
III. Paying for Performance

An important consideration in Chapter 7 was intrinsic motivation. We also saw in Chapter 5 that if employees are given decision rights, aligning their motivation with organizational objectives is essential. In the next few chapters, we build on these ideas by considering pay for performance or “extrinsic” motivation. Chapter 9 covers what is usually the most difficult issue confronting any incentive system: How is performance measured? In Chapter 10 we ask, what should we do with the performance evaluation (how should it be tied to rewards)? Chapters 11-12 cover special topics: incentives from promotions, employee stock options, and executive pay. Before we begin, it is useful to provide an introduction to the overall topic. First, let’s think about why it is important to study pay for performance.

First, the evidence is clear: employees tend to respond very strongly to incentives. This means that if an incentive plan is designed well, it can be an important source of value creation; but if it is designed poorly, it can be an important source of value destruction.

Second, incentives may play an important role even if an employee has strong intrinsic motivation, because the motivation may not be adequately aligned with organizational objectives. For example, two groups that tend to have very high intrinsic motivation are researchers working in corporate R&D laboratories, and medical doctors. In both cases, the employer may have to resort to incentives to redirect their motivation appropriately. A firm may need to motivate R&D researchers to focus on innovations that can be sold profitably, rather than ones that involve cutting edge research. A health care provider may need to motivate doctors to focus more on the difficult tradeoff between quality of care and cost.

Third, it is easy to underestimate the importance of incentives. Psychologists say that people tend to make a “fundamental attribution error” in evaluating human behavior. That is, they tend to over-attribute behavior to a person’s psychology makeup, and underestimate the extent to which the behavior is caused by the environment—constraints, rewards, and group influences. In other words, people tend to underestimate the extent to which incentives (broadly construed) cause behavior. In many cases, subtle incentives drive what may at first seem to be puzzling actions by an employee.

From the perspective of motivating employees, this is even more important. Roughly speaking, there are two factors that drive worker behavior: psychology and incentives. The psychology of your workforce is, generally, quite difficult to change. Some change can be affected through recruiting and job design, but largely speaking the psychology of your employees is hard to alter, at least in ways that are easy to direct and control. By contrast, incentives are relatively easy to alter. Thus, pay for performance and other forms of extrinsic rewards are the most important motivational levers that a manager can pull.

Fourth, rewarding performance improves many human resource objectives for the firm. We have seen this several times in earlier chapters. For example, deferred pay based on performance can improve self-selection in recruiting. Similarly, performance pay may increase the return on investments in human capital, motivating greater investments in skills. Better incentives can improve decision-making, encouraging employees to use their knowledge in the firm's interests. Most human resource policies involve some form of incentive scheme (though many are subtle). More broadly, incentives are what drive modern economies. Understanding basic incentive theory provides you with intuition that is useful in many business contexts.¹

We now sketch out the way that economists formalize their thinking about incentive problems. When you read the discussion of pay for performance, think about "pay" metaphorically. We are not referring solely to monetary compensation. Rather, we are thinking abstractly about *any* rewards the firm can vary with employee performance. Of course, monetary rewards such as bonuses or stock options are the most important in practice, but firms use many rewards (some implicit). For example, good performance might be rewarded with a better office, more flexible work hours, more interesting job assignments, or a promotion. To the extent that these are rewarded based on performance, they are a form of extrinsic incentive, and the principals described in these chapters apply.

The Principal-Agent Problem

In economics, the basic framework for analyzing most incentive problems is called the *Principal-Agent* problem. This literature can often be highly technical. Here we provide a brief sketch to help sharpen your intuition, but our interests are more practical. We will express the ideas in equations, but no advanced mathematical techniques are needed. The equations help force us to make our ideas and intuition more rigorous.

Incentive problems arise when an *agent* (in our context, an employee) acts on behalf of a *principal* (the firm's owners), but has objectives that are different than the principal's. Consider an entrepreneur. In this case, the agent *is* the principal, so there is no conflict of interest and no incentive problem. However, modern corporations usually involve separation of ownership and control – managers are hired to run the firm on behalf of the owners. (This may be so that owners can diversify their investment portfolios to reduce risk, or because owners delegate to managers who have skills that the owners do not.) This suggests that incentive problems for top management are an important concern; we cover that topic in Chapter 12.

How might we analyze such a conflict formally? We need to model the objectives of both the principal and the agent. Assume that the principal's objective is to maximize the discounted present value of the firm. For a publicly traded firm, this is total shareholder value (share price times number of shares outstanding). Virtually all of the intuition would apply if we considered different definitions of the principal's objective (say, if the firm was a government agency or not-for-profit). The important point is that there is a conflict of interest.

The employee provides different kinds of "effort," which affect firm value. By "effort," we mean actions the employee can take that the firm wishes to motivate. It might mean working harder, faster, or longer on various tasks; thinking more carefully when making decisions; cooperating with colleagues; being helpful to customers, and so on.

¹ Much of modern business school curricula, such as corporate finance and managerial accounting, is applied incentive theory.

In the previous section of the book we discussed multitasking as a feature of job design. Multitasking would imply that there is more than one type of effort. We will discuss some implications of multitasking for incentive systems later in this section of the book. For now, though, suppose that there is only one dimension to the employee's job, so that the firm wants to motivate the employee to provide a single type of effort e .

The employee's overall contribution to firm value Q depends on the employee's effort, $Q = Q(e)$. Q is not the firm's total profit, but is the discounted present value of the profit created by this employee, ignoring the employee's pay. Thus, the firm's profit attributable to this employee equals $Q(e) - \text{Pay}$.

There is a conflict of interest only to the extent that the employee has too little, or too much, intrinsic motivation to provide various kinds of effort. The most typical case is where intrinsic motivation is too low – the firm would like the employee to work harder, or more diligently. A way to formalize this is to assume that such actions are costly to the employee in some way: they would prefer to work more slowly, less carefully, etc. In this sense, you can think of the employee as incurring a psychological cost if they provide more effort. This cost is called the *disutility of effort*. Let us refer to this as $C(e)$. Remember that it is a psychic cost to the employee, not a monetary one.² However, even non-monetary concepts can be scaled in monetary terms. For example, we might be able to quantify an employee's cost of effort by the increase in pay that they demand when asked to work harder.

The employee's pay only provides incentives if it depends on performance. Suppose that the firm has a performance measure PM that estimates the employee's contribution Q . If performance can be measured perfectly, $PM = Q$, but generally this is not the case. Then, pay is some function of PM : $\text{Pay} = \text{Pay}(PM)$. If the measure is imperfect, we might model this as $PM = Q + \varepsilon$, where ε is a random variable.

Pay for performance is risky because performance can almost never be measured with perfect accuracy. People are generally risk averse, so there is an additional cost to the employee from working at the firm: the cost of the riskiness of pay. Let us call this cost $R \cdot \sigma_{\text{Pay}}$, where we assume that the appropriate measure of riskiness is the standard deviation. R is a parameter that reflects the extent to which an employee is risk averse. Someone who is less risk averse would have a smaller R , while someone who is more risk averse would have a larger R . Putting this all together, the employee's net value from working for the firm is: $\text{Pay}(PM) - C(e) - R \cdot \sigma$.

As mentioned above, the firm's value from the employee equals $Q(e) - \text{Pay}$. The firm chooses the compensation plan $\text{Pay}(PM)$ to maximize this net profit from the employee. The firm is constrained by the fact that total pay must be at or above the employee's labor market value. For this reason, the firm must also compensate the employee for effort cost C , and risk cost R , generated by the incentive system the firm implements.

What Does the Employee's Incentive Depend On?

What drives the employee's incentive? As in all of economics, decisions are made by balancing the marginal benefits of changing one's behavior against the marginal costs. In this case, the question is whether the employee should work a little harder or not. The mar-

² Typically it is assumed that $C(e)$ rises at an increasing rate as e rises. This captures the idea that effort is costly to the employee, and that the harder they work, the more costly additional effort becomes.

ginal cost is the extra disutility from working harder, $\Delta C/\Delta e$, where Δ is notation denoting an incremental change in a variable.³

$$\text{Employee's marginal cost from working harder} = \frac{\Delta C}{\Delta e}.$$

Since pay depends on the performance measure, and the performance measure depends on effort, the marginal benefit from working harder on some dimension of the job is:

$$\text{Employee's marginal benefit from working harder} = \frac{\Delta \text{Pay}}{\Delta e} = \frac{\Delta \text{Pay}}{\Delta PM} \cdot \frac{\Delta PM}{\Delta e}.$$

Since the employee balances the marginal costs and benefits, anything that increases the marginal benefit will increase the employee's effort. The second equation tells us that we need to focus on two things. First, how does the performance measure vary with effort? If it reflects the employee's effort well, it will improve incentives, and vice versa. Second, how does pay vary with estimated performance? If it does so strongly, incentives will be stronger, and vice versa. These two factors are the focus of each of the next two chapters in turn; they form the heart of this section of the text. Chapter 9 analyzes how the firm can measure the employee's contribution to firm value. Chapter 10 then analyzes how the firm can relate this evaluation to rewards.

Now let us consider the source of the conflict of interest between the worker and the firm. Recall that C and $R \cdot \sigma$ are costs that must be borne by the firm, implicitly. They reduce the value of the job to the employee, so the employee will require higher total compensation if C or $R \cdot \sigma$ are higher, and vice versa. In this sense, the costs of the incentive system, C and R , do not create a conflict of interest between the firm and the employee. They are a cost of doing business, just like the cost of any inputs to production. If a stronger incentive motivates the employee to work harder, $C(e)$ rises, but the firm will have to compensate the employee for this. In other words,

$$\text{Firm's marginal cost from the employee working harder} = \frac{\Delta C}{\Delta e}.$$

Thus, the firm and its employee have the same total costs from the employee working on the job, $C(e) + R \cdot \sigma$, and the same marginal costs. The real source of the conflict is because the employee's benefit (Pay) is generally not identical to the firm's benefit (Q). More formally,

$$\text{Firm's marginal benefit from the employee working harder} = \frac{\Delta Q}{\Delta e},$$

which will generally differ from the employee's marginal benefit from working harder. This can happen because the evaluation does not perfectly reflect performance, or because pay does not fully reflect the employee's contribution. These are the incentive problems that we will see over and over again below.

While we will not pursue this formalized approach to thinking about incentive pay, we will use these basic ideas to structure our thinking and make it more rigorous in what follows. It is worth your time working through the intuition described above, and used below.

³ Assuming that extra effort does not affect the riskiness of pay.

Performance Evaluation

“When you cannot measure, your knowledge is meager and unsatisfactory.” (Lord Kelvin, as carved in stone on the Social Sciences Building, University of Chicago)

“Measure anyway!” (Frank Knight, noticing the carving on returning from lunch at the faculty club one day – as told by George Stigler.)

Introduction

The most difficult part of any incentive scheme is performance evaluation. Imagine that you are a manager, and you want to measure – quantify – the individual contributions of each of your employees to firm value. How can you do so accurately? Employees may work together in groups, so that it is hard to disentangle who is responsible and who is not. Some employees might free ride off of the work of others; other employees might be very cooperative, but you do not always see this since you cannot observe everything they do on the job. In addition, an employee’s performance may be partly due to luck. An employee might be in the right place at the right time, and land a large sales contract when a new customer calls. Or, an employee might lose sales because a key client goes out of business unexpectedly. Finally, some contributions are quite difficult to quantify. How do you measure the employee’s effects on group norms, mentoring of junior staff, or customer satisfaction?

Not only is performance evaluation quite difficult to do effectively, but also the stakes are high. Evaluation can be quite costly. Subjective evaluations generally take a large fraction of managerial time. Collecting accurate performance metrics (including those generated as part of the accounting system) can take substantial resources.

If the evaluation does not accurately reflect the employee’s contributions, it may cause several negative outcomes. The employee might find that the link between performance and pay is uncertain, and require compensation for risk, thus raising costs. The employee might be poorly motivated. Perhaps worse, the employee might be strongly motivated, but to do the wrong things, thus destroying value. Thus, even though evaluations are difficult and costly to do, they are a necessary component of a good reward system, and it is important for firms to strive for effective performance evaluation methods and procedures. In this chapter, we discuss important issues that arise when evaluating performance.

Purposes of Performance Evaluation

Employee performance depends on, among other factors: innate abilities, accumulated skills or human capital, and efforts. Based on our sketch of the principle-agent problem above, for a simple model we might say that Q depends on ability A , accumulated human capital H , and efforts e_i : $Q = Q(A, H, e_1 \dots e_k)$.

This suggests that evaluations might emphasize measuring innate abilities A , skills H , or effort e_i depending on what the evaluation is used for. We will return to this issue below. For most of this chapter, we focus on the use of evaluations to measure and motivate greater effort.

Ways to Evaluate Performance

Modeling the employee's contribution to firm value as $Q = Q(e_1 \dots e_k)$ suggests several approaches to measuring performance. First, we might estimate Q overall (what we will call a very "broad" measure). One example that is important for executives in publicly traded firms is the stock price. Second, we might estimate different dimensions of performance (what we will call a "narrower" measure). In a manufacturing setting, one common performance measure is the quantity a worker produces. Another is the quality (e.g., number of defective parts produced). Third, we might combine several measures of various dimensions of performance. For example, a plant manager might be measured on revenue, or costs, or profits (revenue minus costs).

Notice that all of these approaches are *output* based performance measures, in that they attempted to measure components of Q . An alternative, narrower approach is to measure the employee's *inputs* e_i , such as hours worked or number of routines or tasks accomplished.

Finally, the evaluation might be quantitative or qualitative. We next discuss quantitative performance measurement. In the section that follows we discuss subjective evaluation.

Quantitative Performance Measurement

Organizations often go to great lengths to try to quantify an employee's contribution to firm value. Quantitative measures have several advantages. Because they are numeric, they can be tied to compensation more easily (e.g., by computing a bonus using a formula). In addition, many performance metrics are readily available through the normal course of business. Accounting systems, for example, are large-scale performance measurement systems. Where the accounting numbers accord well with an employee's contribution, they are often used for computing bonuses, as input into promotion decisions, etc. Firms may also use hours worked, customer satisfaction, and other quantitative information as inputs in evaluations.

Finally, quantitative performance measures are often perceived as more objective than the use of judgment in evaluations. Indeed, they are often called "objective" performance measures. However, it is not obvious that quantitative metrics are objective. As discussed below, many metrics can be manipulated by the employee, or the supervisor or firm. Even if a measure is not subject to manipulation, it may not measure exactly what is intended. For example, if a law firm wants to motivate partners to bring in new business, it may give lawyers a credit for any new clients that they bring to the firm. This would seem to be an easy performance dimension to quantify. However, in some cases the lawyer might have received the new business simply because he or she answered the phone when the new client called. For these reasons, we do not use the term "objective" performance measure, to highlight that quantitative measures have their own flaws. Nevertheless, they are certainly likely to be more objective than are subjective evaluations.

Performance Measure Scope

To evaluate the employee for incentives, an ideal performance measure should reflect the employee's total impact on firm value, and nothing else. Accounting texts often say that a performance measure should include whatever is "controllable" by the employee, but exclude whatever is "uncontrollable." As we will see, the definition of controllable and uncontrollable is subtler than it seems. For now, let's think about these issues in the context of the right performance measure for two employees in the same firm: the CEO, and the custodian.

The most common performance measure for CEO compensation (in a publicly traded firm) is the firm's stock price (or stock value, which is the stock price times the number of shares outstanding).¹ By definition, this is firm value. Thus, this measure does capture all of the things that are controllable by the CEO: if there is anything that the CEO can do to improve or reduce firm value, it will be reflected in this performance measure. In this sense it would seem to be a perfect measure.

However, there are also many things that affect the firm's stock price that are beyond the control of the CEO: actions of competitors, macroeconomic factors, currency fluctuations, and so on. For this reason, the performance measure is also *risky* for the CEO. Performance measurement error is caused by uncontrollables.

Now consider the custodian; the same logic applies. Stock price is a good measure in the sense that it includes the effects on firm value of anything that is controllable by the custodian. However, it includes much that is uncontrollable. In fact, it would be ridiculous to use stock price as a performance measure for the custodian because the uncontrollables so outweigh the controllables. Doing so would essentially make compensation into a lottery ticket. Because the custodian is risk averse, this would be an expensive way to compensate, as the firm would have to pay a substantial risk premium to the custodian.

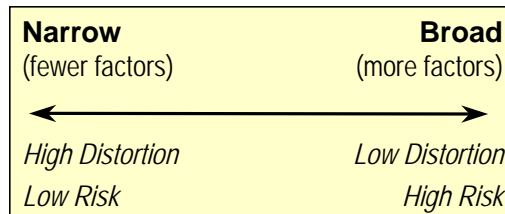
If stock price is a ridiculous measure for the custodian, what metric might we choose instead? We might measure cleanliness of floors, or pounds of trash hauled away per shift. We think of such measures because they are closer to the custodian's job – they focus more on things that the employee can control, and filter out more things that are uncontrollable. In doing so, we reduce the risk of the performance measure. Similarly, for the CEO we might choose accounting earnings as the measure. It is one of the better proxies for profit that an accounting system generates,² so it is a good starting place for trying to quantify the CEO's contribution. Moreover, it is much less risky, since it is relatively more affected by things the CEO can control than by uncontrollables.

Unfortunately, more focused measures cause a new problem – they *distort* incentives. In the case of the custodian, measuring cleanliness of floors does not motivate the custodian to be cost conscious. Measuring pounds of trash hauled away may motivate the custodian to throw away too many things, or only heavy items. Evaluating the CEO on accounting earnings is likely to cause the CEO to take too much of a short-term focus, as earnings is based on performance in only a single period.

¹ Notice that the stock market is, in effect, a large-scale performance measurement system, for top management. This is one of the most important roles of equity markets (though not the only role), and is the basis for much of the analysis in modern corporate finance.

² Of course, it does not measure true economic profit, since accounting numbers are imperfect proxies for economic concepts. There are much more elaborate methods that try to adjust accounting numbers to better reflect economic reality, such as EVA (Economic Value Added).

These examples highlight a common tradeoff that firms usually face in choosing most quantitative performance measures: the *scope* of the measure (Figure 9.1). A *broader* measure is one that includes more aspects of performance. In a publicly traded firm, stock price is the broadest possible measure, since it is firm value. The advantage of broad measures is that they tend to distort incentives less. They distort less because broader measures include more dimensions of the employee's job in the evaluation (more controllables). However, at the same time they also tend to include more uncontrollables, which causes measurement error and makes the incentive scheme riskier.



Tradeoff Between Broad or Narrow Performance Measurement

Figure 9.1

A natural way to reduce risk is to use a *narrower* performance measure, such as accounting earnings instead of stock price. Narrower measures may be chosen because they are easier to measure. Another important reason that narrower measures are often chosen is because they filter out many of the uncontrollables, reducing employee risk. But it is virtually impossible to filter out all that is uncontrollable without simultaneously filtering out some things that are controllable by the employee. Thus, narrower measures tend to be less risky, but distort incentives more.

We thus see two important things to consider in any numeric measure. First, how risky is the measure (how important is measurement error)? Second, how does the measure distort incentives? Virtually any performance measure will have both problems. Sometimes the distortions can be subtle, so it is worth thinking carefully about them before putting too much weight on a measure for incentives.

Since there are several dimensions to most jobs, and to what affects overall firm value, this tradeoff between risk and distortion can play out in multiple ways. Table 9.1 provides examples of performance dimensions and the kinds of distortions that tend to result from using a metric that is narrower on each dimension.

Dimensions of Performance to Consider in an Evaluation	Example
Which tasks to include or exclude?	quantity v. quality
Use available metrics or incorporate qualitative information?	accounting numbers tend to ignore intangibles or opportunity costs
How large a unit should be measured?	individual v. team v. unit v. division v. firmwide performance measures
What time horizon should be used?	last year's sales v. customer retention / growth

Dimensions Along Which Performance Measures May be Broader or Narrower

Table 9.1

Common Distortions in Performance Measures

Intangibles

By definition, intangibles are difficult to quantify. Quality is a classic example, since any incentive based on a quantity measure (like piece rates in manufacturing jobs) distorts incentives away from quality. But there are many dimensions of jobs that are difficult to quantify. In service jobs, customer satisfaction usually can only be gauged imperfectly through methods like customer surveys. Similarly, professional service firms can easily calculate revenues and profits from specific client engagements, but cannot always tell how well satisfied the client is.

Opportunity Costs

An important problem with standard accounting numbers is that they do not reflect opportunity costs – the costs of foregoing other alternatives. For example, if a company owns a factory, and it has been fully depreciated for accounting purposes, it may show up on accounting statements as having no value. Or, it may be listed at book value, which is a measure of the costs of constructing the building at the time. The true value of the building is how much the company could sell the building for to someone else. If the company decides to use the building, they are giving up this value. Thus, decisions about asset use can be severely distorted unless some adjustment to accounting numbers is made.

A similar issue arises when a company requires departments to obtain services only from an internal department. Since the department is granted a monopoly, it can be difficult to estimate the true performance of the department. If the firm allows purchasing from outside vendors, it not only provides some competition (which should motivate better performance by the internal supplier) but also provides an important performance metric – the market price of the service.

Group Size

Firms always face a choice over the group size to use for the evaluation. Since employees are interdependent in production a narrow measure like individual performance will tend to distort incentives. There will be less incentive to cooperate with colleagues. Unfortunately, using a broad measure like group or business unit performance makes the measure much

less controllable, and much more risky. For example, basing an individual's incentives on group performance means that the employee is accountable for actions taken by colleagues, which are not fully controllable. (However, they are partly controllable, as discussed below.) The broader the group used for performance measurement, all the way up to the entire firm, the more will the employee take into account how his or her work affects others in the firm, but the riskier will the measure be. We discuss this further under the topic of employee profit sharing plans below.

Time Horizon

Most performance measures are backward looking: they measure what just happened. This tends to mean that they distort incentives for actions that have long-term consequences. Generally these are various forms of investments; e.g., in new technology, brand name, or employee training. One approach that is sometimes used for these problems is to defer giving the reward for some period of time. This allows the firm to wait and see what long term performance is. A clear problem with this approach is that this is risky in a different way for the employee, who may quit the firm before the reward can be given.

Performance Measure Scope & Job Design

Ideally, the scope of an employee's performance measure should be closely related to the employee's job design. Consider possible performance measures for a divisional manager in Table 9.2.³ The first column shows the performance measure, the second shows typical constraints placed on the manager's decision making, and the third shows typical decision rights given to the manager. The first two measures, cost and revenue, are relatively narrow. The next, profit, is broader – it literally combines the first two measures. For this reason, it reflects all of the controllables reflected in revenue and costs separately. It also reflects all of the uncontrollables from both.

³ Based on Jensen & Meckling (1998).

Performance Measure	Typical Constraints	Typical Decision Rights
Cost Center	output or costs, quality, delivery, product mix & specifications, capital expenditures, "customers"	input mix, sourcing, scheduling, production techniques, personnel
Revenue Center	price, selling costs, product mix & specifications, non-sales support, delivery, capital expenditures	sales effort, selling techniques, personnel
Profit Center	capital expenditures, basic business activity, capacity, corporate policies	all of the above + quality, quantity, product mix, product specifications, capacity, price, sales resources
Investment Center	scale of investment, basic business activity, corporate policies	all of the above + major asset purchases & sales
Franchise	franchise contract, particularly actions affecting brand name	all of the above at local level + ability to sell business subject to ratification by franchiser
Ownership	none	all

↑ Narrower Evaluation
↓ Broader

↑ Less ... More Decentralized
↓

Matching of Job Design to Performance Measurement: Divisional Managers
Table 9.2

An investment center uses a broader concept of accounting profit (such as EVA, Economic Value Added) that attempts to include measures of the opportunity cost of assets (which may be ignored in standard accounting numbers). In addition, while profit measures short-term contributions to firm value, an investment center uses a performance measure that attempts to calculate some version of discounted present value of profits.

The next row, franchise, uses an even broader performance measure, since the franchisee's primary objective is to maximize the resale value of the franchise. The final row, ownership, uses the broadest possible performance measure, firm value.

The important point to note in the table is how the constraints and decision rights change, when moving from narrower to broader measures. Moving down in the rows after the first two narrow measures, decision rights tend to include all of those in the prior row, plus some new decision rights. Similarly, as the performance measure becomes broader, there are fewer constraints. Simply put, a narrower job design (in the sense of both tasks and decision rights) is associated with a narrower performance measure and vice versa.

This makes perfect sense. If an employee is given more tasks or authority, a narrow performance measure will cause greater distortions. Thus, the balance between distortions and risk will be struck by using a broader measure that incorporates some of the additional dimensions of the job, even though this may mean that the measure is riskier.

Indeed, the matching of performance evaluation and job design will tend to happen automatically. Suppose that an employee is evaluated on a very broad performance measure, but given little discretion. This tends to mean that there are many uncontrollables in the performance measure. In order to reduce risk, the employee will request, or simply start taking on, additional responsibilities, in order to avoid being punished for things outside of the employee's control.

Finally, an employee's job tends to evolve over time (typically toward a broader job with more discretion, as the employee's skills increase). For this reason, the evaluation should usually broaden with job tenure. Often this occurs by gradually holding the employee accountable for more and more factors through subjective evaluations.

Manipulation

A final problem with quantitative performance measures is that they may be manipulated or "gamed." Consider again our example of the custodian evaluated on pounds of trash hauled away. This might motivate the custodian to bring trash to work, because this would improve the performance measure – but it would not benefit firm value.⁴

Of course, either side might manipulate the measure. Imagine a joint venture between two firms, in which one is to provide a service to the other, and compensated with a share of the profits from the venture. In such a case, the second firm may be tempted to allocate too many of its costs to the joint venture, resulting in understated profits. In fact, this situation occurred with the movie *Forest Gump*. Writer Howard Groom sued the studio when the wildly successful movie was said by the studio to be a money loser. Groom had been promised a share of the profits. If the contract had been based on revenue, this conflict would have been less likely, since costs are easier to manipulate than revenue in this case.

Manipulation is similar to the issue of distortion in incentives, but is somewhat different. The problem of distortion is that different aspects of the job are given inappropriate relative weights (possibly zero) in the incentive plan, causing the employee to emphasize some things too much, and others too little. Manipulation occurs because the employee or employer has specific knowledge of time and place. This knowledge may be used strategically, after the performance metric is chosen, to improve the evaluation even when such actions do not improve firm value.

The concept of manipulation is related to our notion of performance measure scope described above. As with distortions, manipulation is more likely to occur with a narrower performance measure. Because a narrower measure reflects fewer parts of the employee's job, altering behavior along only one dimension of work may have a large effect on measured performance. By contrast, broader measures tend to be less susceptible to manipulation because the employee would have to change more dimensions of performance to manipulate the measure.

An implication of manipulation is that the quality of a performance metric may *degrade* over time once it is used for incentive purposes. Consider a measure that previously had not been used for calculating an employee's bonus. The firm believes that the measure is

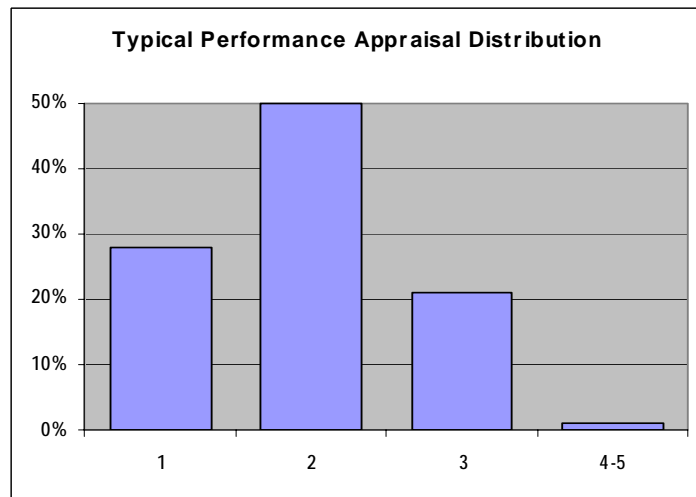
⁴ One of the authors had a similar performance metric as a child, when his mother required him to collect a grocery bag of weeds from the yard, as punishment for poor behavior. The author would take a large bundle of grass cuttings (from the recently mowed lawn), stuff them in the bag, and then pull enough weeds to cover them. Fortunately for the author, Mom never detected the manipulation, even though she has a PhD in Physics. (Or did she, and she just didn't say anything ...?)

correlated usefully with firm value, so it decides to give the employee a bonus based on the measure. The employee now has an incentive to increase the measure's value, possibly in part through manipulative behavior. If there is manipulation, this will tend to reduce the measure's correlation with firm value – making it a less useful performance measure! The longer the employee has had a bonus based on the measure, and the greater the incentive placed on the measure, the more likely is this phenomenon to be a problem. Thus, the firm may find that it ultimately has to choose a different performance measure, which itself may gradually degrade, and so on.

