Contents

Introduction

I: Sorting & Investing in Employees
   1. Setting Hiring Standards
   2. Recruitment
   3. Investment in Skills
   4. Managing Turnover

II: Organizational & Job Design
   5. Decision Making
   6. Organizational Structure
   7. Job Design
   8. Advanced Job Design

III: Paying for Performance
   9. Performance Evaluation
   10. Rewarding Performance
   11. Career-Based Incentives
   12. Employee Stock Options & Executive Pay

IV: Applications
   13. Benefits
   14. Entrepreneurship & Intrapreneurship
   15. The Employment Relationship
Introduction

What is Personnel Economics for Managers About?

Organizations and economies are human enterprises. They are the sum result of the motivations, decisions, and actions of many individuals. These individuals and their actions are combined to create innovation, higher economic growth, more job opportunities, and better products. The process by which this occurs is one of the miracles of modern economies and modern firms. It is also the topic of this book.

One can hardly overestimate the importance of understanding better how firms organize themselves and manage their employees. In large corporations, roughly three-quarters of all costs are human resource related. Similarly, roughly seventy percent of worldwide wealth is in the form of human capital – the skills and knowledge of individuals – rather than in physical or financial capital. Economies grow and change through the creativity and motivation of entrepreneurs and employees. The strategy of many firms today is explicitly human resource driven, emphasizing customization, service, and innovation.

Organization and management are also of fundamental importance to you. The topics discussed in this book take on increasing importance as your career progresses. Those at an early stage tend to focus on specialized areas of knowledge. As careers evolve, however, jobs tend to increasingly depend on supervision and management of others. A broader, general manager perspective becomes essential to coordinate the work of many. As the career evolves further, the ability to set up, structure, and manage the entire organization becomes important. A strategic overview of the organization, and how it relates to the firm’s goals and environment, is necessary.

In order for a manager to be effective at these various stages, it helps to have a structured, rigorous framework for analyzing the issues you will face. Gut instinct, common sense, and wisdom from years of experience can be extremely valuable. However, combined with a deeper understanding of the issues and tradeoffs behind them, you can become even more effective. The purpose of this book is to give you a rigorous framework for understanding organizational design and the management of your employees.

The study of organizations and human resources has not always been rigorous, but that is changing. Economics has proven to be a powerful approach in this area, adding rigor and structure, and clarifying many important issues. This area of economics is often called Personnel Economics, and much of the founding research in this area was done by Edward Lazear. This textbook is based on Lazear’s Personnel Economics for Managers (1997).
It may seem odd to apply economics to human resource and general management topics. In fact, it makes perfect sense. Economics is a methodology that has been applied to many areas of human activity, and has had enormous influence on the social sciences. That methodology is quite flexible, and can be applied to many problems of interest involving human behavior. The ability to apply a consistent framework allows us to develop a useful framework for studying organizational design.

**What is the Economic Approach?**

Economists recognize two elements that drive human behavior. One is pure psychology, or preferences. Understanding such preferences and their formation and evolution is the realm of classical psychology. The second is the environment in which people act to attain their goals. This is the realm of economics. Thus, economics focuses on budgets, prices, constraints, information, and incentives. It also focuses on social interactions, since an employee’s colleagues, manager, and customers play important roles in driving behavior.

This distinction between preferences and the environment is recognized in psychology. The subfield of social psychology generally focuses on the impact of the environment on individual behavior – just like economics. The fields of social psychology and personnel economics study many of the same issues, though from somewhat different perspectives. This also means that what we often think of as psychology is not, in the purest sense.

Because economics focuses on the effects of the environment on behavior, it generally starts with only crude assumptions about preferences of individual employees. This is more of a virtue than it might seem. The more abstract and general the model, the wider is its applicability. Thus, in economics we may assume that employees attempt to maximize their Pay. By Pay, we mean not only compensation, but also benefits, job amenities, work environment, and other things offered by the firm that they might value. A theory of pay for performance then has relevance for using any motivational tool, not just cash.

The key part of the economic approach is to focus on how the environmental variables – information, resources, constraints, decisions, and incentives – affect the outcome. Those are the issues that are analyzed in this book. More often than not, the analysis results in a statement of one or more important tradeoffs between benefits and costs that must be balanced.

Two results of this approach are worth noting here. First, the economic tools that we employ are used to analyze a variety of problems. This allows us to provide a more structured approach to the topics covered in this book. By the end of the book, we will be able to develop a framework for thinking about organizational design as a whole.

Second, economics focuses on variables that managers have a great deal of control over. The primary factors analyzed in this book are information, decisions, and incentives. These are exactly the levers that managers tend to have the most ability to pull to better design their organizations. It is much easier to alter the incentives than to change the psychology of your workforce.

It was mentioned above that economics and social psychology are different fields analyzing similar topics (organizational sociology could be added to this group as well). There is a great deal of healthy dialogue (competitive and cooperative, as it should be) between economists, social psychologists, and sociologists studying the issues covered in this textbook. The new field of personnel economics has grown out of this dialog. It started as a
small subfield of labor economics (the study of labor markets). It then incorporated new insights from informational economics, to start studying management of employees inside firms. Over time, personnel economics became more refined and more successful, and started to incorporate insights, evidence, and topics from social psychology and organizational sociology. (It can be argued that personnel economics is causing those fields to evolve as well.) Thus, while our approach and emphasis is economics, this book is more properly thought of as the result of an active debate between, and mixing of, different social sciences that study management issues.

Of course, this is not to say that this book is the beginning and end of study in this area. A full understanding of human resource management also requires the study of psychology. This text does not pretend to be the final word on organizational design. Rather, it is a strong complement to more traditional approaches, as well as a fresh approach for most students and managers.

**Who is This Book For?**

This book has several natural audiences. Undergraduates would benefit greatly from studying the book (in most cases, more than they would by taking a traditional labor economics course). Not only will they learn and apply ideas from microeconomics, such as incentive theory, but they will also learn principles that will be valuable in their careers.

Though the focus of the book is on personnel policies and organizational design, it is not written for a human resource specialist. Specialist texts focus on detailed examination of how to implement personnel policies, such as design of a pension plan or a performance appraisal form. Nevertheless, the text should be extremely valuable to an HR specialist, since it provides a strategic and analytical overview for human resource policies. It provides the broader perspective that is necessary before focusing on the details.

We both teach MBA students, and naturally write this book from that perspective. It provides a way to think about organizational design overall, as well as specific human resource policies. Since MBAs tend to become consultants, general managers, or run organizations themselves, the issues and approach used here are extremely relevant for MBA students. Executive MBA students should also find the book useful. It should help them take well-earned experience and common sense, and make it even more powerful and effective.

**Overview of the Text**

The textbook has three core sections, followed by a shorter section with applications and advanced summary discussions. The first few chapters may seem narrow, but that is intentional. In order to develop a rigorous approach it is necessary to simplify and build each idea carefully. As the book progresses, more elaborate perspectives are developed. The text builds and cumulates tools from chapter to chapter, and from section to section. By the end, a rich model of personnel and organizational design will be built. We now provide a brief overview of the contents of each section.

---

III
I. Sorting & Investing in Employees

In the first part of the book, employees are thought of somewhat mechanistically: inputs to production that can be rented or managed just like other inputs or assets. In this part of the text, employees have two important characteristics: innate abilities, and skills that they can accumulate in school or on the job. To the extent that employees have skills that are innate, the key issue is to sort them into the appropriate firms and jobs. To the extent that they can learn and improve their productivity, the key issue is to invest in them.

This part of the textbook focuses on the employee career pipeline – recruiting employees, investing in their skills so that they can advance, and managing employee turnover. The first two chapters consider issues involved in getting employees into the firm. Chapter 1 is a short chapter that considers how firms can think about setting hiring standards – the quality of the workforce that they wish to hire. Chapter 2 continues with the issue of hiring, focusing on the extent to which the firm should invest resources in recruiting, and how the job offer can be structured to improve the effectiveness of recruiting.

Chapter 3 begins to consider what to do with employees once they are inside the organization. It provides a framework for analyzing investment in employee skills through training (by the firm or the employee). This discussion also raises the issue of how to think about the complex economic relationship between the employee and the firm. That issue arises throughout the book, and is picked up again at the end.

Chapter 4 applies the economic tools from the first three chapters. These tools are important in understanding how a firm can manage turnover effectively. This chapter also illustrates how simple economic tools can have wide application.

II. Organizational & Job Design

As the textbook progresses, our model of employees becomes more elaborate. In this section, we consider what the employee actually does on the job. Chapter 5 begins the section by analyzing tradeoffs in decision making. An important idea used here is that an organization has to solve the same problems that an economy does, so the metaphor of a well functioning economy is useful for thinking about organizational design. This chapter uses these ideas to consider decentralization and centralization. Chapter 6 extends this discussion to questions of overall organizational structure.

Chapter 7 brings the discussion to the level of the individual employee’s job design. For the first time, employee motivation becomes a key issue. The idea of intrinsic motivation – that the type of work the employee performs has important implications for how, and how hard, they work – is explored. This chapter provides an overview of, and an explanation for, how job design has evolved in recent decades. It also links the psychological idea of intrinsic motivation to the economic idea of decentralization. Finally, Chapter 8 provides an analysis of several advanced job design topics, such as teams, and the effects of information technology on organizational structure.

III. Paying for Performance

The third core section of the text picks up on the theme of motivation from Chapter 7. The focus of this section is on extrinsic motivation – pay for performance. This section adds the perspective of incentive theory from economics to our toolbox. Chapter 9 analyses how an employee’s performance can be evaluated. Chapter 10 considers how such an evaluation
can be tied to rewards to motivate better performance, as well as pitfalls and implementa-
tion issues that commonly arise.

Chapter 11 provides a bridge between this and the first section of the text, exploring the re-
lationships between an employee’s career dynamics (such as promotions) and pay for
performance. Finally, Chapter 12 applies the ideas from this section to two important spe-
cial cases of pay for performance: employee stock options, and executive compensation.

IV. Applications

The first three sections of this book constitute the core material. In the last section, the
ideas developed in the text are applied to special topics that different readers may find of
interest. Chapter 13 analyses employee benefits. Chapter 14 discusses Entrepreneurship
and Intrapreneurship, or how to promote innovation and value creation in both startup and
mature organizations.

Chapter 15 picks up on a topic first raised in Chapter 3 – implicit and explicit methods for
contracting between a firm and its employees. Good organizational design involves both
formal and informal policies (for example, employee evaluation usually includes both nu-
meric metrics and subjective judgments). Here we extend that discussion. Chapter 15 also
pulls together several themes that have been developed in the book, providing a broader
perspective on the issues, and on how we are trying to help you think about organizational
design. Chapter 15 is thus an important part of the overall message of the book, and can
be read after the core chapters if there is not enough time to cover Chapters 13 and 14.

References


Chicago Press.


I. Sorting & Investing in Employees

In this first section of the text, we take a very simple view of employees. It is one that is quite similar to that taken in much of biology: nature versus nurture. In our context, employees bring to the workplace certain innate abilities, such as to think quickly or creatively, or to work with numbers. They also develop new or more advanced skills over time, through education, experience, and on the job training.

The topics of this section are how to sort employees by their innate or accumulated skills, how to invest further in their skills, and how to manage their exit from the organization, as a function of their talents and skills. One can think of a firm’s career policies as a kind of pipeline, bringing employees in, developing and promoting them, and transitioning them out eventually. That is the sequence of this section. In later sections of the book, we broaden our perspective, to consider issues such as the work they perform, their motivation, and the firm’s complex relationship to the employees.

In the process of exploring these issues, several important economic concepts are developed: asymmetric information, investment, and different methods of contracting.

Asymmetric information refers to situations where two parties to an economic transaction (in our case the firm and the employee) have different information that is relevant to the transaction. Problems of asymmetric information are ubiquitous in economies and organizations (e.g., the quality of a new hire; the effort that an employee expends on the job). They also tend to lead to inefficiencies, because incorrect decisions are made, lack of information creates risk, or because one party exploits its informational advantage for personal gain at the expense of overall efficiency.

When we consider recruiting, asymmetric information arises because the employee has more information about suitability for a job than the firm does (the opposite case can also arise sometimes). This presents a challenge for the firm in recruiting. We will see that one way to deal with this is to use the economic principle of signaling, which encourages the employee to use their information in a constructive rather than strategic way. The idea of signaling has applications in many areas of business, and we shall mention a few. This is an example of how the tools used in this book have broad application outside of employment.

The second economic tool used is the idea of optimal investment. Employees and their employers can invest in increasing their skills. In studying this issue, we will use the same ideas that play a key role in finance courses.

Finally, we will see three approaches to thinking about economic transactions or contracts. We start with the simplest – a spot market whereby the firm simply pays an employee’s market price at each point in time. This is the standard view in introductory microeconomics classes. But in trying to improve recruiting, we will soon see the need for more complex, multiperiod con-
tracts between the firm and the employee. These contracts will also be contingent, in this case on employee performance. Finally, in some cases we will see that the contract between the firm and the employee involves implicit or informal elements, because it is not always possible to write complete formal contracts. This gives us a useful framework for thinking about the overall employment relationship, and even issues such as corporate culture. Those ideas will be picked up again at the end of the text.

Now let us provide a small amount of structure to get started. The firm desires greater employee performance (contribution to profits), and lower labor costs. These must be traded off against each other. For the first 6 chapters of this book, employee motivation to work harder is ignored. Instead we will simplify the problem by assuming that the employee’s performance depends on innate ability, and the degree of their acquired skills (which we will call human capital).

In Chapters 1-2, we will be thinking about ways to sort employees on the basis of ability and skills. In Chapter 3, we then analyze how to invest in human capital. Chapter 4 applies the tools from Chapters 1-3 to questions of employee turnover. This not only gives a useful analysis of the turnover issues, but also shows how the tools can provide insight into new questions.
Setting Hiring Standards

“When you’re around someone good, your own standards are raised.” (Ritchie Blackmore)

In this chapter our goal is twofold: to introduce the topic of recruitment, and also to introduce the economic approach used in the textbook. Let’s ease into both by considering an example.

An Example: Hiring Risky Workers

New Hires as Options

Imagine that you are a partner in an investment bank in the City (financial district) in London, and are deciding between two candidates to fill a position as an associate (junior) investment banker. Gupta has the standard background of most of the applicants that you see, including a degree in economics, a few years of experience as a financial analyst, an MBA with a focus in finance, and a summer internship at an investment bank. You feel that his productivity is extremely predictable, and that he can produce £200,000 of value per year. Svensen has a very unusual background compared to other applicants. She has a strong track record and appears quite talented, but does not have much experience related to investment banking. Hence you feel that her success is much less predictable. She may turn out to be a star, in which case she will produce £500,000 per year, but she may also turn out to be a disaster, actually losing £100,000 per year. Suppose that each of these outcomes for Svensen is equally likely (fifty percent odds). Then expected (average) output from Svensen in any given year is exactly the same as the output from Gupta:

$$E_{\text{output from Svensen}} = \frac{1}{2} \times 500,000 - \frac{1}{2} \times 100,000 = 200,000.$$  

If the cost (wages, benefits, etc.) of both employees is the same, which is the better hire? The answer might seem counter-intuitive, but usually the firm should hire the riskier worker.

Suppose that both Svensen and Gupta can be expected to work at your firm for ten years. Suppose further that it takes one full year to determine whether Svensen is a star or not. The salary is £100,000 a year, and for the moment let us assume that this will be the salary for the foreseeable future. In that case, your firm earns a profit of £100,000 per year from Gupta, for a total value of £1 million over ten years. The top branch of Figure 1.1 shows this choice.

---

\(^1\) In this example, we ignore issues of present value, by assuming that the interest rate is zero to keep things simple. When we do this in examples in this book, it is always the case that the intuition that is developed would be identical if we used discounted present values. Similarly, all examples in this text use inflation-adjusted figures, since inflation does not affect the conclusions.
Alternatively, you can hire Svensen. With probability equal to \( \frac{1}{2} \), Svensen is a star, producing £500,000 per year, and your firm earns profits of £400,000 from employing her, for ten years, netting £4 million. With probability equal to \( \frac{1}{2} \), Svensen loses money for your firm. If this occurs, you can terminate Svensen at the end of the year, so £200,000 is the total loss, including her salary. These two outcomes are the remaining branches in Figure 1.1. Thus, the expected profit from hiring Svensen is:

\[
\text{Expected profit from Svensen} = \frac{1}{2} \times 4,000,000 - \frac{1}{2} \times 200,000 = 1,900,000.
\]

Svensen is therefore almost twice as profitable to hire as Gupta! Even though the two candidates have the same expected value, Svensen is worth much more. The firm can keep her if she turns out to be a good employee, and dismiss her if she turns out to be a bad one. The firm has the option of firing poor workers, and keeping the good ones.

This is the argument that is sometimes made for hiring workers with potential over conservative, proven ones. With the more proven worker, the firm gets a solid performer. With the risky worker, the firm may find that it made a mistake, but this can be remedied relatively quickly. It may also find that it has a diamond in the rough.

This simple example can be quite surprising to many students, since it seems to contradict the intuition that, if expected values are equal, risk is always a bad thing. However, risk is not a bad thing in the case of real options like hiring employees. It is a nice example because it illustrates how formal economic analysis can lead to better decisions. Our intuition tends to be the opposite of the correct answer in this case.

**Analysis**

The structure developed above suggests several other factors that are important in deciding whether or not to take a chance on a risky hire.

**Downside Risk**

The value of taking a chance on a risky candidate can be so large that it is often the better strategy even if the safe worker has a higher expected value per year. Even if Svensen might have been a total disaster, destroying £1,000,000 of value with probability \( \frac{1}{2}, \) it
would have paid to take a chance on her. However, the more the potential there is for an employee to destroy value, the less likely is it to be optimal to take a chance on a risky worker.

**Upside Potential**

Svensen was valuable because she could generate high profits if she turned out to be a star. The greater those profits, the greater the option value from a risky hire. Thus, in jobs where small increases in talent lead to large increases in value creation, hiring risky candidates will be even more valuable (as long as there is no increased downside risk as well). Think of an entrepreneur assembling a new management team. There is little to lose, but there may well be much to gain. In such a case, it will make more sense to take a chance on a risky candidate.

**Termination Costs**

The more costly it is to fire a worker, the more costly is a risky candidate. Nevertheless, it may still pay to hire the risky worker, and terminate in the case that the worker does not turn out to be a good fit, even if there are high termination costs. In most countries, firms are prevented from terminating workers at will. Legal or social restrictions can make the option of firing a worker after one year costly. Consider the extreme case where hiring is for life. If the firm is risk neutral (is willing to accept any risks, as long as expected values are equal), as long as Svensen’s expected productivity is equal to or greater than Gupta’s, it will be a profitable bet to hire Svensen. More generally, the benefits from the case where Svensen turns out to be a star are so high that it would often be worth hiring Svensen even if firing costs were high.

**Risk Aversion**

If the firm is risk averse, it may still be optimal to hire Svensen. Svensen will now be costly to the firm in a different way, because she is risky. However, the differences in expected productivity are quite large, and should more than compensate for typical levels of risk aversion.

---

**An Implementation Problem**

The issue of risk aversion raises an interesting side point. When managers and recruiting specialists are taught this example, their typical response is to reject its conclusions, saying that they would be more conservative in hiring. Why is this the case? Is the theory wrong, or are the managers wrong? Quite possibly it is neither. Rather, the analysis is conducted assuming that the firm is relatively risk neutral. However, decision makers are typically risk averse, and this will affect their decisions. For example, they might expect that they will be criticized or receive a poor evaluation if they hire a bad candidate for the job. The more risk averse they are, the more they will make decisions to avoid such an outcome.

To the extent that a manager’s risk aversion is different from that of the employer, this is an incentive problem or conflict of interest. This is a topic that we will address in Chapters 9-12. In the meantime, if those who make hiring decisions are too conservative, a possible solution to the problem would be to try to avoid punishing them when they make mistakes in hiring. Another would be to appoint less conservative managers to handle recruitment.
Length of Evaluation

The time that it takes to evaluate whether Svensen is a star or a disaster affects the value of hiring a risky candidate. If the evaluation takes ten years, there is no value to hiring Svensen, in our example. If the evaluation takes only one year, the firm can limit its cost of a disastrous hire to only one year of pay and poor productivity.

Length of Employment

The value of hiring Svensen would have been even greater if the firm could have employed Svensen for more than ten years. For example, if Svensen was 30 years old when hired, and stayed at your company (for the same salary) until retirement, the profit from hiring Svensen would be £14 million if she turned out to be a star (£400,000 per year × 35 years). This suggests that the value of a risky hire will usually be larger, the younger is the new hire, and the lower is turnover in the company (so that employees tend to stay with the firm longer).

A Counter Argument

Our conclusions are only as good as the assumptions behind them. An important element of the economic approach to personnel is careful consideration of when the assumptions do or do not apply, and of what the effect would be of changing key assumptions. In the model above, the conclusion rested primarily on one key assumption: that we can profit when we find a star employee. Let us reconsider this assumption.

If Svensen turns out to be a star, is it safe to assume that we can continue to pay her £100,000? Might she try to bargain for a better salary? Might other employers try to hire her away from us? What would happen to our argument if these considerations applied?

These questions bring up a crucial consideration throughout this textbook: the firm always has to match an employee's outside market value. More precisely, the firm offers a job package with many characteristics, including the type of work, extent of effort required to do the work, degree of training, pay and other benefits, possibility for further advancement, and job security. The employee will consider all elements of the package in valuing the job, and compare it to alternative jobs offered by competing employers. Firms must make sure that their job offers match those of competing employers in terms of pay and other characteristics.

For now, let us keep things simple, and focus on pay and productivity. Suppose that other employers can observe how productive Svensen is. Moreover, assume that Svensen's productivity as a star or a disaster would be the same at any other investment bank. These are reasonable starting assumptions for investment banking; the work is often quite public, and is similar at most firms.

When this is the case, if Svensen is revealed to be a star, other investment banks will be willing to pay her more than £100,000 per year. In fact, they should be willing to pay as much as £500,000 per year, since that is her productivity. Labor market competition will tend to drive competing employers toward zero profit from hiring Svensen.

If Svensen is a disaster, no investment bank should be willing to hire her. She is likely to find better employment in a different industry where her productivity is not negative.
What is the benefit to your firm of hiring Svensen in this case? There is none. In order to retain her if she is a star, you have to compete with other firms, and should end up paying about £500,000 per year. In other words, our conclusion that it would pay to hire a risky candidate rested on our ability to earn profit from Svensen if she turned out to be a star.

How can we benefit from Svensen? There are two possibilities.

**Asymmetric Information**

Competing firms may not figure out Svensen’s productivity, at least not immediately. Even though investment banking is often quite public work, some of it is not, and the work is also generally done in teams. Outside firms may find it difficult to estimate Svensen’s individual contribution because of these factors. This implies that in industries where productivity is less individualistic, and less public, hiring a risky candidate is more likely to be worthwhile. Furthermore, to the extent that your firm can delay the ability of the labor market to figure out who your stars are, it can profit from the informational advantage. Therefore, your firm may want to avoid public recognition of Svensen’s contribution.

The general issue of asymmetric information is one of the key economic ideas used in this book. When one side has important information that the other does not, interesting issues arise. For example, if a firm has trouble distinguishing between the best and worst job candidates, this can lead to ineffective recruitment, and also to a different approach to recruiting (that we will discuss in Chapter 2). Incentive problems are an important example of information asymmetry; they arise when the firm cannot monitor the employee’s actions perfectly.

**Firm-Specific Productivity**

Svensen’s productivity may be higher at your firm than at other firms. To the extent that this is true, Svensen may be a star at your firm, but less so or not at all at other firms. Then they will not bid up Svensen’s market value as much, and there may be profit to our firm from employing Svensen. In Chapter 3, we will discuss two important reasons why Svensen’s productivity may be higher at your firm than elsewhere: firm-specific job matches or human capital. The more important is either, the more likely is it to be profitable to hire a risky candidate.

One last issue here: even if there is profit from hiring Svensen compared to what she can earn from other employers, there is still a question of how that profit is split between Svensen and the firm. That raises the question of bargaining. We will not focus on that issue in this text. However, we do discuss it briefly in Chapter 3 when we analyze how to share investments in firm-specific human capital.

---

**Google’s Unusual Approach to Recruiting**

Google, a leading Internet search engine, competes with many high-technology companies for talent. It desires talented, creative employees that fit with its geeky and informal culture. To distinguish itself from other recruiters, Google sometimes tries unusual methods.

In one case, it inserted an "aptitude" test in technology magazines. It included questions such as, “How many different ways can you color an icosahedron with one of three colors on each face?” In another case, it placed billboards that read only, “(first
Such tactics serve three purposes. One is to provide information to recruits about the type of candidate that fits at Google. Another is to set the tone for future employees, since Google has a distinctive corporate culture. Recruiting is a crucial point at which to start defining the *implicit contract*, a concept that discussed in later chapters. Finally, the tactics get Google valuable attention in the press – and in textbooks.

*Source: Straits Times (Singapore), Fall 2004*

The example of risky hires is a good introduction to the economic approach to studying personnel issues. We used a very simple model to analyze a complex decision. The model was useful in guiding us to important issues to consider in making this kind of hiring decision. Once the structure was set up, we were able to study the problem formally, even expressing some ideas in a few simple equations. The equations can be an effective way to rigorously express certain ideas. We will use these techniques throughout the book.

Simplification of a complex problem makes it easier to solve the problem, and get concrete answers. Of course, too much or inappropriate simplification can lead to incorrect answers, so one must be careful. But when applied intelligently, simple economic modeling can lead to powerful and practical analyses.

When we analyze organizational issues in this textbook, we will see the same set of economic ideas appearing over and over again. By the end of the book, we will have an economic toolbox that can be used to analyze all sorts of personnel problems. You will see examples in Chapter 4, where we use the principles discussed in the first three chapters to analyze some specific personnel policies. Economic concepts that we used in analyzing risky hires include labor market competition (for employees); prices (salary); asymmetric information; and incentives. For those who have studied economics, this will be familiar territory: it is microeconomics applied to how firms are designed and their employees are contracted with.

### Setting Hiring Standards

Let us now step back and think about what hiring standards the firm would like to establish, before it actually begins recruiting employees. We will focus on a very simple case to generate some useful intuition. In the following, keep in mind that the firm’s objective is to maximize profit. We assume that there are no constraints on the firm’s ability to hire however much labor as it desires. Finally, we also assume that the price at which the firm sells its output, and the price per hour at which it pays employees, are constants.

### Balancing Benefits Against Costs

Managers often say that their goal in hiring is to obtain the best quality workers. It sounds like a good idea, but is it? The most productive workers are also likely to be the most ex-

---

2 An icosahedron has twenty sides, so you can color it with three colors as many as \(3^{20}\) ways (allowing for some that use only one or two colors). That equals 3,486,784,401. The first ten-digit prime number in consecutive digits of \(e\) is 7427466391.
pensive. Should the goal instead be to hire the least expensive workers? A simple analysis resolves this question.

Consider the hypothetical productivity data in Table 1.1. These data indicate that college graduates are about 28% more productive than high school graduates.

<table>
<thead>
<tr>
<th>Worker ID</th>
<th>Monthly sales</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$100,000</td>
<td>HS</td>
</tr>
<tr>
<td>B</td>
<td>$108,333</td>
<td>College</td>
</tr>
<tr>
<td>C</td>
<td>$125,000</td>
<td>HS</td>
</tr>
<tr>
<td>D</td>
<td>$125,000</td>
<td>HS</td>
</tr>
<tr>
<td>E</td>
<td>$133,333</td>
<td>College</td>
</tr>
<tr>
<td>F</td>
<td>$141,667</td>
<td>HS</td>
</tr>
<tr>
<td>G</td>
<td>$166,667</td>
<td>College</td>
</tr>
<tr>
<td>H</td>
<td>$175,000</td>
<td>College</td>
</tr>
<tr>
<td>I</td>
<td>$175,000</td>
<td>College</td>
</tr>
<tr>
<td>J</td>
<td>$183,333</td>
<td>College</td>
</tr>
</tbody>
</table>

Average, HS grad = $122,917
Average, college grad = $156,944

Productivity & Education of Hypothetical Employees
Table 1.1

Now consider the data on monthly wages for high school and college graduates in Table 1.2. College graduates cost more than high school graduates.

<table>
<thead>
<tr>
<th>Year</th>
<th>High School</th>
<th>College</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>$2,163</td>
<td>$2,850</td>
<td>1.32</td>
</tr>
<tr>
<td>1991</td>
<td>$2,140</td>
<td>$2,954</td>
<td>1.38</td>
</tr>
<tr>
<td>1992</td>
<td>$2,131</td>
<td>$3,075</td>
<td>1.44</td>
</tr>
<tr>
<td>1993</td>
<td>$2,128</td>
<td>$3,106</td>
<td>1.46</td>
</tr>
<tr>
<td>1994</td>
<td>$2,151</td>
<td>$3,116</td>
<td>1.45</td>
</tr>
<tr>
<td>1995</td>
<td>$2,088</td>
<td>$3,119</td>
<td>1.49</td>
</tr>
<tr>
<td>1996</td>
<td>$2,087</td>
<td>$2,999</td>
<td>1.44</td>
</tr>
<tr>
<td>1997</td>
<td>$2,173</td>
<td>$3,217</td>
<td>1.48</td>
</tr>
<tr>
<td>1998</td>
<td>$2,170</td>
<td>$3,373</td>
<td>1.55</td>
</tr>
<tr>
<td>1999</td>
<td>$2,208</td>
<td>$3,283</td>
<td>1.49</td>
</tr>
<tr>
<td>2000</td>
<td>$2,203</td>
<td>$3,354</td>
<td>1.52</td>
</tr>
<tr>
<td>2001</td>
<td>$2,202</td>
<td>$3,490</td>
<td>1.59</td>
</tr>
<tr>
<td>2002</td>
<td>$2,254</td>
<td>$3,593</td>
<td>1.59</td>
</tr>
</tbody>
</table>

Wages are expressed in 2005 dollars.


Wages for High School & College Graduates, U.S.
Table 1.2
If the firm has to pay its workers approximately the wages shown in the last row of Table 1.2, then both a high school and college graduate would be profitable to employ. (In a complete analysis we would add in any other costs of employing them, such as fringe benefits, workspace, etc. This example is simplified to focus on the general point.) Moreover, it would be more profitable to employ a single college graduate compared to a single high school graduate:

\[ \text{Monthly profit from hiring HS graduate} = \$122,917 - \$2,254 = \$120,663, \]

\[ \text{Monthly profit from hiring college graduate} = \$156,944 - 3,593 = \$153,351. \]

However, this analysis is misleading. Suppose that your firm wants to hire enough workers to produce $1 million in monthly sales. This would require 6.4 college graduates, at a cost of $22,995, or 8.1 high school graduates, at a cost of $18,257. Employing high school graduates would actually be more profitable. The reason is that high school graduates have lower cost per unit of output:

\[ \frac{W_H}{Q_H} < \frac{W_C}{Q_C}. \]

In this example, high school graduates cost about $18 per $1,000 of monthly sales, while college graduates cost about $23. As long as the expression holds, it is more profitable to hire high school graduates, and vice versa. The most cost-effective employee has the lowest ratio of salary to output. The firm should choose this type, and hire enough of them to reach the desired level of output.

This illustrates two simple but important economic principles. The first is to always think in terms of tradeoffs between costs and benefits. In this example, the desire for high quality workers must be balanced against their higher cost. Many questions boil down to assessing the benefits of a given policy against the costs.

The second general principle is to always compare your approach to your best alternative. In this case, college graduates are profitable, but less so than high school students. Once we take that into account, college students are not profitable compared to our other choice.

**Worker Choice at Days Inns**

Days Inns of America, a hotel franchiser, traditionally staffed its reservation center with young employees who were willing to work at the minimum wage. The staff answered phone calls from potential guests and booked reservations. However, due to a shortage of low-skilled workers, wages began to rise for this age and experience group. Days Inns decided to re-examine their targeted recruiting group for these positions.

Management realized that the sedentary nature of the job was well suited for elderly workers. In addition, senior citizens were a readily available work force that could be hired for only slightly more than young workers, when training and other costs were included.

\[ \text{Do not be troubled by the fact that this may require hiring a fraction of a worker. It is possible to hire a fraction of a worker – by hiring someone part time, or allocating them to other tasks part of the time. Moreover, the larger the scale of the firm, the less consequential are such issues of indivisibility.} \]
What happened? With productivity defined as a combination of average call length and number of reservations booked, older workers talked on the phone longer but made more reservations. The higher proportion of calls that resulted in actual reservations more than offset the additional time used. Elderly workers had a lower salary to productivity ratio, making them the most cost-effective labor. The savings were even higher because the older employees had a significantly lower turnover rate.

Source: McNaught & Barth (1992)

Foreign Competition

This analysis is useful for thinking about globalization of labor markets and the role of foreign competition. It is often argued that countries with low labor costs drive companies in countries with high labor costs out of business. Is that accurate? Table 1.3 shows representative statistics on pay and productivity (GDP, gross domestic product) for several nations. Mexico has the lowest labor costs, while Norway has the highest. However, the real issue is not whether labor is cheaper, but whether it is more cost-effective. For example, Japan's labor costs are among the highest, but its productivity is the highest. In fact, Japan has the second lowest labor cost per dollar of productivity. A firm that had a choice between hiring Japanese or Argentinean workers would prefer the more expensive, but also more productive, Japanese workers.

<table>
<thead>
<tr>
<th></th>
<th>GDP per worker</th>
<th>Annual salary, manufacturing</th>
<th>Cost per $ of GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mexico</td>
<td>$15,964</td>
<td>$5,743</td>
<td>0.360</td>
</tr>
<tr>
<td>Japan</td>
<td>$78,065</td>
<td>$33,573</td>
<td>0.430</td>
</tr>
<tr>
<td>New Zealand</td>
<td>$40,690</td>
<td>$18,067</td>
<td>0.444</td>
</tr>
<tr>
<td>Argentina</td>
<td>$22,399</td>
<td>$9,973</td>
<td>0.445</td>
</tr>
<tr>
<td>Norway</td>
<td>$85,923</td>
<td>$38,447</td>
<td>0.447</td>
</tr>
<tr>
<td>United States</td>
<td>$75,571</td>
<td>$34,682</td>
<td>0.459</td>
</tr>
<tr>
<td>Sweden</td>
<td>$55,680</td>
<td>$27,371</td>
<td>0.492</td>
</tr>
<tr>
<td>Australia</td>
<td>$45,357</td>
<td>$25,266</td>
<td>0.557</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>$54,848</td>
<td>$36,234</td>
<td>0.661</td>
</tr>
<tr>
<td>South Africa</td>
<td>$7,880</td>
<td>$7,828</td>
<td>0.993</td>
</tr>
</tbody>
</table>

Averaged over 2000-2, in 2005 dollars.

Source: United Nations

Productivity & Pay of Manufacturing Workers, Selected Countries

Table 1.3

These numbers are meant to be illustrative, rather than definitive. Still, they make clear the point that cheap labor is not necessarily low-cost labor. Similarly, high-productivity labor is not necessarily the most profitable labor. You should seek low cost per unit of output, whether that arises from low wages, high productivity, or both.

---

4 The numbers may contain some aggregation error. Wages are from manufacturing, but productivity is for the economy as a whole.
The Method of Production

So far, we have proceeded as if production was independent across employees. In reality, production is interdependent across employees. We now consider three scenarios representing different approaches to production, to see what effect the method of production has on our analysis. In the first, production is independent across workers. In the second, a worker’s production depends on the skills of co-workers. In the third, a worker’s production depends on the capital that worker uses on the job.

1. Productivity is Independent of Co-Workers

A manager describes production in his unit as follows:

My team is a sales force. Each salesperson works independently. The organization consists of my salespeople and me. What kind of worker should I hire?

Here each worker’s sales depend on his own ability and effort, irrespective of the efforts of other salespeople. This fits closely with the situation described above, so the choice between college and high school graduates is exactly the one that we have focused on so far. The simplicity of this case is best seen by comparing it to the next example.

2. Productivity Depends on Co-Workers

A second manager describes her production as follows:

This firm manufactures small appliances. We find it better to have a combination of worker types. High school graduates are cheaper and more cost-effective in the short run, but we find that we can’t keep their skill levels up without some college graduates around. The high school graduates forget what they knew. The college graduates keep the high school graduates sharp. So we like to have both kinds of workers. The problem is that I’m not sure about the appropriate balance.

Here workers interact with one another. This is much more typical than the first case, since in most workplaces many jobs are interdependent with each other. College graduates affect the output of high school graduates, and vice versa. Since college graduates are not only producing appliances, but also act as part-time teachers, part of their output consists of their effect on high school graduates.

