

SEPARATING IDEAS AND BUBBLES IN HUMAN RESOURCE MANAGEMENT

David Lewin, Daniel J.B. Mitchell, and Mahmood A. Zaidi

Question: Why have a handbook dealing with the latest in research and theory in human resource management?

Answer number 1: Because research and theory have something of practical importance to say about the field. At least, that is what academics surely like to think. Perhaps a better answer, therefore, is that this handbook will allow you as the reader to make your own judgment on that issue. As editors, we think the reader will find valuable insights from the academic literature.

Answer number 2: Because the issues in the field are typically more complex than many writing (and especially consulting) outside the research community would like them to be or appear to be. Even if research and theory do not provide definitive guides to "how-to-do-it," they at least provide a note of caution. And caution is needed in a world in which there is no shortage of gurus proclaiming to have found the Truth about human resource practices.

THE POWER OF IDEAS

Economists, especially those based at universities, are fond of quoting John Maynard Keynes's famous observation in 1936 that "practical men, who believe themselves to be quite exempt from any intellectual influences, are usually the slaves of some defunct economist" and that "madmen in authority, who hear voices in the air, are distilling their frenzy from some academic scribbler of

a few years back" (Keynes, 1936, p. 383). Given the madmen then in authority, it is not entirely clear that this influence was something to brag about. But perhaps it is Keynes's further observation that should receive the bulk of our attention today:

I am sure that the power of vested interests is vastly exaggerated compared with the gradual encroachment of ideas. Not, indeed, immediately, but after a certain interval; for in the field of economic and political philosophy there are not many who are influenced by new theories after they are twenty-five or thirty years of age.... But soon or late, it is ideas, not vested interests, which are dangerous for good or evil (pp. 383-384).

Keynes is widely viewed as the father of macroeconomics. He wrote his paean to the power of ideas during the Great Depression when unemployment and inadequate demand in the immediate term was *the* key policy concern. Jump ahead to the 1990s and we find that macroeconomists now worry as much about the supply side as the demand side and about the long term as well as the short. Specifically, a major concern at present is to determine what steps might accelerate the rise of productivity and what institutions might increase the trend rate of economic growth. Some of the conclusions are of great interest for the field of human resources.

Macroeconomist Martin L. Weitzman has pointed out that in the contemporary world it has become apparent that the source of long-term growth goes beyond just adding more and more inputs of capital and labor to the national "production function." Rather the source involves a "combinatoric growth process" of *ideas* which are then *applied* to production (Weitzman, 1996).

As Weitzman sees it, most innovation stems from the combining of existing ideas and concepts into new ones, that is, the obtaining of new insights from existing strands of knowledge. Then there must be a sorting out of the useful new ideas from the nonviable. Such a process of producing innovative ideas has a tendency to proliferate (in the right environment) in chain-reaction style. In the information age it is ideas that ultimately determine economic success. Bursts of success of the type seen in high-tech sectors, such as computer-oriented electronics and bioengineering, are the result of the combinatoric process. As noted, however, the ideas must be good ones if they are to foster growth. And therein lies a critical problem.

THE RISK OF FADS

Sorting out good ideas from bad ones and useful ones from amid the dysfunctional is as essential to the growth process as generating new ideas. But the task is difficult and must be performed in the face of considerable uncertainty. Seemingly promising ideas and innovations may turn out to lead

nowhere. The market ultimately will reward true success. But for a time it may also reward the adoption of notions that seem useful but eventually come to be seen as dead ends or worse.

Financial markets teach us that herd behavior—which now goes under the name “rational economic bubbles”—can lead to (temporarily) profitable departure from fundamentals. A rise in price of an asset (stocks, currency) attracts investors who use recent history as a guide to the near future. Soon prices are rising because they are rising and herd-following market participants (called “noise traders” in the finance literature) all enjoy the resulting gains. Eventually, however, the bubble bursts to the disappointment (and, perhaps, substantial loss) of all.

Although the field of management—and human resource management in particular—has a different set of incentives from financial markets, bubbles of a different sort are always a danger in the marketplace of ideas. What is more commonly called a fad is really an idea bubble. A plausible notion is put forward and produces some apparent successes—or is at least correlated with success. (In the real world of market uncertainty it is often difficult to separate cause and correlation.) Imitators are attracted. Money can be made by appearing to locate the key practices that appear to lie behind success stories.

Managers will want to get on board early on, just as investors want to get “in at the bottom” of a rising market. On the one hand, doing what appears successful elsewhere carries little risk; if the practice ultimately does not produce the desired results, the built-in defense is that everyone else was doing it. And at low cost; relabeling existing practices with the latest nomenclature is not especially costly. On the other hand, not adopting the latest thing and actually being left behind if it does work carries grave career risks. How can a manager justify not doing what—in hindsight—appears to have been a sure thing?

Most fads do start with a useful idea. Scientific management, human relations, management by objectives, quality of worklife, and total quality management all began with good ideas. But all ultimately were oversold. Even fad-following itself became a fad in the late 1980s and early 1990s; what was the “benchmarking” craze other than the idea of imitating practices that seemed to characterize successful enterprises?

Ultimately, therefore, our goal in presenting this handbook is to help the reader examine contemporary ideas and to filter out the current rage from fundamental advancements. We urge a healthy skepticism, but also a receptivity to innovation. Achieving the optimal balance between the two is the key.

HANDBOOK STRUCTURE

Our handbook is divided into seven sections. We begin with two sections on “employee participation,” the first being of the “nonfinancial” variety, that is,

participation in decision making. Such participation can range from the traditional suggestion box—hardly a new idea—to elaborate schemes aimed at creating self-managing work teams. The latter notion, incidentally, is less new than many practitioners and proponents realize.

Participation: Nonfinancial

As *David Levine* points out in chapter 2, the problem for management is to convince employees that participation and the effort it requires will bring about benefits for *them*. He emphasizes the need for “trust” to be developed and sees the failure to develop trust as the key cause of the high failure rate of participatory experiments. Naive orders from on high to implement participation are headed for failure.

In any event, participation takes many forms. The quality circles of the early 1980s were purely consultative. These programs seemed to have had short-run, but not long-run, benefits. Team production is the current rage but Levine notes that studies of the consequences of team usage are ambiguous. To make teams work, the employer must encourage organizational solidarity through job security and narrow pay differentials.

Levine finds that formalized systems—such as European-style works councils—do not seem linked to productivity gains. Foreshadowing the next handbook section, he concludes that financial participation tied to nonfinancial seems the most effective policy. In particular, total quality management (TQM) is likely to be perceived by workers as a newfangled speed up if it is not connected to the reward system. Successful human resource practices come in packages. Managers cannot simply buy them off the guru’s shelf one at a time.

Adrienne Eaton, Paula Voos, and Doug-one Kim further disaggregate the range of participatory practices available. Participation may be *on line* or *off line* and may be *voluntary* or *involuntary* for participants. In off-line systems, small groups of employees are asked to offer solutions to problems but higher management ultimately decides whether the solutions will be implemented. In on-line systems, the employees have direct authority.

With the exception of collective bargaining—which the authors remind us is a form of participatory arrangement—participation is generally voluntary for employers in the United States. (In European countries, there may be legal requirements to establish works councils.) Middle managers and supervisors may be ordered to set up participatory systems by top managers although nonsupervisory participants are typically volunteers. Eaton, Voos, and Kim note that while American management dislikes mandates from government, it seems to feel that mandating participation within its own middle ranks is appropriate. Financial participation is often involuntary for the individual worker, that is, mandated by higher management decision makers.

Although the politically correct view would probably favor completely voluntary arrangements, Eaton, Voos, and Kim do not conclude that such practices are always best. Again, there is much more ambiguity in the research literature about best practice than commercial consultants are likely to suggest to their clients. Cautious experimentation is therefore indicated.