Subjective Evaluation

We have discussed quantitative performance measures and their limitations. We now consider the benefits and limitations of subjective performance evaluation.

Perhaps the most painful job for a manager is giving and receiving subjective performance evaluations. In many jobs, employees receive a subjective rating once or twice a year (often on a scale of 1-5 or A-E). Figure 9.2 below shows the actual distribution of such ratings given to employees in Acme Incorporated (1 is the best rating, and 5 is the worst). The distribution is quite typical of what is seen in most firms, and exhibits several traits that often raise concerns. There is grade inflation: the average rating is well above the middle score. Similarly, managers are very reluctant to give the lowest ratings: only about 1% of employees received either of the worst two ratings. There is very little feedback in the ratings: almost 30% receive the best rating, while another 50% receive the second best. If the goal of ratings is to distinguish performance, and to highlight the best and worst performers, these kinds of rating distributions do not seem very effective.



Source: Gibbs (1995)

Performance Rating Distribution at Acme, Incorporated

Figure 9.2

When asked why they are reluctant to give negative feedback or poor performance appraisals, managers often say that they are worried that this would reduce employee motivation. This seems hard to understand, since a good incentive system should provide both positive and negative feedback. We will provide one explanation for this phenomenon in Chapter 11 when we discuss promotion-based incentives.

Managers are reluctant to give poor ratings partly because it is no fun to give someone bad news. Thus, there may be leniency bias, helping explain why so few poor ratings are given. Moreover, employees may pressure managers to change the rating, which is unpleasant for the manager.

Of course, employees dislike subjective evaluations too. A chief concern is that they are more subjective than numeric evaluations (though as noted above, those are often not entirely objective either). They may worry that the evaluation reflects the supervisor's personal opinions and biases, and that the supervisor is playing favorites. Of course, this will reduce incentives, since the perceived link between effort and reward is weakened.

Despite all of these flaws, essentially every job uses subjective evaluations in important ways. Subjectivity is often necessary in hiring, promotion, and termination decisions. For middle managers doing qualitative knowledge work, good quantitative measures of performance may not exist, so that raises and bonuses are usually based on subjective appraisals. Subjectivity may play an important role even for salespeople, where performance is often easiest to quantify. For example, sales prospects or training opportunities may be allocated subjectively. Finally, one of the most important roles for a Board of Directors is subjective evaluation of the CEO's performance.

Why Use Subjective Evaluations?

Consider a manager who runs a factory. The manager's annual bonus, which averages about 40% of salary, is calculated as a percentage of the factory's annual profits (that is, the plant is a profit center). The factory is 40 years old, with a sheet metal roof. One day, a sudden tornado rushes through town, and tears the roof off.⁵ Because of the substantial damage, the factory cannot return to operation for quite some time, and there is a loss rather than a profit for the year. If you were the plant manager's boss, when you arrived at the plant the day after the tornado, how would you evaluate her performance?

One common reaction to this story is that a tornado is uncontrollable, as it is an act of nature, and thus the manager should not be punished at all. However, another common reaction is that the manager should be severely punished, if not fired, because she has ultimate responsibility for the plant. Which view is right? Without more information, it is difficult to say. Either view could be justified under some circumstances.

For example, if this was the first tornado in the area in 50 years, or if the manager was very new to the plant, it is more likely to make sense to not punish the manager. On the other hand, tornados might occur often in the region, and the manager may have worked at the plant for many years. Perhaps the roof was in need of repair, but the manager had deferred maintenance. In such cases, she is at least partly responsible for the effects of the tornado on firm value.

There are also intermediate cases. Suppose, for example, that on the organization chart responsibility for roof maintenance was given to a plant engineer, who oversaw the structures of all factories for the company. This suggests that we should not hold the plant manager responsible. However, it might be that the plant manager had some information about the condition of the roof, and neglected to tell the plant engineer. If so, the plant manager should clearly be punished.

⁵ This is a real example that happened to an Executive MBA student of one of the authors.

Finally, even if the manager had no way to foresee this event, we may still need to punish her for not reacting properly to the event. We want to motivate the manager to address immediate safety issues, obtain second-sourcing for production, and get the plant up and running again as soon as possible.

This discussion illustrates a couple of points:

What Do “Controllable” and “Uncontrollable” Mean?

In this example, there are many cases in which we might punish the manager for an act of nature, because the manager could have taken actions before, during, or after the event to reduce the damage to the firm. Thus, whether an event is random or not is not a useful definition of what is controllable or uncontrollable. A better way to think about it is:

An event is at least partially controllable if the employee can have some effect on the impact of the event on firm value. In the case of adverse events, the employee can prevent or mitigate some of the damage. In the case of positive events, the employee can prepare for or exploit the opportunity.

Given this definition, very few events are completely uncontrollable by an employee. There may be many circumstances in which we want to punish – or reward – an employee for things that were not the employee’s “fault.”

When Should an Employee Be Held “Responsible”?

Even if the manager did not have formal responsibility for roof maintenance, the manager may have had some specific knowledge about the roof’s condition. As we saw when we discussed decentralization, when an employee has specific knowledge, we may want to give him some decision rights (responsibility) so that such information can be acted upon, and some incentives to act upon that knowledge. Thus, much of what we are doing with a careful subjective performance evaluation is defining what the employee is, and is not, responsible for. We should consider holding the employee at least partially responsible for events any time that they are partially controllable by the employee, because the employee has some relevant specific knowledge of the situation. As we’ve seen, the employee might have this information before, during, or after an event. If so, the employee should be given some responsibility for foreseeing events and making contingency planning; for reacting to events in real time; and for following up events after the fact. This is what we mean when we say that we want to motivate the employee to take initiative.

How Do You Conduct a Subjective Evaluation?

This discussion also gives us a useful way to think about *how* to evaluate employees. Suppose that you are evaluating a subordinate in order to give a year-end performance rating. As in the example above, it is natural to begin by thinking *retrospectively*: looking back on what the employee did and what happened over the last year.

When you do so, it is important to avoid what psychologists call “hindsight bias.” You are likely to know more after events have unfolded than the employee did when they occurred. In our example, we now know that the tornado occurred and the roof was weak. However, did the employee know these things at the time? Thus, a typical first step is to figure out, along the lines of the famous question by U.S. Senator Howard Baker about President Richard Nixon during the Watergate Hearings, “What did he know, and when did he know

it?" You do this to decide whether the employee's actions were appropriate given the circumstances.

This will tend to lead into a broader analysis of the extent to which events were *foreseeable* or not. It is important to evaluate employees on whether they planned for foreseeable events – and also, to a reasonable extent, whether they developed contingency plans and procedures for unforeseeable events.

So far, the evaluation has focused on the past, and what that means for the employee's rewards (rating, bonus, raise, promotion, firing, etc.). However, unless the employee will be fired, the most important outcome of a good subjective evaluation comes from thinking *prospectively*. The process of discussing with the employee what happened, what they did and why they did it, and what they should have done instead is a way of communicating what the employee's responsibilities are *in the future*. The evaluation in effect defines the job, and sets precedents about what will be rewarded and punished in the future. In complex work environments, it is usually difficult if not impossible to completely define an employee's responsibilities in writing. Discussions about subjective evaluations are an important time to clarify them. The benefits include better decision making, and better alignment of the employee's incentives with firm objectives.

Finally, a good evaluation will move from the "here's your performance rating" stage to a constructive discussion of the future. The manager should clarify what the employee is expected to focus on, what he is supposed to be responsible for (and not). It is an excellent chance to think about any skills the employee lacks but needs to perform the job well, thus leading to new training. Similarly, the debriefing about last year's performance may suggest additional information or resources that the employee might need in order to perform the job effectively.

How Do You Receive a Subjective Evaluation?

It is also useful to briefly think about how you can best *receive* an evaluation from your supervisor. First, bosses hate to hear the phrase, "It's not my fault." That is tantamount to saying, "The event was uncontrollable" (though using less academic jargon), and as we saw above few events are completely uncontrollable. Instead, your boss is looking for you to take initiative. Consider discussing your mistakes and setbacks. Admit them, and describe to your boss what you have learned from them and how you will change your future work as a result. Your boss is likely to be relieved that someone is taking responsibility for them self!

Second, try to use an evaluation as an opportunity to improve your job. Ask your manager for advice and suggestions about how to perform better. Request new training, information, or resources that will improve your performance. The evaluation can be an excellent opportunity if you take initiative.

The Benefits of Subjective Evaluations

The plant manager's quantitative performance measure, profit, fell to zero (or negative) in the tornado example. As a result the manager's bonus vanished. It is possible that this outcome is exactly the one that we would arrive at after a considered judgment of all relevant factors, but only by an astonishing coincidence. In other words, holding the manager rigidly to the quantitative bonus plan would almost certainly result in the wrong outcome: either the performance metric is wrong, or the weight given to it for calculating compensa-

tion is wrong, or both. The only way to fix this is to introduce discretion into the incentive system in some way.

We now return to our question above about why subjective evaluations are so important, despite the many difficulties that they cause. The discussion of the tornado example suggests that if done well, a subjective evaluation can have many benefits:

Improve on Quantitative Performance Measures

Subjective evaluations may be used to avoid the typical flaws in quantitative measures. Not only does a careful subjective evaluation result in the manager being held accountable for controllables, properly defined. It also can filter out uncontrollables, lowering the risk to the manager. In the tornado example, good judgment can remove from the plant manager's evaluation the effects of things that were truly unforeseeable and not the manager's appropriate responsibility. In this sense, performance measure error may be lower than by sticking with quantitative metrics.

Similarly, subjective evaluations can reduce distortions in incentives. Some dimensions of the job are hard to quantify. Giving those dimensions adequate emphasis during performance appraisals can motivate such tasks. A classic example is where jobs involve quality, creativity, or other intangibles. These are usually difficult to put into numbers, so in order to motivate them, such tasks generally must be evaluated and rewarded using judgment.

Finally, subjectivity may be able to reduce manipulation of the incentive system. If an employee games the quantitative metrics, the manager may detect this ex post (or at least have strong suspicions of such behavior). Subjectivity allows the manager to adjust for this. To the extent that this might be anticipated, the employee will be deterred from too aggressively manipulating the numbers.

Improve Incentives for Risk Taking

As just noted, a good subjective appraisal can reduce risk by filtering out true uncontrollables from the employee's overall evaluation. It can also improve incentives for risk taking, since it makes it easier for the manager to reward good results without simultaneously punishing mistakes. In effect, it gives the manager greater flexibility to reward the upside without punishing the downside.

Give the Incentive System More Flexibility

Incentive plans that are put in place at the beginning of the year may no longer be ideal if circumstances change. When this occurs, the firm can change the incentive plan. However, when it does so it runs the risk of this being perceived as unfair (like the "Ratchet Effect" we discuss in the next chapter). Effective use of subjectivity means that changes to incentives are more likely to be accepted, since the employee already expects judgment from the supervisor. This makes it easier for the manager to tell the employee to change emphasis during the middle of the year, in effect changing incentives.

Improve Decision Making

As we saw, it often makes sense to at least partially reward or punish employees for random events. This is important because it motivates the employee to use (and even to develop) specific knowledge of time and place. There are three ways that this motivates better employee decision making: more effective preparation, real time response, and ex post reaction after the event is over.

Expand Communication

If you go back and reread the discussion about how to give a subjective evaluation, you will realize that what it describes is simply good day-to-day management. The best subjective evaluations occur implicitly every day, as the manager works with the employee. The manager monitors what the employee does and why, makes adjustments, and suggests improvements. Instead of waiting until the end of the year to do the evaluation, having these conversations all year long will improve the employee's effectiveness, and the working relationship with the supervisor. It will also clarify the terms of the implicit contract. Moreover, clear communication makes it more likely that employees will trust subjective evaluations, making them more effective.

Improve Training

A manager can use a thoughtful subjective evaluation to provide lessons for the employee from the manager's experience. If done well and regularly, it can be an excellent form of day-to-day training.

It is clear that there are many benefits to subjective evaluations if they are done properly. These are many of the benefits to effective day-to-day management as well. Given these benefits, it is not surprising that subjectivity is so important in practice, despite the many difficulties that it raises, and costs that it implies.

Sweating the Details on Evaluations

Lincoln Electric Co. in Cleveland, Ohio in the U.S. has perhaps the most famous incentive schemes in business history. One important component is an annual profit sharing bonus paid out to all employees. While profit sharing schemes usually do not motivate well, Lincoln Electric's does, for two reasons. First, the stakes are very high – the bonus can double an employee's pay in a typical year – reducing the "free rider" problem. Second, the bonus is based on a measure of *individual* performance, rather than the more typical (and essentially uncontrollable) firm-wide measures that most profit sharing plans use.

The performance measure for this plan is a subjective performance rating. Lincoln views its incentive system as the key to its success, and the subjective performance rating as one of the crucial components. Every rating is checked by one of the firm's top executives, so managers take them very seriously. The ratings are given twice per year, and according to the company a typical manager spends about three weeks doing ratings for subordinates – that is, six weeks per year. Doing effective subjective evaluations is hard work, but can have enormous benefits. Arguably it is one of the most important tasks for a manager.

Source: Berg & Fast (1975)

Practical Considerations

Who Should Evaluate Whom?

There are many risks to the firm from decentralizing evaluations to an employee's manager, because the manager's incentives will be imperfect. These include reduced employee motivation, poor promotion decisions, and distorted decision making. In addition, a

firm can open itself up to risks of legal liability, since subjectivity makes favoritism and discrimination more likely. Despite these problems, evaluations are not centralized to control the risks. The reason is simple: evaluations are an example of subjective, experiential knowledge that is very costly to communicate to others. In order to make use of this knowledge, most evaluations – especially in complex work environments where good numeric performance metrics are unavailable – are inevitably decentralized to the direct supervisor.

Some firms make use of “360-degree” evaluations. This practice involves the subordinates in evaluating and giving feedback to the *supervisor*. What is the purpose of this practice? In theory, it can improve management, since the recipients of the manager’s treatment are giving feedback. However, there are clear problems, since the subordinates face a good possibility of retribution if they criticize their boss.

For this reason, 360-degree evaluations are almost always done in secret; the names of employees who make specific comments are not told to the supervisor. While this can help, in small work groups the supervisor can often guess who gave which particular feedback. Thus, the effectiveness of 360-degree evaluations can be quite limited. Even so, many firms do use them as one tool to improve supervision, communication, and the general work environment. They are more likely to be effective in organizations that have cultural norms and job designs that emphasize open communication and employee participation in decision making (what we called decision management).

Fairness, Bias & Influence Costs

Subjectivity makes it easier for a manager to discriminate, play favorites, and generally be biased in allocating rewards. Of course, this reduces incentives, since it means that factors other than performance affect rewards, and can impart a subtle form of risk into the evaluation. Another way to put this is that there is an additional layer of incentive problem when subjectivity is involved: the firm has to worry about the incentives for the manager to implement the system in accord with the firm’s interests instead of his own.

This suggests immediately that if the supervisor’s incentives are well designed, we should expect fewer problems at lower levels. In addition, constraints might be placed on the evaluator. For example, some firms impose various types of “forced curves” on evaluators. We will discuss these systems under the topic of relative performance evaluation in Chapter 11.

An obvious point is that the greater the trust that the employee has of the evaluator, the more effective will subjectivity be as part of the incentive system. Thus trust considerations play a key role.

The manager’s reputation can have a substantial impact on how incentives play out in practice. A manager who is capricious or biased will induce certain behavior, and certain types of employees who want to work for him. If a manager has a reputation for fairness and careful evaluation, subjectivity will be easier to use effectively, improving the reward system in the ways described above. Thus, in jobs where such judgment is important, a manager can try to invest in a strong reputation as an effective manager. Similarly, the firm should attempt to put “judicious” managers into positions where these considerations play a major role.

In addition, organizations usually put in place formal policies to try to establish a greater degree of fairness in the evaluation system. For example, employees usually have the

right to ask to be re-evaluated if they disagree with their evaluation. In some firms, the supervisor's boss checks evaluations. The purpose of such oversight is to provide some incentives for the supervisor to do appraisals fairly. In some cases, it is possible to have multiple evaluators of an employee; this will usually reduce the likelihood that the ultimate evaluation is biased, since different managers tend to have different biases.

Of course, how such policies are implemented in practice determines whether they are effective or just "window dressing," so there is an important element of corporate culture here as well.

Supervisor judgment can also distort employee incentives. Employees will have some incentive to try to improve their evaluation not through effort, but through influencing the supervisor in other ways. For example, they might spend time and resources trying to lobby their boss for a larger raise, more resources, and so on. They might flatter their boss, take up similar outside interests, and so on. To the extent that such activities are engaged in instead of productive effort, and to the extent that they change the evaluation, they impose *influence costs* on the organization. What are the costs? Reduced or distorted incentives, and promotions that are based less on ability than is desirable.

A more subtle cost of bias and influence activities is that it may distort decision making. When a manager's opinion affects a worker's reward, the worker may distort what he says to the manager. Thus, the *quality of information* on the job will suffer. In principle, a good manager will want to hear the truth from subordinates, and will attempt to establish a corporate culture in which this occurs. For example, the manager might establish cultural rules that encourage employees to speak freely, and perhaps even reward subordinates for criticizing the manager's analyses.

While this can help, it is unlikely to completely alleviate the problem. Suppose that a manager is presented with an opinion from a subordinate that differs from his own. Either the new opinion is correct, or the manager is correct (or neither). As a statistical matter, the manager should place some weight on each possibility. But workers realize this. Because there is some chance that a worker's dissenting opinion will be viewed as a mistake, even if correct, workers have some incentive to shade their reports in the direction of the manager's initial opinion. This creates a "Yes Men" phenomenon. This highlights the importance, and complexity, of proper analysis and decision-making procedures. Such procedures need to take into account (and be co-designed with, to the extent possible) the corporate culture and explicit incentives of analysts and decision makers.

Different Roles of Evaluation

As noted in the introduction to this chapter, employee performance depends on ability, accumulated human capital, and efforts. Thus an evaluation can be used to measure any of these three. In practice it can be very difficult to disentangle the effects of these three factors. Suppose that an employee is performing very well. Is this because of the employee's raw talent for this type of work? Or is it because of skills and experience? Or is it because of hard work? Different purposes of performance evaluations imply different weights on evaluating innate ability, human capital, and efforts.

One purpose of a performance evaluation is to decide who to hire, or which probationary worker to keep. For such decisions, it makes sense to try to measure the employee's innate abilities A , rather than H or e_i . If further skills and effort are needed, the employee can then be trained and motivated in the job they are assigned to. Similarly, evaluating an employee's innate skills is more important when promotions or changes in job assignments

are being considered. Of course, accumulated skills H will also be relevant for such decisions.

Poor evaluations are more likely to be taken seriously for sorting in earlier periods. This is because there is little information at first, so each evaluation is more informative. As more data is accumulated about the worker, new evaluations are less informative.⁶ By the same token, however, probationary evaluations should last longer when the value of small differences in ability is more important to the job, and when the job is more complex and difficult to evaluate. Thus, probation for secretarial jobs may be very short, but for professional jobs may be very long indeed. For example, in professional service firms and universities, the first up-or-out promotion may only come after several years of work, and partnership and tenure decisions can take six years or more.

A second purpose of evaluations is to measure the extent to which the employee improves human capital. Especially early in the career, supervisors often provide extensive training to subordinates. In such cases, they may emphasize changes in human capital (the rate of growth in H) in performance evaluations.

A third purpose of evaluations is to motivate employees to work harder. When this is the case, the evaluator will want to try to measure how much effort of various types the employee put forth on the job, not how much talent the employee has.

In some cases, different goals of evaluations can conflict. For example, feedback about current performance may give the employee clues about their long-term prospects for advancement, which may actually reduce motivation (see Chapter 11). To avoid this, many firms attempt to separate out evaluations for current rewards from those for “coaching and development,” for example by giving two different types of evaluations at six month intervals. However, in practice it is always difficult to separate the two.

How Frequently Should Evaluations Be Made?

An additional purpose of an evaluation is to measure the employee's value to the firm in order to decide the amount that the firm is willing to pay as compensation. This is useful for deciding whether to match outside offers, for example. Suppose that a worker is worth \$1000 per week to the firm, but is only paid \$800. He is worth \$900 at another firm, which offers \$875. If the current firm offers him \$900, both the firm and the employee are better off than if he quit. The general rule is that the firm should offer enough to encourage the worker to stay, if the worker's productivity is highest at the current firm.

Which workers are more likely to have productivity at the current firm that exceeds their values elsewhere? Those with more firm-specific human capital. For this reason, evaluations are less likely to produce information that will result in a separation when the worker has more firm-specific human capital. This implies that costly evaluations should occur less frequently when firm-specific human capital is more important.

The frequency of the evaluation should also decline with the worker's experience in both the firm and in the current job, for two reasons. One is the effect of sorting. The longer that a worker has been with a firm (and similarly in a specific position), the better informed are both the worker and the firm about whether the worker is a good match for the job. Thus, the less likely is an evaluation to lead to a change in position. The second effect is that the

⁶ Statistically, the firm updates its priors with each new observation about performance, but gives increasing weight to prior performance as data accumulates.

feedback and training aspect of an evaluation becomes less important, the longer the worker has been with the firm and in the current job.

Summary

Performance evaluation is the most difficult, and probably the most important, part of a well-designed incentive plan. Quantitative performance measures such as accounting numbers are an important component of evaluations. An equally important component is subjective evaluations.

An ideal performance evaluation for incentive purposes is one that captures all of the effects of an employee's actions on firm value, but nothing else. These are often called "controllables" and uncontrollables," but the concepts are rarely defined carefully. A good way to rigorously think about the terms is to say that something is at least partially controllable if the employee can affect the event's impact on firm value. Thus, even purely random events are often controllable to some extent; most events are partly controllable and partly uncontrollable.

A basic tradeoff that is often faced with quantitative performance measures is the *scope* of the measure. Broader, more inclusive measures tend to include more that is controllable, and more that is uncontrollable; narrower, more focused measures tend to have the opposite properties. The first effect means that broader measures usually distort incentives less because they ignore less. However, it also means that broader measures are riskier as they have more measurement error. Risk is costly to an incentive scheme, because employees are risk averse.

Because of the riskiness of very broad measures like stock price, most performance measures used in practice are much narrower. Performance measure choice involves a careful balancing of risk against distortion. This virtually guarantees that any measure will distort incentives in some way. Thus, an important consideration in managing an incentive system is to carefully watch for distortions (and the related problem of manipulation of measures), and use additional incentives or discretion to reduce such problems.

Performance measures and subjective evaluations should be chosen to match the employee's job design as closely as possible. This makes it more likely that the measure captures the most important controllable dimensions of the job, reducing the distortion problem. Subjective evaluations can be viewed as a way to *define* the employee's job and responsibilities.

Subjective evaluations are an alternative way to evaluate performance. Since they require judgment, they can cause a host of problems if the manager does not take them seriously, or if the manager does not have adequate motivation. For example, discretion makes it easier to play favorites, and for biases to creep into evaluations. Discretion also requires that managers make tough decisions about good and poor performers, and give constructive feedback, even in the face of complaints and lobbying from subordinates. If done effectively, however, subjectivity is an effective way to improve virtually all parts of an incentive system.

Review Questions

1. In macroeconomics, there is a debate over “rules versus discretion.” The debate involves the extent to which monetary policy should be governed by the discretion of the head of the central bank (e.g., the Chairman of the Federal Reserve), or by relatively fixed rules that cannot be altered. Using the concepts from this chapter, can you provide an argument for each approach? Can you name other situations in business where there is a similar dilemma?
2. In law firms, litigation work is done mostly by lawyers who work independently, and who become famous for their work. By contrast, corporate law is usually done by teams of lawyers with different specialties, and the law *firm* develops a brand name. What conflicts do you see in having litigators and corporate lawyers in the same firm? Does your analysis suggest anything about the relationship between the quality of a performance evaluation and a firm’s organizational structure?
3. Many business experts advocate systems such as “Management by Objective” (MBO) for managing employee incentives. Under MBO, the supervisor negotiates a set of mutually agreed upon objectives for the employee to work on during the year. At the end of the year, rewards are based on the extent to which the objectives were successfully achieved or not. What costs and benefits do you see in “negotiating” objectives with your subordinate?
4. Still other firms use “360° (Degree) Evaluations.” Under this system, the firm asks a manager’s subordinates, colleagues, and customers for feedback on the manager’s performance. What advantages do you see to such a system? Do you see any disadvantages? What other policies might make such an approach less prone to the problems that you see? What kind of culture would be necessary for 360° appraisals to work?

References

- Baker, George (2002). “Distortion and Risk in Optimal Incentive Contracts.” *Journal of Human Resources* 37(4): 696-727.
- Berg, Norman & Norman Fast (1975). “Lincoln Electric Co.” *Harvard Business School case #376028*.
- Gibbs, Michael, Kenneth Merchant, Wim Van der Stede & Mark Vargus (2004). “Determinants and Effects of Subjectivity in Incentives.” *The Accounting Review* 79(2): 409-436.
- Gibbs, Michael (1995). “Incentive Compensation in a Corporate Hierarchy.” *Journal of Accounting & Economics / Journal of Labor Economics* joint issue, 19(2-3): 247-277.
- Jensen, Michael & William Meckling (1998), “Divisional Performance Measurement.” In *Foundations of Organizational Strategy*, ed. Michael Jensen, Boston: Harvard University Press.
- Lazear, Edward (1990). “The Timing of Raises and Other Payments.” *Carnegie-Rochester Conference Series on Public Policy* 33: 13-48.

Lazear, Edward (2000). "Performance Pay and Productivity." *American Economic Review* 90(5): 1346-1361.

Milgrom, Paul (1988). "Employment Contracts, Influence Activities, and Efficient Organizational Design." *Journal of Political Economy* 96: 42-60.

Murphy, Kevin J. (1993). "Performance Measurement and Appraisal: Motivating Managers to Identify and Reward Performance." In *Performance Measurement, Evaluation, and Incentives*, ed. William Bruns. Boston: Harvard Business School Press.

Murphy, Kevin J. & Paul Oyer (2005). "Discretion in Executive Incentive Contracts." Working paper, USC Marshall School of Business.

Prendergast, Canice (1993). "A Theory of 'Yes Men'." *American Economic Review* 83: 757-770.

Prendergast, Canice & Robert Topel (1996). "Favoritism in Organizations." *Journal of Political Economy* 104: 958-978.

Rewarding Performance

“Without merit there should be no reward.” (Chinese proverb)

Introduction

Now that we have discussed performance evaluation, we consider the next logical question: how should the firm use the evaluation to motivate the employee?

Before we consider that, however, we mentioned earlier that pay for performance helps an organization improve many human resource objectives, not just motivation. This idea is so important that it is worth illustrating before we turn to motivating effort. Therefore, for now assume that there is *no* motivation problem for employees. Is there still any benefit from tying pay to performance?

In fact, we already saw that there is. In Chapter 2, we analyzed sorting of workers. We argued that some form of probationary or deferred pay, *based on performance*, might improve the firm’s ability to recruit a better workforce. The more general idea is easy to illustrate formally. Suppose that the firm bases pay in some abstract way on performance, not necessarily through some kind of deferred pay or probationary period. It should still expect to improve recruiting – and investments in skills.

Assume that potential employees differ in their ability A or accumulated human capital H , and that performance (and the performance measure PM) are a function of both, $PM = PM(A, H)$. If pay is a function of performance, then it is also a function of ability and human capital. Those with higher ability or skills can expect to earn more, while those with less can expect to earn less:

$$\frac{\Delta Pay}{\Delta A} = \frac{\Delta Pay}{\Delta PM} \cdot \frac{\Delta PM}{\Delta A} > 0; \quad \frac{\Delta Pay}{\Delta H} = \frac{\Delta Pay}{\Delta PM} \cdot \frac{\Delta PM}{\Delta H} > 0.$$

Clearly, those who believe that they will be most productive at the firm are more likely to apply for or stay at a job there. The probation / deferred pay example discussed earlier is a special case of this more abstract idea. Similarly, it is easy to see that employees will have greater motivation to invest in skills, since the return on skills will be higher if their performance is tied more strongly to pay.