The analysis above still holds, but output must be defined carefully. When measuring the output of college graduates, the number of high school graduates must be specified. Table 1.4 provides an example of the kind of information needed.
<table>
<thead>
<tr>
<th>Number of College Graduates Employed</th>
<th>100</th>
<th>110</th>
<th>120</th>
<th>130</th>
<th>140</th>
<th>150</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output</td>
<td>63.1</td>
<td>66.8</td>
<td>70.4</td>
<td>73.9</td>
<td>77.2</td>
<td>80.5</td>
</tr>
<tr>
<td>Number of High School Graduates</td>
<td>110</td>
<td>64.9</td>
<td>68.8</td>
<td>72.4</td>
<td>76.0</td>
<td>79.5</td>
</tr>
<tr>
<td>Employed</td>
<td>120</td>
<td>66.6</td>
<td>70.6</td>
<td>74.4</td>
<td>78.0</td>
<td>81.6</td>
</tr>
<tr>
<td></td>
<td>130</td>
<td>68.3</td>
<td>72.3</td>
<td>76.2</td>
<td>79.9</td>
<td>83.5</td>
</tr>
<tr>
<td></td>
<td>140</td>
<td>69.8</td>
<td>73.9</td>
<td>77.9</td>
<td>81.7</td>
<td>85.4</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>71.3</td>
<td>75.5</td>
<td>79.5</td>
<td>83.4</td>
<td>87.2</td>
</tr>
</tbody>
</table>

**Productivity (Per Worker) of High School & College Graduates Working Together**

Table 1.4

It is easy to see that output of a college graduate depends on the number of high school graduates working. For example, if 100 of each type are employed, total output is 63.1 units. If the number of college graduates is increased from 100 to 110, the gain in output is 3.7 units. However, if 150 high school graduates are employed, the gain in total output in going from 100 to 110 college graduates is 4.2. The gain from adding 10 college graduates is larger when there are more high school graduates around. Since college graduates train high school graduates, their services are more valuable when the firm has more potential “students” for them to teach. The larger the number of high school graduates in the workforce, the higher the value of adding college graduates to the workforce.

Similarly, the more valuable that high school graduates are, the more college graduates are employed. High school graduates are more valuable when the “classroom” in which they learn is less crowded. Thus the firm wants a balance of college and high school graduates. This example reveals the importance of interaction. This can be stated as follows: when workers interact on the job, a worker’s contribution to output includes the effect on co-worker output. As a result, it pays to hire better qualified workers when output is interdependent.

**3. Productivity is Independent of Co-Workers, but Depends on Capital**

A third manager describes the production process as follows:

*We are a large clothing company that has our men’s dress shirts produced by a factory in Malaysia. Each worker uses a sewing machine, which costs us $7.50 per day to rent. We can use skilled labor, which produces an average of 4 shirts per day, or professional labor, which produces an average of 6 shirts per day. Skilled labor costs $7.50 per hour, and professional labor costs $12 per hour. The sewing machine company says that it will rent us a new machine that doubles output per worker, but the better machine costs $16.50 per day to rent. Should I rent the new machine? What kind of labor should I hire?*

The analysis is easy once the relevant data are compiled, as in Table 1.5. First, consider the old machines. Without looking at the table, the manager might be tempted to not rent the new machines, since they double productivity but cost more than double compared to the old machines. But this ignores the fact that producing a shirt involves both machines and labor. Adding a new machine more than doubles the capital cost, but it does not double the total cost. There is no doubt that the firm should use the new machines.
Analysis of Productivity Using New or Old Machines

<table>
<thead>
<tr>
<th></th>
<th>Output</th>
<th>Labor Cost</th>
<th>Capital Cost</th>
<th>Total Cost</th>
<th>Cost / Output</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Old Machines</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skilled</td>
<td>4</td>
<td>$60.00</td>
<td>$7.50</td>
<td>$67.50</td>
<td>$16.88</td>
</tr>
<tr>
<td>Professional</td>
<td>6</td>
<td>$96.00</td>
<td>$7.50</td>
<td>$103.50</td>
<td>$17.25</td>
</tr>
<tr>
<td><strong>New Machines</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skilled</td>
<td>8</td>
<td>$60.00</td>
<td>$16.50</td>
<td>$76.50</td>
<td>$9.56</td>
</tr>
<tr>
<td>Professional</td>
<td>12</td>
<td>$96.00</td>
<td>$16.50</td>
<td>$112.50</td>
<td>$9.38</td>
</tr>
</tbody>
</table>

Furthermore, given that the firm is using the new machines, it should hire professional rather than skilled workers. When old machines are used, the cost per shirt is higher with professionals than with skilled labor. But when the new machines are used, the cost per shirt is lower with professionals than with skilled labor. When expensive capital is employed, it may be cost effective to use it intensely. Professionals use the machines more efficiently, which leads us to conclude the following: a firm should improve the quality of workers that it employs as it increases the amount or quality of its capital stock. More specifically, the optimal level of skill rises as the use of capital relative to labor increases.

This helps explain why the president of a firm should be very highly skilled. His or her labor is combined with the entire capital stock of the firm, in a sense. It makes no sense to waste the capital by placing it under the stewardship of a low-skilled individual.

We will see later in this book that the labor market has valued highly skilled workers relatively more over time. One explanation for this is that firms have made increasing use of valuable, and very productive, capital in the form of new information technology.

How Many Workers to Hire?

The answer to this question is straightforward. The firm should continue to hire workers so long as the incremental profit from hiring an additional worker is positive.

In our example at the beginning of this section, both college and high school graduates were profitable to employ, but high school graduates were more profitable – they produced more output for the same amount of compensation. We can combine the two decision rules: hire the type of worker that provides the greatest output per dollar of pay, or least cost per unit of output. (When doing so, take into account the effects of interdependencies with co-workers or capital, of course.) Continue to do so until the point where hiring more of that type of worker is no longer profitable.

This approach implies that there is a limit to the number of workers the firm should hire, because of the principle of diminishing marginal productivity. As more and more workers are added to an organization, the value of an additional one falls. Why would marginal productivity fall as you hire more workers? The main reason is that workers are combined with other resources: computers, machines, your time as their manager, etc. The more workers that you hire, holding other resources fixed, the more thinly are those resources spread across each worker. For example, if you have a small office with you, your staff, and three computers, as you hire more workers, each one gets less time on one of the computers, and less supervision by you, which tends to reduce their productivity. This logic holds for any resource that is increased while keeping another resource fixed.
Consider Table 1.6. As more workers are hired for the office, the marginal productivity (extra sales) from each additional worker declines. That is a typical pattern in any business. The table also illustrates the principle that you should hire workers up to the point where they are no longer profitable; that is, when marginal productivity is less than or equal to marginal labor cost.

<table>
<thead>
<tr>
<th>Number of employees</th>
<th>Total sales</th>
<th>Marginal productivity of employee</th>
<th>Total labor cost</th>
<th>Marginal cost of employee</th>
<th>Profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
</tr>
<tr>
<td>1</td>
<td>$100,000</td>
<td>$100,000</td>
<td>$14,404</td>
<td>$14,404</td>
<td>$85,596</td>
</tr>
<tr>
<td>2</td>
<td>$141,421</td>
<td>$41,421</td>
<td>$28,808</td>
<td>$14,404</td>
<td>$112,613</td>
</tr>
<tr>
<td>3</td>
<td>$173,205</td>
<td>$31,784</td>
<td>$43,212</td>
<td>$14,404</td>
<td>$129,993</td>
</tr>
<tr>
<td>4</td>
<td>$200,000</td>
<td>$26,795</td>
<td>$57,616</td>
<td>$14,404</td>
<td>$142,384</td>
</tr>
<tr>
<td>5</td>
<td>$223,607</td>
<td>$23,607</td>
<td>$72,020</td>
<td>$14,404</td>
<td>$151,587</td>
</tr>
<tr>
<td>6</td>
<td>$244,949</td>
<td>$21,342</td>
<td>$86,424</td>
<td>$14,404</td>
<td>$158,525</td>
</tr>
<tr>
<td>7</td>
<td>$264,575</td>
<td>$19,626</td>
<td>$100,828</td>
<td>$14,404</td>
<td>$163,747</td>
</tr>
<tr>
<td>8</td>
<td>$282,843</td>
<td>$18,268</td>
<td>$115,232</td>
<td>$14,404</td>
<td>$167,611</td>
</tr>
<tr>
<td>9</td>
<td>$300,000</td>
<td>$17,157</td>
<td>$129,636</td>
<td>$14,404</td>
<td>$170,364</td>
</tr>
<tr>
<td>10</td>
<td>$316,228</td>
<td>$16,228</td>
<td>$144,040</td>
<td>$14,404</td>
<td>$172,188</td>
</tr>
<tr>
<td>11</td>
<td>$331,662</td>
<td>$15,434</td>
<td>$158,444</td>
<td>$14,404</td>
<td>$173,218</td>
</tr>
<tr>
<td>12</td>
<td>$346,410</td>
<td>$14,748</td>
<td>$172,848</td>
<td>$14,404</td>
<td>$173,562</td>
</tr>
<tr>
<td>13</td>
<td>$360,555</td>
<td>$14,145</td>
<td>$187,252</td>
<td>$14,404</td>
<td>$173,303</td>
</tr>
<tr>
<td>14</td>
<td>$374,166</td>
<td>$13,611</td>
<td>$201,656</td>
<td>$14,404</td>
<td>$172,510</td>
</tr>
<tr>
<td>15</td>
<td>$387,298</td>
<td>$13,132</td>
<td>$216,060</td>
<td>$14,404</td>
<td>$171,238</td>
</tr>
</tbody>
</table>

**Marginal Productivity & Marginal Cost of Hiring An Additional Worker**

**Table 1.6**

The next-to-last column shows the marginal cost (compensation and other benefits) from hiring an extra worker. If this is less than the marginal productivity, profit rises from hiring more workers. If marginal productivity is below marginal cost (the lowest rows), profit can be increased by laying off some workers.

The general result is familiar to anyone who has studied economics: profit is maximized by using any resource, including employees, up to the point where the marginal benefits just equal the marginal costs.

**Other Factors**

**Availability of Workers**

In many communities, more high school graduates than college graduates are available. Does this mean that a firm should have a bias toward hiring high school graduates, as they are cheaper to hire? In most cases the answer is no. Most employers, even very large ones, employ a small part of the local labor force, so the total availability of workers is irrelevant. There are two exceptions, both of which arise when the firm employs a suffi-
ciently large part of the relevant labor market that its hiring has some effect on the market price. \(^5\)

The first case is where a firm employs a very large fraction of the local labor force (say, a factory in rural Thailand where there are few other employers). In that case, hiring more of a certain type of worker drives up the wage. The analysis is as before, but the firm needs to take into account the rising wage in analyzing output per dollar labor cost.

The second case is more important. When the type of labor being hired is very specific, the market for it may be thin (there are few buyers for this type of worker). If so, there may be significant search costs to finding a worker with the right skills. The wage must build in these amortized search costs, since it is a cost of employing this type of worker. Once that is done, the analysis is as above.

**The Firm’s Financial Condition**

Suppose that the firm is in financial distress. How should this affect its hiring decisions? Similarly, if the firm is having a very successful period, should this affect hiring? Once again intuition can be misleading here. None of the analysis above makes mention of the firm’s financial condition. Choosing the wrong kind of labor will only make the financial condition worse.

A firm in financial distress may have trouble paying employees because of cash flow problems. However, this is a financial problem, not a labor problem. The best solution to such a problem would be to arrange financing to cover the short-term cash flow problems, so that the firm can hire workers when it is profitable to do so. In fact, creditors should encourage this, if it increases profits, since it makes it more likely that the debt will eventually be paid off.

**Making Decisions With Imperfect Information**

Throughout this chapter, we have conducted our analyses based on data that either exist or were assumed to exist. Unfortunately, the information required are often not immediately available or are expensive to obtain. What can a manager do in such circumstances? There are three possibilities: (1) make a decision independent of analysis; (2) estimate the relevant information; (3) conduct an experiment.

**Make a Decision Independent of Analysis**

A frequent temptation is to conclude that the data are too difficult to obtain. The solution is then to simply guess an answer, using gut instinct, experience, or standard practice. Implicit in the guess are a number of calculations that are not made explicit, but are there nonetheless. This approach is easiest, but least likely to lead to an effective decision. Even a little formal thinking is likely to lead to a better result, and if that thinking is supported by some estimation of the tradeoffs involved, the decision is even more likely to be made effectively. This book is intended to guide you to more structured, and therefore better, decisions about organizational design.

\(^5\) This is the economic condition where the firm has monopsony power.
Estimating the Relevant Information

Rather than guessing at the answer, a manager may estimate the key numbers to determine the appropriate course of action. Such an approach is likely to lead to better decision making than by guesswork. Moreover, this book can aid such an approach.

Suppose, for example, that you develop a formal analysis of a personnel issue using the concepts in this book. This analysis will help you uncover the important pieces of information needed to make a good decision. The conclusions depend on certain pieces of information, such as the effect on productivity of high school graduates of working side-by-side with college graduates. When such information is not available, an estimate – even an informed guess – may be appropriate.

The structured approach also makes it easier for you to consider how robust the conclusion is, by varying estimates up and down. In some cases the correct decision may be the same for a wide range of values for the estimated information. In those cases, the right answer is clear. In other cases it may be that the correct decision depends critically on the specific value of the information. In those cases, it would be worth further expenditures to get a more accurate estimate before making any decision.

This approach may involve educated guesses, but the decision can often be improved further by using available data to estimate the tradeoffs under consideration. This is becoming increasingly viable with the dramatic fall in costs and rise in capabilities of computers. In the old days, firms kept personnel records on computer tapes that were difficult and slow to access. Data entry was costly and cumbersome, and data were rarely used, so firms had little incentive to maintain detailed databases. Now, firms often have detailed personnel records readily at hand, and database and other software to analyze the data easily. In addition, human resource professional associations and consulting firms can often provide additional data on practices, costs and effects across a wide sample of firms. It is becoming more and more feasible to provide estimates – sometimes rough ones, and sometimes fairly sophisticated ones – of the impact of organizational policies on desired outcomes such as turnover or profitability.

Experiment

The third option is to experiment. Sometimes this is easy to do and carries little cost. When data on relative productivity of different types of workers is unavailable, the firm might hire some of each type (perhaps even part time, or as temps) and measure their output. Similarly, when trying to figure out the correct commission rate in a sales force incentive plan, the firm might experiment with different rates in different locations, before rolling out a plan for the entire organization.

Sometimes experimenting is difficult and potentially costly. There are five questions that a manager can consider beforehand to determine whether experimenting is a viable option:

1. What are we trying to learn and why do we want to know it?
2. Will obtaining the answer have a large or small effect on profit?
3. What kind of data are necessary to answer the question?
4. How costly is it to obtain these data?
5. Will the data that we are likely to collect provide a reliable answer to the question?
Question 1 must be answered before any experiment can be run. Otherwise the experimenters may get so caught up in running the experiment that it becomes a purpose in and of itself.

In order to justify a major experiment, the answer to Question 2 must be that the potential effects on profits are large compared to the costs of the experiment.

Question 3 must have a well-defined answer. If it is difficult to specify in advance the kind of information required, experimentation is likely to result in money spent without useful results. Managers should be able to say in advance that if the results turn out one way, the decision will be made in a certain way. If they turn out the other way, the decision will be different. If this statement cannot be made in advance, it makes no sense to gather the data.

Question 4 must be asked to complement Question 2. If the cost of obtaining the data is large, it may not be cost-effective to undertake the experiment, even if the results will have a significant effect on profit.

The data are most valuable if they given an unambiguous answer to the question posed. If the data obtained contain a great deal of error, or if they only roughly approximate the kind of information needed to answer the question, the experience is less valuable.

Of these approaches, the first is almost always the least effective. If the analysis is complex, and information is incomplete or unavailable, you may be tempted to use intuition or gut instinct to make a decision. Your intuition and instincts are usually based on your experience, so they are not worthless. However, we have seen examples in this chapter where the analysis led us to some counter-intuitive conclusions. The point of the kind of formal analysis described in this book is to improve your decision making, by making it more rigorous, and clarifying the important (and unimportant) issues. Moreover, more formal thinking can help you realize the situations in which your experience guides you correctly, and incorrectly. Unfortunately, too often in this area managers rely on intuition, because the problems can be very hard to analyze formally and with structure. By the end of this book, you will have tools that will increase the effectiveness of your decisions about personnel and organizational design.

Summary

This chapter provided a simple, short introduction to the topic of hiring. We follow this up with consideration of the job offer in the next chapter. The primary purpose of this chapter was to get you started in thinking about organizational issues using economic tools. A little formal thinking can go a long way in clarifying issues, sometimes leading to surprising conclusions.

The chapter raised several issues. We started with a scenario of choosing between two job candidates, one with relatively predictable job performance and one more risky. To the extent that turnover costs can be avoided, this actually may be a good thing for an employer. This is because risky job candidates may have an option value. If she turned out to be a poor choice, the loss is limited as the firm can terminate the employee. The firm may be able to capture a substantial return if the employee turns out to be a star in some circumstances.
We then considered an employee’s output to wage ratio. The best worker is not the cheapest, nor the most productive, but the one with the highest ratio of productivity to cost. We should hire as long as the marginal productivity of the last worker hired is greater than or equal to the cost of the worker.

The chapter introduced some important economic ideas that you should begin to incorporate into your everyday thinking. First, never forget that you are constrained by market competition. In the case of personnel management, the job package that you offer your employees must be adequate to attract and retain the kind of employees that you want—especially if they are known to be stars. Second, always try to think in terms of tradeoffs and alternatives. When you are analyzing a decision, consider not only the benefits, but also the costs, and balance them against each other. The costs will be determined primarily by labor market pressures, which act as a constraint on optimal personnel policies. The benefits of employees depend on the production process; how they work, whom they work with; and what capital they work with. Some benefits and costs may be subtle or intangible, but they may still be important. One good example is that you should always include your best alternatives when making a decision. A choice may be profitable, but it may be less profitable than other options.

---

**Study Questions**

1. You are recruiting to fill a position in your firm. Should you try to “sell” the job to applicants by describing it in the nicest terms that you can? Should you make the job sound as desirable as possible? As a job applicant, should you try as hard as you can to convince the recruiter that you are the perfect candidate for the job? Think about these issues over the next couple of chapters.

2. Potential employees can be unusual in many ways. Can you think of any attributes of job applicants that would make them risky hires, but in a way that might suggest they have high option value as employees? In what ways might candidates be risky that would suggest their option value is not high?

3. Capital (including advanced information technology) can serve as a substitute for employees in producing a firm’s output. It can also serve as a complement, making workers more productive at their tasks. Which effect do you think is likely to be more important in practice? Why? Which kinds of jobs are most likely to be ones in which computers or machines can replace workers? Which are most likely to be ones in which they cannot replace workers, but can help workers perform their jobs better?

4. Many employment issues are very complex, and involve interpersonal relationships, psychology, or qualitative considerations. For this reason, they are often difficult to quantify. If you cannot quantify some of the issues that we discuss, are the tools developed in the text irrelevant? Why or why not?

5. After reading the introduction and this chapter, how would you characterize the approach of personnel economics? What distinguishes it from other approaches to studying organizational design that you are familiar with?
Appendix

The formal theory behind the conclusions in the second section of this chapter is the standard economic theory of production. Assume that firm output $Q$ is a function of high school graduate labor $H$, college graduate labor $C$, and capital $K$:

$$Q = f(H, C; K).$$

The firm maximizes profits by deciding how much of each type of employee to hire. For a given amount of output, the firm would choose between $H$ and $C$ to minimize costs:

$$\min_{H,C} W_H \cdot H + W_C \cdot C + \lambda [Q - f(H, C; K)],$$

where $W$ represents wage rates, and $\lambda$ is a Lagrange multiplier. The first-order conditions with respect to $H$ and $C$ are:

$$W_H - \frac{\partial f}{\partial H} \lambda = 0,$$

$$W_C - \frac{\partial f}{\partial C} \lambda = 0.$$

The multiplier $\lambda$ reflects the marginal cost of output for a given $Q$. Once $\lambda$ has been determined, the firm sets marginal cost equal to marginal revenue to determine the amount it wants to sell, which determines the optimal output $Q$. We now model the three scenarios described in the chapter.

1. **Productivity is Independent of Co-Workers**

The independence is easily modeled by an additive production function:

$$f(H, C; K) = [aH + bC]^z,$$

where we assume that $0 < z < 1; 0 < a < 1; \text{and } 0 < b < 1$. In this case,
\[ \frac{\partial f}{\partial H} = az[aH + bC]^{z-1}, \]

\[ \frac{\partial f}{\partial C} = bz[aH + bC]^{z-1}. \]

Since from the equations above we know that:

\[ W_H = \frac{\partial f}{\partial H}, \]

\[ W_C = \frac{\partial f}{\partial C}, \]

collecting results we have that:

\[ \frac{a}{b} = \frac{W_H}{W_C}. \]

All four of these terms are exogenous parameters (unless the firm is a monopsonist, in which case the W's are endogenous). This means that the first-order conditions cannot be met except by coincidence, so that a corner solution is the optimum. Either \( H > 0 \) and \( C = 0 \), or \( H = 0 \) and \( C > 0 \). If the left-hand side of the last expression is larger than the right-hand side, high school graduates should be employed, and vice versa.

2. Productivity Depends on Co-Workers

A good way to model this case is:

\[ f(H,C;K) = zH^aC^b. \]

Analysis similar to the previous case yields the condition:

\[ \frac{H}{C} = \frac{aW_C}{bW_H}. \]

Here we can have an interior solution, and the optimal amounts of \( H \) and \( C \) depend on each other.

3. Productivity is Independent of Co-Workers, but Depends on Capital

This case could be modeled similarly to case 1, but with the marginal productivities of \( H \) and \( C \) depending positively on \( K \). In this case, a corner solution would again apply, but the optimal amounts of \( H \) and \( C \) would be positively related to the level of \( K \). A similar approach could be applied to case 2, of course.
Recruitment

“I sent the club a wire stating, ‘Please accept my resignation. I don’t want to belong to any club that will accept me as a member.’” (Groucho Marx)

Introduction

In this section of the book we are interested in how firms bring employees into the organization, and the patterns of careers they have once they are there. Table 2.1 presents some data on these questions. The data are from confidential personnel records for all management employees in a U.S. firm, over a 20-year period.\(^1\) Since the data are confidential, we will refer to the firm as “Acme Incorporated.” We will present data from Acme in several chapters to help illustrate concepts.

<table>
<thead>
<tr>
<th>Hierarchical Level</th>
<th>% of total employees</th>
<th>% hired at this level</th>
<th>Number of years in</th>
<th>% who stay at Acme</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Current position</td>
<td>Only 1 year</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Acme</td>
<td>Only 2 years</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5-10 years</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>More than 10 years</td>
</tr>
<tr>
<td>1</td>
<td>25.4</td>
<td>99.0</td>
<td>2.3</td>
<td>2.4</td>
</tr>
<tr>
<td>2</td>
<td>26.2</td>
<td>31.0</td>
<td>2.5</td>
<td>4.5</td>
</tr>
<tr>
<td>3</td>
<td>25.4</td>
<td>31.0</td>
<td>3.0</td>
<td>6.0</td>
</tr>
<tr>
<td>4</td>
<td>20.5</td>
<td>27.0</td>
<td>4.1</td>
<td>7.9</td>
</tr>
<tr>
<td>5-8</td>
<td>2.5</td>
<td>19.0</td>
<td>4.0</td>
<td>9.7</td>
</tr>
</tbody>
</table>

Career Patterns at Acme, Incorporated

Acme is in a service industry. Its management ranks have 8 hierarchical levels, from entry (Level 1) to CEO (Level 8). Most management employees are in the first four levels. Level 1 is what is often called a port of entry; virtually all who work in jobs at this level were hired into Acme at this level. Since it is the bottom of the management job ladder, this should not be surprising, as demotions are very rare in most companies. At Levels 2-8, most employees were not hired from the outside, but instead were promoted from within.

Because of promotion from within, managers in upper levels have substantial experience at Acme on average. For example, Level 4 managers have almost eight years of experience in the company. It also appears that movement between levels (promotion) is more

---

\(^1\) The Acme tables are based on Baker, Gibbs & Holmstrom (1994a,b). The Acme data are from a single firm, but the patterns that we illustrate with the Acme data in this book appear to be fairly representative of the policies in many firms, in different countries. See the papers cited in Gibbs & Hendricks (2004).
rapid at lower levels, since the average number of years in the current job is longer in upper levels.

The last four columns give some sense of turnover and career length for Acme’s managers. There are two patterns. First, many leave Acme very quickly after being hired. For example, about 11% of those who are hired at Level 1 leave within the first year, while another 10% leave the next year. Conversely, if a manager stays at Acme beyond the first year or so, he or she has a good chance of staying for many years. For example, about a quarter will stay for between five and ten years, while about a third will stay for more than ten years.

Thus, there seems to be evidence of sorting in the first few years on the job. Almost a quarter of new hires will leave within two years, either because Acme decided it did not want to keep them, or because new employees decided they did not want to stay. Second, if employees survive the sort, they often enjoy careers at the company that last for many years. This suggests there is some value to having employees remain with the firm. These are issues we explore in this and the next three chapters.

With this introduction, let us now return to the issues we raised in Chapter 1. Once your firm has decided which types of workers it would like to hire, it must recruit for those types. There are two general issues. First, how can it weed out undesirable applicants? In some jobs, hiring the wrong type of employees can cause major problems, disrupting output and costing the firm not only wages but lost profits. Second, how can the firm attract the right types of applicants? Attracting the right type will reduce workforce problems, as well as recruitment and turnover costs. Put another way, your firm must sort new hires, just as Acme apparently does. How can you think about sorting for the most effective workforce?

**Screening Job Applicants**

One strategy for attracting good quality job applicants is to offer a high level of pay or benefits. This will generate a large pool of applicants, and higher quality applicants will be more likely to apply for the job than if pay was lower. Unfortunately, so will low quality applicants. The personnel office will be flooded with resumes, and only a small portion may be qualified. Some undesirable workers will slip through the hiring process and become employees, while some desirable workers will get lost in the shuffle and never hired. This is not a very useful approach just by itself.

The problem of the wrong type of applicant applying to the firm is called adverse selection.\(^2\) This is a general problem in economics, not just in employment. The problem arises because of asymmetric information. One party knows what type they are (in this case, a high or low quality job candidate), and the other does not. The one that knows uses this information strategically to personal advantage. A classic example involves used car sales. Owners know the quality of their used car. Those with good quality used cars are more likely to keep them, while those with low quality used cars are more likely to sell them. This implies that the quality of used cars is lower than it would otherwise be. It also implies that owners of high quality used cars may find it difficult to get a good price for their car, since buyers worry that it is a low quality car.

---

\(^2\) In 2001, George Akerlof was awarded the Nobel Prize in Economics for analyzing the problem of adverse selection. He shared it with Michael Spence, who was awarded the prize for analyzing the problem of signaling that we discuss later in this chapter. Joseph Stiglitz also shared the prize that year, for analyzing problems of asymmetric information.
Adverse selection arises in our case when the wrong kinds of workers are attracted to the firm. A number of approaches can be used to mitigate the problem of adverse selection in recruiting. Let us begin with a simple case first, the use of credentials.

**Credentials**

An obvious approach to weeding through resumes of job applicants is to look for credentials that distinguish some applicants from others. The most important ones are generally the type of experience (job and promotion history) the applicant has, the type of training (e.g., college major or MBA), and the quality of school the applicant attended. Indeed, these are almost always the most important lines on anyone’s resume. What makes a credential useful for hiring? Here are some considerations:

**Informativeness of the Credential**

Ability to perform well on the job must be positively correlated with ability to obtain the credential. For example, a university degree is a useful credential only if university graduates tend to be more productive at the job in question. A credential can be informative in two ways. First, it may mean that the holder of the credential has knowledge or skills that apply directly to a job. This might be the case with a CPA or MBA. Second, it may mean that the holder of the credential has innate abilities that tend to make one more productive on the job. An example might be a high score on an aptitude test, or winning a scholarship.

**Cost of Obtaining the Credential**

A valuable property of the credential is that it is relatively easy for well-qualified workers to obtain, compared to poorly qualified workers. When this is true, the credential is very likely to signal differences in ability. For example, it is not very difficult for a qualified accountant to pass the CPA (Certified Public Accountant) exam, but it is virtually impossible for someone with no training in accounting to pass. Thus, using the CPA as a screen effectively sorts between qualified and unqualified accountants.

On the other hand, a credential that is extremely expensive for all workers to obtain will not do well in sorting them. If a credential is very difficult to obtain, few applicants will have it. For a credential to be effective, it must be that most qualified applicants possess the credential while most unqualified ones do not. If a very small subset of qualified applicants has the credential, or a large subset of unqualified ones does, the credential is not helpful.

**Return on investment in the credential.** If the difference in wages between those who have credentials and those who do not is not very great, small differences in credentials will signal large differences in ability. For example, if the credential is education, and the increase in earnings from obtaining a college degree is small, only the most talented will get the degree. This is because they are the ones for which obtaining the degree is cheapest. When the rewards to obtaining a degree are large, even not-so-able people can be induced to get the degree.

We will see these ideas below when we discuss *signaling*. Signaling is one way of resolving adverse selection problems. In many cases, a high quality type can incur some costs to *signal* to others that they are high quality. If low quality types do not also invest in the signal, then it can serve to distinguish the high quality types from the low quality types. For example, an owner of a high quality used car can offer a warranty. It may also be possible for a high quality job applicant to signal this to potential employers. Before we discuss this idea, it is useful to consider simpler sorting issues first.
Learning a Worker’s Productivity

Suppose that you advertised an investment banker job as in Chapter 1. In response you received a set of resumes from job applicants. You looked through them, and selected a subset with appropriate credentials. In a job like investment banking, small differences in ability, personality, or other employee characteristics may lead to large differences in effectiveness on the job. Unfortunately, the self-selection of applicants to the job, and further winnowing by sifting through their resumes, makes the pool of remaining applicants look more and more alike. In general, the more that a set of applicants has already been sorted, the lower will be the variance between the remaining candidates. What should you do next?

You could hire one at random, and take a chance. However, given the stakes it is likely to make sense to expend some resources to screen them further.

There are a variety of methods that firms use to screen applicants. Some give job applicants tests, to see how they perform on specific tasks. This approach is more likely to work well for jobs with fixed, measurable tasks. It is not likely to work well for an investment banker. Many firms use psychological profiling. Unfortunately, this technique does not tend to work well in practice. One reason is that psychology is a highly inexact science. Another is that job applicants have an incentive to game the test, trying to sound like better employees than they are in reality. For example, one study found that 90% of job applicants who took one popular psychological test were able to inflate their “conscientious” score. Finally, virtually all firms conduct personal interviews of job applicants. Such interviews can vary from simple to elaborate. In the case of investment banking, applicants may be put through several rounds of interviews, eventually being flown to the company’s headquarters to meet with high-level partners over several days. Such a process can be extremely expensive.

All of these examples involve some costs (except when the firm hires applicants without any screening whatsoever). Consider the following example, and think about the extent to which your firm should invest resources in screening applicants carefully.

Screening Bankers

Table 2.2 shows productivity levels for five hypothetical types of job applicants (A through E) in two different firms, an investment bank and a commercial bank. Assume that the remaining applicants (after earlier rounds of weeding out) are paid about £100,000, so each bank expects to pay about the same salary to anyone it hires.

<table>
<thead>
<tr>
<th>Type</th>
<th>Investment Bank</th>
<th>Commercial Bank</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>£250</td>
<td>£95</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>C</td>
<td>125</td>
<td>110</td>
</tr>
<tr>
<td>D</td>
<td>200</td>
<td>120</td>
</tr>
<tr>
<td>E</td>
<td>450</td>
<td>125</td>
</tr>
</tbody>
</table>

Table 2.2

Screening Investment Banking Job Applicants

3 See Paul (2004).
Finding out what type a job applicant is has obvious value to each firm. The investment bank would want to avoid A and B types, because productivity would be lower than pay, while the commercial bank would want to avoid A types. Suppose that applicants can be put through a series of tests that cost £2,000 per person, and give definitive information on which type the applicant is. How valuable is such information? In other words, how much would each be willing to pay to screen workers before hiring? Table 2.3 provides figures to help us answer this (all numbers are expected values rounded to the nearest £100).

<table>
<thead>
<tr>
<th>Screen?</th>
<th>Productivity</th>
<th>Salary</th>
<th>Screening Cost</th>
<th>Profit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Investment Bank</strong> No</td>
<td>£110</td>
<td>£100</td>
<td>£0.0</td>
<td>£10.0</td>
</tr>
<tr>
<td>Yes</td>
<td>193</td>
<td>100</td>
<td>2.9</td>
<td>90.1</td>
</tr>
<tr>
<td><strong>Commercial Bank</strong> No</td>
<td>110</td>
<td>100</td>
<td>0.0</td>
<td>10.0</td>
</tr>
<tr>
<td>Yes</td>
<td>112</td>
<td>100</td>
<td>2.2</td>
<td>9.8</td>
</tr>
</tbody>
</table>

**Profitability of Screening at an Investment Bank & a Commercial Bank**

Table 2.3

With no screening, both banks would have £110,000 average productivity from each new hire, or average profit of £10,000.

With screening, the investment bank would reject types A and B, and accept only 70% of all applicants. The average productivity of the C, D and E types hired would be about £193,000, substantially higher than without screening. The screening cost per worker actually hired would be £2,000·10/7 (since the bank would hire an average of 7 out of every 10 applicants), or £2,857 per hire. Average profit from each new hire would rise with screening to about £90,100. The investment bank would profit greatly from screening applicants.

If the commercial bank screened applicants, it would reject those of type A, hiring 90% of all applicants. Average productivity would rise only slightly to about £112,000, but at a screening cost of £2,000·10/9, approximately £2,222, per new hire. Profit per new hire, net of screening costs, would fall to about £9,800. The commercial bank would not benefit from screening.

Why the difference? There are two reasons. First, the investment bank wanted to screen out three times as many workers as the commercial bank. The point of screening is to avoid hiring the applicants who would not be profitable. Second, the downside from hiring poor candidates was worse at the investment bank; some applicants would have produced nothing, and others would have destroyed value. The investment bank was more at risk from hiring the wrong type of worker.

This example motivates issues to consider when screening (see the Appendix for a formal treatment):

*Screening is more profitable when the test is more effective:* A test can be more effective in several ways. First, it can be cheaper to administer. Second, it can be more accurate. That is, it can correctly distinguish between desirable and undesirable job applicants a

---

4 You should remember from Chapter 1 that each bank would want to hire those with the highest productivity per £ of compensation cost. That is true; the firm will also want to keep hiring as long as expected productivity is greater than expected employee costs (profit from the hire is positive).
higher percentage of the time. No test is 100% accurate. Moreover, as noted earlier, job applicants often try to game such tests so as to appear to be a better candidate than they really are. Finally, an effective test is more discriminating. That is, it weeds out a higher fraction of candidates, recommending a smaller fraction for hiring. In the example above, the commercial bank’s screen was not very valuable because only 10% of candidates were rejected.

*Screening is more profitable when the stakes are higher:* The purpose of screening is to avoid the unprofitable candidates. Therefore, the greater the downside risk from hiring the wrong person, the more value there is to screening. Similarly, the longer that a new candidate can be expected to stay with the employer, the more valuable will be the screen. Firms that intend to hire employees for the long term thus tend to invest more in careful screening before committing to a new hire.

**Is Screening Profitable? For Whom?**

If the investment bank screens workers, productivity is much higher for its employees than for random job applicants. Now the same issue arises that we faced with our risky hire in the previous chapter, Svensen, when she turned out to be a star. The labor market will value our screened employees more highly, simply because we decided to hire them (they passed our screen). Therefore, it is not realistic to assume that we can continue to pay £100,000 at the investment bank, if productivity is almost double that. Other investment banks would bid away our workers, once they realized that we screen our employees carefully.

What will we end up having to pay our employees? It is hard to tell without additional information. It is even conceivable that we would have to pay as much as their productivity, £193,000, if the labor market is very competitive. Screening may not always be profitable to the employer; indeed some firms screen extensively, while others hardly screen at all.

What about job applicants? Why would they apply to a firm if they knew they would be screened? It must be that the potential higher pay, if they pass the screen, compensates for the trouble and risk involved in the screening. If the application process is not too difficult, then the extra compensation need not be too high to make it worthwhile for applicants to try. If the screen is extensive, such as probation (described below), however, job applicants may have to be compensated significantly in order to be willing to undergo the screening process.

When the firm cannot benefit much from the screen because labor markets are competitive, job applicants will have to pay for much or all of the screening. Of course, this happens already in the case of pre-job market screening such as education or professional certification. But it may also happen with on-the-job screening. Workers can pay for the screening implicitly, by their willingness to accept lower pay during the screening period then they would otherwise earn.

In any case, it is likely that both the employer and its employees will share the benefits (and costs) of screening. Firms that screen more extensively will tend to pay more, both because their employees are more productive, and because applicants will require some compensation for the costs and risks of trying to successfully earn long-term employment at the firm.

A further consideration arises when employees have some idea about whether they are high or low ability. Those who are high ability have a better chance of passing the screen,
so they have more to gain from the screening. Thus, they should be more willing to undergo and pay for the screening. We discuss this below under the topic of signaling.

**Probation**

The screening methods described above may be useful, but are imperfect. An important concern is that they are only proxies for what the firm really cares about – how the person actually performs the job. In many cases, the only way to truly tell if a job applicant is a good fit for the job is to have them *perform* the job itself. Thus, a final approach to screening is to have the job applicant do the job for some period, either very briefly during interviews, or more extensively during some testing period. The most elaborate form of this is to hire a worker for a probationary period, and only hire them long-term if their performance during probation is adequate.