An important element in participatory systems is the flow of information that can be two way or one way, as *Morris Kleiner* and *Tzu-Shian Han* point out. There is a long history of grievance systems, attitude surveys, suggestion boxes, and the like in business enterprises. More elaborate two-way systems, particularly in nonunion settings, may run afoul of the Wagner Act's ban on company unions and fall under the recent *Electromation* decision of the National Labor Relations Board. But simpler one-way systems in which management attempts to gauge worker opinion are not likely to present this legal problem. Where there are unions, information may flow through the negotiations process. And, indeed, strikes and lockouts themselves provide (costly) information to both sides about the strength of the other party's position. There is some evidence that managerial failure to react to what it is hearing eventually discourages the workforce from providing new information. Financial sharing arrangements can reinforce information sharing by convincing employees that all benefits will not flow to others.

Participation: Financial

Given the conclusion of those reviewing nonfinancial participatory systems that a program of financial participation can be (should be?) an important adjunct, the next section of the *Handbook* deals explicitly with such programs. Financial participation comes in various forms. *Douglas Kruse* and *Joseph Blasi* consider one version: employee ownership (majority or minority).

As Kruse and Blasi point out, the motives behind employee ownership vary. At the public policy level, where tax benefits create incentives to establish employee stock ownership plans (ESOPs), one motive has been "social," that is, the promotion of wealth sharing. At the firm level the motive is more likely to be micro: providing an incentive for increased productivity. Arrangements may include ESOPs, stock purchase plans, deferred profit-sharing plans (through which the company's stock is bought), stock-option plans, and 401k plans (tax-favored saving plans in which employees may accumulate shares but may also have other assets from which to choose). Other than the fact that larger firms are more likely to implement such plans than small ones (perhaps because of economies of scale in administrative costs), it is difficult to predict what kinds of firms will have them.

Situations in which ESOPs have been used to fend off takeovers or as part of union contract concessions (especially in the 1980s) have received substantial media coverage. But most stock ownership does not stem from such

circumstances. There is some evidence of a link between ownership and positive worker attitudes, but more direct evidence linking ownership to higher productivity and profitability is fuzzy. Still, there does seem to be some productivity connection.

As many prior authors have noted, ESOPs—and other large group incentives—raise the so-called $1/N$ problem. If N is the number of participating employees, and if N is large, the reward an individual worker will obtain from added effort is small. The fruits of the effort are divided among many other workers, an arrangement that could encourage free riding.

Like stock ownership, profit sharing does not necessarily entail employee participation in decision making, as *Derek Jones, Takao Kato, and Jeffrey Pliskin* point out. Gain sharing is more likely to have such an element. Although there is not much empirical research on gain sharing, profit sharing is found in many countries and a significant international literature has developed.

Pinning down the impacts of such plans on firm performance runs into the standard range of statistical problems. Control variables are needed but not always available. Measuring the plan as an explanatory variable itself poses a problem. (Is it mere plan presence, the proportion of workers covered, or the amount of the payout that should be used as an explanatory variable?) Complicated interactions may exist between the plan and other human resource practices.

Particularly with regard to profit sharing, the academic/economic literature departs from the traditional concern of human resource managers—productivity—and considers other effects. Economists, notably Martin Weitzman (quoted earlier), have looked at the impact of profit sharing on employer behavior rather than on employee effort. Because profit sharing creates a gap between the marginal cost of labor (the wage) and the average cost (wage plus the profit-sharing bonus), firms with profit sharing are expected to increase their hiring and to be less likely to lay off their workers in downturns. Whether this theoretical prediction is borne out empirically has been the target of considerable research.

Some evidence has been found for the employment-stabilizing effect of profit-sharing plans, an outcome that could be of social benefit. If there are macro (economy-wide) benefits from profit sharing, they will not be captured by individual employers looking only for micro effects (more productivity). Hence, a case can be made for offering encouragement of profit sharing through such devices as tax-favored treatment. Such treatment was part of the Weitzman proposal to encourage more widespread use of profit sharing.

Beth J. Asch and *John T. Warner* take a more micro approach to the general problem of pay incentives. In a world of imperfect information, employee behavior cannot be completely monitored or monitored without cost. Opportunities arise for shirking, a form of moral hazard since the employee is partially “insured” from detection by information and monitoring costs.

These sorts of problems are the staple of principal-agent theory: How does the principal (employer or supervisor) induce appropriate behavior from the agent (employee)? Various compensation-related devices have been used by employers for this purpose.

Under piece rates (or sales commissions), higher productivity results in more pay. But the devil is in the details. Common problems of these pay plans are a focus on quantity over quality, lack of incentive to avoid wastage of materials, and dynamic games as employers seek to determine performance standards (through time and motion studies) and workers respond by restricting output. Workers are generally assumed to be more risk averse than employers so such systems typically have a base wage (a minimum guarantee) as well as a production-related component. The more that is guaranteed, however, the less the incentive to supply more effort. A tradeoff must be made; no pay system gives perfect results, that is, results identical to what would occur if information were costless.

Of course, by offering variable pay plans employers may encourage self-selection of employees. Workers may be less likely to take jobs with highly variable pay arrangements if they are risk averse; so piece rates may attract risk takers. Or such plans simply may attract employees who know themselves to be more productive than average.

There are many other difficulties with regard to actual design of piece rate systems. They reward individual behavior, not teamwork, and might encourage competition where cooperation is needed. If teamwork is important, employers can install group (rather than individual) piece rate systems. But these can invoke the aforementioned $1/N$ problem if the group becomes too large.

Incentives are important for executives as well as nonsupervisory employees since executives, even top executives, are agents of the firm's owners. Various models of executive incentives have been put forward. In tournament models, for example, promotions are seen as rewards. Executives compete for a limited number of higher-ranking job slots which pay significantly more than their current positions. As they climb up the organizational pyramid, the number of higher opportunities for executives becomes more and more limited. (At the very top, there is only one CEO slot.) However, as with individual piece rates, competition, rather than cooperation, is encouraged. Office politics and sabotage are part of popular corporate lore as a result.

In any event, as with piece rates there is a selection process involved as well as an incentive process. If there is a hierarchy of jobs and salaries, those who are better "matches" with the firm will tend to remain with the employer and so move up the ladder. This upward drift—a learning model—will occur even if no one works harder than they otherwise would without a promotion reward. Of course, direct incentives to work harder can also be included in the executive pay program (such as stock options).

The observation that executive pay has been rising relative to ordinary worker pay has sparked considerable debate as to whether executives are now overpaid, perhaps because they can influence their own pay setting. Of course, the question of whether executives are overpaid today can be stood on its head. We could ask if they once were underpaid, since a clear nonmarket standard of pay adequacy—at least among economists—does not exist. (Ordinary mortals may have notions of pay equity apart from market outcomes!) Asch and Warner cite some evidence that, in the 1930s, the executive/ordinary worker pay gap may have been *wider* than it is today.

Most employees are paid on a time basis (per hour, per week, per month, or per year). Yet employers would be loathe to say that all they want from workers is their physical presence at the workplace. So with time pay comes considerable monitoring. A minimum work standard may be set by employers and those employees who don't measure up are dismissed. On the positive side, some form of merit pay—with merit bonuses or wage increases based on subjective supervisory evaluations—may be provided. Of course, subjectivity has its downside relative, say, to the tangible measurements involved in piece rates. But for many types of jobs, especially professional and managerial, counting "pieces" may not be practical.

One element in all participatory pay systems is that they are not based on new concepts. The idea of having an element of pay based on performance of the firm, plant, work group, or individual is an old one. Yet in the 1980s, the phrase "pay for performance" came into vogue. Was the phrase a fad or had something real happened to make the climate more favorable to such pay arrangements? Certainly, it is possible to point to factors in the market environment that would have led employers to seek more pay flexibility during that era.

In fact, lack of wage flexibility has long been a puzzle to economists from both the macroeconomic and microeconomic perspectives, as *Erica Groshen* and *Mark Schweitzer* point out. At the macro level there are two questions: (1) Why don't wages fall indefinitely when there is excess worker supply (high unemployment)? (2) Would falling wages in fact help cure the excess worker supply? Both questions are based on the micro analogy. At the micro level, excess supply is supposed to trigger price reductions, at least in auction-type markets. The price (or wage in this case) falls to the level that clears the market.