Piece Rate Pay and Sorting at Safelite Glass Corporation

Safelite Glass Corporation is the world’s largest installer of automobile glass (replacement windshields). In 1994, CEO Garen Staglin and President John Barlow changed the compensation scheme for glass installers. Prior to the change, installers were paid an hourly rate. The pay scheme was changed to a piece rate –

installers were paid a fixed amount for each windshield that they installed. After the change, output per worker rose by about 36 percent. How much of this was because installers worked harder? Was any of the increase due to better sorting?

These two effects can be estimated in a relatively straightforward manner. The effort effect can be estimated by taking a given worker and calculating the amount by which output rose after the change in the pay scheme. This incentive effect was estimated to be about 20 percent.

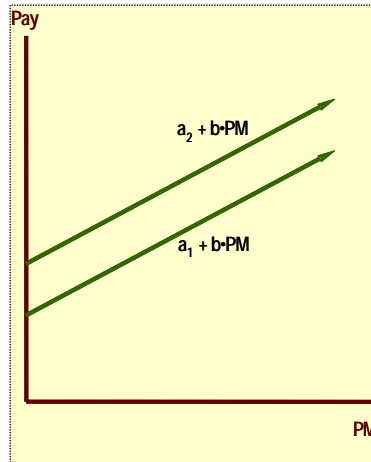
The remaining 16 percent increase in productivity was due to sorting. Safelite was able to retain its high-quality workers, and recruit other high-quality employees, because these employees were paid more (even for the same effort). Indeed, turnover rates fell for the most productive workers, and rose for the least productive.

Source: Lazear (2000)

The importance of incentives cannot be overstated. Modern market-based economies work so effectively because they provide generally good incentives for owners to make use of their assets, run their firms, innovate, etc. In other words, incentives are at the heart of effective economies, as well as effective organizations.

Of course, the most important reason to tie pay to performance is to increase employee effort, and better align it with firm interests, and that is the focus here. The introduction to this section of the text argued that we care about how the employee's evaluation varies with effort, and how pay varies with the evaluation. How should the firm vary pay with the performance evaluation? To analyze this, we first consider a very common form of pay for performance: the worker earns a base salary, a , plus a bonus. The bonus is calculated by multiplying a "commission" rate b times the performance measure. Thus, $\text{Pay} = a + b \cdot \text{PM}$.

Consider Figure 10.1 below. It shows two pay plans that have the same commission rate b , but different base salaries a . Which provides stronger incentives? The answer may not be obvious to you at first glance – to a first approximation, both pay plans should provide the *same* incentives and thus the same performance, even though one pays a higher base salary. To see this, consider the question in a different way. If the employee works a little harder, increasing the performance measure a little (say, installing 1 more windshield for Safelite Auto Glass), what is the reward? It is the same for both plans: the worker gets $\$b$ of additional bonus for each extra unit produced. In other words, the most important issue is *how pay varies with performance*, not the total level of pay.



Incentive Intensity v. Level of Overall Pay

Figure 10.1

The total level of pay may have some effect on motivation. If the worker's performance is low enough, he might get fired. The higher the base salary, the more the worker will wish to avoid this outcome. For this reason, higher base pay might increase motivation, but only to the extent that the threat of termination is a serious one. That is likely to be a very weak incentive in most situations.

To think about incentives correctly, it is much more important to focus on $\Delta\text{Pay}/\Delta\text{PM}$, or the "slope" or "shape" of Pay rather than the level. In this example, that is b , rather than a . The level of pay will largely be a function of labor market competition (which determines the price of employee skills), and the level of employee skills that the firm wants to employ. The slope of the pay-performance relationship is often called the "incentive intensity"; our next question is what determines this intensity.

Thus, if you are designing a compensation plan for an employee, a three-step process is generally most appropriate. First consider performance evaluation issues (Chapter 9). What metrics are available, and what are their properties (risk, distortion, etc.)? Should subjective evaluation also be used, and if so, how? Once the evaluation issues have been thought through, then consider how to tie rewards to the evaluation (this chapter). Only after the incentive issues have been analyzed is it time to think about the level of pay. Indeed, the level of overall expected pay is largely governed by the labor market's value for the employee's skills and the characteristics of the job.

How Strong Should Incentives Be?

Intuition

Although performance may depend on multiple types of effort, for now consider the simplest case: the employee provides only a single kind of effort, e , on the job. The task of the incentive system is straightforward in this case: motivate additional effort, until the point where it is too costly to do so.

Consider a salesperson selling personal computers. Before any output is even produced, the firm must bear a set-up (fixed) cost of \$1 million. Beyond that, each additional com-

puter costs \$900 to produce, so this is the marginal cost. Suppose that the company sells these computers for \$1,000 each, so that the marginal profit on each computer sale is \$100. For this reason, the firm makes incremental profit on each additional sale as long as the salesperson's commission for each sale is no more than \$100.

In addition to the marginal costs of production, the firm must compensate its salesperson. Suppose that it pays a bonus that is calculated as a % of revenue brought in by the salesperson. Revenue is a logical performance measure to consider for a salesperson, since the employee has little or no control over costs, but does have a great deal of control over revenue; it matches well to the tasks that are usually assigned to a salesperson. The firm may (or may not) pay a base salary to the employee. If it does, this is an additional fixed cost. Since the firm must cover its fixed costs to make a profit, assume that it only pays a commission for sales above some amount, say 10 computers sold per week.

There is an additional cost to the firm from each sale, which we mentioned in the introduction to this section of the text. To increase sales, the employee has to increase effort, incurring some disutility. The employee will require compensation for this disutility, which we will call $C(e)$. For the moment, assume that there is no performance measurement error, so that there is no cost to the employee from bearing risk.

Suppose that the salesperson's disutility of effort $C(e)$ is the amount shown in the second column of Table 10.1 (all numbers are calculated in terms of weekly sales). In other words, the worker requires at least \$2 to sell the first computer. Having sold one computer, the worker must be paid at least \$6 more ($\$8 - \2) to be willing to work hard enough to sell another. Similarly, if 22 computers were sold (so that the employee is already working relatively hard), the employee would require an additional \$90 to motivate the extra effort to raise sales by one unit. Given these figures, what is the optimal level of output? Try to answer before reading further.

Required Compensation for Different Levels of Sales		
Computers Sold	Total Disutility of Effort	Marginal Disutility of Effort
1	\$2	\$2
2	\$8	\$6
3	\$18	\$10
4	\$32	\$14
5	\$50	\$18
...
10	\$200	\$38
...
15	\$450	\$58
...
20	\$800	\$78
...
23	\$1,058	\$90
24	\$1,152	\$94
25	\$1,250	\$98
26	\$1,352	\$102
27	\$1,458	\$106

Disutility of Effort for a Salesperson
Table 10.1

The optimal level of sales is 25. Up to that point, the extra revenue exceeds the firm's production costs, plus the employee's marginal disutility of effort, so profit can be earned (and split in some way between the firm and the employee). The extra cost of going from 25 to 26 units, however, exceeds the marginal costs plus effort disutility. More simply, the disutility of effort is below 100 up to unit sales of 25.

What, then, is the optimal *commission*? Suppose first that the salesperson was paid a commission of 8%, or \$80 per computer. The salesperson would provide effort up to sales of 20 computers, but would be unwilling to go beyond that amount. In such a case, profit opportunities are lost, because sales are too low. If the company raised the commission a little, the salesperson would be willing to work a little harder (and be compensated for the extra effort), and profits would increase. In fact, the firm should increase the commission up to about 10%, or \$100 per computer, since all sales result in net profit up to 25.

It should *not* pay a commission above 10%. This would increase sales, but the extra revenue would not cover the extra marginal costs plus compensation to the employee for the marginal disutility of effort. Thus, the optimal commission rate exactly equals the value of the extra output. This is a general principle: it motivates the employee to just balance the marginal benefits (revenue) against marginal costs (from production and disutility of effort). In other words, the optimal commission rate (what we call the incentive intensity below) is set so that the employee equates total marginal costs, including effort, to total marginal benefits:

$$\text{Optimal commission rate} = b^* = \text{Revenue} - \text{MC} = \text{marginal profit on next sale.}$$

If the performance measure is revenue and the optimal commission rate is 10%, this is equivalent to using marginal profits from each sale ($\$1000 - \$900 = \$100$) as the performance measure, with a commission rate of 100%. In other words, our optimal commission gives *all* of the incremental profits from computer sales to the employee. In effect, it rescales the performance measure into units of profit. Thus, our optimal scheme is (with Q denoting units sold):

$$\text{Pay} = a + b \cdot \text{Revenue} = a + Q \cdot (\text{profit per unit sold}) = a + \text{profit from employee's sales.}$$

With this pay plan, what is the firm's profit? It is:

$$\text{Firm's profit} = \text{profit from employee's sales} - \text{Pay} = -a.$$

In other words, the bonus scheme gives back to the employee as a reward all of the incremental profits created from the employee's efforts. It also pays the employee a base salary. The only way for the firm to make any profit in this scheme is to pay a *negative base salary*. So much for economic theory!?

Selling the Job

Economic theory is more useful and realistic than might appear. This example illustrates that to give the employee perfect incentives, the firm must "sell the job" to the employee. In fact, many actual employment arrangements look quite similar to this idea. Consider the following examples.

Taxicab drivers: In many cities, cab drivers rent the car (or the license to operate a cab) from the cab company. They then receive a very large fraction (sometimes 100%) of the revenue that they collect when driving the cab. Usually, they also pay 100% of the incremental costs (gasoline). In effect, they buy or rent the job from the company that owns the asset (the car or license). This scheme gives the driver excellent incentives to maximize the value of the asset while it is being rented.

Securities traders: On stock, bond, options, or futures exchanges, those who trade must own a "seat" on the exchange. These seats often cost several hundred thousand dollars. Buying the seat gives the owner the right to the job, trading, as well as strong incentives to make effective use of the seat.

Wait staff in restaurants: In some cultures, tipping by customers is an important component of compensation for wait staff. Since tipping is based on the quality of service, it is a form of pay for performance. In the U.S., wait staff are often paid below the legally mandated minimum hourly wage (restaurants are a legal exception to the minimum wage). Thus, wait staff "buy" the job – by incurring an opportunity cost since their salary is below what it would be in other jobs. They do so because they expect that they can earn enough by working hard and earning tips.¹

¹ This system works well if the restaurant has repeat business. However, if customers do not expect to return to the restaurant, they may be tempted to leave without tipping, without fear of retribution later. Thus, tipping requires a cultural norm to motivate customers to implement of the incentive system – a form of implicit contract at the level of the society, in effect. This helps explain why tipping practices vary in different cultures.

Restaurant Wait Staff Compensation with a Twist

A famous restaurant in Chicago is reputed to add a twist to the standard compensation scheme for wait staff. It “charges” employees for the food and beverages that they sell to their customers, and credits them for the revenue (sales plus tips). (Presumably, it does so in such a way that it retains some of the profits.) Why might it do this?

One explanation is that imposing some costs on the employee reduces the temptation for the employee to cheat the restaurant (for example, bartenders sometimes pour extra drinks for their best customers without charging them, hoping to earn bigger tips). Another explanation is that using profit as the performance measure distorts incentives less than using revenue: the waiter has a better incentive to try to sell high profit-margin items (such as wine) than it otherwise would have.

Outsourced sales: Some firms sell their products through their own employees, while others outsource sales. In the insurance industry, for example, both practices are common. If a company outsources sales, in effect it is selling the job. The contractor buys the product, and then gets to keep a large fraction (usually 100%) of the profits from reselling it.

In fact, any time a firm uses a supplier for part of its business process, in a sense it is selling a job. One of the primary benefits of outsourcing is that it may be possible to implement stronger incentives. This also gives us a rudimentary theory for when outsourcing is likely to be more effective: the more that a part of the business process can be separated from other jobs in the firm, so that the performance evaluation for that job is as perfect as possible, the more likely is it a candidate for outsourcing. Firms do not typically outsource tasks that are highly interdependent with other jobs in the organization. There can be other costs of outsourcing, of course, including the costs of writing and monitoring the implementation of contracts with suppliers. With employees, firms can develop long term, implicit relationships (as discussed both earlier and later in this text), which may not be as easy to do with suppliers.

Middle managers: The examples above all involved cases where performance of an individual employee is relatively easy to measure. However, the “selling the job” intuition applies, if perhaps more weakly, to virtually any job. Consider an MBA student who is about to graduate, and is choosing between two job offers. The first offers a standard salary and small annual bonus. The second has similar job content, but offers a much lower base salary. However, it also has the possibility of a larger annual bonus if performance is high. If the student accepts the latter job, she will be buying the job, to some extent, since she will incurring the opportunity cost of a lower base salary than she could get elsewhere. However, she will be buying into the opportunity to work hard and earn more through the more generous bonus plan. Note, too, that the higher her ability, the more likely it is that she will accept the latter job with greater pay for performance.

In general, jobs that have stronger pay for performance will tend to have lower base salaries, all else equal. However, it is also the case that jobs with stronger incentive intensities will tend to give higher total pay. This is for three reasons. Can you think of what they are?

The first is that the employee will be motivated to work harder, and will be compensated through the reward scheme for higher effort. The second is that stronger incentives will attract better employees into the job, and the firm will have to pay such employees more because they have higher market values. The third is that stronger incentives imply riskier

pay, so that employees will have to be paid a greater risk premium. We consider this last point in the next sub-section.

This discussion has yielded very simple, but very important, intuition about pay for performance. The employee has perfect incentives, with interests completely aligned with those of the firm, if the incentive scheme “sells the job” to the employee. When this is done, the employee is in effect turned into an entrepreneur of his or her own. This motivates the employee to correctly balance the marginal costs of extra effort against the marginal benefits. This is why entrepreneurship is so important to a dynamic economy: entrepreneurs have very strong incentives, which motivates talented individuals to enter such positions, work hard, and apply their creativity to their fullest extent.

Furthermore, this illustrates why many organizations appear “bureaucratic” and inefficient to some extent. Incentives in most jobs will be imperfect compared to this theoretical ideal. Thus, middle managers may have relatively weak incentives, and even CEOs of large corporations may have weaker incentives than those provided by complete ownership. This does not mean that such incentives are not optimal. Incentive schemes involve trade-offs, and the incentives and efficiencies that arise from them will be imperfect.

Imperfect Evaluations and Optimal Incentives

Measurement Error

As discussed in the last chapter, it is virtually impossible to develop a performance measure that is error free. If pay is tied to the performance measure, the employee is rewarded and punished for some things that are uncontrollable: pay is variable. What is the effect of this risk on pay and optimal incentives?

Individuals tend to be risk averse. Therefore, variable pay imposes some psychological cost on them. One very simple way to model this is as was described in the introduction to this textbook section: the employee is assumed to have a disutility from riskiness of pay, equal to $R \cdot \sigma_{\text{Pay}}$. R is a risk aversion parameter that captures how risk averse the employee is. Employees who are less risk averse will have lower R , and vice versa. Thus, the total cost of effort to the employee is now $C(e) + R \cdot \sigma_{\text{Pay}}$.

Suppose, for example, that the employee’s true contribution to firm value is Q , but that the performance measure captures Q imprecisely, with measurement error ε : $PM = Q + \varepsilon$. ε is a random variable, with standard deviation σ_ε . Thus we have:

$$\text{Pay} = a + b \cdot PM = a + b \cdot Q + b \cdot \varepsilon.$$

Standard statistics then tells us that $\sigma_{\text{Pay}} = b \cdot \sigma_\varepsilon$, and the employee’s effort cost is:

$$C(e) + R \cdot b \cdot \sigma_\varepsilon.$$

Thus, to motivate the employee to work harder the firm now has to compensate not only for the extra effort, but also for the extra risk – the firm has to pay the employee a *risk premium*. Not surprisingly, the less accurate is the performance measure (the larger is σ_ε), the greater will be the risk premium that must be paid. This is one reason why firms may incur substantial costs to monitor employees, measure their performance carefully, and filter “uncontrollables” out of the evaluation. The costs of doing so are offset, at least partially, by lower compensation costs, since they increase accuracy of the performance measure (they are also offset by lowering of distortions in the performance measure).

The other interesting feature of the last equation is that the second part, the risk premium, is increasing in b . The stronger the incentive intensity, the riskier is the incentive plan to the employee (all else equal). This should make intuitive sense. Tying pay to performance more strongly means that the effects of measurement error will be magnified: good luck will be rewarded even more, and bad luck will be punished even more.

Therefore, we now have a tradeoff to consider. Stronger incentives improve the worker's effort, but also raise total compensation costs through a larger risk premium. Because of this, optimal incentives will be lower than those we considered above. Though we may partially sell the job to the employee, in general incentive schemes will not reward 100% of the employee's measured contribution to profits, since doing so is too risky. *The less accurate is the performance measure, the weaker will be the optimal incentives.*

In fact, this is an example of a general fact of economic life: the greater the insurance, the weaker the incentives, and vice versa. This issue arises in many contexts, such as the provision of health insurance.

Distortions & Multitask Incentives

A second problem with performance measures is that they almost always involve some distortion. The more that a performance measure distorts the employee's contribution, the less weight should be placed on it for incentives. The danger here is summed up by the familiar expression, "You get what you pay for": putting a strong incentive on a measure that emphasizes some tasks but not others will motivate the employee to pay too much attention to the tasks that influence the performance measure.

To see the problem with distorted performance measures more formally, let us consider a job where there two tasks or kinds of effort, e_1 and e_2 , with the employee's disutility of effort equal to $C(e_1 + e_2)$. For our salesperson, the first might be selling new computers (quantity) and the second might be providing some installation help to the customer. Suppose that the employee's contribution equals $Q = q_1 \cdot e_1 + q_2 \cdot e_2$. As we will see, when the job is more complex, incentive schemes tend to get more complex as well.

How might we measure performance in this example, and tie it to pay? In many jobs, the firm has metrics that provide some measure of one dimension of the employee's job. In the salesperson example, revenue is very easy to measure. However, customer service is intangible, and difficult to quantify. Nevertheless, the firm may have access to some metrics on service, such as customer satisfaction survey data. This suggests that the firm might have the following three performance metrics for the salesperson:

$$PM_1 = q_1 \cdot e_1 + \varepsilon_1,$$

$$PM_2 = q_2 \cdot e_2 + \varepsilon_2,$$

$$PM_3 = \alpha \cdot PM_1 + \beta \cdot PM_2.$$

The first estimates the employee's contributions to revenue (probably with high accuracy, so that σ^2_1 is low). The second estimates the value to the firm of the employee's customer service activities (probably with much more inaccuracy, particularly since it is an attempt to quantify an intangible). The third is a combination of the first two.

If the firm gives a bonus based only on PM_1 , there is no incentive for the salesperson to provide service. This is a strong form of distortion. Unfortunately the firm is likely to end up

with lots of sales in the short term, but many unhappy customers and little repeat business. A natural response to the problem is to offer a second bonus based on PM_2 . The idea is to provide *balanced* incentives across the different tasks.

Unfortunately this is not likely to solve the problem very effectively. Since σ^2_1 is relatively low, by the arguments we just gave the commission rate on PM_1 should be relatively strong. But since σ^2_2 is relatively high, by the same arguments the commission rate on PM_2 should be relatively weak. Therefore, we are still likely to have unbalanced incentives.

One response to this is to split the work into two different jobs. The firm might put easy to measure tasks in one job, for which it uses strong pay for performance. It might then put difficult to measure tasks into another job, for which it uses relatively muted incentives. It might then invest more resources (and personnel) into the tasks with weaker incentives, so that overall output is balanced across the different dimensions of work. Pursuant to our example, many firms do, in fact, provide service and support in jobs that are separate from sales, and use different approaches to incentives and monitoring for each.

However, altering the job to fit the performance measure seems to be a bit like putting the cart before the horse; it seems more natural to try to alter the performance evaluation and incentives to fit the job. How might the firm do this?

One potential solution is to combine the metrics in some way, as in PM_3 . If the two individual metrics could be weighted and combined in some way, the resulting (broader) metric might distort incentives less. In this example, if $\alpha = \beta$, then PM_3 does not distort incentives for the employee to provide both kinds of effort. (The commission rate can then be re-scaled, by dividing through by α , so that the overall reward is equal to the employee's contribution, perhaps reduced for risk aversion reasons.)

Of course, the correct relative weighting of different performance metrics is not so easy in practice. How much relative weight should the firm put on customer satisfaction, which is a numeric proxy for something that is qualitative? Over time, the firm may be able to obtain reasonably good estimates of the relative value of both dimensions of performance. Or, it may be able to experiment with different relative weights, and arrive at a reasonable balance through experience.²

The world is a dynamic place, though, and the relative weights to give to different dimensions of the job for incentives often change. When this is so, the formulaic approach of picking numeric weights α and β is not likely to work well. What is often the best approach in those cases is to use judgment – subjective evaluation – to provide balanced incentives across multiple tasks. Indeed, the more complex the job, the more likely is it that evaluations are done subjectively, and rewards are tied more informally and less formulaically to performance.

Summary: How Strong Should Incentives Be?

We can now summarize what factors affect the optimal intensity of (explicit and implicit) incentives, and how. If our employee were risk neutral or we could measure performance very accurately, we would set the commission rate so that, when multiplied by the per-

² This approach of calculating several performance metrics (some based on qualitative dimensions of the job), and then combining them to produce a broader performance measure that distorts incentives less, is the principle behind the Balanced Scorecard technique that many firms use. Unfortunately, Balanced Scorecards are often very complex to design and manage.

formance measure, the reward was approximately the value of the employee's incremental contribution to firm value. In practice, employees are risk averse and measures are imperfect, so actual incentives tend to be weaker. The factors to consider:

Value of Employee Effort

The more profitable to the firm is additional employee effort, the stronger should the incentive be. For example, in Table 10.1, if the % profit margin was increased, the firm should increase the % commission rate. For this simple but important reason, incentives are almost always stronger at higher levels in the hierarchy, and weaker in lower levels.

Importance of Sorting

Incentives also generate good self selection. The more important it is to the firm to sort workers by ability or accumulated skills, the stronger should be pay for performance. Thus, incentives tend to be more important for new recruits, those who are new to their job, and in high-skill occupations.

Measurement Error

The more accurate the measure, the stronger the incentive that should be placed on it.

Risk Aversion

The less risk averse the employee, the stronger the incentive intensity should be. When recruiting for jobs with strong incentives, the firm should consider risk aversion as a factor in hiring.

Trust and Subjectivity

A variation on the themes of measurement error and risk aversion arises when subjectivity is used in the incentive scheme (for evaluation, weighting of metrics for rewards, or both). The employee then faces the risk of favoritism and bias of the evaluator. Thus, the more trust the employee has in the evaluator, the better the evaluator is at making judgments, and the more effective the appraisal process is, the stronger the discretionary incentive can be.

Potential Manipulation

The less likely is it that the employee can manipulate the measure, the stronger should the incentive intensity be.

Distortion and Multitask Incentives

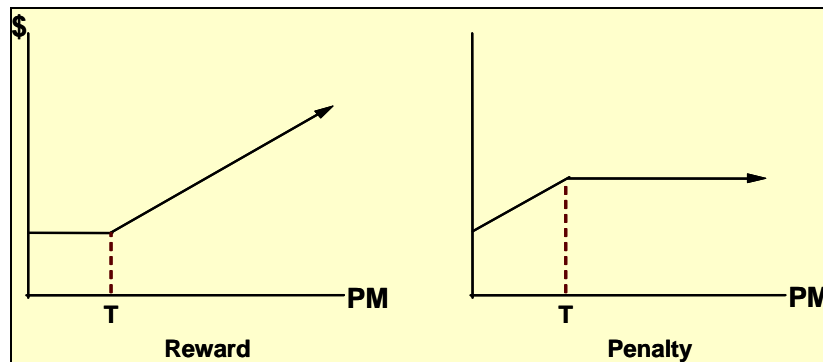
The more distorted the measure, the more the firm should consider adding additional incentives, and think about incentives as a system of reinforcing rewards. This might imply the use of several formal incentives, or the use of a broader, more subjective and implicit approach to incentives.

Paying for Performance: Common Examples

The discussion so far has focused on the simplest case, a linear pay for performance scheme. In that case, the question of incentive intensity boils down to how steep or flat the slope is in Figure 10.1. In this section we briefly consider several other commonly observed pay-performance relationships, to develop some other practical considerations.

Rewards or Penalties?

Figure 10.2 shows two pay-performance schemes. On the left is a “reward” scheme similar to the one described above and plotted in Figure 10.1, except that the employee earns a base salary for low levels of output, only earning a bonus if performance rises above some threshold level T . On the right is a “penalty” scheme, where the employee earns a base salary for high levels of output, but the reward is reduced if pay falls below T .



Reward & Penalty Incentive Schemes

Figure 10.2

Schemes like the one at left are quite common. Why might a firm add a threshold? One reason has to do with risk aversion. Paying a fixed base salary even for relatively low performance provides some *insurance* to the employee. Performance may be low because the employee did not work hard, but it may also be low because the employee was unlucky. Risk averse individuals are most concerned about avoiding the most negative outcomes. Therefore, this kind of pay combined with insurance can cause an employee to effectively be less risk averse.

This has two advantages. First, the firm may be able to increase the incentive intensity to the right of T . If T is not set too high, this means that the employee will have stronger incentives. Second, the employee should be more willing to take risks, because they are less likely to be punished for mistakes or chances they take that do not succeed. This can be helpful in jobs where some amount of risk taking, perhaps to spur innovation, is important. This second effect is important when pay comes in the form of employee stock options. We will see in Chapter 12 that employee options have a payoff structure that looks very similar to the Reward scheme in Figure 10.2 above.

The penalty scheme is not as common in practice. When might it be used? In the penalty scheme, the slope is positive for low performance, but zero for higher performance. The employee has little incentive once performance has passed the threshold T . Thus, the

penalty scheme may be useful in cases where, beyond some point, there is little or no value to the firm when the employee's performance measure increases.

A real-world example is an executive MBA student of one of the authors, who managed the electric power system for a small Asian country. The power company's "up time" (% of time that electricity was available) was 99.96%, a nearly perfect record. Suppose that the performance measure PM was % up time. While it might be possible for the power plant to increase up time beyond 99.96%, it might be extremely costly to do so, since performance is already near perfection. If so, it may not be profitable to increase the performance measure – the measure does not align perfectly with actual profits or firm value. Then the power company might want to put in place an incentive scheme like the penalty scheme above, which motivates the manager to avoid a fall in performance, but does not motivate increasing performance.

Thus, the Reward scheme is useful when there is upside potential in the employee's job – when high levels of performance translate into increased firm value. If there is also little downside, providing insurance of the kind illustrated at left in Figure 10.2 makes sense. An entrepreneur is a perfect example, so stock options may work well in that case. The Penalty scheme is useful when there is downside potential – the employee can damage firm value – but little upside potential. These are sometimes called "Guardian" jobs; a security guard is a good example.

Framing

Psychologists sometimes argue that positive reinforcement is a more powerful motivator than negative reinforcement. In the context of our discussion, that would suggest that rewards are more effective than penalties, since the worker's pay rises if the performance measure increases.

However, this is not so obvious on further reflection. The penalty scheme on the right is graphically identical to a simple bonus scheme with a "cap" (see below) on the bonus. As long as performance is below T, it *also* involves positive reinforcement. Indeed, we could easily have referred to it as a reward with a cap, rather than as a penalty.

To the extent that such labels matter, it may be worthwhile avoiding labels such as "penalties" or "punishments." However, these labels do serve a purpose. By calling the picture at left a reward, and the one at right a punishment, the firm is communicating to the worker something about the expected level of performance and the nature of the job. The scheme at the left is appropriate when the firm expects typical performance to be to the right of T. It also signals to the employee that the firm wants the employee to strive to increase output, and perhaps to be more willing to take risks. By contrast, the scheme at the right is appropriate when the firm expects typical performance to be to the left of T. It signals that it is important to avoid declining output, and to be conservative on the job.