Of course, a problem with probation is that costs of terminating employees can be substantial. In Italy, if the firm is found to have fired an employee who has worked for the firm long enough without legal cause, the firm must rehire the worker, pay lost wages and social insurance contributions, and pay a penalty to the government. In Indonesia, firms must pay workers severance of more than one month of salary for each year that the employee worked for the firm, up to 9 months of severance, plus 15% of salary as “worker’s rights replacement money.” By contrast, turnover costs are usually very small in Denmark. The general trend has been toward increased turnover costs, due to greater employment regulation and litigation over wrongful termination.

Where firing costs are high, employers can often still use probation, under a different form. For example, employees might be hired on a temporary basis through a temporary employment agency. Those who perform well could then be offered regular employment. Those who do not perform well do not need to be fired; they are just not hired from the temporary agency. Indeed, some temp agencies have an explicit strategy of serving as a screening agency of this kind for employers.

A similar approach that firms can sometimes use is to hire applicants on temporary (fixed term) contracts. When the contract term is over, the firm can elect to hire the worker more permanently, offer a new fixed term contract, or not rehire the worker at all. Such contracts are not just used for low skilled jobs; many firms hire high-level consultants in a similar fashion.

Evidence suggests that increasing regulation of employment is one factor behind the recent growth in temp agencies worldwide. For example, in the United States some employment regulations vary from state-to-state. Those states that regulate termination more strongly tend to be the ones where firms are more likely to use temporary workers. In Europe, the employment relationship is more highly regulated than in most other parts of the world, and hiring of temps is quite common. According to one study, 13 percent of all wage earners in the European Union were employed on temporary contracts. In Spain, the corresponding figure was 31%, and roughly half of workers under the age of 30 were employed on temporary contracts.
Reducing Firing Costs in France

In September 2005, the French government passed a new law designed to make it easier for companies with 20 or fewer employees to hire and fire workers. It specified a “New Recruit Contract” allowing such companies to lay off workers anytime in their first two years in the job, for any reason. Those laid off were to be given at least two weeks’ notice, and would be entitled to unemployment benefits, but would not have to be given the level of severance pay that is standard for other French workers.

Labor unions and opposition party leaders criticized the law, which was initially passed by decree under a new “emergency procedure” that allows the French government to enact employment legislation without consulting lawmakers. In April 2006, students, union members, and others marched in protest on the streets of Paris. In response the law was cancelled by President Jacques Chirac.

Source: Associated Press (2005-6)

Implications of Probationary Screening

If the firm uses probation to screen workers, and retains those who are good performers, there are several interesting implications. First, the firm is likely to promote those who survive the screening. They have been revealed to be more productive than average job applicants. Once this is evident, the firm is likely to give them greater responsibility.

Second, the system will generally be up-or-out, since those not promoted will typically not be rehired. This is much like the lower levels of promotion ladders that are seen in most professional service firms.

Third, a large raise in compensation will usually accompany the promotion. The firm promotes those it finds most productive, so a promotion implies that you are more talented than the average new hire, which raises your market value. For this reason, firms typically have to offer raises on promotion, or risk losing those promoted. Moreover, since promotion is based on performance, and performance depends in part on an employee’s effort on the job, the promotion will become a form of incentive pay. We return to these issues in Chapter 11.

Signaling

Most people have a good idea about their skills, work ethic, and ambition – the things that make them a good employee. Let us now assume that workers know what type of employee they will be. If workers know what type they are, and share this information honestly with employers, a firm could recruit employees of a certain type by simply putting up a notice that it is looking for workers of that type. Unfortunately, this is not likely to be an effective approach. Recall our discussion earlier in this chapter about offering high wages to high skilled workers. A firm that tried this would find itself facing adverse selection, since job applicants who were not highly skilled would be tempted to apply anyway. This is why we decided that some kind of screening would be necessary.
When workers know more about their employability than employers do, screening can be used to solve this adverse selection problem. After all, screening works by sorting workers, and keeping those who fit well and are most productive with the firm. It then pays higher wages to those who are screened out. Shouldn’t this attract those who are good candidates to apply in the first place, and deter those who are not good candidates? Let us consider an example to see how this might work.

Consider a simplified example of recruiting for an investment bank. Suppose that simple interviewing allows the bank to weed out Types A through C easily, but that it is much harder to distinguish between Types D and E without further screening. The bank would like to hire E types because they are the most profitable. Instead of screening, can it construct a job offer involving probation, up-or-out promotion, and a raise on promotion that is attractive to the E types, but not to the D types?

To model this situation, we need a little more information. First, let us assume that the bank can figure out what type an employee is after observing them on the job for one year. However, the accuracy of this judgment is not perfect: 10% of the time, the wrong decision is made. Thus, 10% of D types are promoted when they should not be (does this remind you of your boss?), and 10% of E types are not.

We also need to know what each type could earn elsewhere, since we have to offer a more desirable package to E types, but a less desirable package to D types. Assume that D types can earn £175,000 in other jobs, while E types can earn £200,000 in other jobs. Thus, their alternatives working elsewhere for two periods are twice these salaries, £350,000 for D types, and £400,000 for E types.

Finally, we need to know how long those promoted will work for us. To keep things simple, assume that they will work for one year after promotion. Table 2.4 provides these figures (rounded to the nearest £1000), and calculates the expected value of the job offer to each type, for different salaries in the two years W1 and W2.

<table>
<thead>
<tr>
<th>W1</th>
<th>W2</th>
<th>Type</th>
<th>Expected pay</th>
<th>Apply?</th>
<th>Expected pay</th>
<th>Apply?</th>
</tr>
</thead>
<tbody>
<tr>
<td>£200</td>
<td>£200</td>
<td>D</td>
<td>£350</td>
<td>yes</td>
<td>£400</td>
<td>no</td>
</tr>
<tr>
<td>180</td>
<td>225</td>
<td>350</td>
<td>360</td>
<td>yes</td>
<td>400</td>
<td>403</td>
</tr>
<tr>
<td>160</td>
<td>250</td>
<td>350</td>
<td>343</td>
<td>no</td>
<td>400</td>
<td>405</td>
</tr>
<tr>
<td>140</td>
<td>275</td>
<td>350</td>
<td>325</td>
<td>no</td>
<td>400</td>
<td>408</td>
</tr>
<tr>
<td>120</td>
<td>300</td>
<td>350</td>
<td>308</td>
<td>no</td>
<td>400</td>
<td>410</td>
</tr>
<tr>
<td>100</td>
<td>325</td>
<td>350</td>
<td>290</td>
<td>no</td>
<td>400</td>
<td>413</td>
</tr>
</tbody>
</table>

Motivating Self Selection of Job Applicants

Table 2.4

The first offer considered equals what E type applicants could earn elsewhere, £200,000 per year. This obviously attracts D types, but is no lure to E types. The second offer lowers pay during the probation period (W1), and raises pay after promotion (W2). Each row below further lowers W1 and increases W2. Because the promotion is not guaranteed, the firm must offer more than £400,000 in total pay W1 + W2 in order to attract E type applicants. For this reason, and reflecting the risk involved in accepting lower initial pay compared to
what E types could earn elsewhere, each subsequent row involves higher total pay for those who win promotion.

To calculate the actual values to D and E if they apply, note that each earns $W_1$ in period 1. In period 2, D types earn $W_2$ with 10% odds, and their alternative wage with 90% odds. Similarly, in period 2, E types earn $W_2$ with 90% odds, and their alternative wage with 10% odds:

Value of applying for D = $W_1 + 0.9 \cdot £175,000 + 0.1 \cdot W_2$;

Value of applying for E = $W_1 + 0.1 \cdot £200,000 + 0.9 \cdot W_2$.

The first two offers are attractive to D types. This is because pay is higher at the investment bank than they could get elsewhere, even during probation. Our first lesson is that in order to deter some applicants from applying, we must pay less than those applicants can get elsewhere before probation.

Similarly, the last few offers are attractive to E types, because the high pay after promotion is enough to compensate for the low initial pay, given the high probability that E types will be promoted. Our second lesson is that in order to attract some applicants, we must offer more than they could earn elsewhere after probation.

Thus, probation can generate good self selection of job applicants, thereby solving the adverse selection problem, if we pay a sufficiently low amount during probation, and a sufficiently high amount after probation. One way to understand this is to note that in effect the firm is demanding that each applicant *post a bond* – by accepting less than they could earn elsewhere – during probation. In return, if they perform well and are promoted, the firm will *give them a reward* – by paying them more than they could earn elsewhere. Figure 2.1 shows the type of contract that we are considering.

---

**Deferred Pay as a Screening Mechanism**

Figure 2.1
Note in Figure 2.1 that the higher-ability employees of type E receive a smaller reward on promotion, and pay a larger cost during probation, than do lower-ability employees of type D. The up-front bond \( W - W_1 \) is larger for E types, since their outside alternative is larger. Similarly, the deferred reward \( W_2 - W \) is smaller for E types for the same reason. If the D types would pay a smaller bond, and earn a larger reward if promoted, than how can this type of job offer deter them from applying, while motivating E types to apply? The answer is that the firm must put the employees through a rigorous performance evaluation before awarding promotion. The evaluation must result in sufficiently high probability that E types will be promoted, and sufficiently low probability that D types must be promoted. The low success rate for D types reduces the expected value of the job for them compared to E types.

This discussion illustrates the general economic idea of signaling. Signaling is a method that can sometimes be used to solve adverse selection problems. The high quality type signals his or her type to the market by incurring a cost. If low quality types are not willing to incur this same cost, then the signaling is effective: the fact that someone incurs the cost proves that they are the high quality type.

**Who Pays, and Who Benefits?**

Signaling only works if the incentives of job applicants are addressed: Type D's must be deterred, but Type E's must be motivated to apply. Thus, it is the employee who will pay for most or all of the cost of signaling, and enjoy most or all of the benefit of the signal. In Figure 2.1, we see that employees pay for the signal by accepting wages below what they could get elsewhere initially. They also pay in the sense that they incur some risk that they will not get promoted, even if they are high ability (if the screening is imperfect). They are rewarded later by earning more than they could earn elsewhere after promotion.

The employer may also pay for part of the probation, and get some of the benefits. Whether this happens depends on what average wages end up being compared to productivity. The firm benefits during the probation period from pay that is below productivity, but incurs costs when it pays more than productivity to those that are promoted.

**Examples**

An example of signaling is the seller of a used car who offers a warranty. The warranty is costly to the seller, and is the signal. The fact that the seller is willing to offer the warranty, while some other sellers are not, may signal that the car is of higher than average quality. In our employment example, Type E's can signal their type (and confidence in their ability to perform well and earn promotion) by their willingness to accept low pay in the first period. This only works if the D types are not also willing to accept the same contract.

There are many applications of signaling in the business world. For example, venture capitalists typically demand that entrepreneurs invest all of their family's personal funds in a new business venture. They may even demand that the entrepreneur mortgage his or her home, and invest the proceeds in the startup. At first glance this may seem odd – isn't the venture capitalist supposed to provide the funds? Demanding that the entrepreneur put some "skin in the game" is important, however, as it helps the venture capitalist separate out the most confident and serious candidates from the least.

Another example involves joint ventures between two firms. In such cases, it is common for both firms to invest some funds. One reason for this may be so that each firm can signal to the other the seriousness of their intentions to make the joint venture profitable.
As discussed above, probation is a form of pay for performance. It can serve as a signal if high ability job applicants are willing to accept a job offer with a risky but potentially lucrative promotion ladder. The same idea applies to any form of pay for performance. If a new hire is willing to accept more risky pay for performance, this may signal that they are higher ability, and believe themselves to be a good fit for the job. The opposite is true if they try to renegotiate to reduce the pay for performance. Thus, pay for performance helps sort employees in addition to motivating them.

**Education as a Signal**

Education is an important potential example of signaling. Assume for the sake of argument that students learn nothing useful in school. However, suppose that more talented students find it easier to learn the material quickly. If so, then they might be able to signal their talents to the labor market by investing in more education than less talented students. In this view of education, instructors require students to pass through increasingly difficult screens. At each stage, some find the cost of passing the next screen to be too high, so they do not get that level of education. Those who find the screen relatively cheap do enroll for that level of schooling. The labor market recognizes this, paying more to those who have obtained more schooling.

Indeed, as Table 1.3 in the last chapter showed, those with more schooling do earn more. Is this due to signaling? It is possible, but it seems extremely implausible that this is the only explanation. If the only purpose of education is to screen, we could probably find more efficient ways to do so than to have students go to college for four years, for example: we could give them a large test at the end of high school. In Chapter 3, we consider investments in new skills. Education clearly plays an important role in doing this.

That said, there is evidence that education does have some role in screening workers. For example, those who almost complete four years of college earn a little less than those who go a little bit further and complete their degree – there is a discrete jump in earnings associated with earning the formal credential. This is hard to explain by training alone.

**Signaling More Formally: Separating and Pooling Equilibria**

Let us consider a formal example of signaling so that we can see how it works. Suppose that junior accountants can invest in some education or on-the-job training. If they complete the training, they become Certified Public Accountants (CPAs). Assume that there are only two types of accountants, “quicks” and “slows,” depending on their ability as accountants: quicks are more productive, and also find it easier to obtain the training necessary to pass the CPA exam.

To formalize these ideas, denote the present value of the employee’s productivity as $Q$, and the cost of obtaining the CPA credential as $C$. Subscripts refer to the two different types of accountants.

Assume that the labor market pays accountants a salary that exactly equals their expected productivity. The fraction of accountants who are quicks is $\alpha$, so the fraction who are slows is $1-\alpha$. Thus, if the labor market cannot tell the two types of accountants apart (if there is no signaling), pay equals:

$$\text{Average productivity} = \bar{Q} = \alpha Q_q + (1-\alpha) Q_s.$$
On the other hand, if the quicks do succeed in distinguishing themselves, they will be paid their productivity \( Q_q \). Those who do not signal will then be assumed to be slows, so anyone who does not signal is paid \( Q_s \).

The quicks would like to distinguish themselves from the slows, so that they can be paid more. At the same time, the slows would like to be confused with the quicks, so that they can avoid being paid less. This is a general property of adverse selection models: the lower quality types generally attempt to associate themselves with the higher quality types, who conversely attempt to set themselves apart from the lower quality types. Will the quicks be able to signal their ability by obtaining a CPA?

In order for signaling to work, three conditions are necessary. First, if all other quicks are signaling, and slows are not, an individual quick must be better off with signaling too. This requires that pay net of the quick’s cost of the CPA is higher than pay if she decided to join the slows by not getting a CPA:

\[
Q_q - C_q > Q_s.
\]

Second, if all other slows skip the CPA and quicks invest in the CPA, an individual slow must be better off not getting the CPA as well. If a slow decided to infiltrate the quicks, he would be paid \( Q_q \), but it would cost \( C_s \). If he did not, he would earn \( Q_s \). Thus, for slows to skip the credential, it must be true that:

\[
Q_q - C_s < Q_s.
\]

These two together imply that:

\[
C_q < Q_q - Q_s < C_s.
\]

Intuitively, the gain from signaling must be higher than the cost for high ability types, but not so high that low ability types are also motivated to signal.

Third, for all of the quicks to be willing to signal, their profit from doing so must be higher than what they would get if none of them signaled at all. If none signal, everyone is paid average productivity, so this requires that:

\[
Q_q - C_q > \bar{Q}.
\]

This condition is even stronger than the other condition for quicks described above, since \( \bar{Q} > Q_s \). Very large \( \alpha \) implies that \( \bar{Q} \) is very close to \( Q_q \), making it more likely that this last condition cannot be met.

Intuitively, signaling to distinguish themselves from the crowd is more likely to be profitable for quicks the rarer that they are. If there are many high quality types, it is relatively easy for low quality types to hide themselves among the high quality group.

If these conditions are not met, neither has an incentive to obtain the credential, and quicks do not distinguish themselves from slows. In this case there is no signaling. Such a case is called a pooling equilibrium. This illustrates that signaling is not always possible.

If the conditions are met, quicks signal and slows do not. This is called a separating equilibrium, since quicks are able to separate themselves out from slows by investing in the credential.
These points provide a formal illustration of arguments we provided about screening earlier. Signaling involves screening, but adds an important new ingredient: workers know their type, and firms try to structure job offers so that those who are a better match for their firm reveal this by their willingness to accept the offer, while those who are a poor fit reveal this by refusing to accept the offer.

**Which Type of Firm is More Likely to use Signaling?**

Signaling is helpful when employers do not have enough information about job applicants to assess their potential accurately enough. It is useful when differences in talent among potential employees matter a lot to productivity. When differences in talent do not make much difference to productivity, signaling will not be very useful. These ideas suggest when we should expect to see employment practices consistent with signaling.

First, signaling should be more important in jobs where skills are most important. Such jobs tend to be those that are at high levels of the hierarchy, in research and development, and in knowledge work. They also correspond well to professional service firms, such as consulting, accounting, law firms, and investment banks. In such professions, even small differences in talent can lead to large differences in effectiveness on the job, so sorting for talent is very important. For this reason, such firms tend to screen very carefully at recruiting, and usually have promotion systems that correspond well to our probation story above, at least in the first few years on the job.

Signaling is also more likely to be used where there is not much information already available about job applicants. Workers who are new to the labor market (say, having just graduated from college or an MBA program) are more likely to see such policies. New hires with many years of experience, and an extensive resume of past accomplishments, should expect to see less use of signaling policies in the job offers that they receive. Nevertheless, firms can use these techniques even for hiring experienced talent at very high levels, when appropriate. For example, a new CEO is often hired for a fixed period contract, with extensive pay for performance. To the extent that the CEO's ability to implement the strategy is hard to determine, and the CEO knows more about this than the board's hiring committee, such practices can improve CEO recruiting.

**Summary**

An important objective of personnel policies is to sort talent to appropriate employers and positions in order to increase organizational effectiveness. Since there is incomplete and often asymmetric information during recruiting, firms must screen job applicants. When a high salary is offered, unqualified workers may apply for the job if there is a significant chance that they can survive. This is the economic problem of adverse selection.

There are a number of ways that a firm can cut down the number of undesirable applicants. One way is to look for credentials that are good predictors of on-the-job performance. This works well when the credential is easy to obtain for qualified people but difficult to obtain for unqualified ones.

Beyond credentials, firms can invest in more or less extensive screening when hiring. This can include formal testing, psychological profiling, lengthy and multi-round interviews, and trying workers out on the job briefly. All of these can be helpful, though it is likely that they will be far from perfect predictors of job performance.
The most accurate screen is to employ workers for some probationary period on the job. Of course, it may also be the most expensive. The worker must be paid during this period, and costs can be especially severe in jobs where there is a large downside risk in the worker’s job (ability to destroy value). Many firms use some form of probation, either formally or informally, in their hiring practices.

In our analysis workers were offered a contingent contract. In the case of probation, workers receive sufficiently low pay during probation so that only those who believe that they will be successful are willing to apply for the job. A well-crafted probation and post-probation salary schedule can keep undesirable applicants from applying while attracting desirable ones. This is easier to achieve when it is difficult for unqualified workers to sneak past the probationary period and when qualified and unqualified workers have similar outside opportunities.

Screening and probation introduced the economic idea of signaling. Signaling is a method that sometimes can solve adverse selection problems. If workers know their ability, qualified workers would try to make this known to employers by signaling, and unqualified workers would try to hide the fact that they are unqualified. Therefore, firms should design recruiting policies to encourage qualified workers to apply, while making it difficult for unqualified workers to get through probation.

Another possible way to induce good self-selection among job applicants is to use strong pay for performance. In fact, probation does this, since the post-probation promotion and raise in pay is contingent on good performance. More generally, incentive pay of any kind tends to improve recruiting, since better fitting employees are more likely to accept strong incentive pay.

Firms can use contingent rewards after probation to attract higher-quality workers. However, this is not free; it comes at the cost of higher wages. These policies are most likely to be useful in companies where small differences in talent can lead to large differences in productivity of employees. This is most likely at high levels of the organization, and in firms where intellectual work is emphasized. Professional service firms, especially leaders in their respective industries, often employ policies of this type. These include careful recruiting, extensive performance evaluation during the first few years of employment, and up-or-out promotion systems with large rewards to those who are retained. They may also include strong individual pay for performance.

In the classic view of economics, goods are bought and sold on spot markets, with the terms of the transaction consisting of quantity, quality, and price. Our analysis in this chapter has opened up a different view: when firms use probation or contingent pay to screen workers, they offer a multiperiod contract. This contract is contingent on the employee’s performance: how they are treated depends on how they perform. Finally, it also involves a promise by the firm – to reward those who perform well with higher pay later. These complications arise because the quality of the good – ability of the worker – is not readily available information. Thus, the economic relationship between the employer and the employee becomes complex. This idea is developed further in the next chapter, and even further still in Chapter 15 at the end of the text.
Study Questions

1. If workers might also be screened by universities, can a firm earn profit from screening them? Under what circumstances is screening more likely to take place on the job rather than before workers enter the labor market?

2. What kinds of firms are most likely to aggressively sort through workers and use up-or-out career systems? Why?

3. How might a firm signal to potential employees about important characteristics of the job it is offering? Can you give any examples?

4. Can you give any examples of signaling in other business contexts? How about that you have seen in your own life? In each case, what is the cost of the signal? In what way do those who signal differ from those who do not? Does the signal meet our criteria for a separating equilibrium?

5. Think about the French law allowing small firms to lay off workers at lower costs in the first two years on the job. Who would benefit from passing such a law? Would any workers (or potential workers) benefit? Were the students protesting at the Sorbonne in Paris (one of the most prestigious universities in France) likely to be affected by the law? How? What about unions?

References


Appendix

Screening

Here we provide a formal example of the principles of screening discussed in the chapter. Assume that there are two types of job applicants, E and D. Productivity is Q; E types are more productive than D types, so that $Q_E > Q_D$. The probability that a random job applicant is type E equals $p$; the probability the applicant is type D equals $1-p$. The firm pays wage $W$ to those it hires; $Q_E > W > Q_D$. Thus, the firm makes profit from E types, but a loss from D types.

Expected profit from random new hire = $p(Q_E - W) + (1-p)(Q_D - W)$.

The firm has a screen available that costs $s$, with accuracy $q$. That is, $q$ equals the probability that the correct decision is made, while $1-q$ equals the probability of a mistake.

Expected profit with screening = $p·q(Q_E - W) + (1-p)(1-q)(Q_D - W) - s$,

since a fraction $(1-q)$ of E types are mistakenly rejected, and the same fraction of D types are mistakenly hired.

The change in profits from screening compared to not screening equals:

$\Delta \text{profit} = -p(1-q)(Q_E - W) - (1-p)q(Q_D - W) - s$.

The first term is negative. It is the loss from mistakenly rejecting candidates of type E. The second term is positive, since $Q_D < W$. It is the gain from appropriately rejecting D types. Of course, the third term is also negative.

That the test is more effective when it is more accurate, cheaper, or more discriminating follows immediately:

$\frac{\partial \Delta \text{profit}}{\partial q} > 0; \quad \frac{\partial \Delta \text{profit}}{\partial s} < 0; \quad \frac{\partial \Delta \text{profit}}{\partial p} < 0$.

The more negative is the loss $Q_D - W$ from hiring the wrong type, the greater is the gain from screening. Therefore, the test is more effective when the stakes are higher. The longer that the wrong type will remain employed at the firm, the larger will this expression be.

Signaling

We now incorporate signaling into the probation model described above. We show how the wage must be structured each period to ensure signaling. There are types E and D, as defined before. The firm offers $W_1$ and $W_2$ in two periods. In period one, workers are observed on the job. Those deemed to be good fits are promoted and paid $W_2$, while the rest are fired and earn their outside pay. The promotion decision is made with accuracy $q$ as above.

The outside alternatives for each type are $W_E > W_D$. In assuming that there are different outside alternatives, we are taking a different tack than in the formal treatment of signaling.
in the chapter above. In effect, we are assuming that while this firm may be able to induce self selection of E types, we are begging the question about whether or not E types signal in other firms as well. This may not be the case. However, it seems reasonable to assume that E types can expect to have higher average earnings over their career than D types through some means or another, since they have higher ability.

In order to deter D types, but attract E types, we must meet both of these conditions:

\[
W_1 + (1-q)W_2 + qW_D < 2W_D.
\]

\[
W_1 + qW_2 + (1-q)W_E > 2W_E,
\]

The first expression says that D types expect to do worse at this firm. The second says that E types expect to do better. A little algebra shows that this scheme can induce self-selection if:

\[
W_1 < W_D + (1-q)(W_D - W_2) < W_D,
\]

\[
W_2 > W_E + (W_E - W_1)/q > W_E.
\]

In fact, the optimal wages (that minimize compensation cost) are:5

\[
W_1 = W_D - \left(\frac{1-q^2}{2q-1}\right)(W_E - W_D),
\]

\[
W_2 = W_E + \left(\frac{2-q}{2q-1}\right)(W_E - W_D).
\]

These imply that \(W_1 < W_D < W_E < W_2\), of course, so the scheme involves deferred pay. By inspection of either of the last two sets of expressions, it is easy to see that:

- The more accurate the test (larger q), the larger is \(W_1\), and thus the smaller is the bond posted by E types. Similarly, \(W_2\), and thus the reward on promotion, is smaller the more accurate is the test. Intuitively, a more accurate test makes E types more willing to accept a smaller reward for accepting the risk of signaling, since there is less chance of a mistake.

- The smaller is \(W_D\), the smaller is \(W_1\). The larger is \(W_E\), the larger is \(W_2\). Thus, the larger the differences in productivity between the two (reflected in their outside market values), the larger will be the reward on promotion.

---

5 Technically, there are infinite optimal wage pairs. For example, the firm could subtract $1 from \(W_1\) and add $1/(1–q) to \(W_2\). The wages provided here minimize the employee’s spread in wages from \(W_1\) to \(W_2\) among the set of wage pairs that minimize the firm’s compensation costs.
Investment in Skills

“An investment in knowledge pays the best interest.” (Benjamin Franklin)

Introduction

This chapter is about what you are doing right now: investing in skills and knowledge. Is Ben Franklin correct that this is a good investment for you? How can you tell? Should your employer pay for your education? Offer on-the-job training?

When asked about turnover, most firms express concern that they will lose their investments in employee skills. This suggests that they do offer some training, and would like to avoid turnover when they make such investments.

The data in Table 2.1 showed that turnover is high among Acme’s new hires, but that many others stay for a relatively long time. One interpretation is that Acme sorts new hires through probation. Those who do not fit leave quickly, and those who do fit stay for a long time. But could long careers for some employees be a result of on-the-job training as well?

Another interesting finding was that Acme tends to promote from within. This too might be explained by sorting: upper level positions are filled with current employees who have already passed the screen. In other words, new hires are more uncertain than internal candidates. Once again, though, could on-the-job training have anything to do with promotion from within?

Table 3.1 tests the idea that promotion from within is solely due to sorting. It compares the future performance of new hires and internal candidates that Acme uses to fill positions in Level 2. If internal candidates have already been sorted, we would expect less variation in their career performance after entering Level 2, compared to outside hires.

The data are consistent with our hypothesis (patterns are also similar in higher levels of Acme’s hierarchy). For example, outside hires leave Acme with higher probability than do internal candidates. This is strong evidence that new hires still need to be sorted. Of those who stay, new hires are more likely to be demoted, and less likely to be promoted. However, when they are promoted, they advance further on average than do internal candidates who are promoted. In other words, new hires are more likely to have extreme outcomes: demotion, exit, or rapid promotion. They are more variable in their value to Acme compared to those Acme hired at Level 1 and promoted to Level 2.
Performance of New Hires & Promoted Employees at Acme

New hires may be valuable to Acme because they are risky and have an option value, as described in Chapter 1. But do they differ from Acme’s internal candidates in other ways as well? Table 3.2 provides some evidence on this question, for outside and internal candidates in Levels 2-4. Acme’s new hires tend to have about a half to one more year of education, and several more years of work experience, than those who are promoted to similar jobs from within Acme. In other words, their average education and experience is higher.

Human Capital of New Hires & Promoted Employees at Acme

What can explain this observation? One possibility is that Acme is risk averse. In order to be willing to hire a risky candidate from the outside, Acme might require that their credentials be superior to internal candidates. However, even risk averse employers might well be willing to hire risky candidates because of their option value.

An alternative explanation is that employees who have been at Acme for a few years have a different advantage: they have been trained by Acme in ways that increase their productivity. If the training is specialized to Acme’s business, outside hires would not have this knowledge. This would give internal candidates an advantage in filling job slots. If so, new hires would have to be better along other dimensions, like general work experience, in order to be considered for the job. Thus, some of the patterns that we are seeing may well be due to training.

Until this point we have assumed that workers have fixed talents, in order to analyze implications of sorting your workforce. We now add a new and important consideration: workers learn over time, both through formal education and on the job. Moreover, the Acme data suggest that such training might improve a worker’s productivity at their current employ-
er more than at other firms, at least in some cases. Here we develop a framework for thinking about training overall, and how training might have different effects on productivity in one firm compared with others.

### Matching

Before we consider investments in skills, there is another explanation for the career patterns we have seen that only involves sorting: *matching*. Suppose that, because every firm is different in its business, organization, and corporate culture, employees with similar abilities will not fit equally well at the same employer. If so, employees and firms need to seek out good matches with each other. Two employees of similar ability may be more productive in two different firms. Perhaps one firm has an aggressive culture that expects employees to work long hours and on weekends, and one worker is well suited to that environment while the other is not.

If such matching is important, workers must be sorted, just as when they differ in ability. However, the sorting would be on whether or not a worker’s attributes other than skills match well with a firm’s (or job’s, within a firm) attributes. ¹ This would include factors such as the worker’s portfolio of different skills, personality (and how it fits the company or workgroup’s culture), or locational preference. These factors are much of what is meant when people talk about their fit with an organization.

Matching would imply high turnover early in the career, as workers and firms test the relationship to see if there is a good fit, and low turnover later. It might also imply that workers experience increases in pay once they are found to be a good fit, if pay rises after probation.

Furthermore, this is just another form of sorting, so outside hires would have greater variance in employment outcomes than internal candidates. Finally, internally promoted candidates would, on average, have a better match with the firm than would outside hires, since internal candidates have already survived screening. This would put them at an advantage, which external candidates could only overcome if they had stronger credentials along other dimensions, such as more work experience or education.

The sorting story and its more subtle counterpart, matching, are based on the assumption that the employee’s productivity does not change on the job. Nevertheless, people do acquire new skills through education and on the job training, so trying to explain career patterns without such considerations is unrealistic.

### Investments in Education

Economics, and increasingly the business world, view education and training as investments which can be modeled just like any other kind of investment. The analysis of investments of this kind is called *Human Capital* theory. It is such an important part of mod-

---

¹ Given this description, it should not be surprising that the metaphor of dating and marriage is often used by economists when analyzing the topics that we discuss in this chapter. In fact, similar principles have been employed by economists to study marriage and related topics.
ern economics that two Nobel Prizes in economics were awarded, at least in part, in rec-

ognition of this theory.\(^2\)

One argument stated earlier is that education does not provide any actual learning, but only signals about student's abilities. That view is far-fetched. The human capital analysis of education provides a much more realistic way to think about education.

Human capital can be acquired in many ways. Examples include investments in exercise and health care. The two most important for our purposes are education or other pre-labor market training, and on-the-job training. Here we analyze education; in the next section we analyze on-the-job training.

In capital theory, investments are made if the present value of the cash flow or other benefits generated by the investment exceeds the present value of the costs of the investment.\(^3\) Let us formalize our thinking. Suppose that an individual is choosing whether to drop out or finish college this year, which we will call period 0. Future years will be denoted years 1 \(\ldots\) T, where T is the last year of her career.

If the student drops out now, her earnings in future years will be \(H_t\), where the subscript \(t\) refers to future periods. If she continues on in school, earnings in future years will be \(K_t\). Given this, the increased earnings from finishing school are \(K_t - H_t\) each year.

Education provides many benefits beyond increased earnings. One is the pure joy of learning. Education may also make you more effective at home or leisure activities, or increase your enjoyment of travel or literature. We focus on earnings because it is the most important benefit in the labor market. However, the arguments easily incorporate any non-pecuniary benefits from learning. If there are non-pecuniary benefits, their value should be included in \(K_t\). They would simply increase the return on the investment, and should be included in any investment decision.

Suppose that the interest rate is \(r\) per year. This means that an investment of \$1 made today would be worth \((1+r)\) next year, \((1+r)^2\) after two years, and so on. Similarly, \$1 received next year is worth \(1/(1+r)\) this year.

With these assumptions, the present value of the return on the education investment (again, focusing only on financial gains from education) is:

\[
\text{Return on Education (present value)} = \sum_{t=0}^{T} \frac{K_t - H_t}{(1+r)^t}
\]

There are two costs of investments in education. The first is the direct cost of tuition, this textbook, supplies, etc. Denote this by \(C_0\). The zero subscript calls attention to the fact that direct costs for training investments are generally borne up front and do not need to be discounted.

---


\(^3\) For those who are not familiar with the term, present value refers to the value, from today's point of view, of costs or revenues that accrue later. Income earned next year is worth less now, because you are not able to use the funds for a full year. Interest rates are the way that economies price cash flows in future periods. For example, if you invest \$100 dollars today in a CD (certificate of deposit) earning 5% interest, it is equivalent to \$105 next year. Put another way, the present value of \$105 next year is \$100.
The second cost is the opportunity cost of the time spent on education. For example, typi-
cal (full time) MBA students quit relatively high-paying jobs to go back to school for eigh-
teen months. When they do, they give up salaries that in many cases are greater than the
direct cost of tuition. Even part-time students can incur important opportunity costs: they
give up evenings and weekends to study and attend classes, and have less leisure and
vacation time than they would otherwise. If the student could work at home, on the farm,
or in the family business, there would still be opportunity cost even if there is no salary.
This is because the work has value to the family, and the family may need to pay to get
someone else to do the work. Any proper analysis of an investment, including in training,
should include opportunity costs in addition to direct costs.

Therefore we need to include the (additional) earnings that a student would receive if she
were to drop out before completing college. If she stays in college, she will not receive
those earnings. Let us call that \( F_0 \). The total cost of the remaining investment in college
equals \( C_0 + F_0 \).

The decision rule for any investment is that it should be made as long as the present value
of the return on the investment exceeds the present value of the cost of the investment.
This net present value equals:

\[
Net\ Present\ Value\ of\ Investment\ in\ Education = \sum_{t=1}^{T} \frac{K_t - H_t}{(1 + r)^t} - (C_0 + F_0).
\] (3.1)

When the expression in equation 3.1 is positive, finishing college is a good investment.
When the expression is negative, college is not a good investment. Put another way, if the
costs are larger than the increased income, she would be better off by dropping out of
school, going to work, and investing the tuition cost \( C_0 \) and extra earnings \( F_0 \). The interest
earned would exceed the increase in earnings by finishing school, in present value.

For early years of schooling, the returns to schooling exceed the costs. There are two rea-
sons. First, there is much to be learned when an individual knows very little. A little bit of
school can affect productivity dramatically, but gradually diminishing returns usually set in.

Second, the costs of going to school are very low during the early years of schooling. With
public subsidies to education, direct costs \( C_0 \) are virtually zero up through high school or
college in most societies. Furthermore, foregone earnings during the early years of school-
ing \( F_0 \) are very low (though they are not zero, because even children could work in a fami-
ly business or in household production).

Eventually, however, the reverse must be true. Costs of schooling could exceed the (fi-
nancial) returns. Consider an actual former Executive MBA student of one of the authors.
He had a college degree. He was both a doctor and a lawyer, each of which requires its
own advanced degree. After earning his MBA, he decided to enroll in a PhD program. The
extra degree was unlikely to improve his earnings, and would never allow him to recoup
the direct and opportunity costs of the PhD. At that point, the educational benefits were
purely consumption. In his case, the non-pecuniary benefits of education were very high.

This logic implies that it pays for almost everyone to invest in some formal education, but
that there is also an optimal stopping point for each individual. The stopping point is the
year when the net present value of investment in education calculated in equation 3.1
switches from positive to negative.
Effects of Costs and Benefits

Expression 3.1 has several other implications that are borne out in practice.

Costs

Increases in tuition or other costs reduce enrollment. The reason is that any students who were close to the margin (net present value of education close to zero) will now find that the costs exceed the benefits.

A related point is that those who already have high-paying jobs will be reluctant to go back to school, all else equal. For this reason, universities and MBA programs generally have rising application rates when an economy is in recession, and falling application rates when the economy is doing well. Education is a better investment when your labor market opportunities are weaker, and vice versa, because this is the opportunity cost of schooling.

Interest Rates

Increases in interest rates mean lower optimal levels of schooling, just as with any investment in which the returns are realized in the future. The higher the interest rate, the more that future earnings are discounted from today's point of view.

That said, interest rates do not have dramatic effects on schooling decisions, for two reasons. One is that schooling is an investment with a long payout period, so long-term interest rates are more relevant than short-term rates, and they are generally less volatile. Moreover, the rate that is often important for schooling decisions is the implicit borrowing rate that parents charge their children to finance schooling. What parents extract in-kind or through direct transfer from their children later (if anything) may not be closely tied to the interest rates that were in effect when the child was in school.

Career Length

An additional set of implications of Equation 3.1 involves the term T. The longer the work life, the larger is the optimal investment in schooling. Thus, people are more likely to invest in schooling when they are young, because they expect to be able to enjoy the return on their investment for longer.