Until the 1970s some form of wage rigidity was simply built into macro models. Classical models assumed real wage rigidity; (old) Keynesian models assumed nominal wage rigidity. Given some degree of wage rigidity, would a move to lesser rigidity help the economy? The answer is ambiguous; in some models rigidity has a stabilizing effect by keeping up wage incomes and consumption. In others, since the rigidity is a major factor in creating unemployment when demand falls, less rigidity helps avoid it. Finally, in some of the original "real business cycle" models, wage rigidity was not postulated at all, although more recent models include it.

If there is macro wage rigidity, it must stem from the actions of myriad micro-level wage setters. But why should employers not cut wages (and pass up potential profits) when there are people lined up for job vacancies? Sometimes such rigidity is attributed to unions, not employers. However, in the modern U.S. economy—where unions represent only a small fraction of the private workforce—that explanation is implausible. Moreover, some of the more dramatic examples of wage cutting in the 1980s occurred in the union sector.

Numerous alternative explanations were put forward in the 1970s and 1980s involving job insurance and implicit contracts, efficiency wages, insider-outsider employee conflicts, and worker norms of fairness. Most of these explanations, examined closely, are better at explaining real wage rigidity than nominal. Some element of “money illusion” must be added to produce a nominal explanation. As Groshen and Schweitzer note, there is no shortage of theories in this area. But there is a shortage of empirical tests capable of differentiating them. In this case, therefore, it is unclear that theory has been able to provide much guidance to employers concerning appropriate wage policy in a labor-surplus market. Still, employers might well reconsider their wage policies if they cannot rationalize for themselves what they do.

Employer Flexibility

Flexibility in wages is an area that has intrigued economists. But other forms of flexibility can be exercised by employers. *Richard Belous* points to the rise of “contingent employment” in the U.S. workforce (and some evidence of similar developments abroad). The notion of career (internal) labor markets is fading, he argues. Labor is no longer treated as a quasi-fixed cost. He cites as causes pressures on management from such influences as leveraged buyouts, hostile takeovers, and demands for performance on behalf of shareholders which arose in the 1980s. Originally, the move toward contingent employment was based on division-level and business-unit decisions. But more recently, Belous argues, the move has been based on strategy from the top.

Belous's estimate of 27-32 percent of the workforce as contingent is much higher than the 5 percent estimate of the U.S. Bureau of Labor Statistics. Part of the difficulty is in defining exactly what is meant by “contingent.” If only temps hired through temporary help supply agencies are included, the figure actually would be closer to 2 percent, for example.

What may be occurring, however, is that employers are backing away from job guarantees for most employees, even if employees are not being pushed across some arbitrary line separating contingent from non-contingent. The shift to reduced job security has put strain on the social insurance system which was built up after World War II on the basis employer-provided insurance and pensions. But are there some limits to the shift toward contingency? There

may be limits, according to Belous, since issues of quality and productivity growth over time are hard to manage with a contingent workforce.

Flexibility can take the form of hiring and firing without constraints (with temps being the most flexible employment arrangement). But as *Susan Houseman* points out, the United States has long been more reliant on layoffs (and hiring) to deal with business fluctuations than Europe or Japan—particularly with regard to blue-collar workers.

This tendency is reinforced by the American unemployment insurance system, certain aspects of which subsidize layoff behavior. European and Japanese employers absorb more of the risk of business fluctuations and seem to seek long-term productivity growth through training and retraining of redundant workers. Dismissals are made difficult by law in various European countries and there may be laws discouraging use of temps. Employers may also have to make substantial severance payments to laid-off workers. (Government may also have to make significant outlays in the form of various retraining programs and adjustment payments.) In Japan, where lifetime employment is often depicted by foreign researchers as based on national culture, there are in fact also legal constraints.

The United States has put some minor limits on employers, most notably in a requirement enacted in the late 1980s requiring 60 days advance notice of mass layoffs and plant closings. There has also been some legal erosion of the at-will employment doctrine. So the United States has shown some signs of moving toward the European model. But at the same time, high unemployment rates in Europe have led to a softening of legal restrictions there. Policymakers hope that more employment will result if employers do not fear being “stuck” with unwanted workers.

Since employers in Europe and Japan are less willing or able to rely on layoffs, they are less likely to utilize narrow job classifications. Workers need to be given sufficient skills and capabilities so they can move from job to job within the firm. Elaborate seniority-related layoff rules are less common. An interesting question, however, is whether market pressures will cause substantially more erosion of the European and Japanese models than has already occurred.

Even if there is such slippage, the consequence need not be more rigid job designs, since market pressures may also be pushing toward more flexibility in that area. *Eileen Appelbaum* and *Peter Berg* argue that there are two possible strategies for employers to follow: lowering labor cost or enhancing productivity. Either would produce lower unit labor costs—the ratio of the wage to productivity. Although the contest between the two strategies has usually been posed in the manufacturing context, the same dichotomy exists in services.

Appelbaum and Berg trace the history of work design from turn-of-the-century Taylorism through the challenge posed by early behavioralists to the Socio-Technical Systems and Quality of Working Life movements of the 1960s

and 1970s. A dichotomy remains in modern workplaces; lean production involves worker input but that input can be rejected by management. In contrast, self-managing work teams make on-line decisions.

As have other authors, Appelbaum and Berg conclude that modern arrangements need reinforcement through compensation practices, job security, and training investments. One might add that resistance to such arrangements may arise from workers, unions, and managers. Thus, mechanisms to overcome such resistance also need to be developed and adopted. As is often the case, there is more rhetoric surrounding the newer arrangements than actual implementation.

Unions and Collective Bargaining

As noted earlier, collective bargaining is a form of employee participation in decision making. However, it is not a form management generally prefers, partly because with unions may come higher wage and other costs, but also because the "advice" received via the bargaining process is not always easily rejected. Ultimately, the union's bargaining power in a collective bargaining relationship stems from its ability (whether actually used or not) to impose costs on management if its demands are not heeded.

Although various mechanisms for imposing costs are possible (such as recent "corporate campaigns"), the strike is the highest-profile union weapon. And strikes are the easiest weapon to measure and track, which makes them most amenable to academic research. Of course, not all bargaining need be an adversarial division of the economic pie (distributive). There can be integrative bargaining—now popularly called "win-win"—in which the pie is enhanced. Still much of modern bargaining remains adversarial with the strike threat always in the background.

Exactly how powerful the strike weapon is should vary with the "structure" of bargaining. As *Dan Gallagher* and *Cynthia Gramm* note, bargaining can occur in various structural combinations (single employer, multi-employer, single union, multi-union). In the United States most contracts involve a single employer and a less-than-firm-wide contract, but there are many exceptions. There is evidence in the United States and abroad that bargaining is becoming more decentralized.

Decentralization is often seen as a cause or effect (it is not clear which) of weakened union bargaining power. A reduced scope of "pattern bargaining"—situations in which one settlement is imitated by others—is viewed as the primary symptom of such decentralization in the United States. In other countries decentralization may take the form of a lesser tendency for there to be economy-wide deals or social accords involving union and management central organizations and government.

Is there a substantial amount of strike activity or just a little in the United States? Most contract negotiations do not end in strikes, although size of contract (number of employees covered) is positively linked to strike risk. Gallagher and Gramm put the proportion of "major" contract negotiations (those involving 1,000 or more workers) which end in strikes at 13 percent. Although this percentage is often characterized as a small strike risk, a one in eight probability is not so low if the costs of a strike are considerable.

For years economists have wrestled with the issue of why strikes occur. Even if the strike threat is important, there need not be an actual strike so long as both sides understand the costs and act "rationally." Of course, this situation is no different than is found in other conflicts. If two enemy countries perfectly understood their own respective military strengths and that of their adversary, no war should occur. (They would just settle peacefully for the inevitable outcome and avoid the costs of war.) Law suits and criminal cases should not go to trial if both sides have perfect knowledge of the future outcome. Sadly, we do have strikes, wars, and trials so better models are needed to explain them.

Not surprisingly, current models rely on faulty (imperfect) information. The parties blunder into impasse because of misperceived bargaining strength. Often models assume that workers have the least reliable information and push their leaders into conflict. It may seem unfair to blame one party to a dispute more than others, but economic analysis suggests that *someone* has to bear the blame and that blame need not be borne equally. A random distribution of misperceived information across all parties should produce random strikes uncorrelated with other economic events (such as the business cycle).