One important issue with both the Reward and Penalty – and any scheme where the slope changes in some way – is how to set the threshold T. Consider the Reward scheme. To the left of T, the slope is zero; to the right it is positive. If T is set too high, then it is highly unlikely that the performance measure will be above T even if the worker gets lucky (measurement error is large and positive). When that is true, $\Delta PM/\Delta e$ will be approxi-

mately equal to zero, and the worker will have little or no incentive. Similarly, in the penalty scheme, if T is set too low the worker will have little or no incentive.

In practice, it is often difficult to set the correct level of thresholds when first implementing an incentive plan. Moreover, circumstances change. For example, the worker may learn on the job, or production methods might change. These would make it easier (or harder) to produce a given level of performance, in which case T should be changed. There are many good reasons why a firm might want to change T , and the most typical case involves raising T as skills and methods improve.

Changing the threshold, however, can be tricky. In the reward scheme in Figure 10.2, if T is increased but nothing else is changed, it becomes more difficult to earn the bonus, and the bonus will be smaller for any given level of performance. Not surprisingly, the worker will not be happy. The worker may perceive that the firm is trying to reduce compensation.

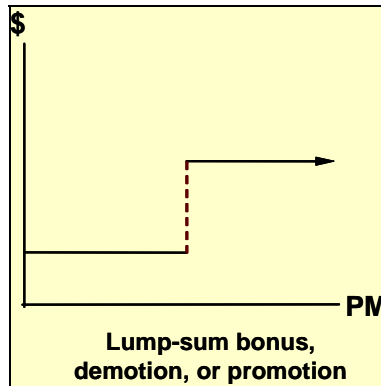
Moreover, managers sometimes design a pay plan, and find that workers produce and earn more than was expected (perhaps because the manager underestimates the power of incentives). A natural response is to increase T , reduce the commission (slope), or reduce the base salary. While this reduces compensation costs, it may have negative consequences. In both of these examples, there is a real risk that the worker feels the firm is renegeing on implied promises about the reward scheme.

This illustrates a general point about incentive systems: simplicity is a virtue. Where possible, straightforward linear schemes tend to work best. Thresholds, changes in incentive intensity, and lump sum rewards often create problems. Moreover, complex pay systems make it harder for employees to understand how they will be rewarded for their performance, which may reduce their incentives since the perceived pay-performance link will be weaker. Finally, there can be a subtle issue of trust. Complex pay systems can cause some employees to worry that management is trying to take advantage of them in some way, though they are not sure exactly how.

Continuing with the implicit contracting idea, consider the dynamic issues raised once the firm responds to high performance and pay by “ratcheting” T higher, or the commission rate b lower. The worker may conclude that he was, in effect, punished for high performance. If this happens, incentives will be *reduced*. Therefore, the firm needs to be very careful about how it implements changes to incentive plans. There should be good reasons to do so, and these should be communicated clearly to the employee. Moreover, when the incentive plan is introduced, the firm should carefully reserve the right to evolve the plan in the future. Clearly, the better the degree of trust between the firm and its employees, the less likely are such ratcheting issues to be a problem. The Appendix discusses this “Ratchet Effect” in more detail.

Lump Sums, Demotions, or Promotions

Figure 10.3 shows what may be the most common pay for performance relationship of all. Here, if performance is above the threshold, there is a discrete jump in reward. Why might this be the most common form of incentive in practice? Because one example of such a pay-performance shape is a promotion. Most promotions come with large increases in compensation (and perhaps other job amenities). If the promotion is based on performance, as most are, then it can be an important incentive. In fact, this case is so important that we spend most of the next chapter discussing it. Also note that the threat of being fired or given a demotion will look similar (if the employee suffers a loss from either).



Lump-sum Rewards

Figure 10.3

Another example of the kind of pay-performance relationship shown in Figure 10.3 is a lump sum bonus. Sometimes firms award a fixed amount if an employee meets a target. For example, an auto dealership might offer a \$1000 prize to an employee who meets a sales goal, or to the best performing employee of the month.

One problem with reward structures like Figure 10.3 is that incentives can be brittle. This is similar to the problem of a threshold that is too high in the Reward scheme in Figure 10.2. The slope of the pay-performance relationship is zero if the employee's performance is below T ; *infinite* if performance is just at the margin between winning and losing; and zero again if performance is above T . This scheme generates very strong incentives if employees are right near the threshold, but if they are too far below or above, they will tend to slack off. (As an example, consider what sports teams do if they are far ahead – they slack off by putting in the second string.) Unless this is desirable, a smoother pay-performance relationship would make sense.

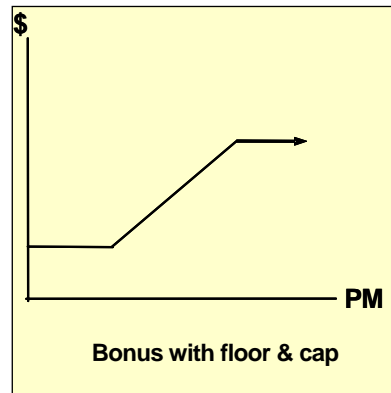
One case where such reward schemes may be useful combines our arguments about Reward and Penalty schemes. Sometimes it is important that employee performance falls into a narrow band. If performance is too low, *or* if it is too high, firm value may be lower. One example is an assembly line worker. Working too slow *or* too fast might cause coordination problems. Another example is a plant manager. The firm may choose production methods to optimize costs based on a certain level of planned output. If the plant is too far below or above that level of output, average costs may rise dramatically. Finally, when coordination and control is very important (say, the firm makes extensive use of budgets), predictable performance from employees may be desirable. In all of these examples, an incentive system that motivates the employee to perform near a target level can make sense.

When incentives change dramatically in incentive systems that involve thresholds, and large changes in the slope or incentive intensity, employee behavior can be problematic. In Figure 10.3, the all-or-nothing nature of the reward is not only likely to motivate the employee to work harder if near the threshold. It is also more likely to motivate the employee to manipulate the performance measure if possible. Manipulation of performance measures is always a problem in incentive systems, especially when the stakes are higher. But it is even more of a concern if the employee faces an abrupt change in the shape of the pay plan, so that small changes in performance can lead to large changes in rewards or the potential for further compensation. By contrast, smooth pay-performance relationships tend to give more continuous incentives for employees to exert more effort.

Thus, incentive plans with thresholds and lump sum rewards are more likely to lead to employee malfeasance. This is why many corporate accounting scandals arose in cases where top management had extensive stock option grants.

Caps on Rewards

Our final scheme to consider, in Figure 10.4, is the Reward plan, but with a cap. A cap is a maximum amount of bonus or other reward that the employee is allowed to earn. Some bonus plans have caps, and others do not. Why might a firm use a cap?



Floor or Cap in an Incentive Scheme

Figure 10.4

Managers sometimes put in caps to ensure that employees do not earn “too much” pay. However, one should be careful in making such an argument. First, the level of pay might be reduced without imposing a cap simply by lowering the intercept (shifting the pay plan downwards in Figure 10.4). Second, if the employee is earning a lot, this must be because the employee is performing well. If the performance measure is a reasonable proxy for the employee’s contribution to firm value, then the firm is probably profiting from this extra performance as well (since optimal commission rates often give the employee less than 100% of their contribution, as discussed above). A cap would reduce incentives (beyond the second threshold), reducing firm profits. In fact, in some cases managers seem to impose caps out of such misguided motives, including a reluctance to allow their subordinates to earn more than they do. But these kinds of motives are very similar to the Ratchet Effect issues described above and in the Appendix.

There can be justifications for a cap on rewards. To see why, remember that the bonus is based on a performance measure, which is an imperfect proxy for the employee’s actual contributions to the firm. The performance measure is a function of the employee’s effort, but also the employee’s luck. And it may also be manipulated or distort the employee’s incentives. In some jobs, it is implausible that extremely high measured performance is due to the employee’s efforts and talents. In those kinds of jobs, the higher is the measured performance, the more likely is it that the measure reflects either luck or manipulation. When this is so, the firm should consider a cap, to avoid rewarding luck or motivating manipulation of the incentive system.

One example of this was Michael Milken, the head of Drexel Burnham Lambert’s “junk bond” group in Beverly Hills. Milken’s compensation scheme had a very strong incentive intensity – a steep slope. It also had no cap. One year, his measured performance was so high that his annual bonus exceeded half a billion dollars. Unfortunately, it became in-

creasingly difficult for Milken to generate high quality, profitable new deals for Drexel. Instead, his group began engaging in transactions and business methods that many viewed as unethical. Eventually, he was prosecuted and served time in jail, and Drexel went bankrupt and was closed down. Arguably, if Milken had had a cap on his incentives, this might not have happened (though it is plausible that he had such strong intrinsic motivation that it might have anyway).

There is a problem with a cap, however. To see this, consider the story of Ross Perot. Before becoming an unsuccessful U.S. Presidential candidate in 1992, Perot was a very successful businessman. Perot's first job was to sell mainframe computers for IBM. This was in the early days of mainframe sales, so it was a very good time to have such a job. Still, Perot was not only lucky, he was also a very talented and hard worker. In fact he was the highest seller in all of IBM. In one year, he met his annual sales quota by January 19th – and January was the first month of IBM's fiscal year! Perot soon became frustrated and anxious to put his talents to better use. He came up with a new product idea for IBM – selling computer systems, including all of the necessary software and installation, instead of selling the parts individually. IBM considered but rejected his idea.³ Perot then quit, founded Electronic Data Systems (EDS), and competed directly with IBM. In the process he made billions of dollars, and IBM shared none of it.

Perot's story illustrates the danger with caps and general "Ratchet Effect" type attempts to limit rewards for very high performers. If the firm does so, it risks losing its best employees. Perhaps that is why Drexel paid Milken the way they did. They knew that in investment banking, it is relatively easy for a highly talented banker to set up their own competing firm. A firm must compensate its employees for their market value. For the most valuable employees in highly skilled occupations, this often means that, in effect, the firm must give most or all of the residual profits created by an employee as the reward – they have to sell the job. For this basic reason, most organizations in very knowledge intensive industries (investment banking, law, consulting, and in some ways academia) are organized as professional partnerships, with the more productive partners earning the most.

Applications

Profit Sharing, ESOPs, etc.

Many firms offer some form of broad-based "incentive" plan such as profit sharing, gain sharing, or employee stock ownership. Managers often argue that such plans give employees a "sense of ownership" and a feeling that "we are all in this together." Do such plans make sense? From the perspective of the theory we have just described, they do not. Let us consider the arguments against such plans, and then consider a couple of reasons why they might – sometimes – make sense.

To analyze an incentive system, use the principles of this and the last chapter. Ask two questions first: What are the properties of the performance evaluation? How is the evaluation tied to performance?

³ Perot reports that IBM "took the idea to the top of the organization, but they ultimately rejected my plan" (1996, p. 72). Of course, IBM was a very hierarchical, conservative organization, as described in Chapter 6, so this is not surprising.

Performance Evaluation

The performance evaluation for such plans is a very broad one. In the case of profit sharing at a factory or other organizational unit, the measure is unit profits, revenue, etc. In the case of firm-wide profit sharing, the measure is firm profits. In the case of employee stock ownership, it is stock.

Already incentive alarm bells should be going off in your head. These are *not* good performance measures for typical employees. Though employees can help improve firm value, the effects of a single individual are not going to show up in such measures unless the employee is a key figure in the firm. These measures are essentially completely uncontrollable by most employees. Thus, it is implausible that they provide *any* incentives at all. Indeed, most studies find that such plans have no detectible effect on productivity or profits.

Pay-Performance Relationship

A further problem with these pay plans is that even if the performance measure were a good one, the incentive intensity tends to be very small. The reason for this is the “Free Rider” problem. Remember the last time you worked on a group project (perhaps for this class?). There may have been a member of your team who didn’t seem to do their fair share of the work. Yet if all in the group received the same grade, then this person shared in the reward anyway. They got a “free ride.” The problem is that the *change* in reward varies little, if at all, with a *change* in employee effort.

This is virtually inevitable in group reward schemes that simply share the reward equally. The reason is that, if there are N people in the group, the reward for each person equals $1/N$. Of course, $1/N$ is a hyperbola, which approaches zero very rapidly. Thus, unless the group size is small, the incentive intensity has to be close to zero.⁴ As an extreme example, the giant German company Siemens has employee profit sharing, and approximately 400,000 employees (in 2005). In this case, the “commission rate” on profits is:

$$b = \frac{1}{400,000} = 0.0000025.$$

That’s a pretty tiny incentive intensity! Although clearly this is an extreme case, it does illustrate the Free Rider problem, and how difficult it is to justify such compensation plans purely on incentive grounds.

Counterarguments

So why do so many firms make use of such broad-based plans? One explanation is that they misunderstand the theory, and to some extent that seems likely. But there are some counterarguments.

First, peer pressure can, conceivably, cancel out the Free Rider effect. If all share in the same reward, then all have some incentive to pressure their colleagues to work harder. To the extent that this is true, then the optimal size of the group for incentives can be larger. Moreover, such effects might increase productivity for other reasons as well. In cases where work is interdependent and the firm makes extensive use of teamwork, group-

⁴ Social psychologists often argue that the optimal group size is about 5 or 6 members, at most. It is likely that one of the important reasons for this is that a larger group would rapidly encounter the Free Rider problem.

based incentives are more likely to make sense, and may also reinforce any cultural rules about the importance of cooperation. Indeed, the few studies that find a positive effect of ESOP or profit sharing plans usually find that they are more likely to be helpful in firms that use team production.

However, anyone who has ever worked in a group (or tried to get everyone to share a dinner bill at a restaurant) knows that Free Rider effects are ubiquitous, so it seems implausible that peer pressure can overcome free riding in most cases.

Another explanation that is sometimes given is that compensation becomes more of a variable cost, rather than a fixed cost: pay rises as performance rises, and declines as performance declines. This may lower the financial risk of the firm, leading to a lower cost of capital. However, this explanation does not make sense for the following reason: while it may lower the cost of capital, it does so by raising the cost of employee compensation by even more. The reason for this is that employees are typically even more risk averse than investors, and require a larger risk premium for accepting this risk in their compensation than do would shareholders. We will discuss this problem more extensively when we consider employee stock options.

One explanation involves public relation concerns. There are good reasons to consider giving large grants of stock and options to top executives, and to reward them generously if the firm performs well. But firms often face pressure from some shareholders, unions, the press, or other groups when they give large payouts to executives. Some of these criticisms might be blunted if the firm also gave stock or profit sharing to lower level employees. Thus, such plans might be bad incentives, but good public relations. (At a cost to lower level employees, whose pay becomes risky.)

A final explanation for stock ownership plans is that top management adopts such plans as a way to increase the demand for their stock, in the hope of raising the stock price. For example, some companies invest their employees' pension funds heavily in the company's own stock. Such practices are far from in the interest of employees, however, since their pension funds become as undiversified as possible: largely invested in a single company, and highly correlated with the employee's human capital.

Organizational Form & Contracting

As with most of the principles in this text, the incentive intuition developed in this module has implications for virtually all aspects of business, not just employment. Here we provide two brief illustrations of how the principles can be applied.

Franchising

Franchising is an unusual organizational form that combines principles about incentives with those about decision making and use of specific knowledge. For the franchisee, it is intermediate between true ownership and more typical employment. For the franchiser, it is intermediate between outsourcing and internal production.

Clearly a franchise is a very broad performance measure (see Table 9.2 in the last chapter); it is almost pure ownership. A typical franchisee has to pay a large up-front fee for the right to run a franchise. In exchange, the franchisee is allowed to run the operation (subject to limitations). Since the franchisee can sell the franchise at its current market value, the performance measure is very close to that of an owner of an independent store. This

means that there are very few distortions in the performance measure, for intangibles, investments, and long-term decision making.

However, the franchise stops a little before pure ownership. The franchiser retains some decision rights. For example, a franchiser typically specifies the product line. It may require the use of specific suppliers (e.g., specific sources for beef in McDonald's hamburgers). It often requires employees to wear certain uniforms, and has control over the design of the store.

In doing so, this organizational form makes effective use of centralization and decentralization. The decision rights that are retained by the franchiser are those that affect the overall brand name: product, quality control, customer experience, marketing, and so on. The desire for consistency of the product means that there are important benefits to centralizing decisions that affect consistency.

By giving other decision rights to the franchisee, this form allows decentralized decision making to make use of local specific knowledge. For example, the franchisee usually handles most personnel matters, from hiring to training to compensation and incentives. Since many important implementation decisions are left in the hands of the franchisee, the broad performance measure is appropriate for such a broad job design.

Cost Plus v. Fixed Fee

Suppose that your firm needs to have a building constructed, and hires a builder. What kind of contract should be written? Two types are common in construction, cost-plus and fixed-fee. A cost-plus contract pays for the cost of all materials and labor used to construct the building, plus an additional fee or percentage to provide some profit margin for the builder. A fixed-fee contract specifies the type of building and materials, and perhaps other aspects of the project, and pays a fixed amount on completion. A fixed-fee contract might be more elaborate, specifying various payments for completing various stages of the project.

Since the cost-plus contract reimburses for all inputs, there is no incentive to reduce quality of construction. Indeed, since the profit margin is usually specified as a percentage of costs, the builder may well have *too much* incentive for quality. Many government contracts are cost-plus, and it is common to see "overbuilding" in such cases. The higher the margin, the higher the quality (and the longer the duration) that can be expected from a cost-plus project.

By contrast, a concern with the fixed-fee contract is that it will give weak incentives for quality. The builder's incentive is to get the job done for the minimum quality that it can get away with while honoring the terms of the contract. (This will be mitigated to the extent that the builder cares about its reputation and the awarding of future contracts.) A fixed-fee contract may also motivate the builder to complete the project very quickly (by contrast to a cost-plus contract, where the incentive is to bill as many hours of work as possible).

However, some fixed-fee contracts specify that the customer makes payments as the work progresses. Therefore, in some cases it can be difficult to get a fixed-fee contractor to complete a job, if too many payments have been made, so that there is inadequate compensation for completion of the project. One way to address such concerns in both types of projects is to specify rewards or punishments for meeting specific time goals.

Which is better? If quality is easily observed and verified, then it is probably better to go with a project-based payment where providers are penalized for deviations from contracted quality. If quality is not easily observed, but the appropriate amount of time to complete the job is known, it is probably better to use a cost-plus system, with penalties for running over the contracted amount of time.

The general point, though, is that the more you think about incentives in complex and subtle ways, the better you will be able to understand economic behavior in a wide variety of settings. Moreover, the better will you be able to apply the principles effectively.

Motivating Creativity

A frequent criticism of pay for performance (especially from social psychologists) is that it might destroy the employee's intrinsic motivation. The exact mechanism by which this occurs is not always clearly stated. One explanation that is sometimes given is that the employee will feel "controlled" if given pay for performance, and will thereby withhold effort that would have been otherwise provided because of intrinsic motivation.

There is a much more straightforward way to think about this issue, and one that explains the examples that are usually given. Intrinsic motivation usually is most important for tasks that are complex and intellectually challenging to the employee. These are often most important in jobs that involve a substantial amount of creativity and learning. By their nature, it is often difficult to develop good performance measures for these kinds of tasks. One reason is that such intellectual tasks can be difficult to quantify. Another is that in creative work, it can be hard to specify the desirable features of output in advance. The metrics that might be used will then distort incentives substantially. Moreover, the tasks that *can* be quantified will tend to be those that involve less creativity. (Imagine what would happen in a research university, for example, if professors were told that they would be given tenure if they published a specific number of articles – the performance measure was purely quantity.)

Putting a strong incentive on such measures will, of course, focus the employee more on what is measured and rewarded, and less on the creative aspects of the job. But this is not because of psychology; it is simply a problem of inadequate performance measurement.

In some cases it is possible to use pay for performance to motivate creativity, because reasonable (if imperfect) metrics are available. For example, some companies reward divisional managers based on the percentage of products sold that were newly developed in the last two years. This may work reasonably well at spurring innovation because products not only have to be new, but they also have to meet with approval from customers. In many cases, however, the best alternative is to perform careful subjective performance evaluations. Research universities do exactly this: professors are typically evaluated only every couple of years, but the evaluation is highly subjective, attempting to assess the creative contribution of the professor's research.

Summary

Incentives are not only the essence of economics; they are an essential part of organizational design. In order to understand organizational design and employee behavior, a thorough understanding of incentives is crucial.

We analyzed incentives by considering an example of a formal bonus plan for a salesperson. However, the intuition developed is much more general. It applies to *all* kinds of incentives, both formal and informal, designed and accidental. Incentives in practice can be very subtle; you need to develop a trained eye for uncovering these subtleties and how they affect behavior and organizational performance.

Abstractly, the goal of pay for performance is to replicate the spirit of ownership and entrepreneurship for employees. We saw that a perfect incentive plan essentially sells the job to the employee, so that they are “mini entrepreneurs.” This is in accord with the metaphor of a market economy that this book uses to think about organizational design.

However, incentives in practice are often quite different from this ideal. The most important reason for this is performance evaluation problems, which is why we devoted an entire chapter to such issues. Any imperfections in the evaluation drive a wedge between the incentives of the employee and those of an owner-entrepreneur. To the extent that there is performance measurement error, firms will invest more resources in monitoring and careful evaluation. But it is also true that optimal incentives will be weaker, so that employees will not exert as much effort as they otherwise would.

This logic illustrates why organizations often seem relatively inefficient. If it were possible to actually replicate markets (that is, have a price system to serve as an excellent performance measure), then the firm would outsource anyway. It is just those situations which require long term *employment* relationships between the firm and the worker that also tend to make performance evaluation imperfect. For example, the worker may be engaged in complex, multitask work, some of the effects of which are intangible or not realized in the short term. Or, the work may be highly interdependent with colleagues, so that it is difficult to disentangle the effects of a single worker from that of the group.

Other performance measurement problems, such as distortions or potential for manipulation, may also lead to weaker incentives. They tend to make design of incentive systems more complex than the simple example of the salesperson. For example, multitask jobs often yield distorted performance measures (these also may arise from the desire to reduce risk in the evaluation). Thus an important intuition for such jobs is that the incentive system must provide *balanced incentives* across the different aspects of the work. This may require multiple rewards based on different factors, a different (broader) approach to performance measurement, or careful subjective evaluation and implicit rewards.

Potential manipulation of performance metrics may also require that the supervisor spend additional time monitoring the employee, to try to detect the manipulation. It may also lead to the use of subjective evaluations and implicit rewards.

Putting these ideas together, the incentive system often ends up being a complex *system* of interrelated parts: monitoring, various performance measures, subjective evaluation, explicit and implicit rewards, and so on. Designing and managing such a system can be something of an art, and is an important part of the manager’s job.

Review Questions

1. Suppose that you use all of your savings to buy a struggling company, and now must turn it around. Should changing employee incentives be one of the first tools that you use to manage change? If so, why? If not, why not? If you do use incentives to drive organizational change, what else is likely important for you to change?

2. How might you try to detect if your employee is gaming the incentive system? Try to think of concrete examples.
3. Consider important medical care decisions. Should such decisions be made by the doctor or the patient? What factors are important? Given your answer, how would you think about structuring incentives in order to balance quality of care against cost?
4. Do the incentive principles described in Chapters 9-10 apply to not-for-profit organizations? Why or why not?

References

Gaynor, Martin, James Rebitzer & Lowell Taylor (2004). "Physician Incentives in HMOs." *Journal of Political Economy*.

Gibbons, Robert (1987). "Piece-Rate Incentive Schemes." *Journal of Labor Economics* 4: 413-429.

Holmstrom, Bengt & Paul Milgrom (1991). "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization* 7:24-52.

Lazear, Edward (1986). "Salaries & Piece Rates." *Journal of Business* 59: 405-431.

Lazear, Edward (2005). "Speeding, Tax Fraud, and Teaching to the Test." Working paper, National Bureau of Economic Research.

Perot, Ross, 1996. *My Life & The Principles for Success*. Arlington, TX: The Summit Publishing Group.

Roy, Donald (1957). "Quota Restriction and Goldbricking in a Machine Shop." *American Journal of Sociology* 67(2): 427-442.

Appendix: Formal Analysis of Optimal Incentives

Optimal Commission Rate

Here we formally derive the result that the firm maximizes profits by paying a commission rate equal to 100% of net revenue, if the employee is not risk averse. This provides a more rigorous treatment of the general intuition of the idea of "selling the job" to the employee.

The problem is broken into two parts. First, the worker's optimal behavior is analyzed. Then the firm's optimal commission rate is derived, taking into account the worker's behavior. Assume for simplicity that e is scaled so that 1 unit of e produces \$1 of incremental profit for the firm, so that $Q = e$. The worker chooses effort to maximize utility:

$$\max_e a + b \cdot e - C(e) - R \cdot \sigma_{Pay}^2.$$

If the employee is risk neutral, $R = 0$. Thus, the optimum is where $C'(e) = b$. This is the worker's effort supply; it tells how responsive effort is to a change in the piece rate. It simply states that the worker sets the marginal cost of effort equal to b , which is the marginal return to effort.

The firm chooses a and b , but has two constraints. First, the choice of b affects the worker's choice of e as just derived. Second, whatever the worker's choice of e turns out to be, denoted e^* , the firm must ensure that the total compensation exceeds $C(e^*)$ or the worker will not accept the job. This means that:

$$\text{Pay} = a + b \cdot e^* = C(e^*).$$

The firm maximizes net revenue minus the worker's pay. Net revenue equals e , so the firm's objective is to maximize $e^* - a - b \cdot e^*$. Solving for a from the previous equation and substituting into this expression yields the firm's simplified maximization problem:

$$\max_b e^* - C(e^*),$$

subject to $C'(e^*) = b$. Note two things at this point. First, base pay a does not affect the choice of e^* by the worker, so it is not part of this expression. Second, this expression is the *net surplus* created by the firm and employee: it is net profit, minus the additional costs to the employee. In effect, the best policy for the firm is the one that maximizes the *total economic value*. This is a formal illustration of one of the themes of this text (see the Summary at the end of the text). The base salary a serves to share this value between the worker and the firm. The first-order condition for the firm is:

$$(1 - C'(e)) \cdot \frac{de}{db} = 0,$$

so b must be chosen so that $C'(e^*) = 1$. Since from above we know that $C'(e^*) = b$, this implies that the optimal $b^* = 1$, which gives 100% of net revenue to the employee. Finally, e^* is determined once we know that $b^* = 1$. The firm then sets a^* so that the worker is just indifferent between this job and the next best alternative:

$$a^* + e^* = C(e^*).$$

It should be easy for you to see that the employee chooses the same level of effort as if he or she owned the firm. When there is no risk aversion, there is no conflict of interest in this model.

In the computer salesperson example in Table 10.1, $C(e) = 2 \cdot e^2$. As a check, try to prove to yourself that in this case, $b^* = 1$, $e^* = 1/4$, and $a^* = -1/8$. (Note: we use calculus here. The third column of Table 10.1 is an *approximation* ΔC rather than the derivative version used below.)

Risk Aversion

Suppose now that $R > 0$ so that the employee is risk averse. Then the worker's value from the job equals $a + b \cdot e - C(e) - R \cdot b \cdot \sigma_e$. This does *not* change the worker's optimal level of e^* , since the risk premium does not vary with e .

What changes is the firm's optimization. It must now compensate the employee for both effort and risk, and risk depends on the level of b . The firm must set pay at least so that:

$$\text{Pay} = a + b \cdot e^* = C(e^*) + R \cdot b \cdot \sigma.$$

Then, the firm's optimization problem becomes:

$$\max_b b \cdot e^* - C(e^*) - R \cdot b \cdot \sigma.$$

The first-order condition for the firm is now:

$$(1 - C'(e)) \cdot \frac{de}{db} - R \cdot \sigma = 0.$$

Putting that together with the employee's first-order condition gives the new optimal commission rate:

$$b^* = 1 - R \cdot \sigma.$$

Several implications follow. First, the commission rate is lower when there is risk aversion. The reason is that stronger incentives (larger b) means larger risk, an additional cost of incentives that the firm must balance against the benefits. Second, the less accurate the performance measure, the lower the commission rate. Third, since the incentive intensity is lower, the effort e^* supplied by the employee will be lower with both effects.