The same logic predicts that women will tend to invest less in schooling than men, even though women tend to have longer life expectancies than men. The reason is that the average woman spends less time in the labor market than the average man, which reduces the return on investment in education. This point can be extended. The primary reason that women tend to spend less of their career in the labor market is fertility. This often leads to careers that are interrupted for a few years, and then continued. Thus, women have some incentive to focus their education relatively more on occupations in which their skills depreciate less quickly over time.

---

4 Women reading this book are not likely to be very different from the men reading it, in terms of labor force behavior. Women who specialize in advanced fields have already made clear by their investment behavior that they plan to participate actively in the labor market. Furthermore, their high wages will help fuse them even more to the labor market.
**Specialization of Human Capital**

Most students eventually specialize their education, for example by focusing on a major area of study in college. Further education beyond college is usually even more specialized, with almost all classes focused on one field. Why is that the case? Education, like most investments, tends to have diminishing returns. That is, one more year of study in a particular field tends to have less impact on mastery of the field than did the previous year. This would seem to indicate that people should invest in a portfolio of skills rather than focus, to avoid the problem of diminishing returns. Indeed, we do, at lower levels of education. Almost all education systems require every student to develop a general education with a little knowledge of many different subjects. It is only at relatively advanced stages of knowledge, in education and on-the-job training, that specialization becomes important.

People generally specialize advanced education because of one of the most important factors in an economy: *comparative advantage* and *gains from trade*. We already mentioned this idea in Chapter 1. If individuals focus on one area and become relatively expert, they can trade their output with those who specialize in other areas of expertise. We all benefit from the fact that advanced chemists develop new products and pharmaceuticals. They benefit from the fact that we specialize in other areas of work. Neither group has to learn about all topics.

In other words, it is usually true that the effective learning in one field gradually falls as more time is devoted to studying that field. However, the economy tends to reward relatively highly incremental advanced knowledge in many fields. Thus, it usually pays to specialize one’s advanced education.

Specialization is an important issue inside firms, with implications for both organizational structure and job design. We will return to this topic in Chapters 6-7.

**Effectiveness of Learning**

The final set of implications revolves around differences in K. K – H is the difference in earnings due to education. It depends on how much is learned, and how much those extra skills are valued by the labor market. When K – H increases, the net present value of education rises and schooling should rise.

It seems plausible that those with more innate ability tend to learn more efficiently in school, leading to higher K. If so, the benefits of schooling are higher for those who are already highly talented, and smarter students should invest in more schooling. Of course, this increases inequality in skills, and earnings, in the economy.

Improvements in school quality should have a positive effect on K, and vice versa. Technological innovation in education can be expected to increase investments in education by citizens. Similarly, changes in the effectiveness of teaching methods or quality of teachers (which could be positive or negative) change the returns to education investments.

An important factor is the level of technology associated with the average job. Although college education may be valuable to farmers, it is not likely to have as much value to a farmer as to an accountant. Education is complementary with a technologically advanced society. Being uneducated, unable to read, or unable to do simple mathematics is more of a hindrance in a society that has a majority of white-collar jobs than in a society of farmers.

---

5 This is one reason why education is a good credential.
Thus, K and overall levels of education are higher in advanced societies than they were in 1900. This logic can also help explain current patterns of education across different societies, and trends in returns to education over the last several decades, which we describe now.

**Was Benjamin Franklin Correct?**

The quote at the beginning of this chapter suggests that education is a good investment—generally it is. Economists have estimated the *internal rate of return* (implied interest rate on the investment) from education in a variety of countries, and it is generally quite high. For example, most students reading this textbook are investing in college or an advanced degree. In the U.S., Asia, or Europe, the rate of return on higher education is usually estimated to be about 11% per year or more. This tends to be better than the stock market, once we adjust for relative risk.

Not only is education a good investment, but it has become an even better one in recent decades. The labor market has valued skills more and more in recent years. Figure 3.1 illustrates this for the U.S. labor market. The series marked with circles (which uses the scale on the right) plots the ratio of average hourly wages for workers with a college degree, divided by average hourly wages for those with a high-school education, over time. It shows a strong trend over the last three decades toward relatively higher earnings for those with a college degree. In 1970, college graduates earned about 50% more than high school graduates. By 2003, they earned roughly twice as much. Similar patterns exist for more advanced degrees, such as an MBA or MD.

![Returns to Skills, 1970-2003](image)


**Returns to Skills, U.S.**

*Figure 3.1*

The series marked by diamonds (the left scale) shows a different measure of the value the labor market places on high skills. It was created by calculating the 10th and 90th percentiles.
tiles of the distribution of hourly wages among U.S. workers each year. The 10th percentile is the wage level that 10% of U.S. workers earn less than; it is a measure of what relatively low-skilled workers are paid. The 90th percentile is the wage level that 90% of U.S. workers earn less than; it is a measure of what relatively high-skilled workers are paid. Comparing these two each year gives us an idea of how the distribution of overall pay has evolved over time. The figure plots the ratio, which shows how pay of relatively high skilled workers evolved compared to pay of relatively low skilled workers.

The pattern is quite similar to the college / high school comparison – there has been a dramatic rise in the returns to investments in skills in the U.S. economy over the last several decades. Earnings of those in the 90th percentile rose from about four times that of those in the 10th percentile in 1970, to about five and a half times as much by 2003.

Similar patterns show up in most advanced economies worldwide. If anything, some estimates suggest that returns to advanced skills may be even higher in less advanced economies.

**Outsourcing**

Outsourcing of work is a controversial topic these days. There are two elements to outsourcing. The first is assigning some tasks to suppliers outside of the firm (obviously, this is greatly facilitated by the Internet and advanced communications technology). The second is the use of lower-wage workers overseas.

In advanced economies, outsourcing has traditionally been focused on the lowest skilled jobs, such as manufacturing. In recent years, however, outsourcing of middle and even some highly-skilled jobs has risen substantially. Many customer service centers now source work to call centers quite far away. Perhaps more interestingly, software engineering is now being outsourced in many cases, to places like India and Russia. In the 1980s, software engineering was a very prestigious and highly compensated occupation.

There are two reasons why outsourcing has crept into higher-skilled jobs. One is that some of what used to be high skilled is actually not that highly skilled now. Modern software engineering techniques such as object-oriented software enable relatively low skilled programmers to develop more advanced applications than would have been possible ten or twenty years ago.

The second reason is the high returns to skills. As any resource becomes expensive, buyers try to find other sources for the resource. Thus globalized labor markets, through outsourcing, temper the rising returns to skills in occupations for which outsourcing works well.

Figure 3.2 presents similar evidence for a more specific sample. It plots average earnings for U.S. engineers by level of responsibility. Level of responsibility was determined for a random sample of engineering jobs by professional analysts; think of it loosely as corresponding to how senior the engineer is in his or her firm. Thus, a Level 6 Engineer has greater engineering or other valuable skills than a Level 5 engineer, and so on.
As can be seen, inflation-adjusted salaries have not changed much for lower-level engineers in the last two decades. However, the relative pay of engineers at higher levels has increased markedly over the same period. Once again we see that more highly-skilled workers are paid relatively more now than they were before.

Why has the labor market valued high skilled workers so much more than low skilled workers in the last few decades? Several explanations have been examined by researchers. The most important seems to be the increasing use of advanced technology, including computers, in the workplace. (We return to this issue in Section III of the book.) As discussed in Chapter 1, capital tends to be complementary in production with a skilled workforce. Greater use of technology, and more effective technology, increases the value of having highly-skilled workers. This increases the demand for skilled workers, which increases their labor market value.

**Investments in On-the-Job Training**

Let us now turn to on-the-job training. This investment is like education in many ways. It increases worker skills and raises productivity. This benefits both the employer and the worker. There may be direct costs (books or other resources, compensation for trainers, etc.). There may also be indirect costs, of two kinds. First, formal on-the-job training may take the worker’s time and attention away from regular duties, lowering productivity. Second, putting a worker who is not fully trained into a job, and having them learn as they do the job, lowers productivity compared to putting a fully-trained worker in the job.\(^6\)

\(^6\) However, this ignores the potential effects on intrinsic motivation and continuous improvement that we discuss in Chapter 7.
Thus, the implications are similar to our discussion of investments in education. For example, firms and workers have an incentive to invest more in on-the-job training for younger workers, and younger workers will tend to be more interested in applying for jobs that offer extensive training opportunities. Since the investment is being made on the job, additional interesting implications arise, which are derived now.

We first start by considering when a possible investment in on-the-job training would be economically profitable. That is, will the gains in productivity exceed the costs of the investment? We will distinguish between the gains in productivity in the worker’s current job, or in jobs with other employers. For the moment we will defer the question of who pays for the investment and receives the return (the worker or the firm). Once we decide what investment should be made, we can return to the question of how the worker and firm might contract to make that investment.7

A real example from the personal experience of one of the authors illustrates the basic idea. A small Silicon Valley startup provides enterprise software that does tax optimization. A typical employee in this company must know something about tax laws, as well as programming in Java. This is an unusual combination of skills. There are many firms that value both skills independently. However, few other employers who would value an employee with the same combination of expertise in tax law and Java as this firm.

Therefore, an employee who leaves the startup will have a difficult time finding a firm that can make use of all of the skills that he acquired in the first firm. The second job will value one of the skills, and perhaps even both, but not to the same extent or in the same proportion.

A similar question faces you in your job. To what extent should you invest in skills and knowledge that help you in this job and a career in this firm? To what extent should you instead invest in training that would improve your job prospects in the outside labor market?

Consider an employee at the startup whose current productivity (in period 0) equals $10,000 per month. Imagine three options for on-the-job training, outlined in Table 3.3. One focuses completely on Java, another completely on tax, and a third splits the training time between the two in proportion to how the employer values those skills in this job. (These are meant to illustrate the general points. Of course there are many possible ways to mix the training.) Assume that the total costs of training (both direct costs and indirect through lost productivity while training) are the same for all three choices, $5,000. Ignore discounting, just to keep things simple.

---

7 For students of economics, this is an application of “Coase Theorem” logic. We first focus on how economic value can be created. Only after that do we consider how this value is split between the worker and the firm. Of course, this approach will not work if there are bargaining costs, as we will see.
For example, consider the first option, which raises the employee’s proficiency in Java. If he worked only on Java projects, productivity would rise by $8,000 per month. However, in this firm he does not spend all of his time on Java. Instead, his job involves spending about 40% of his time on Java, so productivity would rise by only $3,200. His productivity on tax-related tasks would not change, since he receives no tax training.

How would the training affect his productivity if he were to find a job at a new firm? That depends on the kind of job that he can find elsewhere. Since the combination of Java skills and knowledge of tax law is quite unusual, and Java skills are more highly sought after in Silicon Valley than tax skills, it is more likely that the best job he could find elsewhere would also emphasize Java. However, even if a new job is 100% Java programming, it probably involves a focus on somewhat different Java techniques than the ones he emphasized in the training offered by the tax software startup. After all, the on-the-job training will probably focus on how he can program the tax software more effectively.

Therefore, it is unlikely that he can get the full benefit of the Java training in some other job, though he may be able to come close. Assume that the Java training would have about 80% as much effect in a new job as it would in his current job, so that his productivity in an alternative job would rise by $6,400 on average.

Now consider the opposite type of training, which emphasizes only tax. This would raise his productivity at tax-related tasks by $8,000 per month. However, in his current job, he spends about 60% of his time on such work, so productivity would rise by $4,800. If another job that he might find is likely to emphasize tax work only about 20% of the time, the training would raise his productivity at other jobs by about $1,600 per month on average.

The third alternative is to spend some of his time on each form of training. In this example, we assume that the training is more effective if he does a little bit of both kinds, so that the total increase in productivity would be $5,000 for each skill, or $10,000 total per month. The reason for this is the familiar idea of diminishing marginal productivity: the more that you focus your study on one topic, the less you learn with each new hour of study. However, this assumption is not important for the conclusions.
If he obtains mixed training, productivity rises in both tasks. Total productivity rises by $5,200, more with his current firm than in the outside market. That is because the training is designed to match his current job’s skill mix, not the average skill mix required by the labor market.

Our question to answer: What is the best on-the-job training for this employee? The answer depends on where he is most likely to be employed after the training, because the jobs emphasize the two skills differently. The optimal investment maximizes expected productivity. If he expects to stay at the current firm, he should train in both tax law and Java programming. If he expects to quit this firm, he should focus only on Java. The answer depends on the likelihood that he will switch firms.

Suppose that he believes that there is some chance that he will stay with the current firm, but also some chance that he will leave (or his firm will go bankrupt, etc.). What is the best kind of training then? If the odds that he will stay with this firm are high, it is a mix of tax and Java programming. If the odds that he will leave are high, it is 100% Java programming.

This illustrates a very natural way to think about investments in skills on the job. The optimal skills for you to invest in will be quite different if you expect to stay with your current firm for a long time, or if you expect to leave. If you expect to stay at the current firm, your best strategy will be one that focuses training on the skills that your employer values the most. If you expect to leave, your best strategy is to invest in skills that the labor market values the most.

**General v. Firm-Specific Human Capital**

In our discussion of the programmer, Java and tax skills had value with both the current employer and other employers. However, the value of the skills learned on the job was more valuable with the current employer than elsewhere. There are two extreme possibilities: the training is equally valuable inside and outside the firm, or it has no value at all outside the firm. These two cases are often called **General Human Capital** and **Firm-Specific Human Capital**. Most training falls somewhere in between, as in the example.

General human capital is skills or knowledge that a worker can acquire that raise productivity equally at both the current employer, and with many other employers. In other words, there is a thick labor market for the skills. Most skills are closer to this type. An MBA is GHC because the ability to be a good manager has value to thousands of possible employers. Knowledge of a foreign language such as Mandarin would be another example. A good rule of thumb is that most skills that you can acquire outside of the workplace, such as at a university, are general human capital.

Firm-specific human capital is the opposite of general human capital: it raises productivity at your current employer, but does not raise your value to other firms at all. It is more difficult to come up with examples of purely firm-specific human capital. Most knowledge that improves your productivity at one firm is very likely to help you in at least some other jobs at some other employers. Nevertheless, there are some examples of training that is largely firm-specific in value. If your firm has an unusual machine that they have designed for their own use, knowledge of how to operate the machine raises your productivity in this job, but would be of no value if you switched employers. Any idiosyncratic process or method might be firm-specific human capital.
Many examples of training that is more firm-specific involve intangible knowledge. If your firm has a strong and unusual corporate culture, knowledge of the culture helps you in this job, but is probably useless elsewhere. An understanding of informal networks and power relationships inside your firm is quite similar. Finally, if you have developed close working relationships with clients, and deep knowledge of their particular organizations, that may be primarily firm-specific human capital – unless you can get a job with your client, or take your clients with you to a new job.

As our discussion of the Java programmer illustrates, however, the concepts of general and firm-specific human capital are usually not clear-cut in practice. Many skills have a value both inside and outside the firm, though the values may be different. For example, knowledge of Java raises productivity somewhat more in the current job compared to alternative jobs. To the extent that it raises productivity equally inside and outside the firm, it is more like general human capital. To the extent that it raises productivity more inside the firm, it is more like firm-specific human capital.

A better way to think about this distinction is to ask whether the value that the employer places on the employee's training is relatively idiosyncratic or not. To the extent that the firm values the worker’s particular portfolio of skills similarly to other firms, that set of skills is largely general human capital. By contrast, to the extent that the employer values a set of skills idiosyncratically, the skills are largely firm-specific human capital. The startup’s desired skill mix of Java and tax knowledge is quite unusual. Therefore, on-the-job training at that firm is more specific to that firm and less general. This distinction will be useful shortly when we consider who pays for the training.

Special Case: Intellectual Property

Suppose that you are a research chemist. Your firm provides you with a very expensive lab, supplies and equipment, and lab staff. These resources are superior to those that you could use in other jobs that are available to you. They require that you do research in a particular, obscure type of polymers that few other labs are studying, because they already have special expertise in manufacturing products that use this type of polymer. Is your investment in learning about this obscure type of polymer more like general human capital, or more like firm-specific human capital?

At first glance it would appear that this investment is largely firm-specific, because you and your employer probably shared the costs of investing in the intellectual property. Your relatively obscure skills may be less valuable in any other job that you can get. Moreover, your employer may have required that you agree to grant all patent rights to the firm, so that you cannot take any of your patents with you if you leave.

However, in some ways the investment is more like general human capital, because you may be able to take some of the benefits with you to a new employer. In general it is very difficult to fully assign the intellectual property rights to the employer. Even if you cannot take any patents with you, you may be able to take many of the insights and ideas, and they may be valuable to competing firms.

In other words, intellectual property has elements of both general and firm-specific investing. Like firm-specific human capital, it is typically a shared investment, with the hope of a shared benefit. The intellectual property often has a higher value if the employee stays at the current firm, since the skills and knowledge were designed to match with that firm’s strategy. Therefore both have an incentive to try to maintain the working relationship. But
like general human capital, the worker may be able to capture some of the benefits by quitting and going to a competing firm.

**Non-Compete Agreements**

Firm-specific investments, particularly intellectual property, sometimes cause firms to include employee non-compete agreements in employment contracts. These agreements attempt to prevent an employee from taking intellectual property with them if they quit the firm. Such agreements try to restrict the employee’s next job in some way, typically for one year after quitting. Examples are clauses that state that the employee cannot work in a very similar job with a competitor, or cannot take clients to the new employer for one year.

Non-compete agreements are often quite difficult to enforce in court. Most courts frown on them, because of the long-standing principle (since the abandonment of practices like indentured servitude or slavery) that people should be free to work for whomever they wish. In order to make a non-compete agreement more likely to be enforced, the firm should make sure that the restrictive clauses are not too onerous, and do not last for too long a period.

Some courts have also imposed a requirement that employees be compensated in some way for signing a non-compete agreement. In fact, this may be necessary and appropriate, if an agreement is added to the employment contract of an existing employee after hiring, because the agreement reduces the value of the job to the employee.

*Possible clauses:* Some clauses that courts are more likely to approve:

- require that the employee give adequate notice and describe their new job duties before leaving, so that the firm has time to react;

- require that the employee train his or her successor and introduce them to important clients before they leave;

- prohibit the employee from recruiting colleagues to leave with him;

- require that some benefits vest gradually after the employee quits, contingent on the former employee acting in accordance with the non-compete agreement.

*Alternative approaches:* To the extent that non-compete agreements and legal property rights do not fully protect against an employee quitting and taking intellectual property with them, the first has several options to improve matters. First, pay for performance, especially if tied to the value of the intellectual property that the employee is developing, can better align incentives. It not only motivates the employee to stay, but also to increase the intellectual property’s value. Second, the firm might offer deferred pay, rewarding the employee for remaining with the firm. Conceivably, the firm might even offer to pay bonuses to employees a year or two after they have left the firm, as a reward for not competing with the firm in their new job. However, such an approach would seem to have practical limitations.
Who Should Pay for Training?

Who will pay for, and benefit from, training investments? We will consider two cases: education, and on the job training. Education is general human capital, while on the job training can be either general or firm specific human capital. The conclusion is straightforward in the case of general human capital: the worker should pay for these investments. The case of specific human is more complex.

**Education**

Some part-time students have employers who pay for their schooling. Is this a good investment for the employer? In general the answer has to be no: most academic training has wide applicability at many different employers. This makes it general human capital. Imagine, for example, that a firm pays for our student considering finishing her college degree. Once she earns the credential, her market value will rise. In order to keep this employee, the firm will have to raise her salary. In other words, the employer is highly unlikely to capture most of the benefits from the schooling investment. By contrast, the employee almost certainly will enjoy most of the benefits from the schooling. For the rest of her career, her earnings will tend to be higher.

For this reason, it almost never makes sense for an employer to pay for schooling. Instead, the typical solution is for individuals (or their families) to pay for schooling, and to invest in it before entering the labor market if possible. The vast majority of students do not have their tuition paid for by an employer.

That said, some students do have part or all of their tuition paid for by their employer. These are the exceptions that prove the rule. However, it is worth briefly trying to explain these exceptions as well. There are several reasons why a firm might pay some of the tuition costs of some employees.

*Implicit Cost to Employee & Benefit to Employer*

It is possible that the employee is, in fact, paying for the tuition, by accepting lower salary at the job in order to get access to the tuition benefits. In fact, it is not unusual to see employers impose a contractual obligation on employees to pay back the tuition if they quit the firm within a few years of graduation. This cost of quitting makes it possible for the firm to capture some of the benefits of the investment, by paying the employee less than market value for a few years once they are educated.

*Matching*

If the firm offers the tuition benefit only to a few select employees, an explanation might be that these employees are a strong match for this employer. The firm expects that these employees will stay at the firm for many years, and may wish to groom them for key positions. In such a case, the firm can expect to recoup some of the benefits of the schooling if the worker has some incentive to stay with the firm. Essentially, the firm and the employee split the profits from the investment in education. This is similar to our conclusion for firm-specific human capital below. However, the specificity is not in skills, but in the match.

*Recruiting*

We discuss Benefits in Chapter 13. Briefly, offering a certain benefit may generate useful self-selection in recruiting. UPS, for example, offers tuition reimbursement to its em-
ployees. Most of them do not attend expensive universities, so the program may not be very costly. A benefit may be that UPS attracts a harder working, more ambitious workforce. Another benefit is that the workforce will tend to be young, and UPS needs employees that can haul heavy packages as part of the job.

**Arbitrage**

If there are tax benefits to paying for education or training, the firm might have a cost advantage in paying for schooling compared to the employee.

Other than these kinds of unusual cases, in general firms do not, and should not, pay for investments in education of their workforce. Note that this is the same logic that we used in the last two chapters, when thinking about risky hires, screening, and signaling. The firm always faces pressure to match an employee’s outside market value. Anything that raises the outside value will tend to force the firm to raise compensation. Education and general human capital are two important examples.

This is why firms generally do not run extensive formal education programs; rather, schools are organized as independent institutions. The Control Data Institute example illustrates this nicely. Wipro Technologies provides a contrasting approach that is not likely to work in most countries.

---

**Control Data Institute**

Control Data Corporation was one of the first builders of supercomputers. In the 1960s, their computers were some of the fastest in the world. In 1965, they established a division called the Control Data Institute to train operators for their computers. CDI provided some of the best training in the industry, which is not surprising since CDC built the computers the training was for.

CDC found that a substantial fraction of employees who had trained at CDI quit to work for competitors or clients. That is because the training was largely general human capital, since it had wide applicability in the labor market. In 1989, Control Data decided to spin off the Control Data Institute as an independent training company.

**Wipro Technologies**

Due to explosive growth in technology companies and extensive on-the-job training in software design, which is largely general human capital, Indian software companies find it very difficult to retain employees. Most require new hires to sign contracts pledging that they will stay with the company for a specified period of time. Wipro Technologies in Bangalore takes that a step further.

Wipro requires that new hires provide a deposit of Rs 75,000 (Indian rupees; Rs 75,000 equals about $1,400) before they can receive their employment letters. The money is deposited in a bank. Employees who are unable to place the deposit may borrow it from the bank.

The deposit, with any interest earned, is refundable to engineering employees who work for Wipro for at least twelve months after completing Wipro’s three month training program. Employees with science degrees are given six months of training, and are required to work for Wipro for eighteen months before their deposit is returned.
Wipro reports that this program has not adversely affected its ability to recruit on college campuses.


On-the-Job Training

General Human Capital

Now think about who pays for on-the-job training. Consider first the extreme case where the skills are pure general human capital. That is, the new training is valued equally by both other employers and the current employer. In this case, the logic is the same as for education investments. When the worker gets the training, his or her market value rises. The firm will thus have to pay the worker a higher salary once the training is completed, or risk having the worker quit. For this reason, the general rule is that if skills are completely general human capital, the worker should pay for 100% of the investment, and receive 100% of the benefits.

Human Capital that is Partly or Fully Firm-Specific

The more realistic case is where on-the-job training is valued differently (and more) by the present employer than by the labor market. Our programmer investing in Java or tax skills is one such example. In that case, the worker’s outside value is lower than his inside value after the training, even though the labor market does value the training. Most on-the-job training has this flavor; it tends to be at least somewhat focused toward the current job’s skill requirements. What happens in this case?

To think about this, consider the investment for our software programmer as graphed in Figure 3.3. Assume that the probability that he will stay at the current firm is high enough that the best investment choice is the middle option, which provides training in both Java and tax laws. There are two periods, training and post-training, and we have kept things simple by ignoring discounting. If the worker gets no training, productivity is $H = \$10,000$ per month, the dashed line, in both periods.

If he receives the training, it costs $\$5,000$ in direct and indirect costs ($C + F$) during the first period. Thus, his net productivity will be $H – C – F = \$5,000$. This is represented by the solid line during the training period. The cost of the training is represented by the red shaded area between the dashed line and the solid line during that period.

After the training is received, the worker’s productivity rises to $\$15,200$ at his current firm. It rises to $\$14,400$ in the labor market as a whole. These figures are calculated as the initial productivity, $\$10,000$, plus the increase due to training. In Figure 3.3, these two productivities are the solid lines above the dashed line in the post-training period. The return on the investment if the worker leaves the firm is represented by the lighter green shaded area between the solid line at $\$14,400$ and the dashed line at $\$10,000$. The return on the investment if the worker stays at the firm is represented by that area plus the darker green shaded area above it: it is the total area between the solid line at $\$15,200$ and the dashed line. Because the return on the education investment it is larger if he stays, this investment is somewhat specific to this firm.
Investing in General Human Capital

Figure 3.3

This is a profitable investment for the firm only if there is strong reason to believe that he will stay at this firm. There is an incentive for the worker and the firm to figure out a way to make this investment, and have the worker stay at this firm in the second period.

Suppose that, like education and purely general human capital, the worker pays for the investment, expecting to earn the returns in the second period. In other words, assume that the firm agrees to pay the worker a salary equal to his productivity in both periods. Under that contract, the firm earns no profit or loss in either period. In the first period, the worker suffers a loss (pay is $5,000 less than he could earn elsewhere). However, in the second period he earns a profit (pay is $5,200 more than he could earn elsewhere).

Consider the situation after the investment is made (so that it is now a sunk cost). What might the firm do now? If it pays $15,200, it pays more than it has to, to keep the employee. The employee can only earn about $14,400 elsewhere. Therefore, the firm may be tempted to only pay a little more than $14,400 after the investment is made. The worker may have to seriously consider this lower level of pay, since the firm can threaten to fire the worker, and he can only get $14,400 elsewhere.

In other words, the firm may be tempted to renege on its promise, and renegotiate after the investment is made. Why? Because if it does pay less than $15,200 it is able to capture some of the profits from the investment that the worker made!

If you are a worker deciding whether or not to accept the contract in the first place, you may foresee this risk. If you do, you will be unwilling to make the investment. That would be a shame, because this could be a profitable investment. Yet you may choose not to invest since you are worried that the firm will try to take some of the profits after you invest.

This is a general concern that arises in many investment contexts, which economists often call the Hold-Up Problem. The problem arises if one party makes an investment and expects to earn the benefits later, but a second party is tempted to renegotiate after the in-
investment is made. If this risk is foreseeable, the investment may not be made, for fear of losing some or all of the returns if forced to renegotiate later.

If you are unwilling to pay for the investment, can we solve the problem by having the firm pay for the investment and earn the reward instead? In other words, what would happen if the firm agreed to simply pay the worker what he would get if there was no investment ($10,000 in each period), but provide the training? Then the firm would pay the cost, since productivity would be less than pay during training. The firm would also capture the benefits, since productivity would be higher than pay after training.

You can already answer this for yourself: it runs the same risk of renegotiation. Once the investment is made, the employee may be tempted to try to renegotiate pay that is higher than $10,000. After all, his market value has risen to $14,400, so there he can credibly threaten to quit. Moreover, his value to this employer is $15,200, so he might even ask for pay that is close to that amount. The firm will be tempted to negotiate with him, since they would lose $5,200 in profits if he left. But if they do, the employee will get some of the profits from the firm’s investment!

No matter who makes the investment and hopes for the return, the other side has an incentive to break its promise and try to renegotiate after the training has been paid for. The investor may be forced to renegotiate, because if the relationship is severed, the investor has more to lose. Unfortunately, this renegotiation risk will lower the expected returns on the investment, possibly to the point where neither is willing to make the investment.

How can we solve this problem? There are two general possibilities. One is to rely on the trustworthiness of one or both parties; we discuss that below and in Chapter 15. A second is to split the cost and the return on the investment. An example of how to do so is shown in Figure 3.4. In this case, the costs are split by paying a wage $W_1$ during training that is somewhere between actual net productivity and what the worker could get elsewhere (if the split is 50-50, $W_1 = 7.5$). Splitting the costs reduces the risk of the investment in the first period, since there is less to lose. The benefits would then be split by setting $W_2$ after training somewhere between what the worker could earn elsewhere, and actual productivity (if 50-50, $W_2 = 14.8$). Splitting the benefits reduces (though it does not eliminate) the temptation to renegotiate. Moreover, since both would have something to lose if the relationship was broken, both have some incentive to avoid renegotiation.
Therefore, investments in on-the-job training that are specific to the firm are likely to be made differently than investments that are purely general human capital, or education: they are likely to be split by the worker and the firm. This means that pay would be less than what the worker could earn elsewhere, but greater than net productivity, during the investment period. Pay would then be greater than what the worker could earn elsewhere, but less than productivity, after the training was complete.

**Implications of on-the-job training**

On-the-job training has a number of important implications for the employment relationship. Now that we have analyzed the investment, and considered who pays for it and receives the return, we develop these implications.

Before we delve into these implications, you need to remember that on-the-job training will vary along a spectrum from training that is purely general – has equal value at other jobs outside this firm – to training that is purely firm-specific – has no value outside this firm. Education obtained at a university is almost always purely general human capital. On-the-job training is almost always a mix. Even training that is highly focused on the current job will tend to benefit the employee if they leave the firm, just not as much. Therefore, we will talk about training as being relatively more or less general or specific.

*Turnover*

One of the most important implications has to do with turnover. If training is completely general, the firm does not care about turnover (ignoring the costs of replacing the worker). The firm has made no investment, and is earning no return on the training, so it has nothing to lose if the worker quits. Similarly, the worker takes the full investment with him, so he also has nothing to lose by switching to another employer.
By contrast, the more specific to the firm is on-the-job training, the more the firm and the worker care about turnover. With shared firm-specific investments, both lose if the worker leaves. The larger the difference between productivity at this firm and elsewhere, the higher the loss. Thus, when employers say that they are concerned about losing their investments in their employees, they must be talking about human capital investments that are relatively specific to that firm.

This has an important effect on how a firm thinks about its employees. If on-the-job training is general or not important in the firm, then, excepting sorting considerations, employees are largely hired through spot-market type transactions. However, once specific investments play a role, a relationship arises between the worker and the firm. Both have some incentive to invest in and maintain the relationship. The term that is often used for this is that such firms emphasize internal labor markets. The more idiosyncratic the skill mix that a particular firm requires of its workforce, the more does this view of employment become important. These ideas are discussed further in Chapter 15.

**Investment**

The lower is turnover in a firm, the more will workers tend to invest in a mix of skills that is a strong fit for the current job and employer. The higher is turnover, the more will workers tend to invest in skills that can be easily applied at other firms. Thus, firms that desire an unusual mix of skills will generally adopt policies to try to reduce turnover.

Investment patterns should change with tenure. The longer that an employee has been with the firm, the more likely is she to have already invested in skills that are closely matched to that firm. This increases his incentive to stay with this employer. That reinforces his tendency to invest even more in firm-specific skills. Thus, as job tenure rises, workers tend to become even more invested in their current employer.

**Compensation**

These ideas have implications for compensation patterns. First, pay will tend to rise with one’s labor market experience, because most jobs provide some on-the-job training. Second, beyond this effect of total experience on earnings, those who have higher tenure in a firm will tend to have higher pay than those who do not, because their skills will tend to be more heavily oriented toward the mix that is appropriate for that firm, and because they will be earning some returns on their prior investments in firm-specific skills.

Second, the more firm-specific are employee skills, the more pay will they tend to lose if they switch jobs, because the skills will be more tuned to the current job instead of a potential new job. The higher the expected loss in compensation from leaving the firm, the lower the likelihood that the employee will leave the firm. This is because workers tend to invest more in skills that are specific to their current employer, the more that they expect to stay in that job.

**Labor Market Thickness**

Labor markets are sometimes characterized as thick or thin. A thick market is one in which the worker finds it relatively easy to get a new job that values his or her skills well. A thin market is the opposite. To some extent, the thickness of the labor market depends on the occupation of the worker. There are many more jobs for lawyers than for academic economists in most cities. Thickness also depends on the business cycle. If an economy is in recession, few firms are hiring, so it is more difficult to obtain a job offer with pay similar to one’s current job. The opposite tends to be true when the economy is improving, especial-
ly at the beginning of an upturn. Thickness is affected by the size of the local economy, if there are costs to the worker of moving elsewhere for employment.

Finally, thickness also depends on the employee’s skill mix; the more idiosyncratic the skill mix, the thinner is the labor market for that particular employee. This logic illustrates that the concept of general v. firm-specific human capital is endogenous: it depends on how thick the labor market is. In markets that are thicker, all else equal, a worker’s skills tend to be less firm-specific.

**Firm Size**

Workers in larger firms will tend to invest more heavily in a mix of skills that is specific to their employer, for two reasons. First, empirically larger firms tend to have lower turnover than smaller firms. Second, larger firms are more likely to be able to find alternative work for their employees who wish to change jobs. In fact, very large firms sometimes have relatively formal internal labor markets, in which the human resources department actively posts jobs and searches for internal candidates to fill open positions. To the extent that the skill mix is similar across jobs in the same firm, this in effect thickens the market for an employee’s skills when they work for a larger firm.

**Rent Sharing and Compensation**

The concepts of education investments and signaling allow us to now briefly discuss the overall level of compensation. By compensation, we mean cash and other benefits paid to the employee. Benefits have value to the employee, so an employee will be willing to accept some amount of lower salary in exchange for a given benefit.

How is the overall level of compensation determined? Abstractly, perfect competition between firms should imply that firms earn zero profits on their employees. This does not mean that they do not earn accounting profits, shareholder returns, and so on. It only means that the level of accounting profit that they earn from hiring employees is about the same as that earned by other employers.

Similarly, perfect competition between employees should imply that employees earn about the same at any firm.

When these two conditions hold, an employee and a firm would be largely indifferent about how they work for, who they hire, and turnover. This is analogous to the case of pure general human capital investments. The level of compensation would be the same at firms that have similar jobs, and would just equal the marginal value of the worker’s output to the firm.

Obviously, this is not realistic, but it is a useful theoretical base case. In reality, of course, workers do tend to lose earnings if they switch employers. And, employers usually do not want to lose most of their employees. These observations suggest that both are earning some profits (in formal economics lingo, rent) from working with each other, compared with working with another employee or employer. This raises two questions. First, what are the sources of these rents? Second, do they imply that labor markets are not perfectly competitive?

In this chapter we have seen two reasons why a worker and firm might earn some extra profit if the worker stays at this firm. The first was matching: if a worker is, for any reason, a
particularly good fit for one employer, then there is some benefit to the employee staying at that firm. That is, in effect, a violation of perfect competition, because it implies that either the worker or the firm cannot find a perfect substitute for their employer or employee. In that case, some monopoly profits would be accruing.

Firm-specific human capital is the second condition in which a worker and firm generate extra profits if they stay together. However, in this case labor markets can still be perfectly competitive for both workers and firms. The bidding would occur at the time the initial job offer is negotiated. For example, firms might compete with each other over the training opportunities that are offered to job applicants. Job applicants might compete with each other over the salaries or other job aspects that they are willing to accept from a given employer. Once the firm-specific investment is made, there is some profit from working together. However, there can be competition over the terms of the total investment.

There are other reasons why a worker and / or a firm might enjoy some rent for working with each other. For example, either party may have expended resources searching for a job or a new employee. Finding a new job or replacing the existing worker would be costly. Therefore, there is something to lose from ending this employment relationship, for both. This is quite analogous to matching and firm-specific human capital. Again, the profits from search and recruiting might be driven to zero by competition. These are another investment, similar to signaling and training. Once the costs are incurred and the job match is found, the return on the investment is earned. If search or recruitment costs are high, this would be another reason that there is some surplus for the worker and the firm to share, once they have agreed to work together.

The important word here is sharing. In cases where there are some rents or joint surplus to be earned from working together, there arises a question of how those rents will be shared between the worker and the firm. As we have seen, these rents are largely split at the time of hiring, when the explicit and implicit contractual terms are determined. How the rents are split depends on the outcome of this negotiation. That will depend to some extent on the bargaining sophistication of the worker and the firm. It will also depend on some economic factors that we have already seen.

One consideration for how the costs and benefits are shared is to improve incentives for either side to take the right action. Workers generally pay for signals, to motivate their efficient self selection in applying for jobs. By contrast, both share specific investments, to reduce incentives to renegotiate later.

Another consideration is the bargaining power of each side. If there is a lot of competition between firms, employees will tend to capture a larger share of any potential profits. If there are many workers who are very similar, employers may be able to capture a larger share of any profits, because the workers will compete with each other.

A third consideration is the reputation of each party; we discuss this briefly below, and more extensively in Chapter 15.