To obtain systemic correlation with economic events, the distribution of information must not be random. Many variations of asymmetric information (or information processing) are possible. Alternatively, one could conceive of periods in which information is systematically more hidden to all parties. If such periods are correlated with economic events, there could be a blundering into conflict tied to cyclical indicators.

An alternative view is that both parties work jointly to minimize strike costs over the long term (which extends beyond any particular contract negotiation). In such models, inconsistency (saying "yes" after saying "never") could cloud the implicit information exchange in subsequent negotiations. A short-term accommodation could lead to longer-term frictions. Hence, it is better to hold one's ground, even if a concession might seem better in a one-period perspective. (Strikes over small amounts of money seem to fit into this category. The costs of lost wages due to the strike may not be recouped in the subsequent contract but a "lesson" will have been taught to the other side.)

Behavioral models often differ in assumptions about the "rationality" of the parties. Often, empirical behavioral studies do not deal directly with the kinds of problems posed by economic reasoning and instead focus on worker attitudes toward striking. (For example, researchers might ask what personal

characteristics make workers more militant.) This dichotomy between economic and behavioral analysis lead Gallagher and Gramm to call for future attempts at interdisciplinary research integration.

Where conflict arises in a union-management situation, it is not necessary to have the unresolved issues end in a strike, even if an impasse is reached. The parties could turn (or be compelled by law to turn) to a third-party neutral person or body, either to resolve the dispute or to help the parties find their own resolution. As *Victor G. Devinatz* and *John Budd* demonstrate, use of mediation, factfinding, and interest arbitration are hardly new; their usage in the United States can be traced to the late nineteenth and early twentieth centuries. Similar histories apply in other countries.

There have been waves of optimism and pessimism about such intervention techniques. Proponents see them as avoiding strike costs and thus encouraging union-management harmony. But critics see them as fraught with various "chilling" and "narcotic" effects that ultimately will make the parties unable to settle their own problems. The terminology is obviously loaded; no one would choose to be frozen or addicted! However, it is reasonable to analyze the effectiveness of third-party intervention to see if such dysfunctional impacts actually occur.

Recent models of interest arbitration—a technique that is comparatively rare in the private sector but common in government—provide new insights into the process. Older models depicted the arbitrator as a passive "split-the-difference" agent whose decision process would encourage the parties to make wild demands and proposals. These models led to policy proposals such as use of final-offer, rather than conventional, arbitration.

But the more recent models are based on nonpassive arbitrators who have their own norms of fairness. In such cases, the parties will array their demands around the arbitrator's fairness norm, giving the eventual decision an illusionary split-the-difference appearance. Arbitrators may also play a useful role where information is asymmetric, especially if workers are particularly ill informed. Union officials can then blame the arbitrators for imposed settlements they secretly know to be realistic.

As in the strike case, there is an empirical literature that is not always tightly tied to economic modeling. For example, interest arbitration has been found to favor workers slightly in the sense of providing higher wages than might otherwise have resulted. And wage dispersion may be reduced. Presumably, these results stem from arbitrator norms of fairness although it is not clear how these norms are formed. An interesting question is whether arbitrator training in the United States is appropriate for interest arbitration. Most arbitration involves interpretation of "rights" under union-management contracts (typically grievances), a task suited to the lawyers who have come to dominate professional arbitration. But is training also needed in business, economics, and industrial relations for those handling interest disputes?

As noted, interest arbitration has been much more likely to be used in the public sector than in the private. *Leo Troy's* chapter on unionization suggests that public sector unionization will become larger as a proportion of all unionized employees due to continued shrinkage of private sector union membership. Private unions will end the twentieth century with about the same low percentage of the workforce covered by collective bargaining that they had at the beginning of the century.

Unions will remain significant in certain manufacturing industries and in other selected sectors, such as air transportation. But even in manufacturing there may be further erosion due to foreign competition. And even if *Electromation*-type constraints on employer-sponsored employee representation plans are relaxed, the bulk of private workers will not be unionized and may not have any form of employee involvement. Individual employer-worker "contracts" (mainly implicit) will be the norm in private employment.

Public-sector bargaining and representation is the "New" Unionism (which grew rapidly starting in the 1960s). But public successes for unions will be difficult to duplicate in the private sector ("Old Unionism"). The problem for U.S. unions in private employment is not unique. Troy notes that in other countries, such as Canada, where the legal system is more union-friendly, private-sector unionization rates are also declining. Only in Scandinavia does the trend not appear.

Troy argues that employee preferences and product-market competition, rather than employer resistance, are the chief causes of worldwide union decline. (Employers may be led to resist by foreign and domestic market competition, however). Unions have also been hindered by a tendency to stay in old industries and occupations whose employment shares are shrinking.

After a palace revolution in 1995, new leadership was elected in the AFL-CIO and promised to devote new resources to union organizing. Troy does not think leadership per se can halt the shrinkage in private sector unionization but does think there is room for improvement in union leadership quality. An interesting question—if his prediction of continued private decline is correct—is the degree to which unions will turn to political channels, rather than traditional organizing, to exert influence.

Although the union share of employment in the United States has declined since the mid 1950s, academic interest in unions and bargaining has not waned proportionately. Indeed, as the review of *John T. Addison* and *John Chilton* indicates, academic modeling of union behavior has advanced in recent years. But the recent modeling has echoes of previous debates.

In the 1940s there was debate between Arthur Ross and John Dunlop concerning the appropriateness of applying economic analysis to union wage setting. Dunlop suggested that unions "valued" both employment (membership) and higher wages and he provided early models that incorporated these two objectives. Ross argued that the union should be viewed

as a political entity whose behavior was too complex to be modeled using simplistic economic analysis.

Modern economic modeling tends to follow Dunlop but tries to incorporate political analysis. Median-voter models, for example, add a political dimension to understanding the union objective function. Many models depict the union as a "monopoly" that sets a wage after which the employer reacts by setting employment. The passive management role assumed in such approaches, however, seems out of line with the phrase *collective bargaining*—which implies a process involving two parties.

Nonetheless, such modeling does raise an important question. A process in which the union values employment as well as wages, but only negotiates about the latter, is bound to be "inefficient" in an economic sense. The two parties should mutually bargain about—and determine—both variables of concern. (Indeed, with an eye on the long term, there should be bargaining about wages, employment, and future investment plans). Yet the notion of unions bargaining over wages and leaving management to set employment (and investment) squares with institutional reality in many cases. Either there is more to bargaining than the common perception suggests or union leaders at least should worry about the traditional narrow scope of bargaining.

Modeling of union behavior has led to empirical research on union productivity impacts. Management, citing restrictive work rules, tends to see unions as productivity-diminishing. However, if unions have a wage-raising impact—which empirical research confirms—there should be a productivity-enhancing effect resulting from pushing the firm up its labor demand (and marginal product of labor) curves. (This prediction assumes that the employer chooses employment in response to the negotiated wage.) Some researchers, however, see still further productivity enhancement through other effects.

The main such effect is the creation of a channel for employee "voice" which can reduce turnover (and hiring/training costs) or improve morale. Empirical evidence for this proposition is ambiguous—which is itself surprising given the management view that unions surely lower productivity. What may be occurring is that unions have some appreciation for the concept of efficient bargaining. That is, they ought to bargain about wages and employment. But they focus on wages and then discover belatedly that management reacts by substituting capital for labor and lower-paid occupations for higher-paid ones. Formal work rules then result as attempts to limit such substitutions.

Workplace Dispute Resolution

At one time, settling disputes in the workplace was seen as largely a union-sector topic because of the near-universality of grievance-and-arbitration provisions in union contracts. But as *David Lewin's* chapter points out, machinery for dealing with individual employment disputes exists in the

nonunion sector, too. In both sectors, of course, much dispute settlement is handled informally by immediate supervisors. But the nonunion sector has been developing more formal institutions for dealing with unresolved individual employment disputes.

To some extent, the external legal system may play a role in pushing such disputes—union and nonunion—toward formal resolution within the enterprise. In the union sector the legal “duty of fair representation” may lead to processing of grievances, even some that are frivolous. Unions may fear lawsuits by employees claiming inadequate representation. In the nonunion sector, the erosion of the at-will doctrine and other sources spurring growth in use of courts to hear employment complaints provide an incentive to create respectable, internal review systems.