Ratchet Effects

We now show that the Ratchet Effect that arises when a firm makes next year's target a function of this year's performance can be offset by the appropriate multiperiod incentive scheme.⁵

The problem can be analyzed in a two period model. The firm commits to paying a particular commission rate in period 1, but the worker assumes that despite promises, the firm will take advantage to the extent possible next period. (In other words, we are assuming that no effective means of implicit contracting exists in this case.)

The firm can take advantage of the worker only to the extent that the worker can earn at least as much at this firm as elsewhere.

As above in the Appendix, let output $Q_t = e_t$ in each period $t = 1, 2$. The worker has disutility of effort $C(e_t)$ in each period. The worker's cost of effort is unknown to the firm in advance, but the worker's choice of effort in period 1 gives it information on which to base the compensation scheme in period 2.

Since period 2 is the last period, the incentive scheme that the firm chooses then is identical to that for the one-period problem solved for earlier in the Appendix. That is, it will set $b_2 = 1$, and set a such that:

$$a_2 + e_2 - \tilde{C}(e_2) = 0,$$

⁵ See Lazear (1986) and Gibbons (1987).

where we write \tilde{C} to reflect that the firm views C as random, and forms an estimate \hat{C} based on period 1 effort. It is this effect that motivates the worker to slack off in period 1: harder work in period 1 brings higher pay that period, but also reduces a_2 in period 2.

How does the worker behave in period 1? The worker knows that the firm will base its estimate of C on period 1 output, and that the greater output in period 1 will cause the firm to infer that the job was relatively easy (low cost):

$$\frac{\partial \hat{C}(e_2)}{\partial e_1} < 0.$$

In period 2 the firm chooses a_2 such that $a_2 = \hat{C}(e_2) - e_2$. Therefore,

$$\frac{\partial a_2}{\partial e_1} < 0,$$

since \hat{C} is declining in e_1 . The worker's maximization problem in period 2 is:

$$\max_{e_2} a_2 + e_2 - \tilde{C}(e_2),$$

so the worker sets $\tilde{C}'(e_2) = 1$. This is what the firm wants, since doing so maximizes period 2 profits. The problem arises in period 1, since the worker reduces effort, knowing that working hard cuts compensation in period 2. The worker's period 1 maximization problem is:

$$\max_{e_1} a_1 - b_1 e_1 - \tilde{C}(e_1) + a_2(e_1) + b_2 e_2 - \tilde{C}(e_2),$$

subject to $\tilde{C}'(e_2) = 1$. The first order condition is:

$$\tilde{C}'(e_1) = b_1 + \frac{\partial a_2}{\partial e_1} < b_1.$$

The second term after the equals sign is the Ratchet Effect. Effort is lower in period 1 because of the implicit penalty it causes, of lower compensation in period 2.

To maximize profits, the firm must induce the worker to behave efficiently in period 1 as well (this already happens in period 2); that is, induce the worker to set $\tilde{C}'(e_1) = 1$ and $\tilde{C}'(e_2) = 1$. To get the worker to set $\tilde{C}'(e_1) = 1$, we need:

$$b_1 + \frac{\partial a_2}{\partial e_1} = 1, \quad \text{so} \quad b_1 = 1 - \frac{\partial a_2}{\partial e_1} > 1.$$

Thus, the firm must "overpay" performance in period 1 in order to induce efficient effort that period. This reverses the loss in incentives that the worker has from the reduced period 2 base salary implied by high performance in period 1. Thus the commission rate falls over time.

Finally, the firm must set a_1 sufficiently high to attract workers to the firm. Workers are attracted to the firm if:

$$a_1 + b_1 e_1 - \tilde{C}(e_1) + a_2 + b_2 e_2 - \tilde{C}(e_2) \geq 0,$$

given that $a_2 = \hat{C}(e_2) - e_2$. Workers differ in cost of effort, in this model. The higher is a_1 , the more workers (and of lower effort cost, which is analytically equivalent to higher ability) are attracted to the firm.

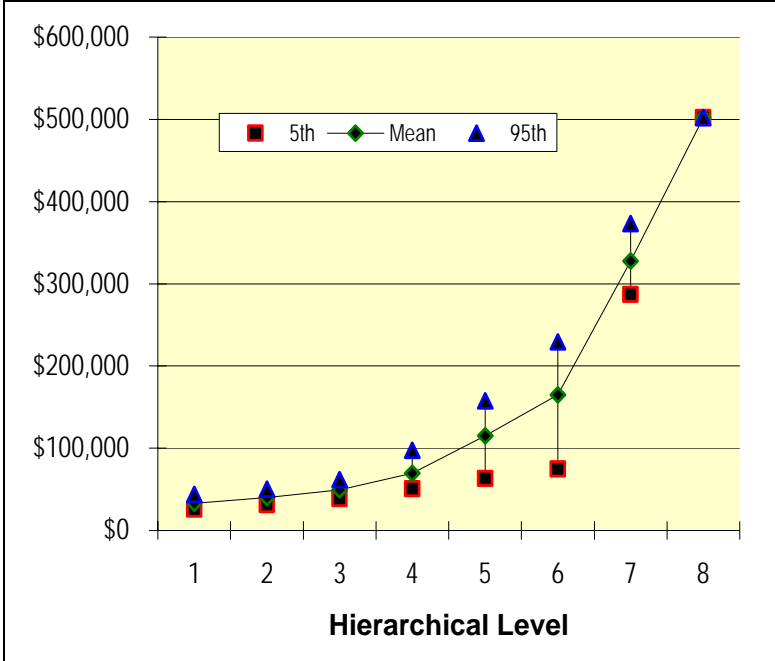
Career-Based Incentives

“In a hierarchy every employee tends to rise to his level of incompetence.” (Laurence Peter)

Introduction

So far, we have largely treated pay for performance as designed with respect to a given job. We now consider another important source of extrinsic motivation: long-term incentives to advance a worker’s career. Most employees experience increases in their earnings over their career through raises and promotions. To the extent that these are based on performance, they are a type of incentive scheme.

Figure 11.1 presents data on salaries of employees at different hierarchical levels in Acme, calculated at a single point in time. In this company, there are eight levels from entry-level management to the CEO. There is more than one kind of job in each level, but jobs within a level have very similar pay, and presumably similar levels of responsibility and skills. The plot shows the average level of salary. It also shows the 5th and 95th percentiles of salary among those in each job level.



Pay by Hierarchical Level at Acme, Incorporated
Figure 11.1

Several observations can be made. First, there are relatively tight bands of pay at lower levels, but the bands widen at higher levels. Second, the difference between average salaries at different levels can be quite large; even at lower levels it appears to be important compared to the width of salary bands. This suggests that the hope of winning a promotion could be an important source of incentives for these managers. Third, there is a remarkable rise in average salary with hierarchical level, and this tendency is especially notable at the top. The rise would be even larger if we included bonuses, stock, and other forms of incentives that tend to be more important for senior executives.

Table 11.1 provides a little evidence on how salary changes when an employee's job changes at this company. For each hierarchical level, the middle columns show the percentage salary increase (adjusted for inflation) for those who do not change level (stayers), those who were demoted, and those who were promoted. The last column also shows the percentage difference between the average salary in that level and in the level below. For example, there is a 48% difference between the CEO's salary (which ignores many other forms of compensation that the CEO probably earned) and salaries of employees in Level 7.

Level	% Real Salary Increase on:			% Difference in Mean Pay Compared to Level Below
	Stay	Demotion	Promotion	
1	-0.5%	-0.7%	—	—
2	-0.4	-0.2	5.1	18
3	0.1	-3.2	5.6	23
4	0.8	0.4	7.4	47
5	-0.1	0.5	8.7	64
6	0.1	—	4.5	40
7	-0.9	—	22.3	107
8	0.0	—	14.8	48
1-8	0.0%	-0.9%	5.8%	

Raise by Type of Job Transition at Acme

Table 11.1

Not surprisingly, demotions are usually associated with a decrease in real salary. Sometimes a demotion actually comes with a raise. However, not much weight should be placed on the demotion data, as they are so very rare in this (and most) companies as to be very unusual cases: out of about 53,000 observations over 20 years, there were only 157 demotions.

Promotions, Demotions, & Lateral Transfers

Demotions are extremely rare, and lateral transfers (movements across jobs within the same hierarchical level) also tend to be less common than promotions. Why do job movements tend to be upward in the hierarchy over one's career? There are several reasons.

First consider demotions. Since they involve a penalty, we might not be used except in cases where the employee has more potential to destroy value than to increase it. Demotions (and firings) may also be more common than is apparent, but are im-

implicit: the firm “eases out” a poor performer, encouraging them to search for a new job. Such an arrangement can benefit both. The firm avoids firing costs and potential litigation, while the employee avoids the stigma of a demotion.

Demotions may be rare because firms may be conservative about placing employees into important positions until their skills are proven. We can view a promotion ladder as similar to project evaluation (Chapter 6). Those who are “accepted” are promoted, while those who are “rejected” are not. Careful upward recruiting makes it more likely that the firm only puts talented managers into top positions.

Demotions might also be rare because of accumulation of human capital. If workers improve skills over time, performance increases. Career movements will tend to be upward, with more promotions and fewer demotions, if it is appropriate to put those with more skills into higher positions.

Lateral transfers generally involve placing an employee into a job involving new skills. This violates the principle of specialization. It can make sense when the employee was initially placed in the wrong job ladder given their talents. We would expect to see such “fixing of mistakes” occurring relatively quickly after an employee is put into the wrong job.

Lateral transfers might also be offered to a talented employee who is in a job ladder with few promotion prospects, to avoid losing the employee.

Finally, lateral transfers might be used to provide a general manager with broad experience in different areas of the business, if their job will involve coordinating across those areas.

More interesting is that if an employee does not get promoted, the average salary increase is *zero* (see the last row). At least in this firm, the only way for an employee to earn increases in salary beyond inflation is to get promoted. Doing so provides about a 5.8% increase in real salary. Moreover, the long-term increase in pay from promotion is much larger than 5.8%, as indicated in the last column. The average differences in salary between levels are much larger than the increase at the time of promotion.

This is for two reasons. First, those who are promoted tend to get further salary increases in their new jobs, while those who are passed over for promotion tend to have slowing (even negative) real salary growth. Second, those who are promoted are then eligible for promotion at the next stage of the hierarchy, and as both the figure and table indicate, promotions become more lucrative as one moves up the hierarchy.

The evidence clearly indicates that an employee’s long-term career prospects, especially in the form of promotions, can be an important source of incentives. The rewards from promotion are large, and promotions are more likely to be given to employees who earn the highest performance ratings. In fact, for middle managers in most firms, where subjective performance evaluations must be used, it is probably the case that promotions are the most important source of extrinsic motivation.

In this chapter we analyze career-based incentives. We first consider promotions. It is shown that the structure of the hierarchy, and of pay across the organization chart, is important for understanding incentives when promotions are a major factor. We then provide a brief discussion of more general career-based incentives. Finally, we consider the use of other long-term incentives within the firm, such as pay based on seniority.

Promotions & Incentives

Should Promotions be used as an Incentive System?

Dual Roles that Sometimes Conflict

You have a position to fill in your management hierarchy, and want to promote someone from the level below. Who should you give the position to? Firms often award a promotion to the best performer in the lower-level job. When they do this, they are using the promotion as a form of incentive. A different view is that the promotion should be given to the employee who has the most potential to perform well in the higher-level job. In other words, there are two important roles to promotions: sorting people into the appropriate jobs given their talents, and incentives.

The sorting and incentive roles will be in conflict if the best performer at one level is not the best performer at the next higher level. For example, in R&D organizations, it is quite common to find that the best researchers are not the best managers. Promoting the best researcher to manage the group may result not only in poor management, but also in reduced research effectiveness.

In cases where this conflict between sorting and incentives is particularly severe, it is probably a good idea to avoid using promotions for incentives. For example, a research group could reward the best scientist with higher pay, more flexibility over choice of research projects, or a higher research budget, but keep the scientist in the same job. It could then identify the researcher who seems to be a good candidate for manager, and promote that person.

In many cases, fortunately, the conflict is not too severe. A manager typically needs to have a good working knowledge of the work that subordinates do. Such knowledge improves the manager's ability to direct, supervise, and evaluate the staff. When there is such a conflict, however, it is not at all obvious that the firm should explicitly try to use the promotion ladder to drive incentives. However, there is an additional complication. The firm may not have complete choice about this issue, since promotion systems often generate incentives automatically.

Intentional or Accidental Incentive System?

The conditions under which a firm can design its promotion system and compensation structure across the hierarchy with only consideration of incentive effects are rather strict. Compensation in different positions is constrained, at least to some extent, by the external labor market. If firm-specific human capital is not very important, employees may easily obtain similar jobs with other employers, constraining the ability of the firm to offer low wages to those who are competing for a promotion.

Second, the ability of the firm to alter its hierarchical structure, and accept variation in ability of those it promotes (in the case of a tournament, as we will see below), are likely to be limited. Because of such considerations, in many cases it may be inappropriate to view the firm's hierarchical and wage structure as *designed* to optimize incentives. Instead, promotions in may often be an "accidental incentive system" that arises on its own and is impossible to avoid, even if the firm does not want to use promotions for incentives.

The logic is simple. Suppose that firms promote the best performers, and have better information about employee abilities than the labor market does. If you are promoted, the labor market should immediately infer that your ability is likely to be better than was previously thought – after all, your employer just signaled that. Because of this, your value should go up on the day you are promoted. In order to retain you, your employer will be forced to offer you a raise on promotion.¹

This means that the hope of earning a promotion generates incentives. Since performance is affected not just by ability but also by effort, employees will try to work harder to earn promotion. In effect, they provide effort to signal their worth to the external market, through earning a promotion.

In this view, the firm does not explicitly design the hierarchy and set pay levels to optimize incentives. Rather, the incentives from promotion are a side effect of the underlying sorting, labor market pressures on pay, and employee signaling. This view may well be appropriate for many firms.

There is some evidence that this view may be valid. Many large Japanese companies have offered lifetime employment to their “core” employees in the post-World War II period (though the practice appears to be gradually eroding now). This meant that such workers rarely, if ever, changed employers mid-career. Such firms should be much less constrained by external labor market forces. In fact, Japanese firms are much less likely to tie pay directly to hierarchical rank, and less likely to offer large raises for promotion, for their core employees. Instead, many such firms give employees two different ranks. One is for hierarchical position, and the other is for compensation. There is no necessary tie between the two, so that pay is not directly associated with position. Such practices might be impossible in economies where employees change firms more often.

Even if this view is correct, however, it does not mean that the theory of promotion-based incentives developed below is not important. On the contrary, if promotions have incentive effects, it is important to understand how they operate and what their implications are. However, it does turn the question around. If promotions are accidental incentives, then the firm should use the theory of promotion-based incentives to identify where the system generates stronger and weaker incentives. This is useful knowledge, since it tells the firm where it needs to focus its efforts on addressing employee motivation.

For the rest of this section, we model promotion-based incentives as though we are unconstrained in the design of the promotion system. Remember, though, that most firms are constrained, so that incentives from promotion may not perfectly match the ideal. When that is the case, the firm may use other forms of incentive, such as bonuses, to adjust incentives appropriately. Or, it might consider altering the structure of the hierarchy to change promotion rates or other parameters.

Promotion Rule: Tournament or Standard?

If your firm has complete flexibility in how to compensate workers at different levels of the hierarchy, how might you do so to optimize incentives? (Most firms will not have such complete flexibility, but this is a useful case to consider first, as it helps us understand how promotion-based incentives work. We return to this issue at the end of this section.)

¹ The firm will also have some incentive to delay your promotion, since you are cheaper before the promotion. This will be offset by the loss of not having you in the best position for your skills for some period of time.

A first question is the rule used to decide whom to promote. Two extreme rules come to mind. In the first, the firm promotes a fixed number of employees (often, only one) who have the best performance. This is a competition or *tournament*. In the second, the firm promotes any (from zero to all) whose performance meets some fixed threshold. This is an absolute *standard*. Thus, the promotion rule is a question about how performance will be evaluated. What are the properties of each approach?

Controlling Structure or Quality

Suppose that the firm has a rigid hierarchy in which job slots are fixed. For example, the firm may be deciding whom to promote to regional manager, and there are a fixed number of regions. Similarly, practically speaking there can only be a single CEO. In such cases, workers are automatically competing against each other for promotion, and the firm must run a tournament if the job is filled by an internal candidate. More generally, the more costly it is for the firm to alter the structure of the hierarchy, the more likely is a tournament to be desirable.

However, a potential problem with a tournament is that quality may be more variable. If the firm commits to promoting the best performer, in bad years it may promote someone who does not have adequate skills for the higher-level job, putting a poor manager in a position of responsibility. In good years, it may fail to promote some employees who are high quality (but not the best), leading to turnover or inadequate use of skilled workers. When such sorting considerations are important, the firm may instead choose to use a standard for deciding promotions, because a standard tends to give better control over the quality of those promoted and passed over for promotion. For this reason, in firms where quality of workers is important (e.g., the best law firms or universities), promotion is more likely to be based on a standard.

Of course, most firms probably use a blend of these approaches, since they face costs from altering the hierarchical structure too much, and also from sorting workers ineffectively. Thus, they may have workers compete against each other for promotion, but when the talent pool is unusual they may relax the rules, promoting more or less than in typical years, or filling a slot from the outside labor market.

Relative v. Absolute Evaluation

Another key difference between tournaments and standards is in performance evaluation. When the firm uses a standard, performance is evaluated for the individual worker. When workers compete, performance is evaluated relative to competitors. This is a special case of a general approach to performance evaluation, relative performance evaluation (RPE). Because a tournament is a good example of RPE, this issue was deferred from Chapter 9 until now. Note, though, that RPE techniques can be used for many types of incentive plans.

Ease & Objectivity of Evaluation

One consideration is that evaluation may be easier in the case of a tournament than a standard. Since a fixed prize is given to a fraction of competitors, the only information that is needed is whom the top performers are. The firm does not need to decide how much better performance was. This is an example of *ordinal* rather than *cardinal* ranking; the order matters, but not the distance between competitors. In many cases, it is very easy to determine who the best performer is, even when the job is complex and involves many intangible dimensions. It is often much harder to determine how much each employee has

performed. (Consider the difference between determining which lump of coal is larger than the other, and determining how much each weighs.) Moreover, because the top performers are often easy for everyone to determine, employees may consider the tournament outcome more objective. These are substantial advantages.

Risk

The evaluations used for tournaments and standards vary in other important ways. One example is risk. Suppose that your firm has two salesmen, one in Denmark and one in Singapore. Employee performance is affected by effort (e), but also by good and bad luck (measurement error). Further suppose that luck is driven by two different factors. The first is local events, such as the state of the Danish or Singaporean economy, and the actions of competitors in the local market (ε). The second is global events, such as worldwide macroeconomic conditions or oil prices (η). Using the subscripts S and D to refer to the two employees, we have:

$$PM_D = e_D + \varepsilon_D + \eta,$$

$$PM_S = e_S + \varepsilon_S + \eta.$$

The term η does not have a subscript, because global economic conditions are assumed to affect both salespeople equally.

If we use a standard for deciding promotions, the Danish employee's performance measure equals PM_D . If we use a tournament, the promotion is decided on the basis of who has better total performance, and the performance measure for the Danish employee is:

$$RPE_D = PM_D - PM_S = e_D + \varepsilon_D + \eta - e_S - \varepsilon_S - \eta = e_D - e_S + \varepsilon_D - \varepsilon_S.$$

(In this example, $RPE_S = -RPE_D$.) This measure is different from the first one in several ways. First, the luck term that was common to both employees, η , dropped out. This reduces measurement error for the Danish employee. However, an additional error term was added, $-\varepsilon_S$. Finally, the Singaporean employee's effort e_S now plays a role. Which performance measure is better?

First, consider the issue of risk. The variance of the two performance measures (assuming that the μ 's and η have zero correlations with each other) is:

$$\sigma_D^2 = \sigma_\varepsilon^2 + \sigma_\eta^2,$$

$$\sigma_{RPE}^2 = 2 \cdot \sigma_\varepsilon^2.$$

RPE can reduce risk *if* measurement error that is common to both employees, η , is more important in determining performance (more variable) than is idiosyncratic risk, ε . In our example, if global factors play a larger role than local factors in determining sales in Denmark and Singapore, relative evaluation can reduce risk, thus improving the incentive plan.

Distortion

A final effect of relative evaluation is that it might distort incentives for workers to cooperate. To see this, let us consider a richer (multitask) model in which each worker can provide two kinds of effort. The first, e^P , increases the worker's performance, while the sec-

ond, e^S , decreases the colleague's performance. This is a simple way to model sabotage. In this case, for workers A and B the absolute performance measures are:

$$PM_A = e_A^P - e_B^S + \varepsilon_A + \eta,$$

$$PM_B = e_B^P - e_A^S + \varepsilon_B + \eta,$$

and the relative evaluation (for worker A) is:

$$\begin{aligned} RPE_A &= PM_A - PM_B \\ &= (e_A^P - e_B^S + \varepsilon_A + \eta) - (e_B^P - e_A^S + \varepsilon_B + \eta). \end{aligned}$$

In the case of RPE, an employee can improve the evaluation in two ways. One is to work harder in the standard sense: increase e^P . The other is to *sabotage*, e^S . By contrast, there is no incentive to sabotage when the evaluation is based on individual performance.

A similar distorted incentive can arise when the employee can engage in influence activities, such as lobbying the supervisor for better evaluations or rewards, or doing actions that the supervisors prefers even when the actions do not improve firm value. To the extent that such actions can improve one's *relative* standing with the supervisor, relative evaluations will increase incentives for employees to engage in such activities.

Similarly, RPE reduces incentives for workers to *cooperate* on the job. This can be a serious downside to RPE, since most jobs are interdependent with those of colleagues to some extent.

In principle, it is possible to use RPE (such as a promotion tournament) and solve these problems, by incorporating measures of cooperation and sabotage into the performance evaluation. For example, a subjective evaluation might be used to encourage cooperation and reduce the temptation to undermine colleagues. Those who were insufficient "team players" would not be promoted. Clearly, firms do take these issues into account in such circumstances. However, it is likely that such "fixes" will be imperfect, since cooperation and sabotage are often hard to detect and quantify. Thus, one drawback to relative evaluation is that it likely to be less effective when workers have more interdependent work. For example, contests may work very well for salespeople who work in different geographical areas, or for assembly line workers who jobs are relatively independent of each other. By contrast, they are almost certainly a poor idea for members of the same workgroup.

Another approach might be to use a broader performance measure, such as $PM_A + PM_B$. In this case, the worker has incentives to cooperate, and not sabotage, since the reward also depends on the colleague's performance. Indeed, many firms use some form of group (or business unit, or divisional) rewards, partly for this reason. Of course, such a measure does *not* filter out common measurement error the way that RPE does; it generally increases measurement error. More concretely, when you are rewarded on the basis of work of a group, you are subject to risk because you cannot control what your colleagues do, and their luck. This is what we mentioned before in Chapter 9: the broader measure can reduce distortion, but also tends to increase risk.

When such distortions cannot be removed from the evaluation, but cooperation is important because of interdependence between jobs, the proper response for the firm is to

change the incentive structure. If promotions will still be used to drive incentives, then the rewards should be decreased. While this will decrease overall incentives, it will also decrease sabotage and increase cooperation. This is a simple application of the idea that incentives should be muted when the evaluation distorts multitask incentives.

An alternative would be for the firm to move toward group-based rewards. The performance measure and reward might both be based on group results. Another alternative would be to base promotion on absolute rather than relative performance, though evaluations may be more difficult to conduct, and the issues raised in Chapter 9 must be addressed.

In addition, where cooperation is important, the firm should not have such workers compete for rewards. This suggests that the composition of the group that is competing can be important. We will have more to say about this below.

Finally, the firm should consider the importance of cooperation or potential for sabotage when recruiting. People differ in the degree to which they cooperate or feel “shame” for sabotage on the job. Where jobs are more interdependent, it is important to try to recruit individuals with more cooperative personalities, who prefer working in groups.

Forced Curves for Performance Appraisals

Recall that a frequent complaint about performance appraisals is that managers tend to give many employees the same rating. An additional concern is that some managers may be lenient, while others may be strict. That is a form of luck that increases the riskiness of the pay system.

To address these concerns, some firms impose various forms of *forced curves* on the distribution of ratings. (Some universities use similar systems for grading.) Some require specific percentages for each rating. Others require that the average be fixed, though the distribution around the average can vary. These methods usually involve some element of RPE, since giving one worker a higher rating forces the manager to give someone else a lower rating.

Such approaches clearly can lead to greater dispersion in ratings across employees, more frequent use of lower ratings, etc. They also can reduce the risk to the employee from evaluators being too lenient or easy. To see this, consider that the effect of the evaluator’s toughness or leniency is to lower or raise ratings for *all* subordinates of that evaluator. This is a form of common measurement error (η) as in our discussion of RPE. RPE filters out this effect.

While forced curves may be tempting because of these advantages, they have their own problems. Just as with any kind of RPE, they do not encourage cooperation, and may encourage sabotage. They impose their own form of risk, since employees may be evaluated against a very good group of colleagues, lowering their rating. (This effect is less likely if the rating is not done when the group size is too small.) And, it may not always be optimal to give employees clear feedback about performance (see below). Thus, many firms do not use forced curves, and some firms seem to switch back and forth between the two approaches, since neither is perfect.

General Electric (GE) is the most famous example of a firm that has successfully used a forced curve; they call it TopGrading. However, GE exerts a great deal of effort, and uses other policies, to make the program effective. For example, they train

managers in how to conduct careful appraisals, and monitor and document appraisals carefully. This reduces the potential legal liability from disgruntled employees. Also important is that GE has a very aggressive corporate culture (see our discussion of “hawks and doves” below). It is well understood that receiving a poor rating for two years in a row puts an employee at high risk of being fired. Finally, GE has a very large, complex organization, so they are often able to reassign poor performers to positions for which they have a good fit, reducing firing costs (which would include litigation).

How Do Promotions Generate Incentives?

Prize Structure & Incentives

As described in the previous two chapters, incentives depend on two things: how effort (the kind that increases productivity, not sabotage) affects the evaluation, and how the evaluation is tied to rewards:

$$\frac{\Delta \text{Pay}}{\Delta(\text{effort})} = \frac{\Delta \text{Pay}}{\Delta \text{PM}} \cdot \frac{\Delta \text{PM}}{\Delta(\text{effort})}.$$

For a promotion, the first term on the right is a constant (the lump sum reward, raise on promotion, etc.), because the employee either wins the discrete prize, or does not. This looks like Figure 10.3 in the last chapter.

The second term is a bit more complicated. In the case of an absolute standard, the threshold for winning the prize is fixed. In the case of a tournament, the contestant must beat a certain number of competitors. Since their performance is unknown *ex ante*, the threshold for winning the prize is uncertain – it is a moving target. Otherwise, tournaments and standard can be analyzed in the same way.

To see how these ideas play out for promotions, let us write down the condition for the employee to receive the reward. Suppose that an employee is paid a base salary of W_1 , and is given a raise to a salary of W_2 if promoted. Define the raise on promotion, the prize, as $\Delta W = W_2 - W_1$. Denoting probabilities by “ $\text{pr}(\bullet)$,”

$$\begin{aligned} \text{Pay} &= \text{pr}(\text{not promoted}) \cdot W_1 + \text{pr}(\text{promoted}) \cdot W_2 \\ &= W_1 + \text{pr}(\text{promoted}) \Delta W, \end{aligned}$$

since $\text{pr}(\text{not promoted}) = 1 - \text{pr}(\text{promoted})$. Therefore,

$$\frac{\Delta \text{Pay}}{\Delta(\text{effort})} = \Delta W \cdot \frac{\Delta \text{pr}(\text{promoted})}{\Delta(\text{effort})}.$$

Viewing the performance measure as binary (performance is good enough to win promotion, or not), then the first term is $\Delta \text{Pay} / \Delta \text{PM}$, and the second is $\Delta \text{PM} / \Delta(\text{effort})$. These are the same two terms that always affect incentives.