The main point of this section is that there are several reasons why, once employees and firms decide to work together, they may want to keep working together. These are sometimes referred to by terms such as rents, quasi-rents, or surplus. They affect incentives for turnover and complex employment contracting. They also mean that the overall level of compensation will often be somewhat indeterminate, since it will depend on a complicated bargain struck between the worker and the firm.
In some cases in later chapters, we will assume that labor markets are perfectly competitive to illustrate the arguments. In no case does this assumption really matter. It is merely a simplification to make the arguments more straightforward. When you see those cases, you can imagine that there may be other sources of surplus for the worker and the firm, and if so there may be some kind of bargain that splits the surplus between them.

**Implicit Contracting**

On-the-job training investments that are relatively specific to one’s current employer, and intellectual property, are special cases of a more general phenomenon. Any time that two parties can make an investment that creates profits only if the parties continue to work together, we have a *relationship-specific investment*. This issue arises in many contexts in the business world. Consider two firms that have a joint venture with each other. If they discontinue the relationship, profits from the joint venture are lost. Similarly, two partners who start a firm together also engage in a relationship-specific investment as they build the firm (if the firm has more value as long as they stay together).

Our analysis of training that is more firm-specific concluded that the firm and employee will share the investment. Splitting the benefits reduces the risk of the Hold-Up problem, which may happen after one party makes an investment hoping to earn benefits later, but the second party tries to renegotiate after the investment is made. Unfortunately, this problem cannot be completely eliminated by splitting the benefits.

We saw a similar issue caused by lack of trust in Chapter 2. Probation periods involve a promise by the firm to give the job applicant pay higher than productivity after probation, as a way to motivate self selection. This means that the firm might suffer a loss on the employee after probation.\(^8\) What we did not discuss then is the concern that the firm might be tempted to renege on its promise to employees once they have been sorted. If there is a non-trivial chance that this might occur, high ability workers would be unwilling to apply for the job in the first place.

Thus, we have another situation where Hold-Up concerns might prevent our solution from working properly. This concern can arise in any situation where there is a relationship-specific investment. Is there anything that we can do to reduce Hold-Up problems?

In the case of a joint venture, there is a direct solution: the two firms can merge. Once they merge, they have no conflict of interest, and the investment will be made.\(^9\) Obviously, merger is not possible in the case of employment, so this is of little help for on-the-job training.

An alternative is to write a formal contract specifying what the firm and employee pay or receive under all circumstances (e.g., severance pay, or non-compete agreements). This could be used to give both incentives to not break their promises.

---

\(^8\) Not necessarily an accounting loss, but an economic loss. They would be paying the employee more than would be paid to employees of similar productivity at other firms that did not use such a probation system.

\(^9\) The example of Fisher Body Works is a staple of MBA strategy courses. The story is that General Motors wanted FBW to build a factory specialized to work with GM – a relationship-specific investment. The story then is that in order to solve the Hold-Up problem, GM eventually bought FBW. Apparently most of these facts are incorrect, and the overall story is a fable (Casadesus-Masanell & Spulber, 2000). Nevertheless it does illustrate how a merger can solve the problem.
A related approach is to rely on government regulations or common law. In most economies, employment is highly regulated. Some of these regulations may protect either the firm or (more likely) the employee from attempts at renegotiation. For example, in most societies firms do not have complete control over the management of employee pension funds. This reduces the risk to employees that earnings they are promised now will be taken away later.

Unfortunately, the employment relationship is so complex and unpredictable that it is usually impossible to write a contract, design a law, or make a judicial ruling that can cover all possible contingencies. What else can be done?

An important way to reduce Hold-Up problems is to rely on implicit contracting. In our examples, the firm in effect promises employees that they will be rewarded later if they perform well or invest in skills that are idiosyncratic to this firm. If the employee has enough reason to trust that the firm will keep its promises, he may be willing to do so.

This approach is called implicit contracting because it is distinctly different from formal contracting or regulation – it is the part of the employment relationship that is difficult or impossible to enforce through the legal system. When the legal system is not available, parties to a relationship-specific investment must rely on trust, reputation, and similar mechanisms to impose some reliability on the relationship.

We will see this issue arise several times. For example, most incentive systems require some element of subjective evaluation. Since subjective evaluations cannot easily be independently verified, implicit contracting becomes an important part of managing incentives. The topic is discussed more extensively and formally in Chapter 15; this is only a brief introduction.

**Summary**

In this chapter we have analyzed investing in worker skills. Education and on-the-job training are some of the most important investments that can be made in an economy. Historically, and especially so in recent years, these investments have paid high interest. We discussed what factors affect decisions about investments in education.

Education increases human capital. There are types of human capital, general and specific. The former is skills or knowledge valued equally at many employers (having a thick market). The latter is skills or knowledge that have unusual value at a particular employer. Training can involve some learning that is more like general human capital, and some that is more like firm specific human capital.

We then discussed the question of who should pay for the training. We argued that a worker should pay for training that is general, accepting lower compensation than could be earned elsewhere during the training period. The worker will then enjoy the return on the investment through raises and promotions received later.

To the extent that this is true, the firm is in effect selling a service – training – to the worker. We saw similar intuition earlier, when the firm sometimes provides a service to the worker by being particular effective at sorting through and identifying the most talented workers. This intuition is interesting, because it flips the employment relationship around. It is not

---

10 Other terms that are sometimes used are relational contracting, and psychological contracting.
just the worker that sells a service to the firm, but often the firm sells something of value to
the worker as well. We will see this again in the contexts of job design and pay for perfor-
ance. This illustrates the fundamental point that a healthy contract between the worker
and the firm is one that maximizes the total benefits, to both the worker and the em-
ployees. This should be the first consideration, with analysis of how to split those benefits
only analyzed second (and depending on issues like incentives, competitive pressures on
the firm, and labor market constraints).

Having the worker implicitly pay for on the job investments in general human capital
means that not every firm should invest in its workforce’s skills. To the extent that there is a
thick market for its employees, workers should invest in those skills instead, and the firm
should care little about turnover.

When human capital is firm-specific, the question of who pays for the training is complex.
If the employee does, she runs the risk that the firm will attempt to renegotiate or renege
on promises of higher pay after the training period. If the firm pays for the training, it runs a
similar risk from the employee. This is an example of the Hold Up problem, which may oc-
cur when one party tries to renegotiate terms after a relationship-specific investment has
been made.

However, employees are more productive if their skill mix closely matches the demands of
the job. Since most jobs are somewhat idiosyncratic, this means that optimal on-the-job
training usually is somewhat firm-specific. When that is so, a host of new considerations
arise. Workers and firms will tend to split the costs and benefits of the investment, in order
to mitigate Hold-Up problems. Turnover therefore becomes costly for each; they have an
incentive to maintain a long-term relationship. The longer that they have worked together,
the more that they will tend to have invested in each other, only reinforcing such effects.

Investments in skills vary in the extent to which there is a thick market of firms that value
them. When the market is thick, the skills are toward the general human capital end of the
spectrum. Most skills of this kind are provided by specialized organizations like universi-
ties, rather than on the job. Such investments are almost always paid for, and the returns
enjoyed by, the worker. (Societies often subsidize these investments.) However, some
skills are most effectively learned on the job, by actually doing the work. When that is the
case, the firm may provide that training to the worker.

Since contracting over complex employment relationships is typically incomplete, when
there are relationship specific investments, reputation and trust become an important way
in which firms and employees can improve the value of their economic relationship. Thus,
where skill investments should be more idiosyncratic and matched to a given employer,
the employer will adopt policies that foster an internal labor where employees are hired at
the bottom and spend long careers working their way up the corporate ladder in that firm.
Where skill investments are more typical of those needed by other employers, by contrast,
a firm may adopt more of an aggressive weeding out approach, since turnover will not be
costly. In short, different contexts require different approaches to managing the employee-
employer relationship.

In the first three chapters of this book, we have seen a sequence, from a simpler to a more
complex economic relationship between the worker and the firm. We started by imagining
that workers are paid on a spot market, with wages roughly corresponding to productivity.
The analysis almost immediately led to thinking about multiperiod contracts, to exploit the
option value of risky hires, and to sort employees. The next step was to make invest-
ments, often jointly, in worker skills. Finally, we add the notion of implicit contracting over
the employment relationship.
The concept of implicit contracting gives us an important piece of the puzzle for modeling organizational design. Much of what is sometimes thought of as the soft side of personnel management falls into this category. While it is difficult to develop a comprehensive formal model of all of the issues that arise, in Chapter 15 we will be able to provide an economic framework that improves your thinking about issues such as reputation, trust, and corporate culture.

**Study Questions**

1. Why is it a good rule of thumb that skills typically taught at universities are general human capital?

2. Think of jobs that you have worked in. Was the skill mix more firm-specific or general? Why?

3. If a firm requires that its employees invest in a mix of skills that is very specific to that firm, will there be any cost to the firm? Should a firm ever try to design jobs so that employees develop readily marketable skills. Explain.

4. Some firms hire employees at lower levels, train them extensively, and nurture long-term relationships with them. Others aggressively weed out employees. What characteristics of firms would push them toward one of these extremes or the other? Why? Try to list as many characteristics as you can think of.

5. What is the Hold-Up problem? Why does it happen? Can you give specific examples (say, from the sports or entertainment industries)? What are possible ways to avoid such problems?

**References**


In this appendix we present a simple model that illustrates some of the points about investments in on-the-job training. For details, see Lazear (2006).

We abstract from bargaining over the investment. Assume for simplicity that the firm and the worker share the costs and benefits of the investment. Furthermore, since we abstract from bargaining, we will treat the investment decision as being made by the employee. Of course, if there is efficient bargaining, the optimal investment decision will arise through the bargaining between the firm and the employee.

A worker invests in skills J (Java) and T (tax), at cost \(\frac{1}{2}(J^2 + T^2)\). Different employers have different relative values for the skills. Let \(\lambda\) be the weight that the firm gives to skill J, and \(1-\lambda\) the weight given to skill T. Thus, the potential earnings that a worker can earn at the current firm equals:

\[
W = \lambda J + (1-\lambda) T.
\]

Wages will be determined similarly at other firms, but the weights \(\lambda\) will vary from firm to firm. There are two periods. In the first, the worker invests in on the job training. In the second he may or may not switch employers, and works with no further investment. The probability that the worker stays at the current firm next period equals \(p\). Thus, the worker chooses \(J\) and \(T\) to maximize net earnings:

\[
\max_{J,T} \left[ \lambda J + (1-\lambda) T + (1-p)(\lambda J + (1-\lambda) T) - \frac{1}{2}(J^2 + T^2) \right],
\]

where \(\lambda\) is the expected weight put on Java skills at potential other employers. The first-order conditions are:

\[
p \cdot \lambda + (1-p) \cdot (1-\lambda) - J = 0,
\]

\[
p(1-\lambda) + (1-p)(1-\lambda) - T = 0.
\]

Investment is a weighted average of the relevant skill-values inside the firm and outside, where the weights depend on the probability of separation. The intuition should be clear. If \(p = 1\) so that continuation in the firm is certain, the only skill value that matters is \(\lambda\), the current employer’s relative valuation of skill J. If \(p = 0\) so that separation is certain, the current firm’s valuation does not matter. In that case only \(\lambda\) matters.

Define the optimal values of skills invested in as \(J^*\) and \(T^*\). Now consider what happens to a worker who switches to another firm after investment. We denote the wage in the second firm by \(W'\), and the weight given to skill J in the second firm by \(\lambda'\). The change in earnings is:

\[
W' - W = (\lambda' - \lambda)(J^* - T^*).
\]

The sign for this equation is uncertain. The typical case would be where the probability of separation is relatively low, and the worker invests with an emphasis on the skills that the current employer values. However, if the probability of separation is very high, or the cur-
rent firm’s relative valuation of skills is not too unusual, the worker’s investment will tend
more toward $\lambda$. If so, leaving the firm might lead to an increase in earnings. In any case, it
can be shown that the change in earnings from switching employers decreases with the
probability that the worker will leave the firm next period (Lazear 2003).

It is straightforward to show that an increase in market thickness leads the worker to invest
in a way that is more consistent with the current firm’s relative valuation of skills. The rea-
son is that when the market is thick, the worker in effect gets additional random draws of
$\lambda$’s from other firms. The worker is therefore more likely to find an alternative job that is
similar to the current job, which reduces the extent to which the original employer’s valuation
of skills is idiosyncratic. In other words, the firm-specificity of human capital is endo-
genous with respect to market thickness.

Finally, we could extend the model to three or more periods. Consider a worker who is two
or more periods away from retirement. They have relatively less incentive to invest in skills
that emphasize the current employer’s valuation. This is because, with more than one pe-
riod remaining, the probability that the worker will move to another employer is higher. This
has an interesting implication: a worker’s on the job training should become increasingly
firm-specific as tenure at the firm lengthens: investments become more idiosyncratic, and
less generally applicable in the labor market.
Managing Turnover

“You are the weakest link. Goodbye.” Slogan of the popular, ruthless British TV game show *The Weakest Link.*

**Introduction**

One theme of this text is that there is no one best approach that works for all firms. In the previous three chapters we developed an economic analysis of recruiting, structuring the job offer (pay, probation, and eventual screening for promotion), and investment in skills. Firms differ quite a bit in their overall strategy for these policies. In some firms, turnover is viewed as healthy, as it brings in new blood and facilitates sorting to find the most talented. In others, turnover is costly, because of investment in employee skills that match the idiosyncratic needs of the firm’s business.

In the process we developed several economic tools that have many applications inside and outside employment. These included adverse selection, signaling, and relationship-specific investments.

In this chapter, we finish this first section of the text by using these tools to analyze some related issues in the management of employee careers. The theme in the first two chapters was bringing employees into the organization. The theme of Chapter 3 was developing their talents so that they can be more productive and advance over their careers. The theme here is turnover – under what circumstances is it desirable, and how can it be effectively managed. While we briefly return to the subject of recruiting when we think about trying to hire employers from a competitor, most of the analysis focuses on turnover.

**Is Turnover Good or Bad?**

There are two different circumstances for thinking about employee turnover. One is the dramatic, hopefully rare need to lay off workers due to downsizing. The other is the general need to manage regular workforce flows in and out of the firm. We will talk about layoffs below. In this section, we consider the factors that affect a firm’s optimal turnover when business conditions are normal. Every firm should have some employee turnover as part of its business; the question is how much, and of what kind.

In thinking about turnover, there is often a question about the appropriate level of analysis. Should we think about turnover at the firm level overall, or differently for different jobs? For example, should we think about the right level of turnover at the organizational

---

1 In fact, this issue is in the background in much of the analysis in this text: should we think about a “one-size fits all” set of personnel policies, or tailor them to different jobs, groups of workers, etc.? This is really a question about centralization versus decentralization, which is a topic of the next two chapters.
level, or should it vary from one job to the next? In general, the answer is the latter. Different jobs within a firm have different characteristics. Some jobs may require substantial turnover, while for others the firm may want to keep turnover to a minimum. There may be patterns across the entire organization, if some of the issues discussed apply to many jobs in the firm, but that does not need to be the case. In fact, most firms have quite different turnover rates for different types of jobs, by occupation, hierarchical level, location, and so on.

Importance of Sorting

One of the most important reasons to encourage some turnover is sorting. Sorting allows the firm to increase workforce quality, by screening more candidates per period of time. The more chances that the firm gets to consider new job candidates, the greater the odds that more talented workers will be discovered. The concept of matching from Chapter 3 also applies, since additional sorting increases the odds that a position can be filled by an employee who is a better match for the firm.

Of course, sorting is only valuable to the extent that differences in ability (or matching) are valuable. One such situation is when there is more to be learned about workers: abilities are more variable, and less is known about them. For example, turnover is especially likely to be useful for new hires that are young and have little track record. Turnover is also desirable, to some extent, for employees who have been promoted to new positions, since there is uncertainty about whether they will fit with the new job. These are ideas that we saw in Chapters 1-2.

Sorting is also valuable when small differences in talent or matching lead to large differences in productivity or costs. Thus, jobs where talent is particularly important are good candidates for higher turnover, to continuously sift candidates to find the best ones.

Putting these ideas together, it becomes clear why turnover tends to be so important, especially early in the career, in leading professional service firms and academia. These are firms that are filled with knowledge workers. Ideas and creativity matter, and small differences in ability can be leveraged effectively. These kinds of firms often have fairly aggressive probation and up-or-out systems so that they can continually sort for the most skilled employees available.

Technical Change

An important benefit of turnover is that it brings new blood into the organization. New employees are more likely to have new insights, bring different perspectives, and understand the latest ideas, technology or other developments.

Thus, turnover should be higher in industries where technology advances more rapidly. Computers and telecommunications are two obvious candidates. Some of the turnover can be the hiring of workers from other firms. This benefits your firm because you can obtain some of the new ideas and innovations from competitors (because of imperfect employee non-compete agreements). There is also likely to be a benefit from hiring younger workers in these settings. Younger workers learned the latest techniques in college or graduate school. Thus, we would expect that wages would not grow as rapidly with tenure in industries where much of the innovation occurs through university research.

There may also be an optimal mixing of younger and older workers. While younger workers bring fresh ideas and technology, older workers have deeper understanding of the business, and are likely to have invested in knowledge that is specific to the firm. They
have more ability to profitably apply the new ideas that younger workers bring. In a sense, there is an opportunity for the two groups to cross-train or collaborate with each other. To the extent that skill requirements are specific and cannot be learned at school (say, because the business is somewhat unusual), there will be greater benefit to matching younger workers with older workers to provide this training.

Organizational Change

Organizational change also generally benefits from turnover to bring in new ideas. Current employees are experts at the firm’s current way of doing business. Unfortunately, if the firm needs to change methods, they are almost certainly no longer the best fit. This is particularly true for senior management.

We can flip this argument around as well. As we will see in chapters 7-8, sometimes firms can become highly optimized to a certain way of conducting their business. When this is the case, they are more likely to hire employees at the bottom and promote from within, to develop specific human capital. However, such firms may face a substantial problem if the industry changes dramatically, because their management is in-bred and have little experience with alternative methods. They may not even realize that they face this problem, because they have been successful in the past and have little exposure outside their current firm. To avoid this inward focus, it can be helpful to continuously bring in at least a few employees – at all levels – with outside experience. Firms that continuously bring in outsiders are more likely to recognize when times change, and adapt effectively.

Hierarchical Structure

Higher turnover may be necessary when the organizational structure dictates that the hierarchy narrows rapidly at some level. Take a quick look back at Table 2.1 at the beginning of Chapter 2. The second column shows the percentage of Acme’s workers at different levels. The hierarchy narrows dramatically between Levels 4 and 5 (roughly speaking, where middle management becomes top management). Some turnover is inevitable here, because there will be very few promotion opportunities for Level 4 managers. Some managers may become frustrated and quit to pursue other opportunities.

In fact Acme may want to encourage this. Otherwise, as Level 4 becomes clogged with managers who cannot be promoted further, promotion slots decline for Level 3 managers, and this effect will eventually trickle down the hierarchy to Levels 2 and 1. This will reduce incentives, because promotions are an important form of pay for performance. Moreover, Acme’s best workers are most likely to be lost if promotions are not available. A promotion system is like a pipe, in which a continuous flow both in and out is the goal.

Specific Human Capital

As we saw in Chapter 3, on-the-job training does not generate turnover costs if the training is purely general. By contrast, the more specific to the firm is the training, the higher are the turnover costs. Typically, these costs will be borne by both the worker and the firm, because of sharing of the investment.

Therefore, firms with more idiosyncratic businesses, methods, or cultures are more likely to want to have low turnover. Similarly, in jobs where valuable intellectual property is de-

---

2 A good example is a company that tries to expand its operations to a different country for the first time. Its management is unlikely to have a sophisticated understanding of the many issues that arise in a business that operates across borders, unless it hires some employees that have that kind of experience.
developed, it is important to try to reduce turnover. Finally, in positions where workers develop strong client relationships, turnover can also be quite costly.

Retention Strategies

A variety of tools can be used to reduce turnover. The most obvious is to increase compensation. Of course, that is simple but expensive. But for key employees, in some cases you may have to respond if they receive outside job offers (see the next section).

For your key employees, consider treating them as partners. These are the small number of employees who create the most value or innovation for your organization. They are the most likely to possess valuable intellectual property or customer relationships that they might take with them if they leave. Losing such employees can be very damaging – especially since they may go to your competitors and compete directly against you.

To avoid such problems, you might offer them stock, options, or other pay for performance directed at their particular area of business. In the extreme, such key employees might be made partners. After all, in some cases those employees are the business, and can take it with them. It is this consideration that explains why so many professional service firms are organized into some form of partnership. The bottom line is that you must pay key employees their market value or you will probably lose them.

Frank Quattrone’s Money Machine

Frank Quattrone was one of the earliest, and most successful, investment bankers in Silicon Valley history. His career began at Morgan Stanley’s San Francisco office in 1981. Quattrone had a fascination for the technology companies of Silicon Valley, and developed close ties to those firms and their executives. He moved his family there, and eventually opened the first investment banking office in the Valley, for Morgan Stanley. With his close ties to the industry, he was able to obtain some of the most famous and lucrative initial public offerings (IPOs) for Morgan Stanley, including Silicon Graphics, Cisco, and Netscape. At the time, Netscape’s IPO was the most successful in history, with a 150% rise in price the first day.

Because his Morgan Stanley office was so dominant in Silicon Valley investment banking, Quattrone became more powerful within his firm. He gradually pushed for greater control of operations at the office. In 1996, when Morgan Stanley refused some of his requests, Quattrone and his entire technology investment banking team quit and set up an office for Deutsche Bank Securities, Inc. By 1998, the team had moved to Credit Suisse First Boston (CSFB).

CSFB offered Quattrone’s team a very generous incentive plan: 33% of the revenues their group generated above $150 million. The result was that from 1998 to 2000, the team was responsible for almost as many IPOs as the next two largest competitors – one of which was Morgan Stanley. Quattrone’s group was the single most important source of growth at CSFB in the late 1990s.

(Quattrone was eventually indicted for securities fraud, leading to the shutting down of his group and substantial business problems for CSFB.)

What other retention strategies can you employ? You may be able to retain a specific employee by tailoring some benefits or characteristics of the work to that worker’s tastes. For example, flexible working hours would allow an employee to pursue outside interests or meet family obligations more easily. If such flexibility is difficult for them to find elsewhere, they may be inclined to stay at your firm. Depending on how costly such flexibility is to you, it may be profitable for you as well.

Offering new opportunities to an employee with talent or a good match for your firm can reduce the likelihood that she will seek a new job. This might involve new training, job enrichment (Chapter 7), or early promotion. There are several reasons why this can help. First, new tasks or responsibilities make the job more interesting. Second, training increases the long-term value of the job. To the extent that the training is specific to the firm, it increases the incentive to stay with your firm (as in the last Chapter). Third, an early promotion can signal to the employee the value that you place on their long term employment at your firm.

One of the reasons employees leave jobs is because they feel that they have not been treated well. This can occur if they believe that their manager did not evaluate them correctly. It might be because they believe that certain promises (for training, promotion, or etc.) were not fulfilled. A healthy firm has a healthy working environment in which these kinds of issues do not arise often, and are addressed effectively when they do. This does not mean that there will not be some complaints and some disappointed employees. However, reducing the extent to which employees are treated arbitrarily can reduce turnover problems.

A simple example will illustrate some issues here. It is tempting when recruiting to overstate the value of the job. Doing so makes the employee more likely to accept the offer. However, overstating the value means that the employee will inevitably end up disappointed. Moreover, the employee may infer that the employer is untrustworthy. This effect can be corrosive to the work environment and raise turnover. Therefore, paying attention to implicit contracting issues, as described in Chapter 15, can be an effective way to reduce turnover of all employees.

**An Unusual Recruiting Video**

Cummins Engine is one of the world’s largest manufacturers of diesel engines. In the early 1970s, Cummins adopted what was at the time a relatively new approach to organizing its Jamestown, New York plant. Workers were put in teams, and given substantially more tasks and responsibilities than traditional assembly line workers. (We will discuss this general approach in Chapter 7). The teams were expected to become relatively self-managed, even playing an important role in the hiring (and potential firing) of their members.

Because the work design was so different from other factories in the area, many recruits found their new jobs to be highly stressful. In some cases, the workers did not fit well with the new system. In one case, a team ended up firing one of their members because of such concerns.

The remaining team members found this incident to be so painful that, on their own initiative, they developed a recruiting video for Cummins to show to job applicants. The first screen on the video showed, in very large letters, the word Stress. The first few minutes of the video consisted of interviews with workers discussing how hard
they found the job when they first started, and the personal problems (in and out of work) that arose because of this.

Why would they design such a recruitment video? The point was to set accurate expectations for job applicants, to avoid painful turnover costs in the future. This was particularly important because of the unusual organization of the Cummins plant – new recruits had never seen anything like it elsewhere.

The second part of the video showed workers describing how they eventually got past the initial adjustment to the job, and found the work to be challenging and motivating (this is useful to remember when reading Chapter 7). But the primary point was to serve as an honest warning to job applicants that they would be better off not applying if they did not fit with this particular job. It was a very powerful recruiting video.

Source: author’s personal knowledge of unpublished video.

Reducing Costs of Losing Key Employees

Some turnover is inevitable, but firms can employ a few strategies to make turnover less costly. In the last chapter we discussed non-compete agreements. They generally have limited effectiveness, because of the reluctance of courts to enforce strong clauses in them, and because it is impossible to control some information and ideas that an employee carries to a new firm. However, there are a few alternative approaches that can help.

First, turnover is most costly when the worker has complex, detailed knowledge that other employees do not share. Going back to our Silicon Valley software company that sells tax management software, if the primary routines in the program were written by one employee, the firm would be in serious trouble if that employee left. Software code can be extremely complex and hard to understand if you did not write it yourself.

This suggests some policies that can help avoid such a problem. First, have workers collaborate on key tasks so that key knowledge is not monopolized by one employee. Second, cross-train to reduce the risks even further. By having each worker train colleagues in what they do, and perhaps switch tasks periodically, each develops a broader knowledge of the product or process. If one person leaves, it is easier for others to fill in, and they will already be somewhat up to speed on the work.

Job design can also affect turnover costs. The more standardized jobs are, the less costly to the firm is losing one employee, because others can step into the void. Of course, not all jobs can be standardized, especially in smaller organizations.

Finally, a firm could have a general knowledge management strategy. That is, some attention can be given to procedures by which the knowledge that is created as part of conducting the work can be documented for reuse. As an example, some consulting firms set up databases to document new methods that their consultants have devised on projects. At the end of a new project each consultant is expected to write a description of new ideas, products, etc. that they have created on the project, and submit them to a manager who is responsible for knowledge management. That manager enters the description into the database, along with a set of keywords. Later, this knowledge can be accessed by others who search on appropriate keywords. They can then apply the ideas to new applications without having to figure out new solutions from scratch. To the extent that this system
works well, it saves the firm from having to reinvent approaches, and instead allows them to leverage what they have already created. And, it allows the firm to capture at least some of the knowledge of employees if they quit – as long as they documented what they learned.

**Embracing Turnover**

As we have noted before, turnover is not always a bad thing for an organization. In fact, some organizations embrace turnover. Here are two quick examples to illustrate why this might be beneficial, and how a firm might usefully encourage turnover through its personnel policies.

The first case is professional service firms that have up-or-out systems. An up-or-out system, like our probation analysis in Chapter 2, is one in which employees who are not promoted must leave the firm and seek a new job. Such systems are quite common in professional service firms (consulting, law, and accounting) as well as universities (professors). In a professional service firm, because employees work closely with clients, it is quite common for an employee who leaves to go to work for their client. This reinforces the working relationship between the firm and its client, benefitting both.

Next consider Hewlett Packard. They are one of the original technology companies in Silicon Valley. As the Valley developed, many more technology companies entered the area, competing for HP's employees. In addition, many HP employees quit to start their own companies, often competing with HP.

For many years, HP's response when an employee quit was to encourage them in their new venture, but also encourage them to apply to return to HP in the future if their new venture was not successful.

Such a scheme would seem to motivate HP employees to use company resources to develop new product ideas, and then quit to profit from those ideas on their own. This was less of a risk at HP than at other firms, because during this period HP had strong internal policies to motivate employees to develop new products from within HP (see Chapter 14). Why might HP have such an approach?

First, such employees may well be some of the best in HP's workforce – that is why they have such good outside opportunities. HP's policy can increase the quality of its workforce because some may return later. Second, like the professional service firm, those employees who leave may bring future business to HP. Third, employees who leave HP and then return may be valuable because they have a mix of inside and outside experience. That is particularly important in an industry that is dynamic and constantly changing.

There is an additional benefit that may accrue to HP from this approach. It is essentially taking the point of view that when its employees do well, it will also do well. By encouraging them to pursue successful careers, HP is probably able to recruit employees that are more talented and ambitious. It is also able to develop a reputation as an employer who cares about the interests of its employees, which is likely to increase motivation and reduce conflict in the workplace. This is a broad theme that we are seeing in this text: the interests of the firm and its employees are not in conflict, when thought about properly. We return to this at the end of the book.
A difficult question is how to respond when an employee receives an outside offer and threatens to quit. We discuss those issues now. First, however, we consider the related question of whether or not you should try to raid employees from competing firms. Both illustrate that firms are engaged in an active auction market, bidding against each other for employees, especially for the most talented ones.

Raiding Other Firms: Benefits and Pitfalls

Sometimes an individual who is working at another firm is a particularly attractive hire. This is generally the case when the individual has some idiosyncratic set of skills. It is the unusual nature of the skills that make raiding another firm, rather than hiring from the pool of self-announced applicants, attractive.

If an individual has skills that are commonly found throughout a large fraction of the workforce, the disadvantages of hiring from another firm may outweigh the advantages of hiring from the pool of applicants. The major disadvantage of hiring from another firm is that the worker’s current employer usually knows more about a worker than does an outsider. Outsiders are usually in a weaker position to judge the quality of a worker.

This is sometimes called the Winner’s Curse problem. More often than not, the workers who are easy to steal are the ones not worth stealing. After all, a worker’s current firm always has the option of raising the worker’s salary to keep the employee. If the outside firm can outbid the worker’s current employer, then maybe the outsider is bidding too much. This is much like the Groucho Mark quotation at the beginning of Chapter 2. It is also another example of the idea of adverse selection: your firm bids against employers who have better information about the quality of one of their employees. Your firm is less likely to be able to hire a high quality applicant in such circumstances, because the other firm will use its information to its own advantage in deciding how to bid against you.

When a worker’s skills are sufficiently rare, and when those skills are a particularly good match with an employer other than the current one, it may pay for an outside firm to launch a raid. Figure 4.1 illustrates all the decision possibilities related to raiding an employee by an outside firm. First, the firm decides whether to raid a competitor for a potential hire. If it does, it incurs the costs of time and other resources involved in the bidding war. If it does not, it incurs no direct costs.

The outcome then depends on whether the employee is more valuable to his current employer, or to the raiding firm. The firm that values him more will eventually outbid the other, and have the employee in the future. These cases are illustrated as the two branches coming off of the Raid decision in Figure 4.1. This leads to four possible outcomes if the outside firm decides to raid, and two if it does not.
Workers generally have specific skills that make them more valuable to the current employer than to an outsider. But such is not always the case. There are situations in which a worker has skills that are so special and so well-suited to another firm’s current situation that the outsider is willing to pay more for the worker than is the current employer.

Consider, for example, the case of Lee Iacocca, who was hired away from Ford by Chrysler. Chrysler was on the verge of closing, and Iacocca was judged by Chrysler’s board as one of the few people with the skills to steer the company back to prosperity. As such, he was worth more to Chrysler than he was to Ford. Ford either underestimated his talents or believed that he was not as valuable to Ford because of Ford’s stronger economic position.
The Iacocca situation is illustrated by box 1 in Figure 4.1. In this case, Chrysler elected to make Iacocca an offer, and Iacocca is more valuable to Chrysler than to Ford. In the end, Chrysler succeeded.

The problem with raiding another firm is in box 2 in Figure 4.1. Because of matching and investments in firm-specific human capital, a worker is more likely to be worth more to the current employer than to the outsider. Under such circumstances, the outsider should not raid. However, the outsider has less information about the quality of the worker than the current employer does, and sometimes overestimates the worker's potential productivity. If it does, it might win the bidding war and hire the new employee. However, this would be a mistake. Given the current employer's natural information advantage, this is often likely to be the case.

When should a firm raid another firm for an employee? The first criterion is that the raider must be very certain that the target worker's value is higher to the raider than to the worker's current firm. The second criterion is that the worker's current firm not overvalue and therefore overpay the worker.

If the worker is more valuable to the current firm than to the raider, then either the raider acquires the worker but pays too much, or more likely, the raider simply fails to attract the worker away and wastes time and money in the process.

Raiding is most likely to yield profitable outcomes when the target worker is worth more to the raider than to the current firm and when the current firm is aware of that fact. If the worker is worth more to the raider than to the current firm, the raider will succeed in outbidding the current employer unless the current firm overvalues the target worker as well.

In which situations are the conditions for a profitable raid most likely to be met? The main condition, that a worker is worth more to another firm than to the current firm, is most likely to hold when recent changes have occurred, either with respect to the worker's skills or in the industry in question. Some examples come to mind.

First, workers who have recently completed a schooling program are ripe for the picking. Chances are that a new degree recipient can be more productive in a job other than the one they currently hold. Their current firm may be able to offer the better job, but it is quite likely that the current firm does not have an opening in an appropriate higher-level position. A raiding firm will probably succeed in attracting the new graduate from the current firm. Indeed, the statistics on this are impressive. Schools that have part-time MBA programs report that the vast majority of their graduates leave the firm that they were with during graduate school within a short time of graduation.

Second, workers who are employed by firms in rapidly changing industries, especially those that are declining, are good targets for raiding. Since the firm in which the worker is currently employed is changing, the expectations that brought the worker to the firm in the first place are probably not being met. As a result, the worker's value at the current firm is probably below what it could be elsewhere. This is exactly the situation in which the worker is more valuable elsewhere than at the current firm.

Third, workers who are employed in industries that are undergoing rapid technical change are likely to be good targets. When change is rapid, it is not neutral, so that some firms experience significantly more rapid increases than others. Workers who are capable, but are located in a firm that is behind the leader, are good candidates for a raid. This explains why there is so much turnover among software and hardware companies. A worker who starts out with an innovative firm may find that firm to be trailing the industry within a period
of six months. The worker is probably worth more elsewhere, and will seek out a new job or will be pirated away by a leading firm.

Is it always better to be a raider? If that were the case, all firms would raid the best employees from each other and nobody would hire unproven talent. Firms that hire directly from the pool of applicants get a random sample of the population. Some of the workers are very able, and some are less able. As long as the firm does not pay more than commensurate with the average quality of a worker, it can survive quite well. But firms must recognize that the average quality of the workers employed by the firm that hires all applicants is not as high as the average for the population as a whole. Raiders are going to steal away a non-random sample of the population. Specifically, raiders tend to steal away the most able. Thus, the wage that a primary employer pays to its workers must be low enough so that it can avoid losses even after the secondary raiding firms have picked off some of the better employees. So, for example, if the average worker is worth $30 per hour, a wage of $30 per hour will cause a firm to lose money. Since the better workers tend to be picked off, those who are left do not provide an average productivity of $30 per hour.

**Layoffs and Lemons**

Firms have discretion over which employees to terminate, which implies that the idea of adverse selection can be applied to workers who have lost their job. From the perspective of the labor market, such employees are like used cars. A potential employer will worry that the worker is not of high quality (a “lemon”), because their prior employer let them go. This (and lack of specific human capital) can make it quite difficult for someone who is laid off to find a new job quickly, and may also imply that the new job is at much lower pay than the old job.

Studies have found that used cars that come onto the market because they were leased, but the lease contract expired, are much less subject to the lemons problem. The reason for this is that almost all cars that come off a lease contract will be sold as used cars, regardless of quality. There is little self selection, so they tend to be of average quality.

A similar result applies in the labor market. One study found that workers who are laid off when all workers lost their jobs – due to a plant closing, for example – were able to find jobs more quickly, and generally earned higher pay in their new jobs, then workers who were laid off in other situations. The workers who suffered from a plant closing were able to explain their job loss in such a way that it did not stigmatize them.

*Source: Gibbons & Katz (1991)*

**Offer Matching**

We now realize that bidding for a worker is a common part of labor market competition. But sometimes, current employers refuse to match offers of outsiders. An announced policy of “no offer matching” is thought to discourage disloyal attempts by employees to raise their salaries by obtaining outside offers at higher wages.

When is it reasonable to match outside offers and when not? First, it is important to determine what affects worker search behavior. Let’s take a stylized example. Suppose that a
worker is currently earning $20 per hour. Suppose further that there is one (and only one) job in his labor market that will pay him more than this; it will pay $20.50 per hour. There are 50 firms other than his own where he believes that he has some prospect of finding the high-paying job. The problem is that he doesn’t know which firm of the 50 is the high payer.

The worker can take time to fill out applications for some of the 50 firms. Each application takes time and effort, which the worker values at some amount, say $X. The question is, When does it pay to search?