Various other explanations have been offered for the growth of nonunion grievance systems. The productivity-enhancing possibilities of having a voice mechanism—originally discussed in the union context—could also apply to the nonunion sector. Procedural justice and fairness of process may be valued by employees. However, it is unclear whether employees will be more impressed by process or outcome. Workers might care about distributive justice and not care very much about how it is delivered.

Research suggests that nonunion employees do use available formal grievance procedures. Typically, settlements are at lower levels of the process. But some disputes end up at the highest level of the process, even if this stage involves the CEO rather than a neutral outside arbitrator. Research also suggests that the win rate for grievants rises with the level at which the complaint is settled.

An interesting question, however, given that the process is used, is the eventual fate of nonunion grievants after settlement. There is evidence that workers who file nonunion grievances have poorer subsequent work histories on average than nonfilers. This finding could mean either that there are reprisals against filers or that filers are poorer performers. Unfortunately, there is more support for the reprisal story. So exercising voice may not cut turnover—at least of the filers themselves. And there are eventual problems of distributive justice, even if the grievant is initially successful.

It is also true that supervisors of filers have poorer subsequent work histories than those of nonfilers. Given that the careers of the supervisors of grievants are placed at risk, the finding that the grievants are subject to later reprisals is not surprising. Supervisors may want to demonstrate to other subordinates that filing grievances is not a favored activity.

There are various types of nonunion grievance plans. Some plans provide no agent to represent workers. But there may be ombudspersons available to help workers who seek to file grievances. Some plans provide an agent only at higher stages of the process. Some plans include hearing panels, which may be of the peer-review type, and which usually form the last step of the

process. Use of an outside arbitrator is nowhere near as common as in the union sector. But where there is use of an outside arbitrator, typically the firm pays all fees. (This is a practice that some critics say can compromise the arbitrator's neutrality).

Given the tendency for nonunion grievance systems to take on some union-like elements, it is not surprising that research on union rights arbitration continues despite union membership decline. *Trevor Bain's* chapter provides a historical review of grievance-and-arbitration systems in the American union sector. He notes that grievance filing rates should have some relationship to the overall climate of industrial relations, but empirical estimation of that linkage is often elusive.

One problem is that the informal settlement of employment complaints, before they become measured grievances, often escapes the empirical researcher. However, with regard to formal grievance filings and awards, there is some evidence that these peaked in the mid-1980s, and then declined. Some linkage to the onset of concession bargaining may be involved.

The arbitration profession differs from some others in that formal certification and training are not required. However, arbitrators ultimately are chosen by reputation. And some experience is needed, in any case, to be listed by such agencies as the Federal Mediation and Conciliation Service. There has long been ongoing conflict between two models of arbitration. One model is an informal mechanism emphasizing dispute settlement while the other is a quasi-judicial review and decision process. The latter has become the dominant model in use but complaints about cost, time involved, and excessive legalism have led to experimentation with less formal processes, such as expedited arbitration and "med-arb." Some research suggests that despite the neutral formality of traditional rights arbitration, arbitrators can be influenced by such irrelevant factors as the sex of the grievant.

Outside the United States, the external legal process may be more involved in the settlement of rights disputes, even where unions are involved. (This observation flies in the face of the usual complaint that Americans are culturally biased toward legal remedies.) There may be special labor courts to hear cases. The presence of such legal arrangements does not mean, however, that informal dispute settlement is not used abroad. In some cases, for example, works councils may become involved informally.

A major difference between foreign and American dispute resolution is the use of money damages, rather than reinstatement, as the remedy of choice abroad when a grievance over a discharge is upheld. Foreign systems seem implicitly to assume that putting the worker back where the initial friction arose may simply increase workplace tensions. The American approach implicitly assumes that reinstated workers can be protected from reprisal because the union that processed the grievance is available for that function.

Modelers of union behavior and students of employment dispute settlement have emphasized the Hirschman choice of *exit* (quitting) versus *voice* as the response to workplace unhappiness (Hirschman, 1970). However, as *Brian Bemmels* notes, the choice is more complicated than this dichotomy suggests. For example, Hirschman suggested that *silence* (suffering without taking action) was yet another option. And he saw *loyalty* as moderating the choice between exit and voice. Loyalty made it less likely a worker would choose exit, even if he or she had a complaint.

Another possible reaction of a disgruntled employee is *neglect* which could be exhibited as dysfunctional behaviors such as absenteeism, carelessness, and so on. Still more extreme might be *destruction*. That is, employees might engage in theft or sabotage.

There is an empirical literature predicting what kind of response is likely, using personal and other characteristics of workers. The choice can be modeled in terms of the costs of selecting response options. These costs can be viewed as constraints on worker choice. As an example, possible retaliation could be the perceived cost of filing a grievance. Apart from personal characteristics, cultural norms can influence the route chosen. These can vary across countries and even among groups within countries.

Opportunities may influence choice of response option, for example, union workers probably have more opportunities to exercise voice than nonunion. Studies have shown that after controlling for pay, union workers are less likely to quit than nonunion (and are more likely to use voice). In contemplating the exit option, unionized workers must consider that there are non-pay elements in union settings which they might have to sacrifice (such as seniority-based privileges). And, of course, the advantageous union pay differential might have to be sacrificed if the worker quits a union job.

It is often assumed that voice carries more information for the employer than simple exit. If so, there is a rationale for creating formal voice mechanisms in nonunion firms. But the assumption may not be correct. In cases where there are many grievances, management may not pay much attention to any of them. In contrast, quitting at high rates may get management's attention if turnover is costly to the firm.

Staffing and Reward

For many years, there was a sharp division in U.S. employment relations between "industrial relations"—which had a union-sector connotation—and "personnel"—which generally applied to nonunion employees. Traditional topics in personnel texts were selection, training, performance appraisal, wage setting, and benefits. As the nonunion sector became dominant in private employment, "personnel" was given the more elevated title "human resource management." With the new title came the implication that the function was

something for top management to consider along with other major functions such as marketing and finance. On the academic side, one result of this title elevation was renewed attention to traditional personnel topics.

Selection is the logical first topic in human resource management since it involves initial hiring. As *Cheri L. Ostroff* and *Teresa Rothausen* point out, both parties, employer and employee, are looking for a good "match." However, how good the match must be will vary with the expected duration of employment. If employment is becoming more contingent, perfect matches may be less necessary than in a career labor market. But if loyalty and specific knowledge are important to an employer, and if turnover is costly, use of techniques which ensure good matches is important, even if there is a trend toward contingency in the overall labor market.

The employee selection process should be tied to other systems of human resource management. Alternative strategies are possible. For example, employers can select people with preexisting skills so that further training is not needed. But another approach is to select people based on ability to be trained and then train them. To the extent that employer-provided training involves acculturation, the resulting employees might be more suited to the strategic goals of the organization. Still, there are many industries that rely on external, preexisting training successfully, for example, entertainment (motion pictures, television, multimedia) and construction.

Selection is affected by legal requirements especially equal employment opportunity (EEO); it is not simply a reflection of internal employer strategy. Under EEO regulations, it is often necessary to show the job relatedness of hiring criteria. Thus, employers need good selection models so they can defend themselves against claims that selection processes were not job-linked. However, just looking at employment skills and tasks is not enough. Employers need to search for a hiring match that fits organizational needs where career employment is the norm.

Ostroff and Rothausen propose five steps for the selection process. First, there needs to be a job and organizational analysis of the necessary knowledge, skills, and abilities (KSAs). Second, there should be a determination of criteria for selection. (What does "doing well" at the job entail?) Third, there needs to be a determination of appropriate predictors. (What characteristics will enable workers to do well on the job?)

Fourth, it is important under EEO rules to validate that the predictors accurately forecast job performance. Fifth, there needs to be a cost-benefit analysis of the selection process itself. Is it worth doing, even if it produces a good selection? Does the process scare away good applicants? Does it enhance or damage the employer's reputation? Despite the considerable academic research concerning selection criteria, many employers depend on unstructured interviews and reference checks that are often not reliable. Are they implicitly saying that more elaborate mechanisms are not worth the bother, despite the

greater accuracy they could provide? Or are employers just unaware of the potential for enhancing the selection process? The answer is not clear.