Level of Salary

One important result is immediate: what matters to promotion incentives is *not* the base level of pay W_1 or W_2 ; it is the change in pay ΔW . This is an application of the point made

in Chapter 10 that the shape of the pay-performance relationship drives incentives. Graphically, as in Figure 10.1 in Chapter 10, the level of pay shifts the pay-performance relationship up or down, to raise or lower the overall level of expected pay. For example, the expected salary in the equations above can be adjusted by changing W_1 , without changing ΔW (thus change W_2 by the same amount as W_1).

This is an important general point: the firm has two different instruments for two different purposes in designing the pay package. The base salary generally is used to make sure that the firm is able to recruit and retain the appropriate quality of employees. It responds to labor market supply and demand for skills. It also adjusts for the overall level of effort and risk that the incentive system implies.

In many cases the firm is much less constrained about how to vary pay with performance, or even over time (as we will see when we discuss seniority-based pay below). It can use this flexibility to drive appropriate incentives, separately from the question of recruitment and retention of the workforce.

Prize from Promotion

The most important point in the last equation above is that promotion incentives are larger, the larger is the raise (and other rewards) earned upon promotion. In any contest, the larger is the prize, the greater is the effort that will be provided. There are many examples of this in sports. Teams tend to exert much more effort in the most important games, where there is more at stake. They tend to slack off in the less important games. The same will tend to be true for employee effort in promotion ladders.

We can apply this idea to the data in Table 11.1 (or the plots in Figure 11.1). The raise and longer-term rewards for promotion tend to be larger in higher levels of the hierarchy. This suggests that incentives will tend to be stronger at higher levels. (However, this is not necessarily the case, since we have not yet analyzed the first term in the last equation.)

To a first approximation, the raise on promotion is a good estimate of the reward from the promotion. It is the immediate consequence of the promotion, and is guaranteed once the promotion is earned. Thus, it is the best starting place for analyzing the incentive structure in your firm's hierarchy.

A more complex estimate of the prize from promotion recognizes that an additional benefit from promotion is that it makes the employee eligible for additional rewards. These usually entail higher raises in the new job (as suggested in the last column of Table 11.1), and the ability to compete for the next promotion. These will also have some value to the employee, though they will be somewhat discounted compared to the immediate raise on promotion, since they require further effort and are not guaranteed.

This implies that compensation at higher levels in the job ladder – or at later stages in a sequential contest – affect incentives at *all lower levels*. The larger is the difference in pay between Level 5 and Level 6, the greater should be motivation at *all* levels from 1 to 5. (Of course, the effect might be small for levels further from Level 6, since the probability of getting all the way from a lower level to Level 6 might be quite small.) In other words, the structure of pay across levels in the organization chart can have important incentive effects for lower levels.

One implication of this idea is that it is more important to give larger prizes at higher levels. The reason is that these prizes provide incentives for more employees, since there are more levels below them. This is one explanation for why salaries rise rapidly with hierar-

chical level, as seen in Figure 11.1, and why executive pay levels are often quite high. The high executive pay levels may serve a purpose beyond compensation for the executives: they may motivate those below them to strive to become an executive as well.

Promotion Probability & Incentives

The first term in the last equation above is how effort affects the probability of winning promotion. In an abstract sense, the performance measure when rewards are lump sum is binary: the employee's performance is either high enough to earn the reward, or it is not. Thus, this term is the same as $\Delta PM/\Delta e$.

Formal analysis of the effect of promotion probability is technical. The intuition, however, is straightforward, and applies to both tournaments and standards. Consider two extremes. In one case, the promotion is guaranteed and the probability equals one. In the other case, the promotion is impossible and the probability equals zero. In both of these extremes, there is no point in exerting effort – more or less effort will make no difference to the outcome. Incentives will be zero. Clearly, the only way that there will be an incentive is if we have an intermediate case, where the promotion is possible, but neither too hard to achieve, nor too easy.²

Intuitively, incentives are driven by the effect of incremental effort on the chance of winning, given that luck will play a role as well. High levels of good or bad luck are relatively less likely to occur. If the chance of winning is low, the odds that incremental effort will make any difference to the outcome are very small, since large good luck is also required to win.

Perhaps less intuitively, the same applies when the promotion probability is very high. In that case, the employee has incentives to slack off, because incremental reductions in effort are unlikely to cause the employee to lose the promotion. They would have to be combined with large bad luck. Thus, for example, sports teams tend to put in their second string when they are far ahead in a game.

In real organizations, the probability of promotion to the next hierarchical level tends to be much less than $\frac{1}{2}$, especially for higher levels. Thus, in practice our result is that for a given reward, promotion incentives are weaker, the lower the promotion rate.

Luck

As with all incentive schemes, luck plays a role for promotion incentives. In our simpler incentive systems considered in Chapter 10, the effect of luck was to increase the risk premium required, and consequently to reduce the strength of the optimal incentive intensity. Luck plays those roles here, as it does in any incentive system. However, luck also plays a different role here – it also reduces incentives.

Suppose that you are playing a tennis match. On some days, the wind is very calm, but on other days it is very strong. When the wind is calm, you have greater control over your shots. When the wind is very strong, you have less control. This means that on windy days, the outcome is less likely to be determined by which contestant played better, and more likely to be determined by good or bad luck. The same applies in promotion settings, for either tournaments or standards.

² Several studies in accounting and psychology have concluded that incentives are strongest when the employee has a 50% chance of meeting the target for the reward. Economic modeling helps explain why this may be so.

Because of this effect, risk reduces the effect of effort on the outcome; it reduces $\Delta pr(win)/\Delta e$.³ This is shown formally in the Appendix. And of course, this in effect means that $\Delta PM/\Delta e$ is lower when measurement error is higher. By now, you know what that means – incentives will be low.

What are the implications of luck for optimal incentives? If measurement error is higher, incentives will be lower, unless the firm increases the size of the prize. One possible response is to incur costs to measure performance more carefully. Another is to change the prize structure. Optimal prize structures tend to be more skewed (toward larger rewards for better performers) when luck plays more of a role.

The point that luck matters and affects the optimal salary structure has implications for how compensation varies by industry or country. Consider, for example, the difference between the United States and Japan. Japanese wage structures are more compressed than those in America. Top executives in Japan get paid less relative to production workers than do their counterparts in the U.S. Some have interpreted this as extravagance on the part of American firms and their top management.

An alternative explanation might be that the American business environment is riskier than in Japan. Promotions in the U.S. may depend more on random factors than in Japan. For example, promotions in Japan come later in the career than in the United States. By the time a Japanese manager has made CEO, the firm has very clear signals about productivity. It is unlikely that measurement error will play an important role in determining the promotion. If promotions in the U.S. are more heavily influenced by luck, then American firms might offset the effort-reducing effects of luck by choosing larger salary spreads.

A similar point can be made with respect to new versus old industries. If luck is more important in affecting an individual's performance in newer industries, then firms in newer industries may tend to have higher variance in their salary structures compared to firms in older, more established industries.

Summing Up

We see that in many ways, tournaments and standards have the same implications for incentive plan design. The most important question is the size of the prize for promotion. A good starting estimate for this is the raise on promotion. A better estimate would take into account the added value to the employee of better career prospects that come with the promotion.

The next important factor is the promotion rate. Higher promotion rates (as long as they are not too high, which they do not tend to be inside firms) generally imply stronger incentives for the same reward.

Putting these two ideas together, if the firm has enough flexibility in compensation across levels to set pay levels solely to generate optimal incentives, then the raise on promotion should be higher when the promotion rate is lower, and vice versa. It should also tend to be larger in higher levels of the hierarchy.

³ This statement is only true if promotion rates are not too close to zero or one. The closer they are to either extreme, the more likely is the opposite to be true. That is because incremental effort only makes a difference in those cases, when combined with very good or bad luck.

Finally, greater luck or performance measurement error not only implies a higher risk premium for the employee, but also reduces incentives in promotion systems.

Tournaments and standards do differ in some important ways. Tournaments are necessary when job slots are fixed; they give better control over the number of employees promoted. Standards are more useful when the quality of employees promoted is more important. Second, tournaments are a form of RPE, so they usually distort incentives away from cooperation, and toward sabotage of co-workers. Standards do not have these negative side effects.

Advanced Issues

Heterogeneity of Employees

The theory above assumed that all employees who are hoping to earn promotion are identical. What happens if they vary in ability or some other important dimension? It turns out that mixing employees of different types can cause a couple of problems for promotion-based incentives.

Variation in Ability

First, if employees vary in their ability, then they vary in the likelihood that they will win promotion. Those who have the highest ability may have a very good chance of winning promotion. As we saw above, they will then tend to have lower incentives. Similarly, those who are performing poorly may have little chance of promotion, and may also slack off. Those employees who believe that their performance is right at the margin between winning and losing promotion will have the greatest incentives.

Consider, for example, the pay structure in Figure 10.3 in Chapter 10. The threshold for winning promotion is T (which could be fixed, as in a standard, or variable, as in a tournament). If the employee believes that his performance is near T , incentives are very high, because the incremental effect of effort on the expected reward is very high. If the employee believes that his performance is too far above or below T , incentives may be very weak. In other words, promotion based incentives tend to not work well when the workforce is heterogeneous.

What can be done about this? If the firm uses a standard for promotion decisions, then it could simply vary the standard T for different employees, imposing a higher standard for better employees, and vice versa. Unfortunately, this would have negative effects on sorting – it would make it easier for low ability workers to get promoted, and harder for high ability workers to get promoted.

In such cases, the firm must expend some resources to pre-sort employees, to reduce the variation in their abilities. In sports contests, athletes are separated into different leagues, so that those with similar abilities compete against each other. Inside firms, the more the firm has already promoted employees, the more homogeneous is the group of remaining employees eligible for promotions. Thus, this concern is likely to be more important at lower levels than at higher levels.

Might workers self-select in the appropriate way, as in Chapter 2? Unfortunately, generally the answer will be no. Low ability workers will tend to have an incentive to try to get access to promotion systems (or athletic leagues) that are designed for high ability workers. The reason is that the base salary will be higher in the higher ability system.

An interesting implication of heterogeneity of performance and promotion-based incentives has to do with subjective performance appraisals. Of course, promotions are usually based on subjective assessments of worker performance by the supervisor. Imagine that you are a supervisor, and are trying to decide what feedback to give your subordinates. The subordinates are hoping to win promotion, but the promotion decision will be made in the future. What will you tell them?

If they are right at the margin, giving them this feedback can only strengthen their incentives. The interesting question is what you tell subordinates who are performing above the threshold for promotion. If the goal is to maximize motivation, giving accurate, clear feedback to employees who are “frontrunners” or “underdogs” with respect to promotion may well reduce their incentives.

Instead, giving somewhat negative (or less favorable than deserved) feedback to a good performer may alter their perceptions about how they are performing, moving them closer to the threshold and increasing their incentives. Similarly, giving more positive feedback than deserved to a poor performer may increase their motivation, since they will be less likely to give up on the hope of promotion.

This implies that, when promotion incentives are large and evaluations are subjective, supervisors have some incentive to distort the feedback that they give to their employees. They may be especially reluctant to give negative feedback to poor performers, out of a concern that this would demotivate such employees. And even if they do not distort feedback, they may be vague and uninformative in the feedback that they do give, to at least make it harder for frontrunners and underdogs to figure out where they stand. These ideas may help explain several facts about subjective performance appraisals: distributions tend to be concentrated and upward biased; supervisors tend to be reluctant to give explicit feedback to subordinates; and subordinates often do not trust that their performance ratings were given fairly.

Variation in Personality

We have already discussed the problem of sabotage and lack of cooperation that tournaments may induce. Now suppose that employees vary in their personality: some are more aggressive, or less likely to cooperate, while others tend to have a personal taste for more cooperation or teamwork in the workplace. If these two types of employees are mixed in a workplace where rewards are competitively awarded, problems can arise.

The following example illustrates the idea. Consider four workers, two who are *hawks* H₁, H₂; and two who are *doves*, D₁, D₂. Hawks are aggressive, while doves are cooperative. There are a number of ways to arrange them together into production teams. Table 11.2 lists some possibilities.

Configuration	Group			
	1	2	3	4
A	H1, H2, D1, D2			
B	H1, D1, D2		H2	
C	H1, H2		D1, D2	
D	H1, D1		H2, D2	
E	H1	H2	D1	D2

Group Assignment by Employee Personality Type
Table 11.2

The two polar cases are A and E. In configuration A, all work together. In configuration E, all work separately. Configuration E loses all advantages of worker interaction. If potential synergies from combining different worker types are large, then the firm will want to consider a structure like A. If so, then as we will see a tournament-type reward system would be a mistake.

The incentive problem that arises from mixing the different types is that the aggressive hawks will tend to cooperate even less, and sabotage even more, when they are paired with hawks. The intuition is that the reward is based on relative performance, which motivates both to avoid cooperation. But the hawk knows that the dove is likely to provide more cooperation and less sabotage than the hawk is. This means that the hawk will tend to damage the dove's performance more, and will be in a better position to win.

What this also means is that their relative performances will now differ. As described above, when employees vary in their relative performance, tournament incentives tend to be weaker. Thus, differences in personalities may make differences in incentives even worse when rewards are given competitively.

This effect would not occur if hawks were paired with hawks, and doves were paired with doves, since they would then be competing against someone with the same incentives and personality. Unfortunately, hawks will have some incentive to compete against doves, so that self-selection will not generally occur. Once more we see a benefit from sorting workers so that similar employees are competing against each other, especially when the firm uses a tournament.

This kind of effect gives one reason why firms vary in corporate culture. Firms with more competition in their reward structures should optimally sort for more aggressive employees (and expect less cooperation), and vice versa. We see a link between several issues discussed in this book: the degree of interdependence in jobs is important for deciding whether or not to have employees compete for rewards. This, in turn, affects the type of employee the firm should recruit, and the corporate culture that results.

Incentives for Losers

One problem with any promotion-based incentive system is that it motivates only to the extent that the employee feels that there is enough chance that a promotion will be earned. Those employees who are not in the running, such as those who have been passed over in previous rounds, will not be motivated. This decline in extrinsic motivation for those who have been in a job for a long time, and who do not have prospects for further advancement, is one reason for the common complaint that such workers are "deadwood" that are relatively unproductive.⁴

There are several things that the firm can do for such workers. One is to encourage them to leave the firm or to find a more suitable position in the same firm (for example, see the sidebar about General Electric above). Another is to provide incentives in some other form. For example, workers who do not have strong prospects for promotion may be offered stronger pay for performance in the form of annual bonuses. Finally, the supervisor may be able to increase intrinsic motivation by offering the employee the opportunity to perform new tasks and learn new skills.

⁴ Another reason for the phenomenon, sometimes called the Peter Principle (that employees are promoted to their level of incompetence) is that the average ability of workers declines, the longer that they stay in the same position. This is because the firm is continuously selecting high ability workers for promotion.

Outside Hiring

Of course, firms often hire employees from the outside, and not just at entry-level positions. What is the effect of outside hiring on promotion incentives?

The first effect is that outside hiring tends to lower incentives for internal candidates. This is simply because it reduces the likelihood that the incumbent employee will earn promotion, which generally reduces incentives. In addition, recall the firm's conflict between the desire to promote the best candidate for the higher-level position, and the desire to promote the best performer. Once employees have provided the effort, the firm may be tempted to promote on potential rather than past performance, even if it had claimed that it would offer the promotion to the best performer (this is another example of the Hold-Up Problem). Of course, if employees foresee this problem, it reduces their incentives in the first place. Hiring from the outside only makes that concern worse.

Therefore, an important cost of outside hiring that firms should consider is that it may reduce motivation for existing employees. Most firms do tend to prefer to fill vacancies with internal candidates, and this is one explanation (another is firm-specific human capital).

Outside hiring does have advantages. Recall that the benefit of using an absolute standard to decide promotions is that the firm has better control over the quality of employees in the higher-level position. The benefit of a tournament is that performance evaluation is easier, since it is RPE and only ordinal ranking matters. By using outside hiring combined with a tournament, the firm can achieve both of these advantages simultaneously. Promotions can be based on relative rankings by using a tournament. However, in years when the quality of the pool of internal applicants is too low, the firm can decide to hire externally instead. The outside option will reduce incentives somewhat, but since the firm only resorts to it occasionally, such an effect should be small. And, the firm is able to protect itself against promoting poor quality employees into higher-level positions. Furthermore, competing against potential outside candidates may reduce the incentive to sabotage internal candidates.

Turnover

Turnover is quite important for an effective promotion-based incentive system. The higher the turnover, the more open job slots will be made available. This increases promotion incentives. Thus, if a firm emphasizes the use of promotions for incentives, a healthy degree of turnover can be very helpful. Conversely, where turnover is low, promotion-based incentives are not likely to function well. Consider a point in the organization chart where the hierarchy narrows rapidly: there are many fewer positions in the higher level than in the lower level. In that case, promotion rates will be very low, and generate poor incentives unless the reward from promotion is very high. The firm has several options. It can try to restructure the hierarchy to open up promotion rates over the long term. In the short term, it can try to promote or terminate some employees at the higher level.

Evidence

It is difficult to observe the effects of promotion-based incentives inside firms, because in such cases individual measures of output are not available. Most of the empirical evidence on the theory of tournaments and standards comes from other sources. For example, several studies have examined whether larger prizes lead to better performance in sporting contests like golf. These results tend to strongly support the predictions, suggesting that professional athletes respond to incentives. Indeed, many professional sports teams

make use of elaborate incentive schemes for their players, indicating that they believe their employees can be motivated in this way.

Another series of tests has conducted laboratory experiments to see if participants (usually college students) behave in ways that the theory predicts. These studies tend to find that larger prizes induce greater effort, greater risk induces less effort, and a lower probability of winning induces less effort. All of these are as predicted. In addition, the amount of effort put forth by students usually converges quickly to the precise amount predicted by the theory. A puzzle, however, is that the variance in output is higher with tournaments, but not with standards, than predicted by the theory. Evidence seems to be accumulating to suggest that different people react to competitions differently, which may explain such a result. For example, when offered a choice of a tournament or a standard, men choose tournaments relatively more than women do.

Some studies have analyzed whether firm evaluation practices and compensation structures vary in ways that are predicted by the theory described above. The evidence on whether firms are more likely to use RPE or an absolute standard for deciding whom to promote is quite mixed. It is probably safe to say that firms vary in their practices (and even vary in which they emphasize for individual jobs in the same firm), depending on the importance of fixing the hierarchical structure or controlling the quality of promoted employees.

Other studies have examined implications of the theory for pay structures (e.g., is the raise on promotion larger, if odds of receiving the promotion are smaller?). Such studies are generally consistent with the idea that firms design their compensation structures in accordance with the theory. Unfortunately, there are other plausible explanations for the findings in most of these studies. For example, if the promotion rate is very low, then the difference in talent between those promoted and those passed over should be larger. This means that the raise on promotion, should be larger. This explanation is based solely on sorting, and has nothing to do with incentives. Thus, it is very hard to say for sure whether or not firms explicitly design their pay structures across the hierarchy to optimize incentives.

Tournaments for Economics Professors?

One study analyzed whether tournaments or standards is a better description of compensation policies in university economics departments. This is a good setting for testing the theory, because university departments are hierarchical, have up-or-out promotion systems, and some data on employee productivity (quantity and quality of published research) is publicly available.

One finding was that junior professors tend to be more productive if their department has a larger gap in pay between assistant and associate professor ranks. This is consistent with the idea that the pay gap generates incentives.

Another interesting finding is that the highest-ranking economics departments do not seem to run either clear-cut tournaments or standards. Rather, they seem to resort to outside hiring when the quality of internal candidates is too low. Because professors have little in the way of firm-specific human capital, they compete with each other in the broader academic labor market, rather than only internally, and move frequently between universities.

Source: Coupé, Smeets & Warzynski (2005)

Career Concerns

In an active labor market, employees may be motivated partly because good performance can lead to better employment opportunities outside the firm. This kind of incentive is often called “career concerns.”⁵ It is most important in industries where human capital is more general, and where other potential employers can evaluate performance. Good examples include scientists (whose research is published), professional athletes, and top executives of publicly traded firms. To some extent, career concerns are likely to operate in all industries.

Career concerns raise several interesting implications. For example, workers should tend to be more highly motivated earlier in their careers. This is because they are trying to establish their reputation with the labor market. As the career progresses, more is known about the worker’s capabilities, and there is less possibility to affect one’s market value.

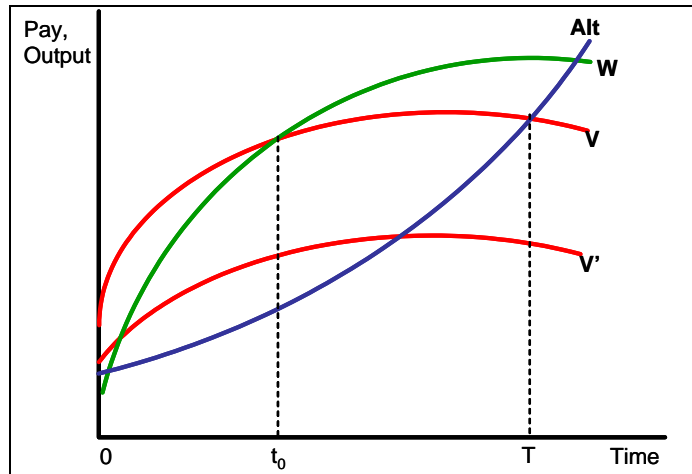
Another implication is that younger workers should tend to be more willing to take risks, such as trying unusual jobs with uncertain prospects. The reason is that if the risks do not work out, they have more time to recover from the bad outcome. Thus, there is a natural tendency for people to become more conservative as they progress in their career.

Seniority Pay & Incentives

The data in Figure 11.1 and Table 11.1 suggest that increases in earnings come not just from promotion, but also from raises in salary over time. Of course, raises can be a form of incentive if they are tied to performance evaluations. In many firms, however, seniority plays a substantial role in salary increases. At first glance, it would seem that tying salary increases to seniority would not generate incentives, since the raise is not tied directly to performance. In this section, we provide a brief argument for why seniority-based pay can also be used as a long-term incentive.

To make things simple, suppose that workers in the firm can choose to work at either a high or low level of effort. An employee who works at a high level of effort will produce output over the career given by the curve V in Figure 11.2. As the worker gains experience, output rises up to some point, after which it may decline. Alternatively, the worker can choose a low level of effort, producing V' , which is lower than V . Assume that high effort would be the efficient choice. In other words, the difference in productivity between V and V' would exceed the marginal disutility of working at the high level of effort instead of the low level. Therefore, the firm and the employee would like to structure a contract in which the high level of effort is provided.

⁵ Career concerns should motivate both investments in human capital, and greater employee effort. We focus on the latter, but the link to human capital should be clear to you after reading Chapter 3.



Productivity & Pay Over the Career

Figure 11.2

In addition, to illustrate the point easily, consider a very simple performance evaluation system. If the worker shirks (produces low output) in any period, there is some probability that the firm will detect the shirking, in which case the worker is penalized (e.g., fired).

The path Alt is the value of the worker's alternative use of time. As the worker nears retirement, the best alternative is likely to be leisure. Thus, T is the date that a worker should retire. Put in other terms, a self-employed worker producing V would voluntarily retire at T .

The path W is a possible wage profile offered to the employee over time. It is drawn such that the discounted present value of W from zero to T is exactly equal to the discounted present value of V over the same period. A worker who is paid V each period would receive the exact value of output. A worker who is paid W would receive less than productivity until time t_0 , and more than productivity thereafter. Over the entire career, the compensation would add up to the value of output, in present value.

Why bother distorting the wage profile in this fashion? The reason is that incentives are not the same along each profile. Incentives are greater along W than along V . In fact, a worker paid exactly V would end up producing less than V . The reasoning is simple.

Suppose that the firm pays wages equal to V at every stage of the career. Consider the worker's incentive on the last day before retirement at T . If the worker shirks, there is nothing to lose, since the worker is not employed tomorrow anyway. Similarly, at any time near retirement, the worker has little incentive, because the loss from being fired, the difference between V and Alt , is very low at this point.

More generally, as long as the outside alternatives for the worker are not very different from those at the present firm, similar logic applies at all points in time. Thus, if human capital is largely general, and the costs of finding a new job are small, the worker has great temptation to shirk, as there is little to lose.

One solution to this would be to increase the penalty from shirking, perhaps by having the employee pay an up-front fee, which was returned at the end of the career if the employee did not shirk. Of course, the scheme W in Figure 11.2 is just a more elaborate version of that, where the firm agrees to tie raises to seniority. At all points in time after t_0 , the present discounted value of W is greater than that of V . In fact, the same is true for *all* points in

time, since W starts below V , and the total discounted present values over the entire period are equal. Thus, the worker has a stronger incentive to provide high effort, the more that pay is deferred.

Note that workers should prefer the deferred pay scheme W to the scheme V (absent the considerations described next). Since the scheme V would generate some shirking, the present value of pay that would have to be reduced. The worker (and the firm) could be made better off if the wage scheme could motivate no shirking.

Practical Considerations

Seniority-Based Pay as an Implicit Contract

Pay need not literally be tied to seniority in order to have this incentive effect; it just needs to be deferred. However, an advantage of formally tying pay to seniority is that it is one way in which the firm can try to pre-commit to the deferred pay incentive scheme. Deferred pay (including the examples we discussed in Chapters 2-3) involves a promise from the firm to the employee. In Figure 11.2, if the pay scheme is W , the firm may be tempted to renege on its promise, lowering pay or firing overpaid workers with high seniority.

Because it involves implicit contracting, seniority-based schemes are more likely to work well for companies that have established good reputations as fair employers (e.g., older, stable firms compared to new startups). Such firms are also more likely to implement policies that signal to workers that they take a long-term interest in their career success, and intend to treat them fairly. See Chapter 15 for some discussion of these issues.

Deferred pay schemes also impose risk on the worker because there is some chance that the firm will default on its obligations to workers if financial performance is poor. Therefore, seniority schemes are more likely to be observed in companies with less risky business environments, such as those in growing and stable industries.

The Worker as Lender

In all deferred pay schemes, the worker in effect is a lender to the employer. If the deferred reward is fixed in advance, the worker acts like a bondholder for the firm. If the deferred reward is variable, such as stock or profit sharing, the worker acts like an equity holder.

Because workers tend to be risk averse, it would not seem to make sense to give employees implicit equity in their firms. However, there is an additional incentive effect that might justify deferred pay correlated with future firm value. In some cases, employees can take actions today that increase firm value in the future. Consider a law firm. The way in which the lawyer conducts business today may have important effects on the reputation of the firm in the future. In addition, the lawyer may be able to bring in new clients today who will generate future work for the company. By tying deferred rewards to future firm profits, the law firm can motivate lawyers to take into account the long-term effects of their actions on firm value.

This is just an example of a distortion in the face of multitask incentives. Long-term effects on firm value are usually very hard to measure today. For this reason, typical incentive plans distort employee incentives toward the short term. Implicit, deferred equity can mitigate this problem. In fact, in companies where employees do have greater impact of future firm value, such as professional partnerships, equity type schemes are quite common.

Mandatory Retirement

One problem with paying W to workers is that they have incentives to stay at the firm after retirement, since at T , $W > V$. This is inefficient, because the value of the worker's leisure Al_t exceeds the value to the firm of the worker's productivity. The firm would make more profit by paying the worker to quit! Of course, that would destroy the pay schedule W .

One way to look at this is that the worker could promise to retire at date T . However, he has incentives to renege on such a promise. The firm could try firing the worker at T , but most economies impose strict regulations on firing workers who are near retirement (probably as a way to encourage firms to not renege on their pension commitments).

Imposing mandatory retirement at date T easily solves this problem. In fact, many firms used to have mandatory retirement policies. However, in the U.S. and some other countries, mandatory retirement is now illegal. Instead, the firm must resort to other incentives, through structuring of pensions, etc. in order to encourage efficient retirement.

Summary

Career considerations are a major source of extrinsic motivation for many, perhaps most, workers. Most increases in earnings over the career come from new jobs, promotions, and gradual raises in salary over time. For white-collar workers in particular, promotions are often the most important form of pay for performance. For all workers, the possibility of raising their reputation and value to the labor market outside their employer can be a large motivator, especially early in the career.

Why do promotions play such a key role in incentives? One answer is that performance evaluations are quite difficult, and inevitably subjective, in white-collar jobs. For this reason, short-term incentive plans may be far from perfect, and firms may defer evaluations until job assignment decisions must be made. Another answer is that promotions generate incentives automatically, because of the signal that the promotion sends to the labor market, which raises the promoted employee's market value. Thus, promotion incentives may be forced on the firm, even if they would rather separate job assignments from incentives.