Suppose that the worker’s current firm agrees to match any outside offer. Then he will search whenever the expected present value of the increased earnings exceeds the costs. The expected present value of searching at the first firm is:

$$\frac{1}{50} \sum_{t=0}^{T} \frac{(2000)(0.50)}{(1 + r)^t}.$$ 

This is like a formula for the return on a human capital investment, because searching for a job is like investing in human capital. The 1/50 comes from the fact that the worker has a 1/50 chance of finding the high-paying job. If he does, because his firm matches the offer, he gets to enjoy an increase in hourly wage for 2,000 hours per year for every year that he works from time zero to T. If this expression is greater than the cost $X$, he searches. Low values of $X$ and long work lives tend to lead to higher return to searching for a better job offer.\(^3\)

The firm is unhappy that the worker is searching for a high-paying job. Under offer matching, the search may result in the firm having to pay higher wages for any given worker. How will a policy of “no offer matching” affect the worker’s behavior?

This depends on whether the worker will actually leave to take another job. If he is willing to leave for the new job, then there is no effect on search behavior of announcing no offer matching. The returns to the worker from search are the same. The only difference is that in order to reap the returns, it is necessary that the worker actually take the job offer at the new firm. With offer matching, the worker need only threaten to leave.

Which situation does the firm prefer? With no offer matching the worker leaves. With offer matching, the worker stays. Although the firm might not always want to match offers, one would think that the firm would always prefer having the option to match offers. If the wage that the firm must pay exceeds the worker’s worth, then the firm can always let the worker go.

Since the policy of “no offer matching” does not seem to discourage search, why announce it? The answer is that it might discourage certain kinds of search under special circumstances.

Suppose, for example, that the worker has a strong preference for his current firm, perhaps because he likes his co-workers, the plant’s location, or its general work environment. The worker might be unwilling to move to another firm, even at a wage of $20.50

---

\(^3\) If the worker applies to the first firm and fails, he will surely apply to a second. The returns to applying to the second are greater than applying to the first, because the odds of finding the high paying firm have risen to 1/49. Of course, in practice a worker will first apply to the firm where he believes the odds of success are highest, which works in the other direction.
per hour. If the firm matches alternative offers that are below the worker’s value to the firm, it will raise wages to $20.50. This provides the worker with an incentive to search for an offer that he has no intention of accepting. If the firm had a policy of “no offer matching,” the worker would not search, because he would actually have to accept the job in order to reap the benefits of his search. Since the worker is unwilling to accept the job at the higher wage, the policy of no offer matching discourages search and saves the firm money.

Now, were the firm to know that the worker was unwilling to accept the $20.50 to work for a competitor, it could simply refuse to match that offer. In essence, the firm could declare that the offer was not a genuine threat, because the worker had no intention of accepting it. The problem is that it is sometimes difficult for a firm to distinguish between genuine threats and offers that the worker would decline. When this is the case, firms can benefit from a policy of no offer matching.

This situation is most likely to arise when the following conditions apply:

1. There is a large non-pecuniary component to compensation.
2. The worker is being paid less than he or she is worth to the firm.

**Compensation with Non-pecuniary Component**

Wages can be compared easily, but compensation packages that include a non-pecuniary component may be difficult to compare. If jobs were identical, except for their monetary compensation, the firm could quickly determine whether a worker was likely to accept an outside offer. If the dollar wage offered by the competitor exceeded that being paid by the current employer, the worker would accept the offer and leave. Under these conditions, the outside offer is a credible threat to the current employer.

Unfortunately, evaluating offers is not so easy. Much of what a worker receives on the job is psychic. Working conditions, status, flexibility, or the ability to work in a particular location may be quite important, but may have different values to different individuals. When the non-pecuniary aspects of a job are important, it will be difficult for an employer to evaluate the significance of an outside offer. Workers are then more likely to search for outside offers that have high monetary rewards that compensate for physic disadvantages unobservable by the current employer.

As such, the gains to a worker from engaging in a strategic search for offers that are high, but which the worker has no intention of accepting, are largest when the non-pecuniary component of compensation is significant. Under the circumstances, an employer benefits from discouraging search for non-serious offers by announcing a policy of no offer matching.

When workers are in the job “for the money,” the non-pecuniary components of the job are less important. This is a situation where little is gained by announcing a policy of no offer matching. Workers will simply move if the firm does not match, and the firm cannot discourage needless search by refusing to match. Investment banks rarely adopt “no offer matching” policies, because money is the driving factor in these industries. At government agencies, non-pecuniary components of the job, including job security, short hours, and perhaps easier workdays, are of significant concern to workers. At these firms, it is more difficult for management to know whether an outside offer at a slightly higher wage is indeed a credible threat. To prevent workers from searching for such offers merely to raise their salaries, public utilities can refuse to match outside offers.
**Undervalued Workers**

A firm that matches offers is most susceptible to disingenuous search by workers when the firm has much to gain from keeping the worker. If the firm were paying the worker exactly what she was worth, then an offer from an outsider that exceeds her current wage – credible or not – would elicit no response from the current employer. It would be better to lose the worker than to increase the wage.

The worker is most likely to succeed in getting her wage raised by an outside offer when the firm makes a large profit on her. This happens when the worker is paid less than she is worth. If there is a surplus to be had, the worker can capture some of that surplus by threatening the employer with a departure. It is under these circumstances that a policy of no offer matching is most likely to be profitable.

To summarize, retaining flexibility is generally better than being locked into a rigid policy. As such, a policy that prevents a firm from matching offers is usually a bad idea. To every rule, however, there is an exception. When a worker’s current employer is uninformed about a worker’s willingness to accept outside offers, a policy of not matching offers may be a good idea. This is particularly relevant when the firm is paying its workers less than the value of their output.

**Layoffs & Buyouts**

Unfortunately, sometimes firms must downsize by laying off large groups of employees. If you had to lay off part of your workforce, how would you think through the issues? For example, should you target the most highly paid workers first? If you decide to offer buyout packages to motivate workers to leave on their own, how can you do so most effectively?

**Who to Target for Layoffs**

Should you lay off your most expensive employees? Those with the worst performance? The answer is not necessarily, but these are good places to start looking for candidates.

As we saw in Chapter 1, pay must be balanced against productivity. One must be careful about laying off the most highly paid employees, because they are often the most productive as well. A better approach is to target those employees from which the firm is losing money relative to other employees. These could be high or low paid workers. That said, some highly paid employees have high compensation because outside offers were met, they are tough bargainers, or for other reasons. Therefore, highly paid but relatively unproductive workers are good candidates for layoffs.

A similar argument applies to employee performance. Again, employees with low performance ratings may have high productivity relative to their compensation, so one must be careful. However, low performance ratings usually imply that the worker is performing poorly compared with others with similar skills and jobs. If that is the case, odds are very good that the ones with low performance ratings are being paid too much relative to more productive colleagues, and thus should be targeted for layoffs.
Specific Human Capital

An important factor in deciding who to target is the degree of firm-specific human capital. As we saw in the last chapter, to the extent that workers have human capital that is specific to the employer, both the worker and the firm are likely to be sharing the costs and benefits of training investments. This has an important implication for thinking about who to target for layoffs.

It is quite straightforward to state the result, but the analysis will be somewhat lengthy: When firm-specific human capital is important, the firm maximizes its profits by laying off from both ends of the age distribution first. These are the workers who have recently started with the firm, and those who are nearing retirement.

The intuition behind this result is presented in Figure 4.2. The top panel shows profiles of a hypothetical worker’s earnings and productivity over the career, with an investment in firm-specific human capital. Productivity at the firm is labeled $K_t$. The wage is labeled $W_t$.

The value of the worker’s best alternative outside of the firm is labeled $A_t$. This depends on two factors. The first is the earnings that the worker could receive in another job. That is the most important factor for younger workers. The second is the value that the worker places on leisure. The older the worker, the more he will tend to value leisure. At some point, a worker’s best outside alternative is retirement. Furthermore, eventually all workers would be better off by retiring. That is represented by a rising profile $A_t$. If the older worker’s outside earnings are less than what could be earned at this firm, then the optimal point of retirement is at $t = T$, where $A_t$ rises above $K_t$. 
In competitive labor markets, the present value of the \( W_t \) profile must approximately equal the present value of the \( K_t \) profile. If \( PV(W) > PV(K) \), the firm would lose money over the worker’s career. If \( PV(W) < PV(K) \), the firm would have difficulty recruiting workers. The present value of both should generally exceed the present value of the \( A_t \) profile, or the worker is employed at the wrong firm.

Though the present values of wages and productivity are equal at the time when the worker is hired, they are not thereafter. Since the worker’s training is partly firm-specific, the worker and the firm will share the costs and benefits of the training. Thus both bear losses initially. Once the training is complete, both earn a return on their investment. At the beginning, \( W_0 > K_0 \) as the training begins. Therefore, at any point after \( t = 0 \), the present value of the productivity profile \( K_t \) is higher than the present value of the wage profile \( W_t \). The difference represents the part of the return that is earned by the firm.

The amount of profit \( P_t \) going to the firm is plotted in the bottom panel of Figure 4.2. Profit is defined as the difference between the present values of \( K \) and \( W \). What is the shape of \( P_t \)? First think about an employee who is about to retire. Although pay is below productivity, there is little remaining profit for the firm to earn since there is little time left in his career. Similarly, the firm loses little from an employee who is very new to the firm, because little training investment has yet been made. In the limit, \( P_t = 0 \) when \( t = 0 \) and when \( t = T \). In general, the firm earns the greatest profits (in present value) on workers who have completed their training and have both high productivity and relatively long remaining careers – these are workers of medium age.

It is this point which provides the intuition behind this discussion. Imagine that productivity falls, perhaps because demand (and prices) for the firm’s products has declined. This is shown as a drop in \( K_t \) to \( \beta \cdot K_t \), where \( \beta < 1 \). At current wages, this corresponds to a downward shift in the present value \( P_t \) to \( \tilde{P}_t \). It no longer pays to make investments in young workers. Similarly, it would be profitable to lay off older workers, since their present value to the firm is lower. The only employees that would be profitably employed in this circumstance would be workers with middle ages.

**Costs of Layoffs**

Laying off younger workers is unlikely to be very controversial. They are typically not legally protected. Moreover, they have invested little in firm-specific skills, so they are likely to lose little when moving to another job. Thus, a policy of FIFO (first in, first out, as in the accounting term) layoffs can effect the goal of reducing the number of young employees.

Laying off older workers is quite likely to be controversial, and may also be illegal. Older workers are protected by anti-discrimination regulations in most countries. Technically, the firm’s lawyers might argue that the layoffs are based on net present values for each employee, but they are not likely to win that case. Thus, the firm may be subject to prosecution by the government, and discrimination lawsuits by older employees.

Moreover, older workers have invested in firm-specific skills, and are now enjoying the returns on their investments that were promised to them (usually implicitly) by the firm. Laying off such workers may be perceived as a breach of trust by the employer, though this may be alleviated to the extent that the firm honors its pension commitments.

---

4 See the discussion of rent sharing near the end of Chapter 3.
Is it truly a breach of trust? That is not so obvious. Any sensible implicit contract would surely provide for the right of the firm to lay off workers when business conditions become severe enough. This might be one of those times. While that may be true, it is still likely that the firm will face criticism (perhaps opportunistic) about breach of trust. Therefore, to the extent that the firm cares about its reputation as a fair employer, it should think carefully before implementing layoffs, either for older workers or more generally.

For example, suppose that the firm has a strong reputation in the labor market, and the current downturn is believed to be only temporary. In that case, the firm will care very much about its reputation, and implement layoffs in such a way that its reputation is damaged least. On the other hand, if the industry is declining dramatically, or if the firm needs to signal about the seriousness of its intentions (say, as a negotiating stance with unions), then dramatic action may be necessary.

Finally, a serious cost of layoffs is the litigation that may ensue. Because many economies provide protection for workers against wrongful termination, employees may sue if they are fired. Such litigation is costly in its own right, as would be any settlement to avoid legal costs. Moreover, if the firm loses the lawsuit, it will have to pay damages.

**Buyouts**

Because of the costs of layoffs, many firms opt to offer employees buyouts instead. A buyout is a contract between a worker and a firm. In exchange for some compensation, an employee agrees to end employment with the firm. Buyout agreements may include other clauses; e.g., the employee agrees to not sue the company for wrongful termination, and to not criticize the company in public forums.

**Amazon’s Non-Disparagement Clause**

In early 2001 Amazon.com laid off 1,300 workers, offering severance pay of six to eight weeks of salary. Employees were required to sign a clause that prohibited them from making derogatory comments about the company, or their severance would be limited to two weeks of pay. After public criticism of this clause, Amazon removed the requirement from the buyout package.

The buyout package had another feature that is very unusual. Amazon set up a trust fund of $2.5 million worth of company stock, to be sold two years later and distributed to laid off employees. In effect, those laid off were given some options.

Why might Amazon have done this? One answer is that this was right before the tech market crash, so there was still great enthusiasm for options (and they were often used without much thought as a compensation tool; see Chapter 12). Another is public relations. A third is that Amazon wanted to tie the interests of those who were laid off to Amazon, both to discourage public disparagement, and perhaps because they hoped the layoffs to be temporary, and wanted to hire some back later.

*Sources: internetnews.com; news.com.*

If a firm does opt to award severance pay, should all employees be eligible? It depends on who is most likely to accept the buyout offer.
One concern is adverse selection. At any wage category some employees are more productive than others. In other words, some are relatively overpaid, and some (the better performers) are relatively underpaid. The more productive are also likely to also have better alternative employment elsewhere. If this is the case, then they lose the least in accepting the buyout, and are thus more likely to accept the buyout.

Consider the attempt by Stanford University in the 1990s to offer early retirement buyouts to professors older than 55 years of age. A number of professors accepted the offer. Unfortunately, in many cases the professors who departed were the most productive because they found it easier to obtain good jobs at other universities than did less productive professors. This suggests that buyout packages should be carefully designed so that they target and motivate the desired groups to leave or stay, respectively. For example, the best performing employees might not be offered buyouts, or might be offered buyouts that are less lucrative, if possible.

Similar considerations apply to how buyout packages might vary with the employee’s age. As described above, the main concern will generally be the most senior workers. Of that group, those close to retirement have little to lose from leaving the firm, since they have already earned most of the return on their investment in skills. They require only small buyout packages. Those further from retirement usually require larger buyouts. To illustrate this, Table 4.1 presents a hypothetical situation similar to Figure 4.2.

<table>
<thead>
<tr>
<th>Age</th>
<th>W</th>
<th>A</th>
<th>K</th>
<th>PV(W)</th>
<th>PV(A)</th>
<th>PV(K)</th>
<th>βK</th>
<th>PV(βK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>$30</td>
<td>$20</td>
<td>$20</td>
<td>$145.5</td>
<td>$99.3</td>
<td>$145.5</td>
<td>$14.0</td>
<td>$101.8</td>
</tr>
<tr>
<td>26</td>
<td>30</td>
<td>20.1</td>
<td>23.2</td>
<td>145.5</td>
<td>99.9</td>
<td>158.1</td>
<td>16.2</td>
<td>110.6</td>
</tr>
<tr>
<td>27</td>
<td>30</td>
<td>20.3</td>
<td>26.2</td>
<td>145.5</td>
<td>100.5</td>
<td>169.9</td>
<td>18.3</td>
<td>118.9</td>
</tr>
<tr>
<td>28</td>
<td>30</td>
<td>20.4</td>
<td>29.1</td>
<td>145.5</td>
<td>101.1</td>
<td>181.1</td>
<td>20.4</td>
<td>126.7</td>
</tr>
<tr>
<td>29</td>
<td>30</td>
<td>20.5</td>
<td>31.8</td>
<td>145.5</td>
<td>101.7</td>
<td>191.5</td>
<td>22.3</td>
<td>134.0</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>20.6</td>
<td>34.4</td>
<td>145.5</td>
<td>102.3</td>
<td>201.2</td>
<td>24.1</td>
<td>140.8</td>
</tr>
<tr>
<td>35</td>
<td>30</td>
<td>21.3</td>
<td>45.0</td>
<td>145.4</td>
<td>105.3</td>
<td>238.6</td>
<td>31.5</td>
<td>167.1</td>
</tr>
<tr>
<td>45</td>
<td>30</td>
<td>22.5</td>
<td>55.0</td>
<td>144.3</td>
<td>110.5</td>
<td>258.7</td>
<td>38.5</td>
<td>181.1</td>
</tr>
<tr>
<td>55</td>
<td>30</td>
<td>23.8</td>
<td>50.0</td>
<td>134.0</td>
<td>109.1</td>
<td>211.3</td>
<td>35.0</td>
<td>147.9</td>
</tr>
<tr>
<td>56</td>
<td>30</td>
<td>23.9</td>
<td>48.7</td>
<td>131.0</td>
<td>105.8</td>
<td>191.2</td>
<td>34.1</td>
<td>141.3</td>
</tr>
<tr>
<td>57</td>
<td>30</td>
<td>24.0</td>
<td>47.2</td>
<td>127.3</td>
<td>103.2</td>
<td>179.6</td>
<td>33.0</td>
<td>125.7</td>
</tr>
<tr>
<td>58</td>
<td>30</td>
<td>24.1</td>
<td>45.6</td>
<td>122.5</td>
<td>99.7</td>
<td>166.8</td>
<td>31.9</td>
<td>116.7</td>
</tr>
<tr>
<td>59</td>
<td>30</td>
<td>24.3</td>
<td>43.8</td>
<td>116.6</td>
<td>95.3</td>
<td>152.7</td>
<td>30.7</td>
<td>106.9</td>
</tr>
<tr>
<td>60</td>
<td>30</td>
<td>24.4</td>
<td>41.9</td>
<td>109.1</td>
<td>89.5</td>
<td>137.2</td>
<td>29.3</td>
<td>96.0</td>
</tr>
<tr>
<td>61</td>
<td>30</td>
<td>24.5</td>
<td>39.8</td>
<td>99.6</td>
<td>82.0</td>
<td>120.1</td>
<td>27.9</td>
<td>84.0</td>
</tr>
<tr>
<td>62</td>
<td>30</td>
<td>24.6</td>
<td>37.6</td>
<td>87.7</td>
<td>72.4</td>
<td>101.1</td>
<td>26.3</td>
<td>70.8</td>
</tr>
<tr>
<td>63</td>
<td>30</td>
<td>24.8</td>
<td>35.2</td>
<td>72.7</td>
<td>60.2</td>
<td>80.0</td>
<td>24.6</td>
<td>56.0</td>
</tr>
<tr>
<td>64</td>
<td>30</td>
<td>24.9</td>
<td>32.7</td>
<td>53.8</td>
<td>44.7</td>
<td>56.5</td>
<td>22.9</td>
<td>39.5</td>
</tr>
<tr>
<td>65</td>
<td>30</td>
<td>25.0</td>
<td>30.0</td>
<td>30.0</td>
<td>30.0</td>
<td>30.0</td>
<td>21.0</td>
<td>21.0</td>
</tr>
</tbody>
</table>

Analysis of Which Workers to Lay Off
Table 4.1

In this table all values are expressed in $1,000s. The table shows the wage (assumed for simplicity to be a constant $30,000), the value of the worker’s best alternative $A_t$, and
productivity $K_t$. The table also calculates the present values of each.\(^5\) Notice that, as in Figure 4.2, the present values of $K$ and $W$ are equal for new and retiring workers. As described above, competition forces this result, all else equal. Finally, the last two columns imagine that the value of the worker’s productivity falls by 30 percent ($\beta = 0.7$) because of a decline in demand for the firm’s products.

In deciding whether or not to accept a buyout of amount $B$, the worker will compare what he can earn if he stays in the firm, $PV(W)$, to what he can earn if he takes the buyout, $B + PV(A)$. This means that the buyout will be accepted only if it is higher than the profit that the worker earns from staying:

$$\text{Profit to the worker from staying (rejecting a buyout)} = PV(W) – PV(A).$$

In the table, the present value of wages is higher than the present value of the best alternative for all workers. All workers would have to be offered a buyout to be willing to leave.

The profit to the firm if the worker leaves is the difference between the present values of compensation and productivity:

$$\text{Profit or loss to the firm from having the worker leave} = PV(W) – PV(K).$$

(After production falls, this expression changes to $PV(W) – PV(\beta K)$ in our example.) If this is positive, the firm would like the worker to leave. In that case, the expression equals the maximum profitable buyout $B$ that the firm can offer. If it is negative, the firm would prefer to keep the worker. In the table, all workers were profitable to the firm before the value of productivity fell. After productivity fell, workers aged 57 and higher or 30 and lower are no longer profitable for the firm. These are the ones that the firm would like to target.

We are now in a position to give the rule for optimal buyouts. As long as the gains to the firm from losing the worker exceed the losses to the worker from leaving, there is room for a deal to be made. The firm can offer a buyout that increases its profits, and also makes the worker better off when they leave. Putting these together, we can make a deal as long as:

$$PV(W) – PV(K) > PV(W) – PV(A),$$

or in other words:

$$PV(A) > PV(K).$$

Our result: a buyout is possible if the present value of the worker’s alternative exceeds the present value of the worker’s productivity at the firm. Indeed, this should make intuitive sense to you. Thus, low output and good alternatives make buyouts feasible.

It is important to recognize that the individuals that the firm would like to lay off are not necessarily the ones for whom a buyout offer is possible. While the firm would like to lay off those aged 57 and older or 30 and younger (after productivity falls), a deal cannot be struck with all of them. Only workers 62 and older have alternatives sufficiently attractive to make an offer feasible. The firm loses money on those between ages 57 and 61, but not enough that a buyout would be a better option, given the amount of buyout that they would require. Similar logic applies to those aged 30 and younger.

\(^5\) The interest rate is about 25%. The principles illustrated are valid regardless of the interest rate.
Implementation of Buyouts

Window Plans

Often, announcements of buyout offers are a surprise, and workers are given only a limited time to accept the offer. Such plans are often called window plans because there is a small period or “window” during which the buyout is available.

There is a very good reason for such practices. Recall that the buyout that the firm is willing to offer the worker depends on the difference between the wage and productivity. The lower his productivity, the more anxious is the firm to be rid of him, and the higher is the buyout offer the firm is willing to make. If a buyout is anticipated, a worker has incentives to reduce productivity. A short fuse prevents the worker from strategically reducing productivity for any significant time period to gain a higher buyout price. It also reduces the chance that the worker can find a suitable outside offer, since there is less time to search.

Threat of Layoff

Another way to increase acceptance rates for buyout offers is to credibly threaten to lay off some fraction of those who do not accept the offer. Suppose that the firm were to announce that it would lay off 50 percent of all workers who did not accept the buyout offer, chosen at random. If you were faced with this offer, how would it affect your incentives? It would make you more likely to accept the offer.

Think about it this way: suppose that if a worker loses his job, he expects that the next job will give him about $10,000 less in present value. Under the analysis given above, this is the minimum buyout that he will accept. However, if the firm threatens to fire half of those who reject the buyout, there is a 50 percent chance that he will lose $10,000 with no buyout. He is then willing to accept a buyout of only $5,000 (or even less, if he is risk averse). In general, greater odds of being laid off lowers the buyout that a worker requires. The Appendix proves this result formally.

Of course, there are costs to the firm of such a strategy – it will have to lay off some workers. We have already discussed these costs above. The firm needs to balance the benefits of lower buyouts against the costs of implementing some layoffs after the buyout window ends. However, threatening to lay off those who reject buyouts has two advantages for a firm: it increases the probability that a worker will accept a given buyout amount, and it reduces the buyout that is required to motivate an employee to quit.

Speed and Extent of Downsizing

An additional benefit from implementing layoffs quickly and by surprise is that it reduces the amount of organizational trauma that is experienced. Downsizing can be highly emotional, and organizations that go through the process tend to find that workers are extremely unproductive while it is ongoing. One of the reasons for this is that workers focus, quite naturally, on who will be laid off, when, and under what terms. This can be quite a distraction from ordinary business. Thus, it often pays to get the pain over with quickly, and unexpectedly.

For similar reasons, a firm should consider laying off more workers than seems apparent at first glance. If it can do so, it minimizes the odds that it will have to do so again soon (many downsizing firms go through several waves of layoffs before they are finished). An additional advantage is that it makes it possible to thoroughly clean house in areas of the
organization that need radical restructuring. The reason for this is that the costs of firing a 
worker tend to be lower when implemented in the context of a larger set of layoffs.

Retirement Bridges

The minimum buyout price necessary for a worker who is close to voluntary retirement is 
relatively small. However, a buyout formula that offers less to 64-year-olds than to 56-year-
olds might run into legal difficulties. A provision that is less likely to encounter legal chal-
lenges, but has a similar effect, is a retirement bridge. A bridge gives a worker seniority 
credit for the purpose of pension calculations as if he had stayed on until the normal re-
tirement date. For example, if the normal retirement age is 65 and a worker leaves at 55 
with 18 years of service, he is treated as if he had 28 years of service for the purpose of 
calculating retirement benefits. Since the number of years awarded by the bridge declines 
with age, older workers are in effect given a smaller buyout award than their juniors.

Job Placement Services

Firms sometimes set up job placement services for workers they lay off or offer buyouts to. 
Is this rational, or does it merely reflect the employer’s guilt about layoffs, or perhaps an at-
tempt at improving public relations?

The practice may not only be good for public relations, but may also result in cost savings. 
Improving the odds that workers can find outside work lowers the buyout price that they 
will require. To the extent that the firm can help severed workers find new jobs, their alter-
natives are better and the firm may be able to offer lower buyout prices.

Whether this logic holds, however, depends on the firm’s efficiency in securing new em-
ployment for its workers. The service should be offered only if the firm can provide it (or 
contract with an outplacement agency) more cheaply than the worker can buy a similar 
service himself. Otherwise, it would be cheaper for the firm to simply pay for outside ser-
vices that a worker purchases (e.g., through some kind of voucher system). In most cases, 
that is probably the best route. Placement agencies specialize in relocating workers, and 
the firm is unlikely to do a better job (especially when distracted by the downsizing).

In either case, however, the point remains. Offering outplacement services in-kind or by 
reimbursement may lower the costs of implementing buyouts.

Summary

There is a fundamental tension in the operation of any company’s internal labor market, 
between the desire to sift through employees to find the best fits, and the desire to build 
employee loyalty. Sifting improves overall quality. Loyalty improves motivation, reduces 
turnover costs, and encourages firm-specific skill investments. A healthy firm balances 
these two desires to achieve some of both goals. The right balance varies from company 
to company, depending on the relative importance of each of these factors.

In this chapter we analyzed this tradeoff using the tools developed in Chapters 1-3. We 
saw that recruiting and turnover are intimately related. Understanding optimal turnover, 
and how to motivate it, requires use of several economic concepts. These include adverse 
selection, investment, and incentives.
Turnover has many benefits. It allows the organization to continuously update its talent. Not only does this increase quality; it also keeps the firm’s skills from depreciating. In changing environments with technological advancement, this can be quite important. Turnover also makes it less likely for a company to fall into the trap of becoming too inwardly-focused. A company may be very good at what it does, but unaware of how the industry is changing, and thus may find it difficult to adapt. Employees with diverse outside experiences make this less likely to happen. A final benefit of turnover is that it frees up promotion slots, allowing the company to advance and motivate its best employees.

Turnover also has many costs. A prosaic one is the costs of recruitment for the firm, and job search for the worker. A less prosaic one is that both will lose the value of any specific skill investments they have made. A more subtle effect is that, if turnover is higher, both will be less willing to make such investments in the first place.

This chapter completes the first section of the textbook. After analyzing the employee career pipeline of recruitment, investment, and turnover, we now consider what the firm will do with the employee. In other words, we turn to issues of job and organizational design in Section II.

### Study Questions

1. Google recently held an IPO (initial public offering). Because of this and the extensive use of employee stock options, many of Google’s employees are now multimillionaires. What retention problems do you foresee for the company? (Yes, it’s a nice problem to have!) What, if anything, might Google do?

2. You and a friend from school are starting a consulting firm together. You want to organize it as a partnership. Realizing that things don’t always work out as intended, how might you structure the partnership agreement to protect you and your friend from future conflict? How will that affect the ways that you each conduct business in the new firm?

3. If you receive a job offer, should you always tell your employer?

4. You are the CEO of Morgan Stanley at the time that Frank Quattrone and his team quit to work for a major competitor. Is this a threat? How can you tell? How would you react? Once the threat is passed, what might you do to avoid such events in the future?

### References


Appendix

Here we prove that if the firm threatens to lay off a fraction $p$ of workers who do not accept a buyout offer, it can offer lower severance payments.

Recall that the profit to the worker from staying (rejecting a buyout) is:

$$PV(W) - PV(A) > 0.$$  

That expression is positive, or she would quit on her own. Now consider the worker's decision (assume risk neutrality). If she accepts a buyout, she receives $B$ plus her alternative value $PV(A)$. If she rejects a buyout, she continues in her employment with probability $1-p$, earning $PV(W)$. However, with probability $p$ she is laid off, and earns $PV(A)$. Therefore, she will accept the buyout if:

$$B + PV(A) \geq (1-p)PV(W) + p \cdot PV(A).$$

Therefore, the minimum buyout that she will accept is:

$$B^* = (1-p)[PV(W) - PV(A)].$$

This expression is largest when $p = 0$, and $dB^*/dp < 0$. A greater threat of being laid off makes the worker more likely to accept a given buyout offer, and reduces the amount of buyout that she is willing to accept.
II. Organizational & Job Design

Section I viewed the firm as a pipeline of skills flowing in, being augmented, and eventually leaving the firm. While this is a useful view, it does not say much about what to do with employees once you have them. That is the topic of this section. Here we consider job design. In doing so, we also analyze overall organizational design, as similar principles apply to both. Chapters 5-6 discuss decision making and organizational structure, while Chapters 7-8 discuss design of an individual employee’s job.

In Chapter 5, we start by briefly stepping back and thinking about important principles for the design of an effective economy. By understanding key problems and tradeoffs that well-designed economies address, we develop important ideas that we can use to think about the design of smaller economic units: firms, business units, or individual jobs. In other words, we will develop and then apply a market metaphor to modeling organization inside of firms.

Our discussion of markets brings out several important ideas. First, markets are powerful information processing mechanisms. They provide collective intelligence difficult to achieve in any other way. Second, markets provide coordination between different economic agents (though not always perfectly). We will see that firms often face tradeoffs between the goals of effectively using information, and coordinating among employees or business units. This tension is the basis for our discussion of decentralization and centralization in Chapter 5.

Chapter 6 analyzes overall organizational structure. This is the point in the text where we take the most macro view of the firm; in most other parts of the book, we focus on management of individual employees. Here we consider issues such as how a firm might break its structure into separate divisions, and how it might coordinate activity across business units.

Chapter 7 moves us back to the micro level of analysis: we now consider how to design an individual employee’s job. In fact we will do so implicitly in Chapter 5, since the decision rights given to an employee (or not) are an important part of job design. In Chapter 7 we extend that argument to analyze how many and which tasks to bundle together into the same job. Once more, classical economics plays a role. Also playing a role is the psychology of intrinsic motivation: how job design might affect a worker’s motivation in performing the job. We will see that there is a nice complementarity between the economic and psychological views of job design. This chapter is one of the places in the text where the interplay between organizational psychology and personnel economics is most evident.

Chapter 8 extends the analysis from Chapter 7 to consider more advanced topics in job design. An important topic is the use and structuring of teams. We also consider “pathology” of job design: as in medical school, looking at extreme cases can sometimes be quite illuminating. This is what are sometimes called high reliability organizations. These are organizations, such as aircraft carriers, in which the organizational and job design tradeoffs are much harder to strike than in more typical firms, because the stakes (e.g., cost of failure) are so high. Thinking about
how such organizations solve their design problems can provide useful insights into how your firm can solve less dramatic, but similar, issues.

The third topic in Chapter 8 is highly relevant today: the effects of information technology on job design and structure. Our discussion in Chapter 5 highlights the importance to firm structure of using information, and of the costs of communicating information. Information technology sometimes dramatically changes the costs of communicating information, which has implications for patterns of decision making in the firm, the use of hierarchies, and individual job design.

In Section II, we will develop a point of view about firm design. As mentioned above, part of the idea is that firms can usefully be viewed as large-scale information systems. In this role they do not just process knowledge, but also create knowledge. Thus, a theme that arises in this part of the text is that a firm can design overall structure and individual jobs to optimize at a point in time, or innovate and adapt. You will see this playing some role in every chapter in this section.

These ideas also tie back to Section I. The importance of knowledge creation and exploitation by the firm helps you understand why skills may be so valuable in firms. It should also help you (especially after reading Chapters 7-8) understand what kinds of skills are particularly valuable in many jobs in today’s economy. And, it will help you see why skills have become even more valuable in recent decades, as discussed in Chapter 3.

Finally, markets provide important incentives to create value. Ownership of assets, and the ability to sell those assets on a competitive market, generates incentives for individuals to use those assets effectively. Prices for different goods and assets become, in this sense, a performance evaluation system. Like markets, firms also need to motivate effort and effective decisions by employees. That is why intrinsic motivation is discussed in Chapter 7. The importance of motivation in our market metaphor also provides an introduction to Section III, which will focus on performance evaluation and pay for performance. As we will see there, it helps to discuss job design before thinking about evaluation and incentives.
Decision Making

“Let no one say that taking action is hard. Action is aided by courage, by the moment, by impulse, and the hardest thing in the world is making a decision.” (Franz Grillparzer)

Introduction

You have just founded a new startup company. In addition to recruiting talent, a basic question that you face is how to structure your firm. Part of that is allocating decision making. Decisions will need to be made on a myriad of questions. Who should decide what? As the leader of the organization, should you make most of the decisions yourself, for consistency and control? Or, would that overwhelm you?

You have probably read in the business press that there has been a trend toward empowering workers in recent years. Should you do the same? If you push decision making downward, what problems might develop? And what exactly does it mean to empower an employee?

More fundamentally, what does it mean to make a decision? Are there different approaches to decision making? What structures are more likely to lead to correct or incorrect decisions? Which inhibit creativity and innovation? In this chapter, we will analyze these questions.

The Organization of an Economy

Before we turn to those topics, let us briefly consider the design of the largest organization: the economy itself. The optimal organization of an economy was one of the most debated issues in the 20th Century. On one side were advocates of centralized economies that were largely run by the government. On the other side were advocates of decentralized economies with much less of a role for government. The ideas raised in this debate are useful for thinking about the topics covered in Chapters 5-8.

By the end of the 20th Century, it was apparent that more decentralized market-oriented economies are much more efficient— they are better at creating economic growth, jobs, and overall prosperity. They are also much more creative and adaptable than economies that are more centrally planned. Why is this the case?

An excellent starting point in answering this question is one of the most famous passages in all of economics: Adam Smith’s Invisible Hand metaphor for decentralized economies:

“... he intends only his own gain, and is ... led by an invisible hand to promote an end which was no part of his intention ... By pursuing his own interest he frequently promotes that of society more effectually than when he really intends to promote it.” (Adam Smith, The Wealth of Nations, 1776)
Smith was remarking on the amazing power of market economies to create economic value without the government playing a major role in their functioning, despite the fact that individual agents are acting in their own selfish interests. Decentralized economies are efficient. They are *self-organizing systems*. That is, they arise and evolve largely spontaneously, and arrive at equilibrium prices and quantities on their own, without any central planner directing the market.

Consider the case of a small, ordinary neighborhood grocery store near the University of Chicago. The products on its shelves come from all corners of the world. The store sells a dozen varieties of coffee beans, from Columbia, Kenya, and elsewhere. It sells tea from Sri Lanka, China, and Japan. At this little local store, you can buy butter from Normandy, smoked salmon from Norway, and prosciutto from Italy. Milk sold there comes from Wisconsin. The breads use wheat grown in Kansas or Alberta, Canada. There are vegetables that were grown on nearby farms, and fruit from California, Mexico, and South America. Finally, you can purchase chocolate bars produced by a small manufacturer in Barcelona.

Somehow, goods from all over the planet were produced, used as inputs into other goods, shipped to Chicago, and placed on these store shelves. Somehow the supply is regular enough that the store’s goods are rarely out of stock, and are reasonably fresh. It is an amazing outcome, given that no one is directly this complex process. If you were designing the internal organization of a firm, a logical approach would seem to be to hire the most talented individuals that you can find, and have them set up and run your organization. It would seem logical that a skilled central planner could allocate resources and run an economy more efficiently than a chaotic, undirected market of individual, selfish decision makers. Yet the opposite tends to be true. Why is that?

**Markets as Information Systems**

Friedrich von Hayek elaborated on Adam Smith’s Invisible Hand argument to provide an answer.¹ His main point was that markets are a form of *collective intelligence* — a powerful *information system* that cannot be replicated by a central planner:

> How can … fragments of knowledge existing in different minds bring about results which, if they were to be brought about deliberately, would require a knowledge on the part of the directing mind which no single person can possess? (Hayek, 1945)

To understand his argument, imagine that you are a central planner running an economy. One of your many jobs is to allocate resources to produce coffee, and then to allocate the coffee to various consumers. What information do you need in order to allocate these resources efficiently?²

First, you need to know the value that different citizens place on coffee. How much do they prefer coffee to tea? To orange juice? How much would they be willing to give up in other resources to get coffee of higher quality? How much coffee will various workplaces wish to have? Hotels and restaurants?

Second, you need to know how to grow coffee. Thus you need knowledge of the technology of farming: fertilizer, weather and soil conditions, and so on. Similarly, as the central

¹ In 1974 Hayek received a Nobel Prize for his writings on this topic, as did Wassily Leontief in 1973.