Once employees are selected and hired, training may be required. Or, as noted before, employees may be selected based on the training they already have. As *Orley Ashenfelter* and *Robert Lalonde* note, interest in training by academics and policymakers has intensified due to various economic trends. These include the productivity slowdown that began in the early 1970s and the rise in income inequality (with more educated workers doing considerably better than those without much education). There has also been concern with the American educational system, especially at the primary and secondary levels. Employers often complain they are forced either to provide remedial training or to sort through large numbers of unqualified applicants.

There have also been concerns that with employment becoming more contingent, and with traditionally higher turnover rates in the United States than in Europe and Japan, American firms will be reluctant to provide "enough" skill-enhancing training. Much depends on the employer's ability to capture training costs, either in the post-training period or during the training itself. Economists have long made a distinction between general and specific training. Employees are thought to pay up front for general training (valuable to many employers), either through formal tuition (as at a vocational school or college) or by working at a low enough wage while being trained so that the employer can collect a kind of *ersatz* tuition.

Specific training—valuable only to the employer—is commonly pictured as collectable in the post-training period. However, the distinction between general and specific is blurred in practice. For example, employers who provide specific training must nonetheless be concerned about turnover (loss of the training investment) and are induced to pay a higher wage to reduce quits. Workers are thus likely to face an upward-sloping pay profile regardless of whether their training is general or specific. The profile slope can be provided via the wage itself or through a deferred benefit such as a pension.

Unfortunately, the empirical study of training is made difficult by the crudeness of available data on training practices. Much training appears to be informal and is provided by co-workers. There appears to be a complementarity between training and formal education. Education may make workers more trainable. Or perhaps education is a signal that training can usefully be absorbed. In any event, the correlation suggests that those who have little formal education will have a hard time catching up on their skills by relying on employer-provided training.

Researchers' estimates of the value of training conducted annually in the United States range from \$150 to \$200 billion. The direct approach to estimating this value consists of adding up costs of known training and schooling. An alternative indirect approach uses differences in pay between trained and untrained workers. But both methods yield similar results. Once

training is seen as an investment, it is important to consider its rate of depreciation. For example, one study estimates that skills depreciate at 10 percent per year.

Various proposals have been made for subsidizing training by those who believe the United States is underinvesting in human capital. However, the correlation between formal education and training suggests another approach. Increased formal education might well lead to more job-related training without the complications involved in trying to target employer training subsidies.

Once employees are selected, hired, and provided with whatever training goes with the job, their performance on the job must be periodically assessed. *Robert L. Heneman* and *Courtney Von Hippel* point out that the literature on performance appraisal at one time focused heavily on the appraisal instrument (typically a form) and on individual appraisal. However, now there is greater emphasis on the entire appraisal system and on group, peer, subordinate, and even customer appraisal—so-called “360 degree appraisal.” Attention has turned toward performance management rather than simple appraisal. Still, without good measurement there cannot be such management. And the measurement system must be reliable, valid, fair, and practical.

Qualities assessed must be in accord with the business plan of the organization. For example, if the organization is based on customer service, then standards must be set for providing that service. Also, assessments must also be in accord with organizational culture. If the organization depends on self-managing teams then it should not adopt a performance assessment approach based on a command-and-control-type organizational culture.

In unionized firms seniority is often important for promotion decisions but some union contracts permit other forms of assessment. Thus, where a union is involved some balancing of seniority and other criteria is likely to be necessary, and tangible measurement, rather than subjective judgment, may be important. Because of the encroachment of the legal system on human resource management, appraisal systems also need to be defensible in court. Having an appeals procedure for employees who feel they have been misjudged can be important from that perspective.

Although systems may look good on paper, actual process is important. Among other problems, raters (typically supervisors) are often reluctant to do honest appraisals of their subordinates. There are various (mis)incentives that account for this supervisory reluctance. In recently downsized firms, supervisors may have substantially more subordinates in their units than before the layoffs, making it difficult to find the time for adequate appraisal. (Note, however, that the ratio of managers and supervisors to other employees has not been falling in aggregate labor-market data despite the substantial media attention devoted to corporate restructuring and downsizing.)

In any event, appraisals are often not done on a timely basis or at all. Even when they are done, common techniques used as part of the appraisal process

such as "management by objectives" are not necessarily reliable or valid. It is important to recall that not all outcomes are in control of rated employees. Thus, employees should not be blamed for failures beyond their control nor credited for successes based on the external environment. Even if not done formally, appraisal studies must control for outside influences.

Although the traditional model of appraisal consists of ratings by supervisors, other possibilities include ratings by peers, customers, or self-ratings. If multiple raters are used, and if the various ratings are in general agreement, management may have greater confidence in the overall appraisal process. Since there is resistance to functioning as a rater, and since rating takes time, the reward system should provide incentives to good raters. If both raters and workers are involved in formulating the rating process, they are more likely to "buy into" the system, thus producing a more accurate measurement process. Still, no matter how it is done, the rating process will produce workplace tensions. There will always be charges that "office politics" were behind the distribution of ratings and rewards. And there will always be a tendency to rate everyone in a narrow, "somewhat above average" range. These problems cannot be entirely eliminated, but they can be managed.

Even if performance appraisal is done well, the resulting ratings must be connected to the reward structure. As *John Fossum* and *Brian McCall* point out, there was a substantial revival of interest in "pay for performance" in the 1980s and 1990s. Yet it was unclear exactly what this phrase was intended to mean since, as noted earlier, even long-standing merit plans were lumped under this label. Since such plans were hardly new innovations, it is unclear that the new label had much significance in practice.

Compensation systems are not generally theoretically driven. But expectancy theory from psychology and agency theory from economics have had an influence. Overall, however, compensation managers basically are imitative. They follow what successful firms seem to be doing on the assumption that compensation practices must have had some role in those successes.

There is a need to define performance before linking pay to it. Many kinds of jobs have difficult-to-measure outputs. Generally, it is easiest to measure performance at top and bottom jobs in the pay scale and difficult in middle. At the top, what the CEO is supposed to accomplish can be defined by financial measures. Although it may be difficult precisely to determine how much success or failure is due to executive effort, comparisons can be made to other firms in the industry. These comparisons at least control for general market conditions. At the bottom of the pay scale, particularly for production workers, output can often be measured easily. Where results are not easily measurable, pay may be linked to inputs, that is, behaviors assumed to be related to outputs.

At any level except the very top, rankings of employees within the comparison group can be used rather than absolute measures. Such forced "grading on a curve" tends to overcome the temptation to rate everyone as

a bit above average. But it may do an injustice to high performers unlucky enough to be in a comparison group of other high performers. That is, both ranking systems and absolute systems have drawbacks.

Pay systems have a variety of elements. There is the issue of average level of pay, typically compared against some external benchmark firms. Such external surveying can be justified in classical labor-market terms. However, during periods of chronic labor surplus or shortage, such surveying may contribute to overall wage rigidity. Nonetheless, following others might be justified on some sort of fairness grounds.

Firms do not mechanically set their average wage equal to the average wage in the market. Rather, a distribution of pay decisions seems to form, with some firms consistently paying more or less than the average. A variety of "efficiency wage" theories have developed rationalizing higher-than-average pay. In the context of less-developed countries some researchers have linked worker nutrition to the wage; higher pay means better health and more productivity.

In the developed world other rationales are applied. For example, a pay premium could discourage shirking. If shirking is detected, the worker risks job loss and thus loss of the pay premium. Turnover costs might also be expected to fall as pay rises relative to the market average. Not all efficiency wage theories fit neatly into economic analysis, however. In some versions, workers simply "appreciate" the "gift" of premium pay and offer in return the "gift" of higher productivity to the employer.

Within the firm, there will be a pay structure, that is, a set of wage differentials across jobs. For many jobs there will not be a clear external market reference that can be surveyed. Hence, the differentials are commonly linked to benchmark jobs through some type of job appraisal technique in which knowledge, skill, and ability are rated and translated into monetary terms.