More generally, this suggests an intuitive hierarchy of pay for performance schemes. When analyzing incentives for a given employee, the firm should first ask whether or not promotion-based incentives are large or small. To do so, it can use the models of tournaments and standards described in this chapter. Also keep in mind the value of additional promotions, if they exist, further up the job ladder.

Where promotion incentives are weak, the firm should consider alternative forms of pay for performance. One possibility is demotions or the threat of termination. However, demotions will generally be rare; as described above, for several reasons it is natural that most employees tend to move upward in the hierarchy over their careers. The threat of being fired can be an important incentive, but it comes at a cost compared to other forms of incentives: the firm and employee incur search and recruitment costs. Thus, if other forms of incentive are available, and if these turnover costs are high (as they tend to be in most of Europe), then the firm may not use frequent termination as a form of incentive. In fact, most firms appear to use terminations only in extreme cases, so this does not tend to be a major source of incentives.

That suggests that the firm would then turn to bonus plans and raises to provide further incentives, when promotion-based incentives are weak. For example, when the hierarchy narrows so that promotion rates are very low, we would expect to see greater use of bonuses. Similarly, at the top of the hierarchy (CEOs and top executives), firms resort to much greater use of stock, options, bonuses, etc. When such approaches are needed, the principles of Chapters 9-10 apply.

Of course, in some cases the firm may choose to use the promotion ladder as an explicit incentive. This is most likely to be the case when the benefits of relative performance evaluation are very high, or the hierarchical structure of the firm is relatively fixed, so that competition for promotions is desirable.

Regardless of whether promotion-based incentives are explicitly designed or accidental, the discussion of tournaments and standards applies. Tournaments and standards have almost identical predictions about how incentives and optimal prizes vary with promotion rates and riskiness of evaluations.

Tournaments and standards differ in two important respects. First, tournaments should be used when the firm desires to fix slots, *unless* outside hiring is feasible (e.g., firm specific human capital is not very important, and hiring costs are not too high). By contrast, standards imply that the number of employees promoted will vary. On the other hand, with tournaments the quality of recruits will be more variable than with standards. In a bad year, the firm might promote a relatively low quality employee, just because he was the best performer. In a good year, the firm might not promote a high quality employee, because the talent pool was so good. Standards allow the firm to have better control over the quality of employees promoted.

An intermediate approach, outside hiring, can be used to balance the desire for stable hierarchical structure against the desire to control the quality of those promoted (or hired) into higher levels of the firm. In practice, it is likely that many firms that use explicit tournaments because of the benefits of RPE may resort to outside hiring in years when the quality of competitors is too low.

The second important way that tournaments differ from standards is in performance evaluation. Tournaments are an important example of RPE, whereas standards use individual performance evaluation. RPE reduces risk, *if* measurement error for different employees contains a common component that is very important. However, if that is not the case, RPE increases risk, since it exposes the employee to the idiosyncratic luck of the competitor.

An important benefit of RPE when the reward is discrete – as it is with tournaments – is that the evaluation becomes ordinal. In other words, the firm only has to decide who performed better, not how much better. This can make performance evaluation much easier, and much more credible, especially when work is more intangible. Of course, this applies to many white-collar jobs. This consideration favors a tournament over a standard.

Finally, tournaments may distort incentives compared to standards. A tournament will motivate less cooperation and more sabotage than otherwise. When work is highly interdependent so that these issues are important, the firm should consider using an absolute standard to decide promotions instead of a tournament. More generally, when teamwork considerations are important, incentive plans that make use of RPE tend to be a mistake. If RPE is used, the firm should compress pay, to balance the desire for strong productive incentives against the incentives for sabotage or lack of cooperation. Moreover, the firm should segregate workers by personality type, avoiding putting more aggressive and indi-

vidualistic employees in competition with more cooperative employees. Finally, the firm can reduce these problems if it can recruit employees with more cooperative personalities (intrinsic motivation).

Review Questions

1. If employees compete for a promotion in a tournament, are they likely to take more or less risky actions on the job? Does the answer depend on their odds of winning the promotion?
2. If promotions are an “accidental incentive system” in your firm, is there anything that your firm can do to avoid the effect?
3. When managers set goals for employees at the beginning of the year, they often “negotiate” the goal. When doing so, how would you recommend they think about how difficult the goal should be to achieve?
4. Suppose that promotions are an important source of incentives for managers in a firm that has eight levels in the hierarchy. Because of reengineering, the firm cuts out several layers of management. How should the firm think about changing its compensation system for managers at different levels?

References

- Bayo-Moriones, Alberto, Jose Galdon-Sanchez & Maia Guell (2005). “Is Seniority-Based Pay Used as a Motivation Device? Evidence from Plant Level Data.” Working paper, Universitat Pompeu Fabra.
- Bull, Clive, Andrew Schotter & Keith Weigelt (1987). “Tournaments and Piece Rates: An Experimental Study.” *Journal of Political Economy* 95: 1-33.
- Chan, William (1996). “External Recruitment versus Internal Promotion.” *Journal of Labor Economics* 14(4): 555-570.
- Coupé, Thomas, Valérie Smeets & Frédéric Warzynski (2005). “Incentives in Economic Departments: Testing Tournaments Theory?” Working paper, Universidad Juan Carlos III Madrid.
- DeVaro, Jed (2005). “Internal Promotion Contests in Firms.” Working paper, Cornell University.
- DeVaro, Jed & Michael Waldman (2005). “The Signaling role of Promotions: Further Theory and Empirical Evidence.” Working paper, Cornell University.
- Drago, Robert & Gerald Garvey (1997). “Incentives for Helping on the Job: Theory and Evidence.” *Journal of Labor Economics*.
- Ehrenberg, Ronald and Michael Bognanno (1990). “Do Tournaments Have Incentive Effects?” *Journal of Political Economy* 98(6): 1307-1324.

Eriksson, Tor (1999). "Executive Compensation and Tournament Theory: Empirical Tests on Danish Data." *Journal of Labor Economics* 17(2): 262-280.

Frederiksen, Anders and Elod Takats (2005). "Optimal Incentive Mix: The Dual Role of Promotions and Layoffs in Firms." Working paper, Center for Corporate Performance, Aarhus School of Business.

Gibbs, Michael (1994). "Testing Tournaments? An Appraisal of the Theory and Evidence." *Labor Law Journal*.

Kandel, Eugene & Edward Lazear (1992). "Peer Pressure and Partnerships." *Journal of Political Economy* 100(4): 801-817.

Knoeber, Charles (1989). "A Real Game of Chicken: Contracts, Tournaments, and the Production of Broilers." *Journal of Law, Economics & Organization* 5: 271-292.

Lazear, Edward (1979). "Why is There Mandatory Retirement?" *Journal of Political Economy* 87: 1261-1264.

Lazear, Edward (2004). "The Peter Principle: A Theory of Decline." *Journal of Political Economy*.

Lazear, Edward & Sherwin Rosen (1981). "Rank-Order Tournaments as Optimum Labor Contracts." *Journal of Political Economy* 89: 841-864.

Lazear, Edward (1989). "Pay Equality and Industrial Politics." *Journal of Political Economy* 97: 561-580.

Peter, Laurence & Raymond Hull (1969). *The Peter Principle: Why Things Always Go Wrong*. New York: William Morrow & Co.

Rosen, Sherwin (1986). "Prizes and Incentives in Elimination Tournaments." *American Economic Review* 76: 701-715.

Waldman, Michael (1984). "Job Assignments, Signaling, and Efficiency." *RAND Journal of Economics* 15: 255-267.

Waldman, Michael (2003). "Ex Ante versus Ex Post Optimal Promotion Rules: The Case of Internal Promotion." *Economic Inquiry* 41(1): 27-41.

Zabojnik, Jan & Dan Bernhardt (2001). "Corporate Tournaments, Human Capital Acquisition, and the Firm Size-Wage Relation." *Review of Economic Studies* 68(3): 693-716.

Appendix

Here we present a brief exposition of the two-competitor tournament model, and compare it to a promotion standard. Most of the intuition applies to multi-competitor tournaments as well, though those are much more analytically complicated. For this part of the appendix, ignore the possibility of cooperation or sabotage.

The Worker's Optimization

The employee's optimization is:

$$\max_e W_1 + pr(\text{promoted}) \cdot \Delta W - C(e),$$

which implies:

$$\frac{\partial pr(\text{promoted})}{\partial e} \cdot \Delta W = C'.$$

This first-order condition has a straightforward interpretation. The left-hand side is the prize, times the change in probability of winning due to marginal effort. The right-hand side is the marginal cost of effort. This equation applies to any standard or tournament.

This equation has an immediate implication: the larger the prize from promotion ΔW , the greater the incentives. The other implication is that the riskier is the performance evaluation, the weaker is the incentive. To prove that, we need to do a little more work.

Nothing we have done so far differs between a tournament and a standard. However, the probability of winning does depend on the promotion rule. Consider the standard first. The worker is promoted if performance is greater than some threshold, z . The cumulative and marginal densities of ε are $F(\bullet)$ and $f(\bullet)$, respectively. We will assume that $f(\bullet)$ is symmetric around and unimodal at zero (e.g., a normal distribution). Thus,

$$pr(\text{promoted} \mid \text{standard}) = pr(e + \varepsilon > z) = F(z - e) = F(e - z).$$

The latter inequality follows because ε is distributed symmetrically around 0. Note that the probability of winning when the rule is a standard is an equilibrium outcome of both the toughness of the standard, and the effort response by the worker. To set the promotion rate to what it desires, the firm must estimate the effort that a given standard will generate.

Now consider a tournament. If the ε 's distributions are symmetric and unimodal around zero, the distribution of $\varepsilon_S - \varepsilon_D$ will be as well. Denote it by $g(\bullet)$, with cumulative distribution $G(\bullet)$. The Danish worker wins if his performance is greater than that of the Singaporean:

$$pr(\text{promoted} \mid \text{tournament}) = pr(e_D + \varepsilon_D > e_S + \varepsilon_S) = G(\varepsilon_S - \varepsilon_D).$$

Since the game is symmetric, we will assume a symmetric Nash equilibrium, which means that they both supply equal effort. We can then rewrite the last expression as:

$$pr(\text{promoted} \mid \text{tournament}) = G(0).$$

Of course, since $g(\bullet)$ is symmetric, $G(0) = \frac{1}{2}$. This makes sense; since we have a symmetric tournament, the ultimate outcome is a coin toss. In order to compare a tournament and a standard, assume for now that the firm sets z so that in equilibrium $F(z - e) = \frac{1}{2}$.

Incentives depend on the change in probability from extra effort. These will be:

$$\frac{\partial pr(\text{promoted} \mid \text{standard})}{\partial e} = f(0), \quad \frac{\partial pr(\text{promoted} \mid \text{tournament})}{\partial e} = g(0).$$

Either of these values can be plugged back into our equation from above:

$$\frac{\partial pr(\text{promoted})}{\partial e} \cdot \Delta W = C'$$

$f(0)$ and $g(0)$ represent the height of the measurement error distributions at its mean and mode, zero. The lower this height, the more variance in the distributions, since they are symmetric and unimodal. Thus, the riskier the evaluation, the weaker the incentive.

The Firm's Optimization

Given the worker's effort supply characterized above, the firm sets the prize ΔW to maximize profits (effort minus average pay):

$$\max_{W_1, W_2} e - \frac{1}{2}(W_1 + W_2),$$

subject to the worker's effort supply, and to the constraint that overall pay is enough to induce the worker to put forth effort:

$$\frac{1}{2}(W_1 + W_2) = C.$$

(Tournament models usually ignore risk aversion, since it is difficult or impossible to derive closed-form solutions. We do so in this appendix.) The firm's first-order conditions are:

$$(1 - C') \frac{\partial e}{\partial W_1} = 0,$$

$$(1 - C') \frac{\partial e}{\partial W_2} = 0.$$

These imply that at the optimum $C'(e) = 1$. In other words, the firm should set pay so that workers exert extra effort up to the point where the marginal cost of effort just equals the marginal benefit (extra output). Thus (ignoring risk aversion), tournaments and standards can generate efficient effort levels, just like standard incentive schemes.

The optimal wage spread is found by substituting $C' = 1$ into the worker's effort supply and solving:

$$\Delta W = \frac{1}{f(0)} \quad \text{or} \quad \Delta W = \frac{1}{g(0)}.$$

Once again we see the analytical similarity of tournaments and standards. This result shows that the optimal prize spread should be larger if the risk in performance evaluation is larger. Note that this applies to the optimal prize, not to the level of pay, which would be reflected in W_1 or W_2 separately. This is an example of the principle from Chapter 10 that the level of pay does not drive incentives; the way that pay varies with performance is what is important.

Options & Executive Pay

“How can you afford to pay your men so well?” (Banker’s question to Andrew Carnegie)

“I can’t afford to pay them any other way.” (Carnegie’s reply)

Overview

In this chapter we finish section on pay for performance by considering two special topics related to executive compensation. These topics are important in practice, and provide interesting applications of the discussions from the last three chapters.

The first topic is employee stock options. Options are an important part of incentive compensation plans for most top executives in publicly traded companies. They are also quite important incentives in many small start-up companies. Moreover, use of employee stock options exploded during the boom of the technology sector in the United States during the 1990s. Many high-tech companies began awarding options all the way down the hierarchy, and the press began publishing stories about secretaries who got rich on their options and drove to work in Ferraris. Finally, the use of employee stock options appears to be expanding in Europe, and especially in some areas of Asia. We discuss the incentive properties of stock options, as well as whether they are a good practical compensation tool, and if so, for which employees.

The second topic is executive pay and incentives overall, with a focus on CEOs. Incentive problems are most important for the firm’s key employees, so good compensation plan design is crucial for top management. The concepts discussed in prior chapters are just as relevant for CEOs and the top management team as they are for other employees.

Employee Stock Options

Stock Options – A Brief Overview

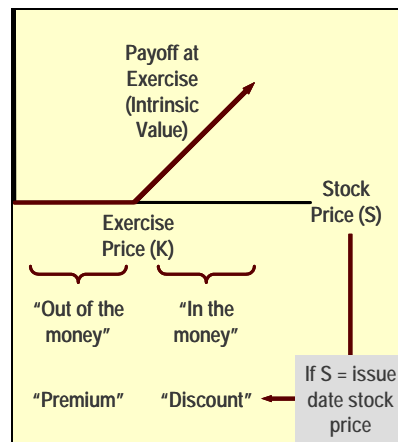
Since not all readers are familiar with stock options, we begin with a brief description of options. If you already know about stock options, you can skip this section and the Appendix.

A “call” option is a financial security that gives its owner the right to purchase 1 share of a company’s stock, at a fixed *strike price* or *exercise price*. For this reason, stock is actually a special kind of call option, with an exercise price of zero. Of course, if the stock’s price is below the exercise price, it would not make sense to exercise the option. If the stock’s price is above the exercise price, the option holder can make a profit by exercising the option, selling the stock, and pocketing the difference between the two prices (after transac-

tion costs). Thus, a call option benefits from stock price increases, but the holder is shielded from stock price declines (to some extent, as discussed below).

A “put” option gives its owner the right to sell 1 share of a company’s stock at a fixed exercise price. Exercising a put would make sense if the stock price fell, just the opposite of a call. Thus, owners of puts hope that the stock falls in value. For this reason, employee stock options are always calls; we will only consider call options in this chapter.¹

Figure 12.1 plots the payoff from a hypothetical call option with exercise price K , and stock price S . If $S < K$, the option is “out-of-the-money,” and should not be exercised, so the payoff is zero. If $S > K$, the option is “in-the-money.” If the in-the-money option is exercised and the stock is immediately sold, the profit equals $S - K$.² This payoff is often called the “intrinsic value” of the option. When an option is issued to an employee, the firm must decide what exercise price to set. Almost all employee options are issued “at the money” (K is set equal to the grant-date stock price). Finally, an option has an expiration date – the last date on which it can be exercised. Beyond that date it has no value.



Call Option
Figure 12.1

Options are often valued using some variant of the famous Black-Scholes formula, which is described in the Appendix. However, as discussed below, one must be careful in applying this formula to employee options, as opposed to those that are traded on exchanges.

Employee stock options differ from those that are traded on exchanges (such as the Chicago Board Options Exchange, CBOE) in several ways. First, when granted to the employee they are usually not vested immediately. Typically employee options vest gradually over 3-5 years. Until vested, the employee may not exercise the option. Even if vested, an employee cannot trade an option to another investor – there is no liquid market for employee stock options. The employee can either hold the option, or exercise it. This matters because modern option theory tells us that it is generally not optimal to exercise an option before its expiration date, if you can sell it to someone else instead. The intuition is simple: the (call) option has value because it is a bet that the stock price might rise above its present value. For this reason, a traded option’s market value is always *higher* than its intrinsic value. Finally, if an employee leaves the firm, all unexercised options are generally lost.

¹ In fact, it is illegal for top executives in US companies to hold puts in their own companies.

² Ignoring taxes. Tax issues can be complex for employee stock options, but are beyond the scope of this text.

Should Firms Grant Employees Options?

As mentioned above, use of employee stock options exploded (in the U.S.) during the tech bubble of the 1990s, particularly in “new economy” firms. Before that, they were not an important part of compensation except in some cases for top executives. Several arguments are often put forward for giving employees options, but they generally do not make sense except for a firm’s key employees.

Source of Firm Financing

It is sometimes argued that options are a cheap form of financing for a company. The idea is that the firm can offer options to employees instead of salary or other forms of compensation, with zero payout of actual cash. Furthermore, current accounting rules result in no expense to the firm’s accounting statements until the options are exercised (unlike cash compensation). Thus, in an accounting sense, options have zero short-term impact. However, we will soon see that this argument is wrong from an economic point of view – options are probably the most *expensive* form of compensation, because they result in an opportunity cost incurred by the firm. This comes from the fact that employees do not value options as much as options traders do.

We will elaborate on this point below, but we can make it in a simpler way for now, that explains why employee stock options are a poor way for a firm to raise funds. A firm’s cost of capital (the expected return on investment it must offer investors) will be lower if the firm can find investors with relatively low risk aversion. This is why most large firms have *separation of ownership and control* – the firm is run by a team of managers, but owned by a separate group of investors – which, of course, causes principal-agent problems that we discuss later in this chapter. Investors reduce their risk by holding a diversified portfolio, rather than investing too much of their portfolio in one company.

Now consider issuing options to employees instead of cash salary. For most employees, this will constitute a substantial portion of their wealth. Some companies encourage their employees to hold shares of stock in their employer, or invest employee pensions partially in the firm’s stock. Moreover, some of their human capital is invested in the firm. For all of these reasons, employees will be quite *undiversified* in the company, and thus more risk averse than typical investors. They will demand a larger risk premium (larger expected total compensation) in order to be willing to hold the options – and the firm’s costs of financing operations will be higher than if it obtained funds from more traditional sources, such as issuing stock or debt.

There is only one case where it might make sense to try to raise investment funds from employees. That is when the firm cannot obtain funds more cheaply through other sources, but still has positive net present value investment opportunities. Such cases are likely to be quite rare, with one important exception: new ventures. In such cases, outside financing may be very difficult to obtain, even from venture capitalists, because of severe problems of adverse selection and moral hazard. Employees inside the company may well have much better information about the prospects of the new company, and thus may be willing to invest in the enterprise. This is one reason why options tend to be an important form of compensation for startups in particular.

Employee Self-Selection

Another possible role for options is to induce better self selection of employees to improve recruiting. As argued above, stronger pay for performance tends to improve self selection,

so to the extent that options provide incentives, this will apply. Options will also be more highly valued by employees who are most optimistic about the firm's prospects. If such employees are also more productive (for example, more enthusiastic about working for the firm), then this will be valuable. Note, though, that this argument applies to all forms of pay for performance, and is not specific to options.

A different form of self selection that options may induce is employee conservatism or will- ingness to take risks. Options are perhaps the riskiest form of pay. Moreover, they tend to be more valuable, the riskier is the underlying stock value.³ Thus, option grants tend to encourage employees to take more risks on the job. To see this, consider Figure 12.1. The option pays off only when performance is high, and the higher is performance, the higher is the payoff. Since there is little downside to option pay from riskier actions, the employee is motivated to take actions that make extreme outcomes – good and bad – more likely. Whether or not this is desirable for the firm depends on the situation, but in many cases it is plausible that such an effect is useful. This is because employees will generally be more risk averse than diversified shareholders, so that incentive plans that do not take into account these differences in risk aversion will be distorted toward too conservative decision making. Finally, note that this kind of self selection only applies to key employees whose decisions can actually affect the riskiness of the stock price, which will tend to be only a few key employees who make major strategic decisions.

Reducing Turnover

A commonly-heard argument for granting employee options is that they reduce turnover, because they are vested gradually, and because an employee typically must give up options if they leave the firm. While this may be so, this is not particular to options. Any kind of deferred pay, such as pay that rises with seniority or gradual vesting of pensions, will have similar effects on turnover. Therefore, this is an unconvincing reason to grant options to employees.

Options as Incentive Pay

The most important argument for granting employee options is to provide incentives. To evaluate this argument, let us consider the incentive properties of employee stock options using the concepts developed earlier.

Performance Measure

Options are similar to stock, in that both have the same performance measure: stock price. This is a very broad measure and thus distorts incentives little if at all. However, it is also a very risky measure. For this reason, incentives based on stock price will tend to require that the firm pays the employee a relatively high risk premium.

More problematically, for all but the firm's key employees, stock price is a performance measure that is largely uncontrollable. Virtually nothing that a lower level employee can do will budge the stock price at all, unless the firm has very few employees. Thus, from a performance measurement perspective, options should generate little or no incentive for most employees. They are more like *giving the employee a lottery ticket*.

³ This is always true for options traded on exchanges. However, it may not always be true for employee options due to risk aversion.

Pay-Performance Relationship

If options are not too far out-of-the-money, then the firm can tie pay more strongly to the performance measure (stock value) for options than for stock itself. This is because options are *levered* incentives: they only pay off when the stock price exceeds the exercise price. Because they do not always pay off, one option has less value than one share of stock. Therefore, for equivalent cost to the firm, a firm can give the employee more options than shares of stock. When the stock price rises, the employee's value will then rise more rapidly with the options grant than with the stock grant. This is the best argument for granting employees options: for that small number of key employees for whom stock value is a sensible performance measure, a stronger pay-performance relationship can be achieved with options than with stock. However, this logic applies *only* for the firm's key employees.

There is a downside to this argument, unfortunately: incentives from options are more brittle than incentives from stock. If the stock price falls so that the option is too far out-of-the-money, then the pay-performance relationship falls dramatically. This is because increased effort by the employee is unlikely to raise the stock price high enough to bring it into the money, if the stock price is currently too far below the exercise price. The argument is exactly analogous to the problem of setting the threshold too high in the "Reward" scheme in Figure 10.2 of Chapter 10.

A related concern is that the value of the pay package to the employee will also tend to be brittle if options are used. An option that is far under the money will have very little value, while the underlying stock will still have some value unless S is close to zero. If options are a large part of employee pay, then the expected value of the employee's total compensation will have fallen significantly. For example, in March 2000 the technology sector of the world economy experienced a sudden, dramatic fall in stock prices (the "Dot Bomb"). Many technology companies had made extensive use of employee options, and their employees now found that much of their pay package was worthless, and often well below market values for similar jobs. These firms found that they had to reprice the options (see below), or offer other forms of additional compensation, or suffer turnover.

Notice an implication of this discussion: employee stock options do have a downside for employees. This point is often missed, since options are profitable when $S > K$, but the employee does not have to pay if $K > S$. However, the option has value because it *might* end up in-the-money before it expires. The lower that the stock price falls, the less likely is that to occur, so the expected value of the option falls. Of course, the fall in expected option pay is lower than the fall in the stock price itself, but employees do have a downside risk from options pay.

Granting Options Over Time

There are several ways to grant options to employees. The most direct is to give all of the options to the employee at once (say, upon hiring). This gives the strongest incentives, immediately. However, this approach is also the most brittle, as both incentives and the value of the pay package fall dramatically if the stock price falls too much, as described above.

As an alternative, many companies issue options over a period of time (say, some options each year). Here there are also two general approaches. One is to grant a fixed value of options each year (e.g., \$2,000 per year). The other is to grant a fixed number of options each year (e.g., 200 per year). Which is better?

To answer this, first remember that virtually all employee options are issued with the exercise price K equal to the stock price S on the grant date (at the money). Also, it helps to know that, for at-the-money options, the option's value rises as S rises (see the Appendix).

Consider first granting a fixed value of options each year. If the stock price rises this year, each option granted next year is more valuable, so the firm will be able to issue fewer options to the employee. Similarly, if the stock price falls, new options issued at the money will have less value, so more can be granted. Thus, under this approach the employee is given more options after the firm performs poorly, and fewer after it performs well, muting incentives compared to giving all of the options at once. However, total compensation is more predictable.

Now consider granting a fixed number of options each year. If the stock price rose last year, this year's option grant will be more valuable, and vice versa. This reinforces (strengthens) the incentives from the initial options grant. However, it also makes total pay more variable.

Thus, there is a tradeoff in granting options over time. A fixed-value grant results in more brittle incentives, but less variability in the value of the pay package. A fixed-number grant has less brittle incentives, but more variability in the value of the pay package.

Subtle Incentives From Options

As we have already noted, options change incentives for taking on or avoiding risk, because they provide insurance against bad outcomes, and reward good outcomes. This can be beneficial if employees would otherwise be too conservative, and dangerous if employees are already willing to take risks. This is one reason why options are used so much in startup companies. In such cases, the firm has little reason to be conservative; it has no brand name or other form of reputation to lose. Instead, its payoffs in good outcomes may be very large, so strategically the firm itself will want to be more innovative and less conservative. In jobs where there is little upside value to good performance, but there is a downside, options would be a mistake.

The payoff structure of an option has a cusp, or change in slope, at the point where the option is just in-the-money. Because of this, incentives may change dramatically for options that are near the money. As discussed earlier, these situations are more likely to tempt employees to manipulate the incentive system. Thus, extensive use of stock options may make it more likely that executives engage in unlawful or unethical conduct to try to bring their options into the range of positive payoffs. This would not be the case if the executives were simply given stock, since the pay-performance relationship is smoother (in fact, linear) for stock.

Should Options be Repriced?

After stock prices crashed in March 2000, many employees in technology companies found their options so far below the money that they were essentially worthless and generated no incentive at all, even for key employees. Some firms repriced their employees' options. Typically, this is done by exchanging the employee's existing options for a smaller number with a lower strike price (equal to that day's stock price). Such a practice is controversial, and many shareholders criticize the practice. What are the arguments for and against repricing options?

The argument against repricing options is that it is, in effect, rewarding poor performance (or at least reducing the punishment). Options are never repriced when the stock price rises greatly; they are only repriced when the stock price falls. A frequently heard criticism is that employees agreed to the incentive plan, and they should be expected to stick with it, even when their hopes for high payoffs are not realized. Even worse, repricing can establish a dangerous precedent; employees might expect further repricings in the future if the stock price declines again.

The argument for repricing acknowledges the points in the paragraph above, but brings in practical considerations. If options are not repriced or employees are not compensated in some other way, then the value of the pay package has declined dramatically, and the firm risks losing employees. Of course, those most likely to leave are those with the best outside alternatives, which tend to be the firm's best employees. Also important is that without repricing, options give little or no incentive. Shareholders can benefit from a repricing if it better motivates employees.

A useful way to resolve these arguments is to apply some of our principles about subjective evaluation. Ask if the stock price fell because of poor employee motivation, or because of uncontrollable factors. If it is the former, then repricing is unlikely to be the best policy, as it rewards poor effort. If it is the latter then the fall in stock prices is largely not the employee's fault. In that case, repricing is a special case of subjective performance evaluation to reduce risk, and improve the accuracy of the evaluation. Thus, repricing may make sense for such unusual events as a stock price crash in an entire sector. It is also more likely to make sense to reprice options for lower level employees, who have little or no control over the stock price, than for top executives. However, the board of directors must be careful not to establish a precedent, so repricings should be rare events, with carefully communicated justifications.

How Do Employees Value Options?