² In the language of economists, you need to know demand curves and production functions for coffee, and for all other goods and services as well.
planner you need the same information about other industries, because you need to know about potential other uses for the resources devoted to producing coffee. To make coffee, you need land, water, labor, fertilizer, logistics, etc. These could all be used to produce other goods, and you will have to make appropriate tradeoffs in the use of those resources. In economic terms, you need to know the opportunity costs of the resources used to make coffee.

These kinds of information require experience and technical expertise, but are largely predictable, systematic and foreseeable. A third kind of information that you need to know is information that is unsystematic and unpredictable. For example, to run the coffee industry effectively you need to know the optimal time to plant the coffee beans, when to provide water or fertilizer, and when to harvest the beans. These may vary from one field to the next. Such information is local and idiosyncratic. By its nature, it is going to be very difficult to communicate all of this knowledge to the central planner. Quoting Hayek again:

If we … agree that the economic problem of society is mainly one of rapid adaptation to changes in the particular circumstances of time and place … decisions must be left to the people who are familiar with these circumstances, who know directly of the relevant changes and of the resources immediately available to meet them. We cannot expect that this problem will be solved by first communicating all this knowledge to a central board which, after integrating all knowledge, issues its orders. We must solve it by some form of decentralization. (Ibid)

Thus, Hayek points to the importance of decentralization to make use of information of “particular circumstances of time and place.” Decentralized markets allow individual coffee farmers to optimize their activities by making decisions using the idiosyncratic knowledge they possess from doing their day-to-day jobs. It would be virtually impossible (too costly) to provide all of that information to a central planner, and this is why the central planner’s decisions tend to be less effective.

At the same time, markets also make use of the more systematic types of knowledge described above. An individual farmer incorporates the value of coffee to consumers, because his decisions are based on the market price for coffee. He incorporates the value of alternative uses for the resources he employs, because his use of labor, land, and other inputs is based on the market prices for those inputs. He does not need to know how inputs could be used in other ways. Nor does he need to know that some of his coffee might be consumed in Chicago, and by whom. All he needs to know is the market prices for his inputs, and for his products. When he uses those prices in his decision making, he takes into account all of these types of information implicitly without having to possess any of that information except for prices.

In other words, prices provide a great deal of information needed to coordinate resource allocation across industries and even countries, without having to provide all of the detailed underlying information that a central planner would require. Prices are an economic information system.

**Markets as Incentive Systems**

Markets provide another powerful benefit in addition to effective use of dispersed knowledge. Consider our farmer again. Because he owns the farm, he has an incentive to run the farm profitably. In a centrally planned economy, he would be more like a bureaucrat, and would not have strong incentives to use the assets efficiently.
Moreover, the incentives that a market economy provides imply that resources tend to be placed in the hands of those with the information or skills that are most valuable when combined with those resources. Suppose that our farmer is not a good coffee farmer. In a decentralized economy, he has three options: he can invest in coffee farming human capital; hire help / rent out the land; or sell the land. If he chooses to hire someone, he has the incentive to hire someone who knows how to use the assets effectively. If he rents or sells the land, who will be willing to pay him the most? Someone who has the skills or information that can most profitably be combined with that land.

**Markets & Innovation**

A final important benefit of market economies is that they are great sources of innovation and adaptation. This arises from the two general principles discussed above. Incentives mean that asset owners are motivated to respond to problems and opportunities quickly and effectively. It also means that they have incentives to invest, and to create new products or services that might be profitable.

Decentralization also helps innovation and adaptation in most cases. It allows the economy to make use of the creativity and ideas of all individuals dispersed throughout society. In a more centrally planned structure, the ideas of those who are far from the central planner are not likely to be considered. Similarly, the ability to respond flexibly to local circumstances allows a market economy to adapt to new situations efficiently and quickly.

**Benefits of Central Planning**

There are cases in which markets may not generate perfect efficiency, and they are relevant when we discuss centralization below. One case is natural monopoly, caused by *economies of scale*. A firm has economies of scale when its average total costs – costs per unit of output – decline as the firm gets larger. If production of a good is subject to economies of scale for high levels of output, then a larger firm has lower per-unit costs than a smaller firm, and can drive it out of business. Such a situation can lead to monopoly, or near-monopoly, in an industry. Economies of scale are usually driven by large *fixed costs* – the costs the firm must incur to be in business, but that do not vary with the level of output. Large fixed costs must be incurred by both small and large firms, but they can be amortized over more units of output by larger firms, putting them at an advantage. Monopolies are usually regulated by the government to reduce the firm’s ability to exploit its monopoly power inefficiently.

In some cases, a government goes so far as to provide a good itself. This is often caused by a *public good* problem. A public good is a good that will not be provided by profit-seeking firms because they are unable to charge enough for the good to cover their costs. One reason for this is that for some goods, consumption cannot be prevented. For example, consider radio. Radio waves can be picked up by any consumer who owns a receiver. For this reason, it is impossible for a radio station to charge listeners to hear their programming. If they were unable to generate revenues from advertisers, there would be no private radio companies, and the government might have to provide such a good.

Another cause of market inefficiency is positive or negative *externalities*. Externalities result when a transaction between a buyer and seller imposes costs on, or provides benefits to, a third party who is not part of the transaction. The classic example of a negative externality is pollution. Because buyers and sellers of steel are not charged for the pollution that they create, a market economy ends up with too much pollution. The government may be
able to improve the situation by imposing limits on pollution, taxes on steel, or other policies.

An example of a positive externality is technology spillovers. In many cases firms can copy the ideas of others, without compensation, because patents and copyright protection are imperfect. This reduces the incentives for an individual firm to innovate. For example, if a pharmaceutical company believes that some of the benefits of its research will be copied by a competitor, it has less incentive to invest in R&D. In this case market economies may invest insufficiently in innovation, and the government can improve efficiency by subsidizing R&D.

A related problem may arise in cases where a standard is valuable. A standard is important when there is value to having more consumers use the same product. This phenomenon is called positive network externalities. A good example is fax machine protocols. If fax machines use different protocols, then the fax machines are less valuable, since there is no guarantee that you can send or receive faxes with someone else. Decentralized economies often lead to competition among firms to establish standards. Sometimes this process works well, but not always (consider the different cellular phone standards in Europe and the U.S.). A central planner could choose a standard for everyone, reducing this wasteful competition.

This last example is not quite so simple, of course. Competition for setting the standard increases innovation, widens the number of possible choices and may improve the technology. This would not happen if the government set the standard. Thus there is a tradeoff between having the standard set by the central government, or set through competition among decentralized firms. However, it is clearly the case that there can be too much competition to set the standard, so that there may be some role for the government to help coordinate in such cases.

### The Market as Metaphor for Organizational Design

A firm’s organization must provide the same important functions as a market. First, markets use knowledge dispersed throughout the economy. A great deal of information is summarized in prices, and in that way communicated cheaply throughout the system. Second, in cases where knowledge is costly to communicate, markets are effective at moving talent and decision making to the location of the knowledge. Third, coordination is achieved despite decentralized decision makers, through the price system. Fourth, markets provide strong incentives for effective decision making, investment, and creativity.

These are the concepts that we cover in the next eight chapters. A key question in Chapters 5-8 is how organizations can develop and use knowledge. We distinguish between information that is cheap or costly to communicate. A key idea is decentralization to make use of local knowledge. However, we also recognize the importance of coordination, either through incentives or other mechanisms, which sometimes leads to centralization. Finally, incentive systems are how firms approximate the role of ownership in an economy.

Therefore, our task is to develop a model of organizational design so that the firm can:

- Use both central and local knowledge effectively;
- Coordinate decisions as necessary;
- Provide strong incentives to make good, coordinated decisions;
- Innovate and adapt.
Collective Intelligence & Decision Making at Hewlett Packard

One of the ways in which markets serve as collective intelligence is prediction and risk assessment. Insurance and securities markets are very effective at pricing risks, based on the private information and assessments of those risks that individuals possess. For example, the discounted present value of a firm’s future cash flows (net of debt obligations) is summarized in the stock price of the firm. The stock price is determined by thousands of individual investors and mutual fund managers, and incorporates their knowledge and assessment of the firm’s prospects.

A similar example arises in gambling industries. The equilibrium odds of one team winning a sports event, for example, are the outcome of bets placed by many individuals, and are in effect the price for that risk.

In these two examples, markets are good at risk assessment not only because prices incorporate and summarize dispersed knowledge of informed individuals, but also because investors and gamblers have incentives to place intelligent bets.

Some firms are trying to replicate this collective intelligence effect inside their organizations. In an effort to improve monthly sales forecasting, Hewlett Packard hired economist Charles Plott to set up a trading system for its employees. A few dozen employees with relevant knowledge were given trading accounts with small budgets (about $50). They used their budgets to place bets on computer sales for the month – in effect, buying or selling futures contracts. Traders kept their profits, and won additional rewards for placing correct bets. HP found that market estimates (that is, the most likely outcomes implied by the trading between their employees) beat the estimates generated by their marketing staff 75% of the time. The company has incorporated the system into its forecasting, and is now experimenting with other applications.

Source: Kivita (2004)

Our first step is to consider whether a decision should be centralized or decentralized. The basic idea is simple. The more valuable is lower level knowledge that is difficult to communicate to top management, the more should we expect to see decentralization. That is what Hayek emphasized as knowledge of “particular circumstances of time and place.” On the other hand, the more important are coordination and control problems, the more should we expect to see centralization.

Suppose, then, that a decision has to be made, and it requires some information in order to be made effectively. The firm can move the information to the decision maker, or it can move the decision to the information.

If the information is not costly to communicate, there is no reason to decentralize the decision. Instead the firm can have the decision made at a high level in the hierarchy, and pass the information to the decision maker. This allows it to use the information, but also to coordinate well (since the most straightforward coordination mechanism is to have decisions made by top management).

If the information is costly to communicate, the firm faces a tradeoff. The more valuable the information, the more likely is decentralization the right answer. However, the greater the benefits of coordination, the more likely that centralization is the right answer.
In principle the firm can mimic the market by using incentives to provide coordination, and then decentralize the decision. However, inside firms incentive systems are nearly always imperfect, so this is not always the best approach. When we think about centralizing decisions, remember that we think of (perfectly typical) cases where coordination cannot be achieved adequately solely through pay for performance. This is the key reason why the market metaphor is an imperfect one for firm design: market prices are better measures of value than the performance measures that firms usually have available.\(^3\)

Instead of centralizing or decentralizing a decision, the firm can also take an intermediate approach. It can put the decision somewhere in between in the hierarchy, say with a middle manager. An employee in the middle of the organization is likely to possess some of the knowledge that the firm is trying to make use of, or the information can be communicated part of the way up the hierarchy at lower cost than communicating it all the way to the top. At the same time, a middle level manager is generally more likely to take into account coordination issues than is a lower level employee. If so, the firm might get a better balance of use of lower level knowledge against the need for coordination by having the decision made somewhere between the lower and higher level of the firm.

For the rest of this chapter, we analyze where and how a single decision should be made in an organization. Imagine a firm that is organized along traditional hierarchical lines. A completely centralized decision would be made by the CEO. A completely decentralized decision would be made by a line worker. Some decisions could be made at middle levels as well. The right to make each decision can, in principle, be given to any employee at any level. Most of this chapter will focus on decentralization. First, however, let us consider the benefits of centralizing a decision.

### Benefits of Centralization

**Centralization at General Motors**

General Motors has been famous for its decentralized structure since the 1920s. GM has been organized into relatively autonomous product divisions and regional sub-divisions. It recently changed its structure to use more centralization for many key decisions.

In particular, it now requires that cars made by different divisions share basic parts, and collaborate on product design. It is doing so to reduce duplication of efforts, and gain economies of scale in production and purchasing. GM is also hoping that this will speed up development, presumably since less time and effort will have to be spent communicating up and down the hierarchy and coordinating across independent units.

*Source: Hawkins (2004)*

---

\(^3\) If we could measure performance perfectly, taking into account all short and long term factors correctly, then we could outsource the employee and write a contract with them. Firms exist in part because it is difficult to price all labor transactions on a spot market. We saw some of the reasons for this in earlier chapters. We will see others in Chapter 9.
Economies of Scale or Public Goods

Different units of the firm may share common assets. These may imply economies of scale or public good effects for the whole organization. For example, different departments may share space in the same corporate headquarters. Divisions may share a common product design group. All departments can share in the costs of designing and implementing an accounting system. Shared assets may also be intangible, such as a valuable brand name, corporate culture, or particularly effective leadership.

Conceivably assets can be shared with decentralized decision making through some kind of incentive system. Most firms do try to allocate overhead expenses through their accounting systems, for example. However, it can be quite difficult to determine how much of the cost of using an asset, or how much of the credit for generating the asset, should be attributed to each unit. When shared assets are intangible, such measurement problems are even worse. In such cases decentralization may lead to distorted incentives and ineffective use of the asset. Departments that are undercharged will use too much of the asset, while those who are overcharged will use too little. Those who are not given enough credit for investing in an asset will invest less than is optimal.

An alternative is to centralize some of the responsibility for the creation, allocation and maintenance of shared assets. For example, a firm with a strong brand name is more likely to centralize decisions about product lines than is a firm with a group of unrelated products. A firm with a strong reputation as an employer is more likely to centralize human resource policies.

Better Use of Central Knowledge

Hayek emphasized the importance of knowledge that is dispersed throughout the system. However, in some cases the most important knowledge resides at the top of the firm. Consider a firm that has operations in many regions. As information from different divisions flows to the central office, that office may see patterns and trends that are not apparent at the local level. Such a big picture view based on the combined experience of the entire firm means that some decisions should be centralized. In many cases, these decisions involve overall strategy.

Similarly, centralization can improve the use of knowledge because it improves the transfer knowledge across units. Unless units communicate directly with each other, they cannot learn from each other's experience, unless central management passes on the lessons.

Coordination

The most important benefit of centralization is to improve coordination. When coordination is more valuable, the firm will centralize more decisions. What are examples of coordination problems?

Coordination can be necessary because output from different units of the firm must be combined in some way. An assembly line is a classic example. An example at the macro level is when one division's product is used by a different division as part of the creation of their product. Of course, whether output of different divisions of a firm must be combined to create the product is endogenous: the firm chooses its divisional structure. It may well do so to avoid such coordination problems. However, these problems do arise in many circumstances.
Coordination problems may arise when different units need to synchronize. Consider the military in wartime. If the infantry goes in before the artillery softens up the battlefield, disaster will ensue. In the case of war it is essential that the units act in a certain sequence. In order to ensure that, decisions about when to act might be centralized. (An alternative would be to let them communicate with each other, an example of what we will call a lateral mechanism for coordination.)

Strategic decisions are often centralized. That is because strategy, by definition, usually involves consideration of most or all business units. However, it is not quite obvious that strategic decisions should always be centralized. In some businesses, the firm’s strategy is, essentially, decentralization. This is the case for firms where innovation is important, but where individual products do not need to work with each other (the different departments of a university is a good example). However, to the extent that the firm’s products do need to be compatible or consistent with each other, strategy must be centralized. This is similar to having the government determine a standard for technology.

**Strategic Planning at Disney**

In early 2005, Robert Iger took the reigns as the new CEO of Walt Disney Co., replacing the famed Michael Eisner.

Disney is a very large, multidivisional firm. For years, Disney’s strategy had largely been in the hands of its Strategic Planning group at the head office. In the past divisions could initiate new products, joint ventures, or acquisitions, but all major strategic decisions had to be approved by the strategic planning unit.

Less than two weeks after taking over, Iger shut down the strategic planning office. The powers that it once had were almost completely decentralized to the divisions. Larry Murphy, who had originally run strategic planning when it was set up by Eisner, stated that, “The disbanding of strategic planning is the natural evolution from a centralized system to a more decentralized control.”

This change probably makes sense for several reasons. First, centralization fit better when Eisner was CEO, because Eisner had long experience at Disney and knew its operations intimately – he built up much of the company personally. While Iger worked at Disney for many years, he could not have the deep knowledge of all aspects of the business that Eisner had. Thus many decisions would be made more effectively by the divisional managers. Second, Disney had emphasized centralized strategy for a long time, so most of the benefits of coordination across product lines (e.g., films, video distribution, and television) had already been realized. Third, the divisions had chafed at the strong centralization of strategy, which limited their creativity.

*Source: Marr (2005)*

More abstractly, coordination problems can arise when there are externalities between different organizational units if they are decentralized. For example, suppose that one division’s R&D investments benefit the products of a different division of the firm. In that case, there is a positive externality. If decisions about R&D are completely decentralized, the first division is likely to ignore the effects of its research on other products. Or suppose that one division launches a new product that competes directly with an existing product in a differ-
ent division. That is an example of a negative externality. If product line decisions are completely decentralized, such problems are more likely to arise.

Benefits of Decentralization

We now consider benefits of decentralizing a decision. The most important consideration here derives directly from Hayek's analysis: whether there is valuable information at lower levels that is costly to communicate. If there is, the firm should seriously consider decentralizing some decisions to make use of that knowledge. We first focus on factors that determine how costly information is to communicate, and then list a few other benefits of decentralization.

Specific v. General Knowledge

Information or knowledge falls along a spectrum from cheap to communicate, to costly to communicate. For example, the price of a pound of coffee is very cheap to communicate. It is a single number, and can be transmitted via information technology instantly. By contrast, the field of personnel economics is costly to communicate. It is complex, and much of it cannot be quantified easily. Even putting it in the words of a text does not completely explain the concepts. Students require the professor to explain and develop ideas more fully. Understanding is also helped by the use of examples, such as case studies, and by the student's personal work experience.

In economics, information that is cheap to communicate is usually called general knowledge. Information that is costly to communicate is usually called specific knowledge. When we say costly to communicate, we mean either costly to transmit, or costly to understand once received. Information may be passed from one person to another, but if the receiving person does not understand the information, it has not been communicated. Thus, here is our spectrum on which to analyze knowledge:

<table>
<thead>
<tr>
<th>General Knowledge</th>
<th>Specific Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>- low cost to communicate</td>
<td>- high cost to communicate</td>
</tr>
<tr>
<td>- more likely to centralize</td>
<td>- more likely to decentralize</td>
</tr>
</tbody>
</table>

Specific Versus General Knowledge

Figure 5.1

Let us take a moment to clarify terms, since this terminology that economists use can unfortunately be confusing. Specific and general knowledge sound like specific and general human capital, and also sound similar to specialization. However, those are different ideas. To refresh your memory, the concepts of specific and general human capital refer to skills that are relatively more valuable at one employer, or valuable at many employers. Think of human capital as training, while what we are talking about in this section is the information that arises as you perform your day to day job. Finally, specialization refers to the extent to which a worker's skills (and, in Chapter 7, tasks on the job) are narrow or...
broad. Choosing a college major means specialization in this sense. An MBA, by contrast, is relatively unspecialized, since it involves studying a little bit about many different fields.]

Specific knowledge is the key idea here, since it will favor decentralization. Below are several attributes of information that make it more specific. These are the kinds of information that Hayek had in mind in his famous phrase about “the particular circumstances of time and place.”

Information that is Perishable

Information may be costly to communicate simply because it must be acted on quickly or it loses its value. A trader on a stock exchange must respond to market movements immediately or lose the ability to exploit new information and buy or sell orders as they appear. For this reason, traders are carefully screened, trained as needed, and then largely given the right to make trading decisions themselves. Similarly, suppose that a firm’s sales force sometimes receives large rush orders from key customers. If the customers will turn to alternative suppliers when they cannot receive a quick confirmation on the order, the firm is likely to give the salesperson some direction (possibly within limited parameters) and allow him to negotiate terms and decide whether to accept the order.

This is an example of an idea you will see again below. Centralization and hierarchy take time. Firms that employ those methods are more deliberative, but also slower to react and adapt.

Information that is Complex

One of the main reasons why information can be costly to communicate is complexity. Complexity can mean that there are many variables (as opposed to just the price of a pound of coffee). More important, however, is complexity in the sense of interdependence between different pieces of information. Consider the difference between a spreadsheet with a simple list of prices, one with many numbers and formulas, and one with many cross-references between cells. The latter is much more costly to communicate, since the recipient of the spreadsheet must understand the formulas and reasoning behind them.

We will use the same sense of complexity when talking about job design in Chapter 7. Interdependence between tasks makes it more difficult to separate tasks into different jobs. It also makes it more likely that the worker who performs those tasks has some specific knowledge about the work, because the work is complex in the sense we describe here.

Information that Requires Technical Skills

An important special case is when information being transmitted requires advanced human capital in order to be completely understood. This is the case with most scientific or technical knowledge. For example, many decisions about R&D must be made by the firm’s engineers, often with little oversight by management, simply because the decisions require technical knowledge that only the engineers possess.

Information that is Unforeseeable / Idiosyncratic

Information can be costly to communicate simply because frequent communication is necessary, which multiplies the costs of communication proportionally. This is the case when the environment is more dynamic and random. However, random environments do not necessarily imply that there is specific knowledge. Suppose that the job circumstances change constantly, but only vary in predictable ways. Perhaps a customer service repre-
sentative handles twelve different types of customer complaints, but the same types of complaints happen over and over again. This does not require much decentralization. Instead, the firm can set up standard operating procedures that dictate what the employee must do in each of the twelve circumstances. Thus, to the extent that the environment is unforeseeable or idiosyncratic, the employee is more likely to possess specific knowledge on the job.

Information that is Subjective / Experiential

There is a famous Supreme Court case in the United States in which a judge ruled that he could not define a concept, but knew it when he saw it. That is an example of information that is subjective or experiential. Such information is usually costly to communicate by its nature. It is called subjective or qualitative because it is impossible to quantify, and difficult to rigorously describe in an economical way. Communicating subjective information usually requires sitting down and talking about the situation, explaining it verbally, and allowing the recipient to ask clarifying questions. In some cases the level of subjectivity is so high that the information is called experiential – in order to really understand the information, you have to experience it for yourself.

In order to use subjective or experiential information effectively, it is almost inevitable that some decentralization will be required. For example, performance evaluations in white collar jobs are usually quite subjective. For this reason firms usually allow supervisors to conduct such evaluations, even though this subjects the firm to the risk of inconsistency and legal liability from discrimination.

Other Benefits of Decentralization

Saves Management Time

Decentralization of some decisions is necessary to prevent overwhelming central management. Less important decisions, and those that require less coordination, tend to be pushed to lower level management simply so that top management can focus resources on the most important decisions. The same applies at all levels of the hierarchy, so that less important and less coordination-dependent decisions tend to get pushed farther down in the firm.

Develops Management Skills

Analysis and decision-making skills are forms of human capital that, to a good extent, must be learned on the job. In order to develop good future managers, the firm must give lower level managers some room to make decisions. As a first step, a new manager will be given discretion over less important issues, and may be constrained (e.g., by limiting expenditures via a budget) to prevent too much damage from mistakes. As the manager’s talent grows, he is given greater discretion, and over decisions with higher stakes. This is an excellent way for the firm to train managers who can then be promoted into higher positions.

Note that this is one reason why it is common for subordinates to be promoted into positions similar to those of their supervisor. As a subordinate gains more experience in the job, the supervisor passes on more tasks and decisions to the subordinate, in effect training the employee in how to do the supervisor’s job.
Intrinsic Motivation

Decentralization is one part of what is often called job enrichment, which we discuss in Chapter 7. The other part of job enrichment is assigning more tasks to the worker. One benefit of both is that the job may be more challenging and interesting to the worker. Therefore, the worker may be more intrinsically motivated to perform the work diligently.

Authority & Responsibility

In the Army, the lowest position is denoted Private. Corporals and then sergeants have authority over these lowest level soldiers. Sergeants report to lieutenants, who in turn report to captains. Captains are below majors, who are themselves below lieutenant colonels and colonels. Generals are at the top of this hierarchy, and the number of stars that a general wears determines rank among generals.

There are a number of reasons for having a strict hierarchy in the military. The most obvious is that in combat situations decisions must be made quickly. Collective decision making is slow and highly impractical during battle. Thus, a clear hierarchy is set up in which orders to subordinates are obeyed without question or delay.

Very few firms are engaged in the kind of production that requires split-second decisions with absolute deference to higher authority. But there are some business situations that do fit into this category. Even when a team may be involved in the negotiations to discuss a deal with another firm or client, one member must speak for the team and must have the authority to make the decision.

When jobs are designed, it is generally important to determine the level of authority associated with the job and how much responsibility is assigned to any given job. More often than not, the degree of authority and responsibility associated with a given job depends on the job holder. Some vice presidents have more authority than others, the difference presumably determined by ability differences.

An important consideration discussed above is the extent to which the job holder has specific knowledge that is costly to communicate to colleagues, but that is valuable for the firm. Another important consideration is the extent to which the job holder’s work must be coordinated with colleagues. We used these ideas to characterize a decision as being relatively centralized or decentralized. While useful, this is a simplification. In this section, we go beyond this basic tradeoff of use of specific knowledge v. coordination, and think about decision making a little more carefully.

Decision Making as a Multi-Stage Process

A decision is not just a binary answer to a question; it is a deliberative process. It can be helpful to think of a decision as having several stages. One way to do so that fits many types of decisions is to characterize the decision as having four stages.4

4 See Fama & Jensen (1983).
1. Initiatives;
2. Ratification;
3. Implementation;

The first stage, initiatives, is the process of coming up with a set of options. This is perhaps the most crucial stage for creativity and innovation. We sometimes think of this stage as brainstorming – what are our possibilities?

Once the possibilities are identified, one of the options must be chosen. This is the second stage, ratification. This stage corresponds to the term strategy in common usage. It is when the basic direction of future actions is decided.

Once the strategy is chosen, there may be many possible ways in which the strategy can be pursued. This is the third stage, implementation. Another common term for this stage is tactics. Once more, there can be quite a bit of creativity, though of a more applied sort, at this stage. As is often said, the devil is in the details.

Finally, it is important that implementation conforms to the strategy chosen at the second stage. Thus, the fourth stage is monitoring of the implementation.

Almost any decision-making process fits this description to a good degree. Suppose that you are a plant manager and have been told to cut costs by 10% in the next year. How would you do so? First, you need to figure out what basic approaches you might take. These would include things like cutting salary; layoffs; obtaining better terms with suppliers; or programs to increase efficiency on the plant floor. Then you need to choose between these ideas (though you can choose more than one option, in this example). Once you have chosen some options, you need to figure out how to implement them. And, finally, you will want to monitor your progress (and certainly the central office will do so) as the year progresses.

It is useful to break decision making into different steps because different steps can be given to different people. In particular, stages one and three, initiatives and implementation, are more likely to be decentralized, while stages two and four, ratification and monitoring, are more likely to be centralized, for two reasons.

The first reason to decentralize stages one and three, and centralize stages two and four, is to balance the goals of using specific knowledge at lower levels against the need for coordination. Often, much of the information needed to figure out what options are available resides with lower level employees. The same is even truer for the implementation stage. Consider again the example of the plant manager who needs to cut costs by 10%. He is likely to ask his staff for suggestions about ways to do so. They, in turn, are likely to turn to their staffs for advice. That is decentralization of stage one, initiatives. At the implementation stage, the manager will again ask his staff to devise ways to implement the cost cutting, since lower level employees will know many details that he will not.

However, the choice of the overall cost cutting strategy is likely to be made by the plant manager, in order to make sure that it is consistent with firm goals. The same is true with monitoring of implementation, to make sure that cost cutting programs do not conflict with each other or with the firm’s strategy (say, by harming product quality too much). Thus, relative centralization of stages two and four provides better coordination.
The second reason to centralize stages two and four compared to stages one and three is that employees tend to have imperfect incentives, so their interests are not always aligned with the firm’s. When that is the case, the firm must safeguard against empowered employees who make decisions that favor their goals over the firm’s. An important way to do so is to limit their discretion, by retaining the right to ratify their decisions, and by monitoring how the decisions are implemented. In other words, when decision makers have imperfect incentives, some system of checks and balances is necessary.

This idea applies not only at lower levels, but all the way to the top of the organization. One of the key roles of a board of directors is to provide oversight for a CEO with imperfect incentives. Note that privately owned companies do not generally need to worry about such governance structures, because their managers are owners, and thus have no incentive problem.

As can be seen, the stark distinction between centralizing and decentralizing a decision that we started with above is too simplistic. In practice, decisions are made over several stages, and a typical pattern is for decentralization of early stages, followed by centralization, and then decentralization for additional initiatives or implementation, followed by centralization for more ratification or monitoring. This allows the firm to get many of the benefits of both use of lower level specific knowledge and coordination simultaneously.

The first and third stages are often referred to as decision management, while the second and fourth are often referred to as decision control. Our basic point is that decision management tends to be more decentralized, while decision control tends to be more centralized.

This characterization is a simple but useful framework for thinking about making many different kinds of decisions. For example, it gives guidance to your thinking about how a company can implement a change program. Furthermore, it clarifies what it means to empower your workers. Empowerment usually means giving decision management rights to employees, but reserving at least some decision control rights (especially the final stage, monitoring), for yourself.

Finally, the distinction between decision management and decision control provides a way to think about what managers and their subordinates do. It is often said that a hierarchy involves the passing of information and decisions up and down. Much of middle management work involves two things. The first is ratifying and monitoring the work of subordinates, and passing them relevant general knowledge as needed. The second is taking their output, processing it in some way, and making suggestions about initiatives and implementation to a higher level manager. Thus, decision control is what we mean by the term hierarchy.

We now present a further refined way of thinking about decision management and decision control. A firm can emphasize the creative stages, decision management. Or, it can emphasize the control stages, decision control. The relative emphasis has important implications for the type of firm that it is – how innovative, how risky, the culture it develops, and so on.

**Flat v. Hierarchical Structures**

A firm has a great deal of leeway with respect to the ways in which it sets up authoritarian relations. For example, it is possible to have a flatter organization in which each individual has more authority over which projects are approved or rejected. Alternatively, a firm can
be set up with a very steep authoritarian pyramid, where each level has the ability to veto decisions made by lower levels. In the language we just defined, a flat structure places more emphasis on decision management, while a steeper structure places more emphasis on decision control.

**A Tradeoff Between Two Types of Errors**

Whether jobs should be designed with flat or steep authority structures depends on the costs of accepting poor projects relative to the costs of rejecting good ones. In statistical parlance, the issue is the trading off false positives (Type I errors) versus false negatives (Type II errors).

To make this more concrete, consider an example provided by a former Stanford student in the women’s apparel business. Hong Kong natives Gladys and Willie run a New York-based firm that imports women’s lingerie and sleepwear. They describe themselves as having a “young and funky” image. Gladys needed to decide whether to branch out into more romantic lingerie. Doing so required some up-front investment in marketing, distribution and, most significantly, in setting up the production line. Major losses would occur if the line did not sell as hoped. She had to decide between going into the romantic lingerie line and forgoing the opportunity. There are two types of errors that she could have made. She might have invested in the line when doing so turned out to be unprofitable, or she could have decided not to produce the line when doing so was profitable. Table 5.1 lists the possibilities.

<table>
<thead>
<tr>
<th>Line is profitable</th>
<th>Produces</th>
<th>Does not produce</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good decision</td>
<td>False negative error</td>
<td></td>
</tr>
<tr>
<td>False positive error</td>
<td>Good decision</td>
<td></td>
</tr>
</tbody>
</table>

**Possible Types of Correct or Incorrect Decisions**

Table 5.1

If she produces lingerie and it turns out to be unprofitable, then she has committed a **false positive** error. False positive error is defined as accepting an unprofitable project. If she chooses not to produce lingerie and the line would have been profitable, then she has made a **false negative** error. False negative error is defined as rejecting a profitable project.

There is a trade-off between false positive and false negative errors. If Gladys adopts a very aggressive policy and accepts every new project that comes along, she will never commit a false negative error. Because she always produces, she is certain to make false positive errors whenever a project is unprofitable. The more aggressive her posture, the higher the likelihood of a false positive error, and the lower the likelihood of a false negative error.

Alternatively, she can adopt an extremely conservative posture, rejecting every new project that comes along. Because she never produces a new line, she can never find herself in the box where she produces, but shouldn’t have. She never commits a false positive error, but now some false negative errors are certain to occur. Whenever a new product would have been profitable, she commits a false negative error because she never accepts it. The more conservative her posture, the lower the likelihood of a false positive error and the higher the likelihood of a false negative error.
Figure 5.2 shows the tradeoff. On the horizontal axis is the probability of making a false positive error – that is, going ahead with the project when it is unprofitable. On the vertical axis is the probability of making a false negative error – that is, rejecting the project when it is profitable. At point D all projects are accepted, so the probability of accepting a project given that it is unprofitable is 1, and bad projects are accepted with certainty. At point C all projects are rejected, so the probability of rejecting a project given that it is profitable is 1 and good projects are rejected with certainty. The tradeoff is shown by the solid line between C and D. If some projects are accepted and some are rejected, then the firm ends up at an interior point, like A. At A, some but not all good projects are rejected, and some but not all bad projects are accepted.

Error Tradeoffs and Authority Structure

How does the firm decide how aggressive it should be when accepting new projects? If it is very costly to accept bad projects, the firm wants a more stringent rule, which moves it toward C. If it is very costly to pass up a good project, then the firm wants a more lenient rule, pushing it toward D.

The goal is to enhance the information set so as to allow for fewer errors of each type. If decisions were better informed, then the tradeoff would be along the dotted curve rather than the solid one. Note that on the dotted curve, the firm can make less of each type of error. Point B implies less false positive error and less false negative error than point A. The firm would always prefer the dotted curve to the solid curve. However, keep in mind that information is costly. Decisions are better along the dotted curve, but the cost of obtaining the dotted curve might be more delay or higher consulting fees.

Three Examples of Authority Patterns

We now return to job design and authority patterns. By structuring the authority relations in different ways, different kinds of errors are made more or less likely. Consider two different decision making structures for Gladys and Willie, shown in Figure 5.3.
Two Possible Authority Structures

Figure 5.3

A hierarchical structure places Gladys above Willie in decision making. Willie comes up with new ideas, and is allowed to reject any project, but does not have authority to accept any project on his own. All he can do is render a recommendation to accept. This is what we previously called decision management; he is recommending either new ideas, or how to implement ideas that were already chosen. Gladys then has the authority to make the final call. This is decision control; either ratifying new ideas, or monitoring implementation. Such a structure works in the direction of reducing false positive error and increasing false negative error.

The second structure, called a flat authority structure, tends to have less false negative errors and more false positive errors. In this case the firm is set up such that both Gladys and Willie evaluate projects and choose whether to accept or to reject individually. Since both spend their time coming up with and evaluating new projects, but do not check each other. There is relatively more emphasis on decision management, and less on decision control, compared to the hierarchy.

Which structure is better? It is easy to show that a hierarchical structure will approve fewer projects than a flat decision structure (see the Appendix for formal derivations of ideas discussed in this section). The hierarchical structure makes fewer false positive errors, but more false negative errors. Fewer bad projects are accepted, but more good projects are rejected. There are two reasons. First, since the hierarchical structure requires two approvals rather than one, the test that a project must pass is more stringent. Second, since two people are required for an evaluation rather than one, fewer decisions are made. Gladys and Willie, working in parallel, can evaluate more projects than each can alone.

If each project must be evaluated by Willie before it goes to Gladys, then half as many projects get an initial screening. Thus, fewer projects are approved. Failure to review a project is an implicit rejection. The bottom line is that a hierarchical decision structure, where low-level jobs are denied the authority to make the final decision on a project, results in a more stringent criterion and fewer project approvals than does a flat, egalitarian authority structure.

There is a third possibility. The structure can be made flat with second opinions required. Rather than putting Gladys above Willie, the firm can simply require that every project reviewed by Willie is also reviewed by Gladys, and vice versa. If both agree, the decision is obvious. If they disagree, then some other rule must be used to reconcile the differences. There are a number of possibilities, but for our purposes the details of reconciliation are irrelevant. It is always true, irrespective of the reconciliation rule used, that a second opinion structure is less stringent than a hierarchical structure, but more stringent than a single opinion flat structure.
Think about it this way. Under the hierarchical structure, when Willie rejects a project, Gladys never even sees it. She only sees those projects that he passes up. In the case of a second opinion structure, Gladys sees even those projects that Willie rejects. If Gladys likes the project, then the two opinions must be reconciled. As long as some of these reconciliations result in positive outcomes, projects that would not have been accepted by the hierarchical structure will be accepted by the second opinion structure. Therefore, although both structures review the same number of projects, the second opinion structure is less stringent and approves more projects than the hierarchical structure.

On the other hand, the second opinion structure is more stringent than a flat, single decision maker structure. This is somewhat less obvious than it seems. It is true that a second opinion can sometimes reverse an initial decision to reject, but it is also true that a second opinion can reverse an initial decision to accept. The main reason that a second opinion structure is more stringent is that when second opinions are required, fewer projects are considered at all. If it takes, say, a week for one person to review a project, then the flat structure with a single decision maker produces two decisions per week: one by each. By contrast, the second opinion structure produces only one decision per week, since both must review every proposal.

Figure 5.4 augments Figure 5.2 by showing the locations of the different job authority structures with respect to false positive and false negative errors. Which structure should a firm choose? Since there is an implicit tradeoff in the choice, the answer depends on the payoffs associated with each outcome. Now that three structures are identified, let us consider three types of payoff regimes, shown in Figures 5.5-5.7.