How individuals react to alternative pay systems will vary. Risk-averse individuals may self-select away from variable pay plans and firms which use them. If job security is put at risk, there may be a reduced incentive to learn specific skills. Because merit pay systems in practice often do not provide substantial pay variation, the risk issue is more likely to be associated with programs such as piece rates and commissions or with "up-or-out" systems, such as law partnerships or academic tenure.

But there can be variations in response to alternative reward systems across individuals which are not easily placed within standard notions of economic rationality. For example, experiments suggest that males respond to incentives by increasing goals but not performance; women—in contrast—set more realistic goals. Hence, the literature on pay often draws on psychological models rather than economic. The two, however, are not always in conflict.

Expectancy theory suggests performance will be higher if the reward is valued, the link to performance is perceived to be high, and if the worker believes he or she can perform. The worker makes decisions on job behavior

based on beliefs about the attractiveness (valence) of expected outcomes. If pay is put at risk, and if the worker cannot control performance, the result may be a perverse "negative coping behavior." *Equity theory* assumes that employees judge their pay outcomes relative to their input and relative to the outcome/input ratios of others. *Justice theory* has a distributive component (division of rewards is fair) and a procedural component (method of dividing rewards is fair).

Even when pay rates are determined, the structure of overall compensation must be determined. Some elements are mandated, for example, payroll taxes for Social Security. But other benefit programs are discretionary with the employer such as pensions, health care, vacations, and other leave arrangements. It is possible to rationalize some of these programs in terms of agency theory and employee preferences. However, the role of the tax code is important for insurance and deferred compensation. Executive pay plans, such as stock options, are also heavily influenced by tax law, since high-paid executives are likely to be in the top income tax brackets. Moreover, the large absolute amount of their pay makes investment in tax avoidance attractive.

Rena Broderick and *Barry Gerhart* extend the Fossum-McCall analysis of the composition of pay by looking explicitly at non-wage benefits. Apart from tax treatment and other economic rationales for the offering of benefits, norms of what compensation should look like become ingrained in the labor market. Workers simply expect that certain plans will be offered and employers therefore oblige by meeting these expectations for "market" or other reasons.

The result of such a process is a path-dependent system in which the original reason for offering benefits can become lost in history. In the United States the growth of unions during and after World War II was associated with expanded benefit offerings. The union interest in certain benefits at that time may have been partly ideological and partly to escape wartime wage controls.

Managerial perception and practice with regard to benefits seems at variance with economic analysis. For economists it seems natural to think of pay in terms of total compensation which is then divided into non-wage and wage elements in response to various incentives. Within firms, however, wage and benefit administration are often organizationally separate.

Economists believe that the incidence of benefit costs and payroll taxes falls partly or entirely on the employee. Put another way, employees "pay" for their benefits by receiving lower cash wages or salaries than they otherwise would (absent the benefits). Managers, however, seem to think of such costs as add-ons to labor costs which come out of profits on a dollar-for-dollar basis. The issue of health care cost containment and the public policy debate over mandating employer health coverage highlights this division. Employers clearly reacted as if they would pay for such mandates through reduced profits.

Employers often complain that employees do not know the true cost of the benefits they receive. (Such statements are partially at odds with statements

employers will sometimes make rationalizing the offering of benefits as being necessary for recruitment.) But it is also not always clear that employers fully understand the eventual cost of their benefit programs. Many employers reacted with shock to proposed changes in accounting rules which required them to put on their balance sheets the implicit liabilities entailed in retiree health benefit plans (which are usually unfunded). In theory, such liabilities should have been known to the firm and reflected in stock prices whether or not they were formally on the balance sheet. It appeared, however, that some firms were surprised at their own liabilities.

As employers have retreated, rhetorically at least, from job guarantees, the basis for employer-specific benefits has eroded. Defined-benefit pension plans penalize employees who change jobs, voluntarily or involuntarily. In contrast, defined-contribution plans and 401k saving plans are more suited to job mobility. Thus, it should not be surprising that defined-contribution and 401k plans have been growing rapidly as compared with defined-benefit pensions. The decline in union membership also may play some role in the decline of defined-benefit plans; unions have traditionally bargained for such plans.

Although employer-provided defined-benefit pensions may be rationalized as a form of incentives, it is more difficult to make that case for health insurance. Offering health insurance attracts and retains sick people! So the widespread offering of health insurance by employers presumably reflects a national health policy as reflected in the tax code (which exempts employer premiums and health benefits from all income taxation).

Given national policy, health care cost containment initiatives are left largely to employers with some government prodding. The result has been a trial-and-error approach involving HMOs and managed care. But some employer initiatives that make sense at the micro level (such as refusal to cover preexisting medical conditions) do not necessarily respond to public policy objectives. They, thus, become targets of federal policy proposals (such as banning refusals to cover preexisting conditions).

Even if benefit offerings are heavily influenced by the tax code and are path dependent, they should be modified as workforce demographics change. That is, employee preferences should play some role in what is offered. *Janet Currie* examines a series of sex-related benefit issues that have become progressively more important as the female proportion of the workforce has risen.

Much has been written about the male/female average pay differential, usually measured in terms of the cash wage only. But less is known about the benefit side. Generally, male workers are more likely to have benefits than females. (Sick leave may be an exception.) Males are more likely than females to receive on-the-job training. If the wage rate is used as a control variable, these gender gaps tend to vanish, an unsurprising result given the high correlation between wages and benefits. However, since wages and benefits are simultaneously determined, such use of the wage as an exogenous control raises analytical issues.

Even if benefits are provided, they may have different value and cost as applied to men and women. For example, a given monthly pension benefit will cost more for a woman than a man because of greater female life expectancy. (On the other hand, a given amount of life insurance will cost more for a man.) Since women tend to earn less than men, and since pension benefits typically are partly based on wage level, the sex impact on a given pension *plan* is more complex. The pension cost for a woman will be lower due to the lower wage but higher due to the life expectancy. The net effect for any particular woman—or man—will depend on the relative impact of these two influences.

Similarly, evaluating benefits by sex is complicated by family issues. For example, women are less likely than men to have health insurance at their jobs. But married women may be receiving health coverage as dependents of working husbands. Similar complications arise for mandated benefits such as Social Security. Since Social Security is a defined-benefit pension, the life-expectancy effect raises the value of Social Security for women but the wage effect diminishes it. However, women may qualify for dependent benefits based on their husbands' earnings. Therefore, the *incremental* value of added Social Security benefits due to women's work may be substantially reduced. (Men may qualify as dependents of wives but it is more likely that a man will receive a larger benefit based on his own earnings.)

Maternity benefits are of special benefit to women. Apart from the direct payment, there is evidence that women who quit their jobs for childbirth suffer large future wage losses. But those on maternity leave remain on the career ladder. Nevertheless, there is some evidence that the direct cost of the leave plan is passed on to women in the form of lower wages.

The economic model of benefits based on employee preferences generally assumes a clearing labor market. When labor markets are in surplus (high unemployment), such models do not readily apply. Presumably, employers in such surplus markets need not pay much attention to worker preferences. And there are a variety of institutional and historical influences on the provision of benefits discussed earlier. Thus, Currie concludes that many of the differences in benefit value and coverage between the sexes cannot be explained purely by employee preferences.

Cafeteria benefit plans are one way to accommodate such preferences. But there are many administrative and legal problems entailed in operating such programs. Hence, employers having trouble recruiting women might do well to examine their benefit offerings and make appropriate adjustments.

The Changing External Environment

The final section of this *Handbook* deals with the external environmental issues affecting the human resource management function. These include government regulation of the labor market, changing product markets, and

macroeconomic conditions. *Barbara A. Lee* and *Donna Sockell* note that there are many unique elements of the American regulatory system. Federal and state jurisdictions overlap in some cases. Where the federal-state division is not clear, courts are left to decide the regulatory boundaries.

Courts themselves can be viewed as having an important independent role in the regulation of the human resource function. For example, the erosion of the at-will employment doctrine is largely a matter of state court decisions. But the degree of erosion varies by state. State courts will give varying weights (or no weight) to such factors as statements made by employers in employee handbooks. Yet only one state has explicitly enacted a just cause standard into law for employee discharges.