As noted above, employee options differ from traded options in important ways. They are restricted: they typically do not vest immediately, cannot be traded (only exercised), and are lost if the employee leaves the firm. In addition, employee far from diversified against the risk in the options.

For these reasons, employees are quite risk averse with respect to options, whereas those who trade options on exchanges are relatively (or completely) risk neutral. Firms always must pay risk premiums to employees if they give them a risky pay scheme, but the risk premium is probably highest when compensation is in the form of options. In fact, while firms sometimes try to value employee options using the Black-Scholes formula, employees generally require risk premiums of 30% or more above the BS value, in order to be willing to accept stock options. In other words, while the BS formula is an excellent approximation of the market value of a publicly traded option, it overstates the value of a similar option when granted to an employee.

An important implication is that employee stock options are not free from the point of view of the firm. While there may be no immediate charge to accounting statements from giving an employee options, the firm incurs a substantial economic (opportunity) cost from doing so. Think about it this way. Suppose that the firm grants option to an employee, reducing salary at the same time. In doing so, the firm is asking the employee to "buy the job" to some extent, giving up certain pay for risky (but hopefully performance based) pay. The

employee will not be willing to pay the BS value for those options, but only that amount minus some risk premium. Thus, the firm incurs the cost of the risk premium by not selling an equivalent option on the open market. Since risk premiums tend to be quite large for employee stock options, options are probably the *most expensive* form of pay for performance, rather than being free. This should not be surprising, since they are one of the riskiest forms of pay for performance that firms use.

Executive Pay

It pays to put greater focus on all of the issues in this text when considering the firm's key employees – those who add the greatest value, have the most important and scarce skills, etc. Typically the most important employee in the firm is the CEO. In this section we consider pay for performance issues for CEOs and top executives. In doing so, we focus on publicly traded firms, and work from the general assumption that the goal of executives in such firms is to maximize shareholder value. This latter assumption can be controversial. Nevertheless it is a good starting point for thinking about the issues carefully. To the extent that other objectives are important, one might reach some different conclusions (for example, about the desirability of layoffs). Nevertheless, the analysis will be relevant in thinking about important tradeoffs and issues in executive incentives.

What is the Most Important Question?

Executive pay stirs great controversy. Most business publications publish some kind of annual roundup of CEO compensation, and these articles receive great attention. The pay of top managers in publicly traded companies is often criticized for a variety of reasons. Many critics argue that executives are paid too much. Others argue that the pay does not reflect performance. Still others argue that CEOs take advantage of their position to cause both of these problems, paying themselves generously from shareholder funds. Certainly, when a CEO earns a \$100 million payout on stock options, or is given a large severance, such concerns are understandable.

Which of these issues is the most important? All matter to shareholders. However, the public criticism seems to focus excessively on the *level* of executive pay. While enormous executive pay packages may seem “too much” and even unethical, if the desired goal is increasing shareholder wealth, then this is an issue of second-order magnitude. After all, for even the most highly-paid CEOs, total compensation is a small percentage of overall firm value.

The more important issue should be the strength of the pay-performance relationship. Turning back to Figure 10.1 in Chapter 10, we saw that incentives are determined little by the overall level of pay. Rather, they are determined by the incentive intensity, the slope (or more general “shape”) of the pay-performance relationship. Thus we will focus primarily on the incentive questions.

That said, there is certainly evidence that suggests that top managers are sometimes able to use their power to extract higher levels of compensation than they would otherwise be able to. How might this happen? CEO and executive pay packages are usually designed by compensation consulting firms hired by the Compensation Committee of the board of directors. These consultants may work closely with employees inside the firm, from Human Resources or other departments. The CEO can often exert influence over board members, and over who is appointed to the Compensation Committee. And they are quite likely able to exert influence over both employees and the consultants. (In this case of the

consultants, consider *their* incentives – they are designing the pay package for an important client.)

One study found that if the CEO is appointed before the chair of the Compensation Committee is appointed, then after controlling for other factors (e.g., firm size, industry, CEO experience, etc.) the CEO's pay is about 11% higher.⁴ The same study found that after controlling for other factors, CEOs of interlocked boards of directors (that is, the CEO of Company A is on the board of Company B, whose CEO is on the board of Company A) have pay that is about 10% higher. Studies such as these strongly suggest that CEOs are, in some cases, paid more than their market value absent these effects.

Executive Pay for Performance

We can quickly analyze CEO pay using the same tools that we used for employee stock options. First consider the performance evaluation. The primary performance measure for executives is stock price, as the primary incentive instrument for them is stock and options. This is a reasonable performance measure to consider for the CEO, since his or her actions can have strong effects on overall firm value. However, the measure is also quite risky. Therefore, many executive pay packages also make important use of narrower performance measures, especially accounting profits (earnings).

What about the incentive intensity? Estimates of the incentive intensity for CEO pay are quite small. If an owner-entrepreneur gets a commission rate of 1.0, the CEO's "commission" rate is roughly 0.0004 in the largest corporations.⁵ That is, for every \$1,000 increase or decrease in stock value, the CEO's compensation (including raises, bonuses, deferred pay, stock, options, and threat of termination) rises or falls by less than \$1. Of course, the optimal incentive intensity should not be based on total firm value, but rather on the employee's contribution to firm value. Thus, these estimates understate the effective commission rate. Nevertheless, their small size suggests that risk and other factors play important roles. More importantly, they suggest that CEOs may well have weak incentives compared to entrepreneurs. The incentive intensity in executive incentive schemes for accounting measures tends to be much stronger – roughly double. This indicates that risk considerations are important, as accounting measures distort incentives compared to stock price as a performance measure, but are much less risky.

It is very difficult, if not impossible, to tell what the right incentive intensity for executives should be. Other research has asked a related question: does the strength of incentives vary with other factors as the theory predicts? By and large, the answer is that executive pay patterns do conform to the predictions, which is reassuring evidence that there is at least some economic logic to executive pay.

For example, several studies have found that CEO incentive intensities are stronger when stock price is less risky, and vice versa. Similarly, executive incentives vary with industry characteristics. In regulated utilities, executives tend to have both much lower overall compensation, and much weaker incentive intensities. Both of these make sense, because the job is much more constrained by regulators – there is less discretion for top managers in regulated industries. This means that talent is valued less than in more dynamic, unpredictable industries. It also means that the incentive problem is less important, since there is less "decentralization" to the CEO and management.

⁴ See Hallock (1997).

⁵ See Jensen & Murphy (1990).

There is a very strong relationship between executive pay and firm size. For every 10% rise in firm size (measured by sales or stock value), executive pay tends to rise by about 1%. This is consistent with the idea that more talented managers are sorted into larger organizations, where their talents are better used and more valued.

By contrast, estimates of the average relationship between executive pay and firm performance are *weaker* in larger firms compared to smaller firms. That is, most research on executive incentives interprets the following measure as an estimate of the incentive intensity:

$$\hat{b} = \frac{\Delta \text{Pay}}{\Delta \text{Stock Value}}.$$

As mentioned above, for the largest corporations this estimate tends to be about 0.0004. The estimated value rises as firm size decreases. If this measure is a good proxy for the strength of incentives, that executive incentives are *weaker* in larger companies. One explanation is that giving an executive the same incentive intensity in a larger firm implies greatly increased risk, so larger firms mute incentives for risk averse executives.

Another explanation is that such estimates are confounding two effects discussed in the last two chapters. Remember that the manager's incentive is determined by:

$$\frac{\Delta \text{Pay}}{\Delta e} = \frac{\Delta \text{Pay}}{\Delta \text{Stock Value}} \cdot \frac{\Delta \text{Stock Value}}{\Delta e}.$$

Thus, the empirical proxy is a good measure of how incentives vary across firms only if we assume that the last term in the equation above is constant. But what if the effect of effort on Stock Value varies with firm size? If it does, then we need to take that into account.

One way to think about this is to suppose that there are, roughly speaking, two kinds of managerial decisions.⁶ The first are "strategic" decisions. Working harder to make a good strategic decision raises firm value in *percentage* terms. This is why they are called strategic, since the decision has implications for the entire operation of the firm. Examples might include overall strategy, product choice, merger and acquisition activity, etc.

The second type of managerial decisions is "operational" decisions. Working harder to make these decisions better increases firm value in *absolute* terms. That is, a decision that improves firm value by \$50,000 does so regardless of the size of the firm. An example might be improving operations at a single factory.

Now look again at the last term in the equation. The measure is the absolute change in stock value from working harder. If all executive decisions are operational, then this will be a constant. However, if all executive decisions are strategic, then this term will vary systematically with firm size. In particular, the effects of effort will be larger in larger firms and vice versa.

Thus, to the extent that some executive decisions have impacts that are strategic and scale with firm size, executive incentives would be stronger in larger firms for a given level of incentive intensity b . In fact, taking this view it appears that overall CEO incentives decline only a little with firm size (probably reflecting risk considerations).

⁶ See Baker & Hall 2004.

Other Incentives & Controls

Employee motivation is driven by factors other than pay for performance. Firms also direct behavior through controls, such as direct monitoring and limits on decision making. And, of course, these also are important for executive motivation.

Four other important extrinsic factors affect executive behavior. One is pressure by outside advocates or shareholders. Another is product market competition. A third is the market for corporate control (hostile takeovers). A fourth is oversight by the Board of Directors.

It is unclear whether the influence of outside pressure groups is a good or bad thing for corporations. To the extent it is driven by informed shareholders, it is likely to improve firm value by pressuring managers to adopt better policies. However, to the extent that groups with other objectives drive it, it may distort CEO incentives (for example, to avoid layoffs when they are important for firm performance, or to spread the use of options or profit sharing inefficiently to lower levels of the hierarchy). One important possibility is that public opinion may limit the ability of firms to design effective pay packages for top executives, because of criticism of high levels of executive compensation. While this is a possibility, research remains to be done to see whether or not such effects are important.

Arguably the most important constraint on executives is product market competition. The more competitive the market, the greater the pressure on a firm to lower costs, increase quality, and innovate in order to survive. Thus, we should expect governance and incentive problems to be more severe in cases where a company has less competition. This includes companies with barriers to entry such as patent or regulatory protection that create monopoly power.

A third potential discipline for managers that has played an important role in the past is the market for corporate control. If a publicly-traded company is poorly managed then management can, in principle, be replaced. This might be accomplished through shareholder proxy campaigns, or by investor groups who buy a controlling percentage of the company. A large body of empirical research indicates that hostile takeovers and related control contests do usually tend to increase firm value.

In the United States in the 1980s, a series of hostile takeovers and other changes in control were used to break up inefficient conglomerates, wring cash hoards out of the hands of incumbent management, and so on. On reflection it should not be surprising that there was a wave of such transactions at this time. In the decade or two before, the business world had begun to change dramatically, including increases in international trade, large-scale deregulation, and the information technology revolution. Many businesses needed to restructure substantially. It appears that many management teams were either the wrong fit for changing their organizations, found it difficult to implement large scale restructuring, or were reluctant to do so. In such cases the ability of outsiders to buy the company, betting that they can manage it better, is an important motivator to management.

Because such contests often involve dramatic changes in a firm's organization, including mass layoffs and selling of divisions, they have been highly controversial. These mechanisms are much more difficult to use in Europe (except Britain), where there are more legal restraints. In addition, there are cultural differences in beliefs about the extent to which the firm has a social responsibility to avoid dramatic organizational changes. Thus, this mechanism has not played such a key role there. Hostile takeovers and other control contests are more common in some Asian economies, though not all. Even in the United States, hostile takeovers are now quite rare. This is because most states enacted laws

making it much harder to successfully complete such transactions. Thus, this disciplinary mechanism constrains management much less now than in the past.

A final incentive mechanism for top executives is oversight by the Board of Directors. There are two primary roles for directors. One is to provide advice and support to the CEO and top management team. In this capacity, a director discusses the firm's strategy and tactics, offers advice, and then when appropriate yields decision rights to the CEO. The second is to provide a last line of decision control, by ratifying and monitoring management decisions, and setting management incentives. Put another way, one of the most important duties of the Board is to conduct subjective performance evaluation, and reward or punish, the top management team. Therefore, in one role the director acts as a supporter of the CEO and management. In the other, he or she acts as the representative for shareholders. Just as it can be difficult for a manager to draw the line between support and discipline of an employee, it can be quite difficult to draw the line between these two roles, which sometimes come into conflict.

One survey of directors found that only about 35% considered performance evaluation to be one of their primary duties as a director. Instead, directors appear to be giving too much emphasis to their role as supporters of top management. Why might this be the case?

One reason is that the CEO usually has quite a bit of leverage over who sits on the board. Many board members, perhaps most, are suggested by the CEO. Naturally, CEOs usually have a strong preference for choosing directors who are personal friends or sympathetic to the CEO's strategy. In addition, directors are often CEOs of other firms, and may be reluctant to pressure management to avoid setting the same precedent on their own board.

Do Executive Incentives Matter?

An amusing example of apparently poor CEO incentives is Ross Johnson, CEO of RJR Nabisco, who used the corporate jet to fly his dog to his vacation home.⁷ More recently, there have been several notorious cases (e.g., Tyco) of CEOs living lavish lifestyles using corporate funds, which suggest lack of oversight by the board. However, such excesses are small in the context of a very large corporation, and might just be viewed as part of the executive compensation pay package – they might not even be inefficient.

More rigorously, what kinds of problems might arise if top management does not have good incentives? Of course, the firm's performance is likely to be relatively poor. Experience with publicly-owned companies (e.g., nationalized postal services or airlines), and those that have subsequently been privatized and faced competition, suggests that this does happen.

Do firms that provide stronger incentives for top managers perform better? This is not as easy a question to answer as might first appear. One cannot simply correlate proxies for executive incentive intensities (e.g., % stock ownership) with stock returns. The reason is that stock markets are generally very efficient: they incorporate all new information about the value of a firm almost immediately. If the firm has a well-designed executive incentive package, this should be incorporated into the firm's stock price right away, with no observable effect on future stock returns (unless the company changes the policy again later).

A few studies have looked at the *abnormal* (unexpected) change in stock price on days that firms have announced *changes* in executive pay plans. These studies find that if ex-

⁷ *Barbarians at the Gate*, 1991.

Executive incentives are increased, say by granting more stock or options, the stock price rises more than expected at the time of the announcement. This is consistent with the idea that better managerial incentives improve firm value.

However, there are other interpretations of such evidence. One is a form of insider trading. Suppose for example that the CEO has private information about the firm's future prospects. Shareholders might infer, from the CEO's willingness to accept proportionately more risky stock and options in the pay package, that the firm's prospects are better than previously believed. This would raise the stock price, but not because of better CEO incentives.

Similarly, a signaling interpretation is possible. Suppose that the board and CEO believe that the stock price is undervalued (based on their private information about the firm's operations). Then they might signal this to the external market, by willingness to accept more options and stock. Again, the stock price would rise. Thus, evidence based on stock price performance is not unambiguous. Other studies have looked at accounting performance, and found similar results, further supporting the idea that executive incentives matter.

An alternative approach is to analyze case studies. For example, management or leveraged buyouts (MBOs or LBOs) are remarkable because they usually implement very strong executive incentives compared to more typical corporations. They also make heavy use of debt (leverage). This severely constrains executives, and puts strong pressure on them to improve cash flow to avoid bankruptcy. Several studies have looked in detail at performance of such organizations, and have generally found that performance increases notably after the transaction.

An important role of extrinsic incentives is to better align interest of the employee to those of the firm, when the employee has intrinsic motivation. So one way to consider this question is to ask what management's intrinsic goals would be, absent financial incentives. One commonly-expressed concern is that managers might be motivated to "empire build." In other words, they may be interested in managing larger organizations. This would suggest that CEOs with poor incentives would pursue strategies of growth and acquisition.

What Are CEO's Intrinsic Motivations?

One fascinating study used plant level data from a large set of U.S. firms to try to analyze the intrinsic objectives that top management would like to pursue, absent the sort of controls and incentives described in this section. The authors exploited the fact that anti-takeover protections vary from one state to the next in the U.S., so that there are 50 different legal regimes that enacted laws at different times. This variation allowed them to study the effects of different levels of management protection from the threat of takeover.

The results suggest that the "empire building" view may not be appropriate. Instead, managers appear to prefer pursuing a "quiet life." When the firm is more insulated from takeovers, it tends to pay higher compensation to employees -- especially to white collar workers. It is less likely to shut down old plants. Contrary to empire building, creation of new plants also declines. It seems that management follows the path of least resistance, being generous to employees and resisting change overall. The study also finds that productivity and profitability decline in such firms.

Source: Bertrand & Mullainathan (2003)

Another possible goal of CEOs is risk reduction and survival. The way that shareholders reduce risk is through diversification of their portfolio – buying shares in companies with unrelated businesses. CEOs might diversify lines of business inside the company to reduce risk in their pay and employment. If the company owns several unrelated divisions, and one division is performing poorly, odds are relatively good that a different division is performing well. The more unrelated the divisions, the less likely that they will all perform badly at the same time, and the less risky are the firm's overall cash flows.

While at first this may sound like a good idea, it is generally not a good idea for shareholders. Investors can produce this diversification on their own by holding shares in a variety of companies. It is difficult to think of an economic justification for a widely diversified firm. If the lines of business are truly different, then there are few or no economies of scale or scope (synergies), while large complex organizations are very difficult to run effectively (for the reasons considered in this book).

Another benefit to management (but not to shareholders) from diversification is that it shields management from outside pressures. The less risky is cash flow, the less likely is it that management will have to borrow funds or issue shares. This allows management to *cross subsidize* divisions that it wishes to invest in for inefficient reasons. For example, a CEO may decide to invest in an aging factory rather than shut it down, because the CEO is personally uncomfortable with layoffs, or because it was where the CEO's career began.

Indeed, until the hostile takeover boom in the 1980s, many large U.S. companies were diversified conglomerates; many firms in Europe and Asia still are.

A related concern is that top management may have incentives to hoard cash. As a dramatic example, until recently Microsoft held \$80 *billion* in cash reserves. Suppose that the firm is making profits, accumulating cash. What should it do with the cash? It should invest the cash *only if* the return on investment is higher, after suitable adjustment for risk, than what shareholders could earn by investing the funds themselves. Any cash available after investing in such net present value opportunities is sometimes called "free cash flow," which in principle should be returned to shareholders.

However, management might invest the cash if there are opportunities with *positive* (accounting) returns on investment, which is not the correct criterion. In fact, management might simply hold the cash for future use. It is not uncommon to hear management talking of accumulating a cash hoard inside the company as a "war chest" for use in future strategy implementation. Since the firm can always raise additional funds later if it has good projects to invest in, these kinds of arguments tend to make little sense. Thus, management may have improper incentives to hoard free cash flow. Some of the hostile takeovers of the 1990s led to payout of large extraordinary dividends to shareholders, apparently removing the free cash flow from management's control.

Finally, top management is likely to have strong incentives to attempt to entrench itself, to reduce the chance of being ousted for misuse of corporate funds, poor performance, or pursuing intrinsic objectives. We have already talked about how management might try to stack the board with sympathetic directors. Management may also pursue measures that make hostile takeovers harder to execute, such as poison pills.⁸ And, if the company

⁸ A poison pill gives current shareholders the right to buy more shares of the company at substantially below the market price. These options are only exercisable if an acquirer buys more than some given percentage of the company, and cannot be exercised by that acquirer. Thus, in order to buy a controlling interest in a company with a poison pill, the acquirer must pay a premium to existing shareholders. The premium is so large that to date there have been no successful hostile takeovers in companies with poison pills.

does receive a tender offer, management sometimes aggressively resists the offer, even when it would give shareholders a substantial premium and a majority of shareholders votes to accept the offer.

Overall, the evidence suggests that there can be substantial loss of value when companies do not pay adequate attention to incentives and governance of top management. And, the opposite also appears to be true: when management incentives are well designed, performance increases, and firms are more innovative and dynamic.

Summary

Employee Stock Options

We discussed several justifications for giving employees stock options. Most have nothing to do with options per se; the same objectives can be achieved through other forms of compensation. The justifications that may make sense (at least under some circumstances) generally apply only to the firm's key employees. For example, options encourage risk taking only for employees who can actually affect the riskiness of the stock price.

The most important justification for giving options is that they can provide stronger incentives than stock, for the same compensation cost, since they are a levered incentive. While the strong incentive is attractive, there are important caveats. Options are an extremely complex form of pay for performance (even without considering the tax and accounting complications that they create, which are also substantial). The incentives and level of pay that options provide are brittle, in that they may decline sharply if the stock price declines. Options may be more likely to motivate manipulation by the employee. And the firm must give careful thought to how to grant options over time, and to possible repricing if the firm suffers unforeseen poor performance. Most importantly, options are a very expensive form of pay for performance, because the riskiness and restrictions that they imply cause employees to demand large risk premiums.

For these reasons, a firm should be very careful about issuing employee stock options. There are almost no good reasons to offer options to all employees. They only make sense for key employees, and only if the stronger incentives outweigh the additional complexity. Indeed, historically firms have granted options to employees other than top executives, except during the "irrational exuberance" of the tech boom of the 1990s.

Executive Pay

CEOs and top management are the most important employees. They have the most ability to create or destroy value. For this reason, the most important incentives for a firm to focus on are those for top management – and the Board of Directors who oversee them.

The evidence suggests that CEO and top management incentives do matter. Better incentives, and better controls such as governance and takeover pressures, do improve firm performance and make it more likely that management will make tough decisions.

Unfortunately, there are also some important reasons why top management pay may not be optimal. Unlike employees at lower levels of the organization, there may be little oversight of the CEO. This arises from the organizational design: separation of ownership and control. Because of the benefits of diversification, shareholders give up much of the control

of the firm they invest in. The mechanisms that exist to control top management are imperfect substitutes – largely because the market for corporate control, which is the way that shareholders are supposed to be able to exert some power over management – does not function ideally.

For this reason, CEOs often try to populate the board with sympathetic directors. Directors do not seem to emphasize oversight of management as much as is desirable. CEOs also have some control over the setting of their compensation packages. Despite these problems, governance and incentives do improve motivation of top management, so these mechanisms are important, if imperfect.

Review Questions

1. During the “Tech Bubble” when high-tech stock prices rose dramatically (before the bubble burst in March 2001), employees may have over-valued stock options because of “irrational exuberance.” In such a situation, should a company issue employee stock options? What are the benefits of doing so? The costs?
2. Should companies whose employees invest in more firm-specific human capital make more, or less, use of stock, options and profit sharing in employee incentive plans?
3. You are on the board of a company that needs to downsize dramatically. Doing so will cause great public criticism. What factors should you consider in hiring a new CEO, and designing his or her incentive plan?
4. You are a new CEO. What kind of people will you ask to be on your board? Why?
5. Some companies give their CEOs “golden parachutes” – large bonuses if the company is sold to an acquirer and the CEO loses his or her job. Does this practice sound like a sensible incentive scheme to you? Why or why not? What are the issues?
6. Are CEOs paid “too much”? In what sense? If incentives are important for motivating CEOs to increase shareholder value, is there any alternative?

References

Abowd, John (1990). "Does Performance-based Compensation Affect Corporate Performance?" *Industrial and Labor Relations Review* 43(3): 52S-73S.

Baker, George and Brian Hall (2004). “CEO Incentives and Firm Size.” *Journal of Labor Economics*.

Bertrand, Marianne and Sendhil Mullainathan (2003). “Enjoying the Quiet Life? Corporate Governance and Managerial Preferences.” *Journal of Political Economy* 111(5): 1043-1075.

Black, Fischer and Myron Scholes (1973). “The Pricing of Options and Corporate Liabilities,” *Journal of Political Economy* 81(3): 637-654.

Conger, Jay, David Finegold and Edward Lawler (1998). “Appraising Boardroom Performance,” *Harvard Business Review* 76(1): 136-148.

Hall, Brian and Thomas Knox (2002). "Managing Option Fragility." Working paper, National Bureau of Economic Research.

Hall, Brian and Kevin J. Murphy (2003). "The Trouble with Stock Options." *Journal of Economic Perspectives* 17(3).

Hallock, Kevin (1997). "Reciprocally Interlocking Boards of Directors and Executive Compensation," *Journal of Financial and Quantitative Analysis*, 32(3): 331-344.

Jensen, Michael (1986). "Agency Cost Of Free Cash Flow, Corporate Finance, and Takeovers." *American Economic Review Papers and Proceedings* 76(2).

Jensen, Michael & Kevin J. Murphy (1990). "CEO Incentives: It's Not How Much You Pay, But How." *Harvard Business Review*, May-June.

Kaplan, Steven (1989). "The Effects of Management Buyouts on Operating Performance and Value." *Journal of Financial Economics* 24(2).

Murphy, Kevin J. (1999). "Executive Compensation." In *Handbook of Labor Economics* 3b, ed. Orley Ashenfelter and David Card, Elsevier Science North Holland

Oyer, Pay & Scott Schaefer, 2005. "Why Do Some Firms Give Stock Options to All Employees? An Empirical Examination of Alternative Theories." *Journal of Financial Economics* 76: 99-133.

Appendix: Technical Aspects of Option Pricing

We first describe the Black-Scholes option pricing formula (Black & Scholes, 1973). Publicly traded stock options are often valued with some variant of this formula. Here is one version:

$$C = S \cdot N(d_1) - K \cdot e^{-rT} \cdot N(d_2),$$

where:

$$d_1 = \left[\ln(S/K) + (r + \frac{1}{2}\sigma^2) \cdot T \right] / \sigma\sqrt{T},$$

$$d_2 = d_1 - \sigma\sqrt{T},$$

and:

- C = value of call option
- S = stock price
- K = exercise price
- r = risk-free interest rate
- T = time to maturity of option
- σ = volatility (standard deviation) of S
- N = standard Normal cumulative probability distribution function.

How does the option's value vary with each parameter? Here is the intuition; largely speaking, similar intuition applies to how an employee values a similar option, though the

employee will demand a risk premium above the Black-Scholes value. All of the intuition is for the effects of changing a single variable, holding all other variables *constant*.

S: the higher the stock price, the more valuable is the option, since it is more likely to be in-the-money, and the payoff is higher if in-the-money. On exercising the option, the exercise price K is paid to buy the share of stock, which has value S , so the payoff if exercised immediately equals $S - K$.

K: the higher the exercise price, the less valuable is the option. The intuition is the reverse of the intuition for S .

r : the higher is r , the more valuable is the option. Think of r as a measure of how fast financial securities are rising per period (above the rate of inflation). If security values are rising rapidly, the option is more likely to end up in-the-money, and pay off more.

T: the higher is T , the more valuable is the option. More time before the option expires implies more chance that the stock price will rise, increasing the odds of being in-the-money and the payoff if that occurs. If you do not wish to hold the option until maturity, you can sell it in the market, if it is a publicly-traded option. Since the extra time has value, you can still capture the value of the higher maturity now by selling it. [Note: many employees exercise their options immediately upon vesting. This is because they are risk averse, and may be worried that the stock price will fall in the future.]

σ : the higher is σ , the more valuable is the option. A call option pays off if the stock price rises, but does not penalize if the stock price falls. Thus, greater volatility is valuable as it increases the likelihood that the option will end up in-the-money with high payoff.

Two other properties of option prices are useful for understanding the valuation of employee stock options. First, virtually all employee options are granted at the money; $K = S$. It can be shown that the value of an at-the-money option is greater, the greater is S , all other parameters held constant. The intuition is straightforward: the formula depends on the *percentage* return (this shows up as the risk free rate of return in the formula). Thus, changing S in the formula while keeping the percentage return constant is equivalent to increasing S *and* increasing the *absolute* rate of return. Of course, the actual value of an option depends on the absolute return, since this tells us the likelihood the option will end up in-the-money, and the payoff (intrinsic value $S - K$) if it does.

The second useful property is that the discounted present value of an option granted to an employee at some period is equal to *today's* value of the option. Thus, in valuing an option package, the employee does *not* need to discount the option to the present period. This intuition is not easy to see at first, since all other forms (except stock) must be discounted to the present. The intuition is easiest to see by thinking about a share of stock that pays no dividends. Suppose that the firm's current stock price equals S , and it promises to grant you one share of stock next year. What is the value of that future grant of stock? It is exactly equal to *today's* stock price S , since you could replicate that share by buying the stock today and holding it until next year. An option is a more complex function of the stock's value, but the same intuition applies.