 1990, 98, 225 + .

Jones, Larry, "A Note on the Joint Occurrence of Insurance and Gambling," unpublished manuscript, 1988.

Jovanovic, Boyan, "Job Matching and the Theory of Turnover," *Journal of Political Economy*, October 1979, Part 1, 87, 972-90.

Kahn, Charles, and Gur Huberman, "Two-Sided Uncertainty and 'Up-or-Out' Contracts," *Journal of Labor Economics*, October 1988, 6, 423-444.

Kahneman, Daniel, and Amos Tversky, "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 1979, 47, 363-391.

Kandel, Eugene, and Edward P. Lazear, "Peer Pressure and Partnerships," Hoover Working Paper E-89-5, 1989.

Kreps, David M., Paul Milgrom, John Roberts, and Robert Wilson, "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory*, August 1982, 27, 245-252.

Lawler, Edward E., *Motivation in Work Organizations*. Monterey, CA: Brooks/Coles, 1973.

Lazear, Edward P., "Job Security Provisions and Employment," *Quarterly Journal of Economics*, August 1990, 105, 699-726.

Lazear, Edward P., "Pay Equality and Industrial Politics," *Journal of Political Economy*, June 1989, 97, 561-80.

Lazear, Edward P., "Salaries and Piece Rates," *Journal of Business*, July 1986, 59, 405-431.

Lazear, Edward P., "The Timing of Raises and Other Payments," *Carnegie/Rochester Conference Series on Public Policy*, 1990a.

Lazear, Edward P., "Why Is There Mandatory Retirement?" *Journal of Political Economy*, December 1979, 87, 1261-64.

Lazear, Edward P., and Rosen, Sherwin, "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, October 1981, 89, 841-64.

Loewenstein, George F., "Anticipation and the Valuation of Delayed Consumption," *Economic Journal*, 1987, 97, 666-684.

Loewenstein, George F., "Frames of Mind in Intertemporal Choice," *Management Science*, February 1988, 34, 200-214.

Milgrom, Paul R., "Employment Contracts, Influence Activities, and Efficient Organization Design," *Journal of Political Economy*, February 1988, 96, 42-60.

O'Reilly, Charles, Brian Main, and Graef Crystal, "CEO Compensation as Tournament and Social Comparison: A Tale of Two Theories," *Administrative Science Quarterly*, June 1988, 33, 257-274.

Pashigian, B. Peter, "Tenure and the Turnover of Faculty," unpublished manuscript, University of Chicago, May 1986.

Rosen, Sherwin, "Income Generating Functions and Capital Accumulation," Harvard Institute of Economic Research Paper No. 306, 1973.

Rosen, Sherwin, "Authority, Control and the Distribution of Earnings," *Bell Journal of Economics*, October 1982, 13, 311-23.

Sah, Raaj Kumar, and Joseph E. Stiglitz, "The Architecture of Economic Systems: Hierarchies and Polyarchies," *American Economic Review*, September 1986, 76, 716-727.

Schelling, Thomas, *The Strategy of Conflict*. Cambridge: Harvard University Press, 1960.

Simon, Herbert A., *Models of Man*. New York: John Wiley & Sons, 1957.

Smith, Adam, *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: Oxford University Press, 1976.

Thorndike, Edward Lee, *Educational Psychology*. New York: Columbia University Teachers College, 1913-14.

Tversky, Amos, and Daniel Kahneman, "The Framing of Decisions and the Psychology of Choice," *Science*, 1981, 211, 453-458.