Some programs—such as unemployment insurance—are designed as federal-state partnerships. In some cases, stricter state laws are explicitly recognized in federal statutes and allowed to override federal requirements, for example, the Fair Labor Standards Act (minimum wage, overtime pay, limits on child labor). Certain laws invite states to enact or operate their own programs, so long as the standards are at least as strict as the federal requirements, for example, the Occupational Safety and Health Act (mitigation of on-the-job health risks) and Title 7 of the Civil Rights Act (equal employment opportunity). Social Security is a purely federal program; but (for most workers) workers' compensation is a state-level program.

In some cases, laws may involve contingent mandates. American employers are not required to offer private pension plans. But if they do, the plans must meet the standards laid out in the Employee Retirement Income Security Act. And some laws provide incentives for certain employer actions, for example, tax incentives for health insurance and pensions.

Employment litigation has been growing. Almost every aspect of the human resource function is subject to legal regulation. Selection and hiring are particularly subject to EEO regulation and compensation is affected by minimum wage laws and tax incentives for certain benefits. Non-wage conditions at work are affected by evolving concepts of sexual harassment. Where unions are involved or for any collective action by employees, a considerable body of law and regulation applies, beginning with the Wagner Act. Layoffs can raise age discrimination issues. Although human resource professionals do not like to think of themselves as regulatory enforcers, they certainly play that role. Being the enforcer does not preclude participating in strategic planning, but neglecting the legal aspect can be a major strategic mistake.

As students learn early on in basic labor economics courses, demand in the labor market is partly a reflection of demand in the product market and partly a reflection of the production function. *Calvin Siebert* and *Mahmood Zaidi* note various influences that have been affecting product markets and production functions, thus reflecting themselves in labor market outcomes.

These influences include international competition, technological change, and financial developments.

The international side for all developed economies is affected by two main factors. First, there are new competitors entering the market place. China was not a significant player in world commerce before the 1980s; its decision since then to move to a market system indirectly adds a large potential labor force to world competition. Similarly, the Asian "tigers" were not major players in world trade before the 1970s. Mexico now hopes to emulate the Asian success stories—all of which involved access to the American marketplace—and has the NAFTA agreement as a tool for doing so.

A common feature of the new competitors is that their wages are substantially below those of the developed world level, raising the possibility that a kind of labor arbitrage is at work. However, the Asian tigers—just as Japan before them—are exhibiting rapid wage increases. (This is not the case for China.) So there may be a convergence process going on.

Second, the move to flexible exchange rates has increased uncertainty about international competitiveness. Erratic moves in exchange rates can render employers uncompetitive or supercompetitive regardless of their internal efficiencies. For industries involved in international trade, this uncertainty makes it more difficult to provide stable or secure employment.

For the United States a third factor might be added, namely the chronic trade deficit that developed in the 1980s. Such a deficit tends to crowd workers out of the trade sector—mainly manufacturing—and into other types of employment. This crowding may also mean loss of union jobs since manufacturing is more highly unionized than many other sectors.

Technology—a change in the production function rather than the product demand curve—is also important. Labor demand is a function both of product-market demand and supply conditions. Many observers believe that the optimum scale of production is falling with a move from mass production to flexible specialization or niche production. Either workers will need to be able to move quickly between jobs within the firm or firms will seek to create more contingent forms of employment. Such arrangements do not accord well with the 1930s model of factory unionization which was based on the Taylorist system. But still older forms of unionization, for example, the crafts in the construction industry, are based on workers moving from job to job.

Financial markets are changing as well. New theories in finance lead to an image of the firm as a collection of assets in a portfolio. Just as portfolios are constantly being reconfigured, so, too, are firms now subject to restructurings, spin-offs, divestitures, acquisitions, and mergers. Such changes in corporate control run up against the older model of a stable employer providing a long-term employment relationship. It is hard for workers to have an implicit understanding with an employer if the identity of that employer is continuously in flux.

In European countries high unemployment and low rates of job creation led to calls for more flexible labor markets, beginning in the 1980s. The term "flexibility" typically meant more ability to hire and fire as freely as American firms do. Some moves toward labor market deregulation were made as a result. However, the theory behind this call for flexibility was often left vague.

As *Christopher L. Erickson, Larry J. Kimbell, and Daniel J.B. Mitchell* point out, if employers are mandated to provide job "insurance" to workers, that insurance has a cost. And if that cost is simply an add-on to wages, fewer workers will indeed be hired. But if that cost is partly absorbed by wages—a macro-level effect involving the aggregate determination of wages—the disemployment effect will be reduced.

This example illustrates that the macro view of labor market processes often is quite different than the micro view as seen by individual employers. If in the case of flexibility just one employer were relieved of the job insurance burden, that employer would be given a competitive advantage and could expand hiring. But if all employers are so relieved, no single firm has a relative advantage and the impact is likely to be much more muted.

Generally, American employers are likely to focus on the micro side—even when faced with collective issues such as proposed government regulation or deregulation—because of the decentralized pattern of decision making in the United States. For example, American employers prefer to operate nonunion. But the lack of employee voice that results may lead to more litigation, court intervention, and government regulation. Faced with those alternatives, employers might prefer collectively to settle matters locally through unions, even if none would choose to do so individually. But such alternatives are not normally posed in the United States.

Unions are also prone to failure to take the macro view. The decision to strike is taken based on local conditions. But strikes may create bad public relations for all unions. A bitter strike, especially one which is eventually lost, may scare away prospective employee-members from joining other unions. Yet there is no way under current institutions for this negative externality to be reflected in the initial decision.

Even government agencies may fail to consider macro effects. For example, training programs are typically evaluated on the basis of their success in finding trainees employment. But such a program might be very successful on that basis without lowering aggregate unemployment. If the number of jobs does not expand, a training program may simply shuffle the queue of those left unemployed at the end of the day. The trainees will have been moved up in the queue but someone else was thereby bumped back.

What impact do human resource practices have on economic performance when viewed in a macro perspective? Erickson, Kimbell, and Mitchell provide the results of various experiments using a large-scale macroeconomic forecasting model. They simulate (1) the impact of making wages more

flexible (in the sense of more sensitivity to the business cycle), (2) making the labor market more efficient through more training and better job matching (thereby lowering the non-inflationary unemployment rate), (3) making employment more contingent (so the ups and downs of the business cycle are quickly reflected in employment variations), and (4) improving the rate of productivity growth (through more training, participation, or use of incentives). All of these changes are shown to have complex impacts on the rate of economic growth, inflation, and interest rates, variables not often considered by human resource managers.

CONCLUSIONS

In the opening section of this chapter, we cited a famous economist, John Maynard Keynes, writing in the mid-1930s in support of the proposition that theories and ideas can be powerful influences on eventual practice. Yet a decade later, another famous economist—Herbert A. Simons—expressed his disappointment about that state of management knowledge and practice:

It is a fatal defect of the current principles of administration that, like proverbs, they occur in pairs. For almost every principle one can find an equally plausible and acceptable contradictory principle. Although the two principles of the pair will lead to exactly opposite organizational recommendations, there is nothing in the theory to indicate which is the proper one to apply (Simon, 1945, p. 20).

In short, at the time Simons wrote management was ruled more by adages than theory and analysis. Today, as our review of the contents of this *Handbook* reveals, there is a shortage neither of theory nor of empirical research. We do not claim, however, that this wave of theory and research has solved Simons' problem of conflicting advice. But it is an advance that rigorous thinking and a search for empirical verification are now more likely to influence actual practice than a half century ago.

Still, as we noted at the outset, the field of management—and human resource management in particular—is fad prone. The next time someone proposes The Key to sound management practice, the lesson to be learned from this review is that the issue in question is probably more complicated than the erstwhile guru suggests. And even if The Key fits, it may simply allow one to advance to the next locked door. Thus, theory and research are continuous processes which, like management processes, can be thought of as holding potential for "continuous improvement."

REFERENCES

- Hirschman, A.O. (1970). *Exit, voice, and loyalty: Responses to decline in firms*. Cambridge, MA: Harvard University Press.
- Keynes, J.M. (1936). *The general theory of employment, interest, and money*. New York: Harcourt, Brace & World.
- Simon, H.A. (1945). *Administrative behavior*, New York: Macmillan.
- Weitzman, M.L. (1996, May). Hybridizing growth theory. *American Economic Review*, 86, 207-212.