**Small Upside, Large Downside**

Figure 5.5 shows a payoff structure that might be appropriate for the Exxon Valdez, a large oil tanker. As you may recall, a few years ago the Valdez was involved in an accident that caused a major oil spill for which Exxon was financially responsible. The losses associated with the oil spill ran into the billions of dollars in cleanup, litigation, and settlement costs. The captain of the ship was blamed for the accident, and there was evidence that alcohol was in part to blame.
Small Upside, Large Downside  

Figure 5.5

The Valdez situation is typical of one variety of payoff structure. Doing the job extremely well results in small gains relative to the expected amount, but making a mistake can be disastrous. The upside of payoffs is limited, but the downside implies losses in the billions.

The Valdez’s captain makes more profit for the company by bringing the oil shipment in somewhat early, but it is not worth the risk of trying to guide the ship while under the influence of alcohol. A mistake is just too costly. When the payoff structure looks as it does in Figure 5.5, the firm wants to minimize false positive error and is willing to accept higher levels of false negative error. Think of the “project” in this case as deciding whether to proceed before being sober, which reduces the expected time to port. “Waiting to sober up when proceeding would not have resulted in an accident” would be a false negative error. “Proceeding while intoxicated, resulting in a crash” would be a false positive error. Because a false positive error is so costly, the firm should adopt a structure that minimizes that risk — a relatively hierarchical decision making process. The captain should not, and likely did not, have the authority to proceed before being completely sober. Had he radioed in for approval, the company would likely have denied permission to proceed, reducing the likelihood of disaster.

Large Upside, Small Downside  

Figure 5.6 has a payoff function with a big upside and limited downside. This corresponds to many new firms. Most of the time new firms fail, producing negative or only slightly positive profit levels. Once every so often the innovators hit it big, earning large profits. Which structure favors the upside? A flat structure with very little supervisory veto power minimizes the amount of false negative error. The startup firm does not want to be too cautious. It has little reputation or capital to lose, and much to gain. It is better to take chances. Those that do not pan out can be abandoned without much harm to the firm, since the firm has little to lose in the first place.
Young firms often do give a great deal of authority to individual workers. It is sometimes argued that creative people do not do well in hierarchical firms. Although true, the problem may not be that different types of people go to different types of firms. It may have more to do with the decision rules used in a given structure. Since hierarchical structures tend to err in the direction of minimizing false positive error, tolerating the rejection of some good projects, a hierarchical firm does not encourage creativity. Flatter authority structures, which allow each worker more choice, also allow creativity to flourish. Risky, wild ideas that would be rejected in a hierarchical structure may be allowed to proceed in the flat, single decision maker structure.

This discussion is another way to see the idea that we discussed above – decentralized structures tend to foster creativity, while centralized structures are better at maintaining control and avoiding large mistakes.

**Symmetric Payoffs**

Most firms are neither in the *Exxon Valdez* or startup category. Payoffs are more symmetric in most businesses, especially established ones. Figure 5.7 shows a payoff function for a local store. Great performance and innovative work are unlikely to generate the large upside that a startup might experience. Poor performance and shoddy work may cost the firm some money, but not the disastrous losses incurred by Exxon when the oil spill occurred. In this case, the firm prefers tolerable levels of false positive error and false negative error, minimizing neither at the expense of the other.
Investing in Better Quality Decision Making

In Figure 5.2 two curves were drawn. The dotted curve lies inside the solid curve. Other things equal, it is better to face the constraint of the dotted curve than that of the solid curve because for any given level of false positive error, the dotted curve implies less false negative error than does the solid curve (except at the endpoints). In other words, fewer errors of either type are made along the dotted curve than on the solid curve.

How does the firm move to the dotted curve? Unfortunately this cannot be done without cost. Fewer errors of either type are made along the dotted curve than on the solid curve. To get to the dotted curve, it is necessary to improve the decision process. There are a number of ways to do so, all of which are costly. The firm can attempt to hire better evaluators by going for more able, higher priced workers. The firm can give an evaluator more time to consider each project. Or, the firm can make more information available to the evaluator, say by hiring outside consultants or buying other data. Whether any of these steps is profitable depends on how much is gained relative to the amount lost by making poor decisions.

An Application: Air Traffic Routes

Consider an example of an airline pilot trying to decide whether to take a quicker route through thunderstorms or a safer but lengthier route around the storms. Some hypothetical information is given in Table 5.2.

<table>
<thead>
<tr>
<th>Probability of crash</th>
<th>Cost of crash</th>
<th>Expected cost of crash</th>
<th>Expected fuel cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Go through storm</td>
<td>$10^{-5}$</td>
<td>$1$ billion</td>
<td>$17,000$</td>
</tr>
<tr>
<td>Go around storm</td>
<td>$10^{-9}$</td>
<td>$1$ billion</td>
<td>$20,000$</td>
</tr>
</tbody>
</table>

Thunderstorms and Airplane Travel

Table 5.2

What should the firm do? If a plane crash occurs, the firm incurs a cost of $1 billion from loss of capital and reputation, higher insurance rates, and litigation. However, fuel costs and the odds of a crash are different on the two routes, so the expected costs are different. The better choice here is for the airplane to go through the storm. This example, with some upside and a large downside, fits the payoff structure in Figure 5.5. Thus firm could set up a hierarchy, but here the situation is somewhat different because there is an incentive problem. The pilot is more likely to care about her own life than about firm profitability. To offset the tendency to be too conservative, the airline might make a rule that pilots take the shortest route unless they receive approval from the company to deviate. Then, when faced with thunderstorms, the pilot radios in to request the longer, more costly route. The company can approve or deny the request.

Given the probability and payoff structure in the table, the airline would always approve the pilot’s request to take the longer route. If the pilot flies through the storm, the expected cost is:

\[
\text{Expected cost of going through the storm} = (10^{-5})(1B) + 17,000 = 27,000.
\]
The first term reflects the probability of a crash, given that the pilot flies through the storm, times the loss to the company from a crash. The $17,000 is the fuel cost. If the pilot takes the longer route, the expected cost is:

\[ \text{Expected cost of going around} = (10^{-9})(\$1B) + \$20,000 = \$20,001. \]

The first term reflects the probability of a crash, given that the longer route is chosen, times the cost to the company of a crash. The $20,000 reflects the higher fuel cost. The expected cost is lower than that of flying through the storm.

Since it will always grant permission when asked, there is no sense in having pilots call in for approval in this case. Under these circumstances, the pilot is given full authority to select a longer route when there are thunderstorms along the shorter route. In other words, a flat structure is preferred.

Note that the probability of committing a false positive error is zero. Since the shorter route is never taken, there is no chance that the short route would be taken when it would have resulted in a crash. Conversely, the probability of committing a false negative error is one. It is certain that the shorter route through thunderstorms will be rejected even when going through the storm would not have resulted in a crash, because the shorter route is always rejected.

Now suppose that the company can buy some additional information. A new device, the Forecaster, can more accurately forecast lightning strikes. By doing this, it can offer a recommendation on when it is safe to fly through the storm and when it is not. Table 5.3 provides statistics on the accuracy of the Forecaster.

<table>
<thead>
<tr>
<th>Probability of recommendation</th>
<th>Go through</th>
<th>Go around</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of crash when following recommendation</td>
<td>0.9999</td>
<td>$10^{-4}$</td>
</tr>
</tbody>
</table>

**Accuracy of the Forecaster**

Table 5.3

The Forecaster recommends that the route through the storm be taken 9,999/10,000 times. The probability of a crash, given a positive recommendation, is only 1 in 100 million. However, when the Forecaster recommends that the storm be avoided, going against the advice will result in a crash in one in ten times. The Forecaster provides very accurate information that allows much better decisions to be made.

If the device is used, the optimal decision from the company’s point of view changes. Whereas before, the company always preferred that storms be avoided, now the company would prefer that the shorter route be taken whenever the Forecaster recommends so. Under a positive recommendation, the expected cost of a trip is:

\[ \text{Expected cost of going through the storm} = (10^{-9})(\$1B) + \$17,000 = \$17,010, \]

while the expected cost from taking the route around the storm is still $20,001. Thus, the company would prefer the shorter route to be taken whenever the Forecaster makes a
positive recommendation, but the longer route to be taken whenever it makes a negative recommendation.

Two questions arise. First, should the firm purchase the Forecaster? The answer depends on the cost of the Forecaster and on its frequency of usage. Second, and more importantly, what kind of authority structure should be used if it is purchased?

Without the Forecaster, each pilot could be left to choose his or her own route. Because the pilot’s incentives were aligned with the firm’s interests, there was no incentive problem from such decentralization. With the Forecaster, the situation is different. The airline prefers that the shorter route through the storm is taken 99.99% of the time. The pilot may have a different view. First, the likelihood of a crash when the Forecaster recommends the shorter route is 1 in 100 million. If the longer route is taken, the likelihood of a crash is 1 in 1 billion. Both numbers are small, but the former is still ten times greater than the latter. Other things equal, the pilot prefers the lower odds of a crash and chooses the longer route, even though the shorter route is more cost-effective. Second, crashes aside, it may be more difficult to fly through thunderstorms than to divert. Both of these considerations suggest that the pilot might not make the same decision that the company would make.

In this case, the authority structure should change if the Forecaster is purchased. The firm now prefers a hierarchical structure, in which the pilot must request permission to divert around the storm and in which the airline has the right to refuse permission.

This example illustrates the interplay between available information, decision making structures, and employee incentives. When central information is available (the Forecaster), centralization through greater use of hierarchy is more likely to make sense. When it is not, decentralization makes sense. However, decentralization only works well when the interests of the decision maker correspond reasonably well with the interests of the organization.

**Summary**

Organizations are more than just a collection of workers performing physical tasks. A more valuable view, especially in modern economies, is that it is a generator and processor of knowledge. Much of an organization’s structure, and (as we will soon see) its job design, has the goal of making use of information to improve efficiency, adapt, and innovate.

Economies themselves face the same issues. The most effective organization of an economy is a market-oriented one. This provides decentralization so that specific knowledge of time and place can be exploited. Using knowledge dispersed through the economy makes production more efficient. It also makes the economy more adaptable, since adaptation usually requires continuous reaction to local events. Finally, it makes the economy more creative, since it encourages all individuals to invest in and make use of their ideas.

Market economies are self-organizing systems – there is little central direction of the system. Yet they are able to achieve a great amount of coordination despite this. The way this is achieved is through prices, which serve as signals of the value of different goods and services. Decentralized decision makers use this information to guide their decisions, without having to know the details of other possible uses of the goods and services, or of who their customers might be and how they will use the product. Thus, prices are an efficient way to transmit general knowledge and coordinate across the economy.
Finally, markets work well because they provide strong incentives through ownership of assets. Because individuals can own, buy and sell assets, they are motivated to maximize the use of those assets. This also improves the matching of information and decisions, since there is an incentives to move information to decision rights, and vice versa, to maximize the value of both. The strong incentives in a market economy are also important reasons why decentralized economies are so innovative and dynamic.

The market metaphor is very useful in thinking about organizational design. While no firm will be able to perfectly mimic a market structure, the basic goal of its organization should be to replicate, as well as possible, the way markets work. Thus, organizational structure should be designed to achieve several key goals: use information, especially specific knowledge dispersed throughout the firm and its customers or suppliers; coordinate as needed across the firm; and provide appropriate incentives to maximize firm value.

A good starting place for thinking about organization design is to identify the most valuable specific knowledge. One way to do so is to ask, “Who / What / Where / When / Why” about specific knowledge. Who in or outside of the firm has knowledge that is valuable to the business, but difficult to communicate to central management? What kind of knowledge is it? Is it tied to a location, or is it perishable? Why is it valuable to the business? Answering these questions will give strong guidance about situations where decentralization of some decision rights is likely to be important.

The next step would be to think about coordination problems that might arise if decisions were decentralized in the ways suggested in that analysis. To the extent that coordination problems arise, three steps could be taken. One would be to try to improve incentives for coordination. This often does not work adequately, however. Another would be to tip the balance toward greater centralization of some decisions. A third would be to implement some other coordination mechanisms (discussed in the next chapter).

At this stage, it would be appropriate to think about decision making processes. Separating out different stages of decision making and giving them to different individuals or units can further improve the use of specific knowledge and coordination. In addition, a firm can choose different structures that emphasis decision management or decision control more. This allows the firm to choose how they approach the fundamental tradeoff that any decision making system has between creativity and control.

Most of our discussion in this chapter has looked at allocation of decision making from the point of view of the entire organization. However, the extent of discretion given to an employee is a major part of the job design. Thus this chapter also serves as an introduction to the topic of job design. After discussing additional issues in overall organizational structure in Chapter 6, in Chapter 7 we take up the topic of job design.

**Study Questions**

1. Define the concepts of specific and general knowledge carefully. Go back and review the concepts of specialization of human capital investments, and of general versus firm-specific human capital. These terms all sound similar, but mean different things, so make sure that you understand the differences.

2. Provide examples of specific knowledge of particular circumstances in a job you held. In other words, give examples of knowledge or information that was important to your
work, but would be costly for you to communicate clearly to your boss. Did your employer allow you to make decisions that required that information? Why or why not?

3. Information technology lowers the costs of communicating many forms of information. What effect do you predict that this will have on organizational structures?

4. Your job is to oversee R&D for a large pharmaceutical company. A new blockbuster drug can mean enormous profits for your firm, partly because of many years of patent protection. New drug development is an extremely large financial investment. Mistakes can be very costly, because new drugs might harm customers, and damage the firm’s brand name. Finally, the government’s Food and Drug Administration imposes its own very stringent oversight (decision control) of your drugs at the last stage, before they are approved as new products. Describe the decision making process that you would recommend. Would it vary with different stages of new product development (e.g., basic exploratory research compared to final drug development)?

5. Do you think that a firm’s decision making methods will evolve as it grows from a small startup into a mature company? If so, how and why? What effect is that likely to have on the company’s workforce? Its corporate culture? Is it likely to alter its effectiveness at product design? How might its emphasis change as the company matures?

6. Your boss is frustrated with the decision making process used at your firm. He asks you to prepare a memo detailing the costs of this process. How would you measure the costs of a given decision making process? What are the dimensions of costs and benefits to a firm from a specific process?

References


Hierarchical, Flat & Second Opinion Structures

Here we formally compare hierarchical, flat and second opinion decision making structures, using our example of Willie and Gladys' firm. In a hierarchy, Willie evaluates new projects, rejects some, and recommends the rest to Gladys. Gladys evaluates the projects that Willie recommends, rejects some, and implements the others. In a flat structure, each evaluates different new projects. Those that are accepted by one or both are implemented. In a second opinion structure, both evaluate all projects. If there is disagreement, the firm uses some resolution procedure that results in a fraction $\lambda < 1$ being accepted and implemented. We will simplify by assuming that they flip a coin, so that $\lambda = \frac{1}{2}$, but it is easy to show that the same results hold for any $\lambda$ between 0 and 1.

Each person reviews $N$ new or recommended projects per period. Thus, in a hierarchy the bottleneck is Willie, and $N$ projects are evaluated per period. In a flat organization, $2N$ projects are evaluated. In a second opinion structure, $N$ are evaluated, and each spends half of their time on new projects and half on those already checked by their colleague. Projects have binary outcomes: they can be good (profitable), or bad (unprofitable).

The probability that a correct decision is made when a project is first evaluated = $p > \frac{1}{2}$. If it were not greater than $\frac{1}{2}$, then the firm would be better off flipping a coin to make decisions. The probability that a mistake is made when a project is first evaluated = $1-p$.

If a project is evaluated a second time, the decision is more accurate. This is because a project that was already recommended at the first stage is more likely to be a good project than if it was not yet evaluated. Therefore, the probability that a correct decision is made if a project is evaluated a second time = $q > p$; thus the probability that a mistake is made = $1-q < 1-p$. Of course, this only applies to the hierarchy and the second opinion model.

Figures 5A.1-3 show flow charts for a random new idea under each structure. In Figure 5A.3, the $\frac{1}{2}$'s note where a coin toss is necessary; half of those projects are accepted and half rejected (more generally it would be a fraction $\lambda$ accepted).
Hierarchical Structure
Figure 5A.1

Flat Structure
Figure 5A.2
Second Opinion Structure

Figure 5A.3

Table 5A.1 collects from these the probabilities of the four ultimate decisions (correct acceptances and rejections, and errors of both types), and shows the results after some algebra to simplify. In the top panel, probabilities are calculated for a single random idea. In the bottom panel, the total number of ideas resulting per period is calculated for each category, by adjusting for the fact that a flat structure evaluates twice as many ideas per period.

<table>
<thead>
<tr>
<th>Rate For One New Idea</th>
<th>Flat</th>
<th>2nd Opinion</th>
<th>Hierarchy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accept Good Idea</td>
<td>p</td>
<td>$\frac{1}{2}(p+q)$</td>
<td>p q</td>
</tr>
<tr>
<td>False Negative</td>
<td>1-p</td>
<td>1-$\frac{1}{2}(p+q)$</td>
<td>1-p q</td>
</tr>
<tr>
<td>False Positive</td>
<td>1-p</td>
<td>1-$\frac{1}{2}(p+q)$</td>
<td>(1-p)(1-q)</td>
</tr>
<tr>
<td>Reject Bad Idea</td>
<td>p</td>
<td>$\frac{1}{2}(p+q)$</td>
<td>1-(1-p)(1-q)</td>
</tr>
<tr>
<td><strong>Overall Throughput</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accept Good Ideas</td>
<td>2N p</td>
<td>N$\frac{1}{2}(p+q)$</td>
<td>N p q</td>
</tr>
<tr>
<td>False Negatives</td>
<td>2N(1-p)</td>
<td>N[1-$\frac{1}{2}(p+q)$]</td>
<td>N(1-p) q</td>
</tr>
<tr>
<td>False Positives</td>
<td>2N(1-p)</td>
<td>N[1-$\frac{1}{2}(p+q)$]</td>
<td>N(1-p)(1-q)</td>
</tr>
<tr>
<td>Reject Bad Ideas</td>
<td>2N p</td>
<td>N$\frac{1}{2}(p+q)$</td>
<td>N[1-(1-p)(1-q)]</td>
</tr>
</tbody>
</table>

Comparisons of Authority Structures

Table 5A.1
Table 5A.2 then summarizes these results in a more convenient form, ranking the three structures against each other. The top panel ranks the likelihood of each outcome for one new idea. The bottom panel ranks the total number of each outcome realized, given the total number of ideas that each structure evaluates.

<table>
<thead>
<tr>
<th></th>
<th>Most</th>
<th>Middle</th>
<th>Least</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rate For One New Idea</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accept Good Idea</td>
<td>2nd Opinion &gt;</td>
<td>Flat &gt;</td>
<td>Hierarchy</td>
</tr>
<tr>
<td>False Negative</td>
<td>Hierarchy &gt;</td>
<td>Flat &gt;</td>
<td>2nd Opinion</td>
</tr>
<tr>
<td>False Positive</td>
<td>Flat &gt; 2nd</td>
<td>Opinion &gt;</td>
<td>Hierarchy</td>
</tr>
<tr>
<td>Reject Bad Idea</td>
<td>Hierarchy &gt;</td>
<td>2nd Opinion</td>
<td>&gt; Flat</td>
</tr>
<tr>
<td><strong>Overall Throughput</strong></td>
<td>Flat &gt; 2nd</td>
<td>Opinion &gt;</td>
<td>Hierarchy</td>
</tr>
<tr>
<td>Accept Good Ideas</td>
<td>Flat &gt; 2nd</td>
<td>Opinion &gt;</td>
<td>Hierarchy</td>
</tr>
<tr>
<td>False Negatives</td>
<td>Flat &gt; Hierarchy</td>
<td>&gt; 2nd Opinion</td>
<td></td>
</tr>
<tr>
<td>False Positives</td>
<td>Flat &gt; 2nd</td>
<td>Opinion &gt;</td>
<td>Hierarchy</td>
</tr>
<tr>
<td>Reject Bad Ideas</td>
<td>Flat &gt; Hierarchy</td>
<td>&gt; 2nd Opinion</td>
<td></td>
</tr>
</tbody>
</table>

**Summary of Results**

Table 5A.2

The bottom panel shows that the flat structure means more mistakes overall than the other structures, when we account for throughput. However, it also implements more good ideas than the other two structures. Thus, flat structures result in more overall change, more successful new projects, and more failures. In addition, since ideas are only evaluated once, they make decisions more quickly. Flat structures are the most creative and turbulent of the three. 2nd opinion structures are intermediate; good ideas are given a second chance to be accepted compared to a hierarchy, but bad ideas are given a second chance to be rejected, compared to a flat structure. Hierarchies are the most conservative of the three structures.
Organizational Structure

“Chaos was the law of nature; Order was the dream of man.” (Henry Brooks Adams)

Introduction

In the last chapter, we focused on allocation of individual decisions. We now concentrate on a closely-related, macro-level question: the structure of the overall organization. This is an enormous topic that could easily be the subject of an entire course in its own right. Here we concentrate on just a few key questions.

Imagine that you are part of the management team for the new CEO of a large aerospace firm. The firm has had mixed success in the past decade. Some very sophisticated new products have been launched with great success, but others have failed. Your firm faces increasing competition from more aggressive, fast moving competitors. A competitor has been plagued with quality problems, and the CEO wants to avoid similar troubles here. Finally, there have been serious conflicts between several divisions, over both product design and marketing. Your CEO asks you to think about the best way to structure your firm to continue past successes and address the concerns. Where do you start?

Clearly, you are going to have to think about how to break the firm up into smaller subunits. Large firms are almost always too complex for a single management team to oversee. One issue that we discuss here is how a firm can break its organization into divisions or other business units.

A second question that you will have to consider is how to set up broad patterns of authority. Your boss is concerned that the competition is fast moving, suggesting that your firm’s decision making is too slow. At the same time, your firm has had some success at innovation, and there are pressures to maintain high quality. The issues discussed in the last chapter are relevant, and they will be useful in this chapter too. For example, you should know from Chapter 5 that you generally face a tradeoff between speed of decision making and likelihood of errors in decision making, which could cause quality problems.

Organizational structures can have an important effect on career patterns for employees. Traditional functional hierarchies imply that most employees spend their careers investing in more in-depth knowledge within their functional area. Other structures may weaken the link between structure and job ladders, potentially weakening skill development. In addition, the choice of structure can have an impact on the effectiveness of performance evaluations. We will mention this at several points in this chapter (Chapter 9 focuses on evaluation in more depth). An aerospace firm must maintain a cutting edge engineering staff, so these could be serious concerns.

Finally, the conflicts between different divisions suggest that coordination should be improved within your organization. What implications do the various structures have for the
severity and types of coordination problems that tend to arise in a firm? What methods are available for coordinating? These are the types of questions covered in this chapter.

**Trends in Organizational Structure**

It is common for the popular business press to say that organizational structures have flattened in the last two decades, but what does this mean?

One study analyzed the structure of the top management hierarchy in over 300 large US firms from 1986 through 1999. They found that the number of hierarchical levels between the CEO and the lowest-level managers had decreased by over 25% during that time. At the same time, the number of managers reporting directly to the CEO increased dramatically. Firms are indeed flattening their structures.

The authors found several other interesting patterns. The evidence suggests that firms have decentralized decision making as they flattened. An explanation for this is that a manager who oversees more subordinates has less time to make decisions, and must delegate more. In addition, firms with flatter hierarchies tended to make greater use of incentive compensation based on broader performance measures (a term that we define more carefully in Chapter 9), such as long-term financial performance or stock ownership.

*Source: Rajan & Wulf (2006)*

**Types of Organizational Structures**

In this section we describe four general types of structures that firms tend to use. In the following section, we discuss the factors that favor use of each structure. If you are already familiar with the basic organizational structures, this section may be skipped or skimmed quickly.

Actual firm structures can be exceedingly complex. They are often a combination of the structures outlined here. For example, a firm may have many divisions, with some using a matrix structure, others using traditional functional hierarchies, and still others using more informal network structures. Even a given unit can combine approaches. A business unit might be organized along functional hierarchical lines, but make extensive use of informal communication and collaboration – a network approach – at the same time. Finally, a unit might organize some decisions and employees in one way, but other decisions in another. Thus, think of these four structures as basic building blocks. The larger and more complex the firm, the more likely is it that its structure will combine elements and be too complicated to describe adequately in an organizational chart.

Before we describe the four general types of structures, we first note that all but three of these structures make extensive use of an important principle that we introduced in the last chapter: Hierarchy.
**Hierarchy**

Figure 6.1 shows the classic organization structure that has played the most important historical role: Functional Hierarchy. A Functional Hierarchy has two important elements, functional structure, and hierarchy. Here we discuss the importance of hierarchy.

![Functional Hierarchy](image)

Hierarch means that most communication, supervision, and decision making occurs in clear linear paths between the bottom of the organization and the top. In Figure 6.1, each functional area has a head (such as the Executive Vice President (EVP) of Sales). The Vice President (VP) reports to the EVP, the Manager to the VP, and the Assistant to the Manager. Each employee is expected to generally work directly with the next levels (above and below) in the chain of command, and not communicate much (if at all) with employees at two or more levels above or below. In this structure, there is no ambiguity about who reports to whom. This structure is exactly what we analyzed in Chapter 5.

In a hierarchy, ultimate authority resides with the CEO. However, many decisions are made at lower levels, for the reasons described in the last chapter. Decentralization is often necessary to make use of lower level specific knowledge, and to economize on the time of top management. Decentralization then leads to some separation of decision management and control, to provide oversight and better coordination.

Virtually all structures have hierarchy in this sense, because of the benefits of having a **single decision maker**, as mentioned briefly in Chapter 5. If it is not clear who an employee should report to, and who is the ultimate decision maker, the organization will incur several costs. Decision making is likely to be slower as the group tries to achieve consensus. There will be more confusion as employees are not certain who they should go to with questions. Finally, group decision making increases the importance of politics in how things get done in the organization. For these reasons, virtually all structures make extensive use of hierarchy in making decisions.

Hierarchy can become costly in larger organizations. To see why, recall the childhood game of “Telephone.” In Telephone, the players sit in a circle. The first player whispers a phrase in the ear of the person next to him. The second person whispers what he heard in
the ear of the next person, and so on. Once the circle of whispering is complete, the last participant says out loud what he heard. Typically, the result is nothing like the original phrase – it has become garbled in the process.

The same may occur in hierarchies, especially ones with many levels. Imagine that a lower level employee has information that the CEO needs in order to coordinate the function with other functions in the firm. The employee communicates the information to his boss, who communicates it to her boss, and so on until the information reaches the top. Unless the information is easily quantifiable, it is likely to be somewhat garbled on retelling. This becomes even more likely if, as is often the case, at each level the information is processed in some way before being passed on further. A similar garbling effect can occur in the opposite direction with respect to implementation of decisions. If the CEO makes a decision that affects the lower level worker, this must be communicated downward.

The first formal organizations developed in the first large institutions: governments, armies, and religious bodies. Hierarchy was an important element in all of these early structures, and remains so to this day.

**Functional Structure**

The second element of classical structure is use of functional organization units. Once a firm grows to a certain size, the overall structure must be broken up into more manageable sub-units, or top management will be overwhelmed. One logical way to do so is to group employees who are similar in some way together in the same unit. A very common way to do so is to divide the organization into units based on related skills and tasks.

In Figure 6.1, the firm is organized into three functions (realistically, there would be more). One group of employees focuses on Research and Development (R&D). Another focuses on production of the product, while a third focuses on sales.

In this structure, an employee’s career tends to be solely within the function. An entry level Production employee would gradually earn promotions, but stay within the Production function. A Sales employee would similarly be promoted to positions of greater responsibility within Sales.

Functional structures are driven by the substantial benefits of specialization. In a functional structure, an employee works almost entirely with colleagues in the same function. Communication between the functions tends to occur at higher levels (in the extreme case, at the level of the CEO and top management). To perform his tasks, a functional employee only needs to have knowledge of the concepts and skills related to his functional area. An accountant would need to have expertise in accounting, but little or none in Production or Marketing. This creates enormous economies in skill investments as stated in Chapter 3.

A second benefit of specialization is an important element of the next chapter: jobs can be specialized, so that workers do a limited set of tasks that are closely related in the knowledge and skills needed to do them. A Functional Hierarchy makes it easier to design what we will call Narrow jobs with a limited number of tasks. And, of course, narrowly designed jobs tend to be quite compatible with narrowly focused human capital.

A Functional structure has additional benefits in that the hierarchy will tend to work more smoothly. Because employees usually start at the bottom of the hierarchy and are promoted upward, a boss in a Functional structure usually supervises a job very much like one she once worked in. This means that she will be a much more effective manager. She
will be better able to process information provided by, and give direction to, her subordinates. (In other words, both decision management and decision control are likely to be more effective.) Communication will be easier because she has the same skills as her staff. Performance evaluation will be more accurate, because she understands the context of the job and how to determine the extent to which performance is due to the employee’s efforts or other factors.

As can be seen, there are tremendous advantages to a Functional structure, in terms of training, job design, decision making, and communication. However, there can also be an important disadvantage. Specialization implies that workers usually have little understanding of how their work affects employees in other parts of the organization, so their work may end up poorly coordinated with that of other functions.

There are two sources of this problem. The first is that specialization in skills and tasks makes it unlikely that employees will take into account the perspectives of colleagues in other functions. This is partly due to ignorance of other functions, and partly to distorted incentives (in a Functional structure, performance evaluation tends to focus on functional expertise at the expense of coordination with other functions; Chapter 8 should give you some tools to understand why). The second is the kind of garbling of communication and decisions that can occur in a Hierarchy, as described above. In a Functional structure, most communication is up and down within the function, rather than across functions.

**Divisional Structure**

The larger the firm, the more will the CEO need to break the structure into manageable sub-units. Authority for most major decisions within each unit is then delegated to the unit manager. One way to break up the organization is into functions. This takes advantage of the compelling economic advantages of specialization. However, a simple functional hierarchy is usually inadequate for larger organizations, because the functional units may be too large to be effectively managed (for example, the garbling phenomenon described above can become too severe).

Moreover, larger firms tend to be more complex firms. Their product lines expand. They sell in more regions. They begin to use a greater variety of techniques and technologies. All of these imply that coordination problems across functions are likely to become even more severe. For these reasons, most mid to large sized firms also break their structures into some form of divisions. An example of a Divisional structure is shown in Figure 6.2.
In this example, the firm is divided into three divisions. Within each division, the firm uses Functional Hierarchy. This illustrates that firms can and do combine elements of all of the structures described in this chapter.

In addition to ameliorating problems that arise from a too-large Functional structure, a Divisional structure allows each division to focus on a narrow part of the overall company business. This is, in effect, a large-scale version of the gains from specialization.

For example, consider a firm that designs and sells high-end computers. It sells desktop, laptop, and handheld computers to educational customers (schools and universities) and corporate customers. Some computers use the latest technology and are very powerful, while others use commodity parts, perform only basic tasks, and are intended to be inexpensive.

This firm has a somewhat complex business, with important variation in customers, product type, and technology. If there is one large unit to oversee all of these activities, top managers will continuously have to make difficult tradeoffs between how they focus their efforts and resources.

Instead, the firm might consider splitting the firm into different divisions. Each division would be assigned to a particular area of focus. This would simplify the mission and operation of each division.

Suppose, for example, that the firm decided to organize into product divisions. The three divisions in Figure 6.2 would be for Desktop, Laptop, and Handheld products. An employee working in the Laptop Products division would focus his efforts on how he can best improve the design, manufacture, or sale of laptop computers. His work is greatly simplified, since he does not have to take into consideration how his work affects desktop or handheld computers. The same is true throughout the Laptop Products hierarchy, all the way up to the President of Laptop Products.
In other words, specialization can occur along many lines. The ones usually emphasized in economics (and this text) involve investments in human capital, and narrow job design. At the level of organizational structure, however, the same principle comes into play, and the specialization could be along many different dimensions.

For example, our computer company could just as easily organize into divisions based on customer types. In that case, there might be two divisions for Education and Corporate Products. Employees in each division would structure their work to emphasize success of the product line that their division was responsible for. Notice that their focus would be slightly different. Now an employee might work on desktop, laptop, and handheld computers, but emphasize how they can be designed or sold to Educational customers.

A third possible divisional structure would be to organize along technological lines. Our computer firm might then have divisions for high-end, advanced technology computers, and a different one for low-end commodity products. This type of divisional structure is less common than the others, but certainly occurs frequently. A final common example would be to organize into regional divisions. The firm might have one division for the Americas, a second for Europe, and a third for Asia.

Any time workers are specialized, there is the potential for coordination costs to arise between workers with different specializations. This also applies to Divisional structures. Organizing a firm into divisions means that, to some extent, the firm becomes a set of autonomous mini firms. Typically, the President of a division is evaluated and rewarded largely for her division’s performance. Specialization of incentives, strategy, skills, and job design within each division may mean that the divisions do not adequately consider the effects of their work on the other divisions.

Using our Market Metaphor for organizational design from Chapter 5, coordination problems arise because different organizational units impose positive or negative externalities on each other. In the case of positive externalities, there is inadequate cooperation. In the case of negative externalities there is too much competition.

Consider our computer company once more. If we organize the firm into Product Divisions, the Laptop and Desktop Products divisions will compete with each other for sales. This competition may reduce overall profits, since each division has inadequate incentives to consider the negative externality it imposes on the other. Thus, much of the CEO and top management’s job in a divisional firm is to oversee the activities of each division to improve coordination between them. This can involve retaining some decision rights over strategy of divisions, establishing incentive schemes that reward cooperation, and settling any disputes that arise.

**How Should a Firm Define Divisions?**

How does a firm decide which divisional structure to use? That is difficult to say without knowledge of the specific business. The question largely boils down to deciding along which dimension the firm needs to differentiate its activities more. For example, suppose that in this firm similar technology is needed for each of the three products. If so, then breaking the firm into technology-based divisions would be a mistake. Each division would attempt to focus R&D on its particular product type. This would lead to duplication and incompatibility of R&D across the divisions. If, on the other hand, very different technology is required for each type of product, little would be lost by dividing R&D into three different product areas.
It often makes sense to organize sales and marketing into geographical, product, or customer divisions. A geographical structure may make sense if the same sales team can effectively sell to all types of customers, in which case they can be assigned to all customers in a specific area. In addition, geographical divisions make it easier for the firm to vary marketing, product information, and selling techniques across regions with different languages or cultures.

Organizing sales by product or customer may make sense if different types of products or customers require different sales and marketing techniques. For example, corporate or high-end customers may be less price sensitive, but demand more sophisticated and specialized products and service.

One important principle for organizing work at any level of the firm is modularity. A system is modular to the extent that it can be broken up into largely distinct, functionally separate parts. The principle of modularity has applications in many areas, from software design to evolutionary biology to sociology. It also can be applied to organizational structure and job design.

Breaking work up into different jobs, or work groups, or business units, or divisions creates coordination costs across those units. To the extent that the work lends itself to modularity, coordination costs tend to be lower. Therefore, a helpful rule of thumb when thinking about how to define divisions is to look for modularity. How can your business be broken into largely self-managing groups? Above we discussed a case where it made little sense to break up R&D – this function is not easily modularized in the computer firm. By contrast, it is often relatively simple to modularize a sales organization by region or product.

Modularity is possible to the extent that externalities or coordination costs are not too large across units. In software design, this principle is sometimes called low coupling and high cohesion. When tasks require high cohesion, it is important to put them together in the same organizational unit, or develop strong coordination mechanisms (see below). Tasks that require only loose coupling are good candidates for being assigned to different modules (jobs, units, or divisions). We will use the principle of modularity again in the next chapter.

### Software Design at Microsoft

In early 2007 Microsoft released the much anticipated Vista upgrade to the Windows operating system – over a year later than initially announced. The reason was apparently inadequate modularity in design of Windows – a very complex, large software program (technically, a set of interconnecting programs).

Individual programmers were tasked to produce their own parts of the overall code. These parts were then pieced together into the overall program. However, Microsoft has found that the development process gave inadequate attention to the principle of modularity. There were over 4,000 engineers working on Windows, and they were all managed together as one seamless project.

Microsoft altered the development process in two ways. First, it broke the project up into more sub-projects, much like divisions in a firm’s structure. When managers drew up a chart of the overall Windows project, it was “8 feet tall and 11 feet wide and looked like a haphazard train map with hundreds of tracks crisscrossing each other.” The leaders of the project redesigned the project into sub-projects which could be added or removed without compromising the whole operating system.
Second, managers pushed for better modularity by requiring higher quality in the code of individual parts before they were added to the larger project. Engineers were expected to more completely debug their projects, and design them to be more like independent Lego blocks responsible for single functions.

Source: Guth (2005)

As a final note on Divisional structure, in principle a firm can organize different sets of tasks into different structures. For example, R&D might be centralized into one group for the entire firm. This would maximize the benefits from economies of scale and standardization of parts and products across the company. The same firm might then break Marketing and Sales into customer divisions, to maximize flexibility in selling techniques. Production might be organized by region, to minimize distribution costs. However, when different areas of the firm are organized along completely different lines, complexity escalates rapidly. Production will be focused on regional variation in its methods, and not coordinate well with Sales or R&D. R&D may focus on developing cutting edge LCD screen technology, but not on how to vary its designs for different types of customers. The job of the CEO and top management in coordinating these groups that are given different focuses becomes only harder. For this reason, there are benefits from keeping the structure simple, and grouping employees into divisions that are relatively consistent with each other.

**Matrix or Project Structure**

One drawback of a Divisional structure is that the firm loses some of the gains from specialization and economies of scale that a simple Functional structure would provide. If different divisions have their own Sales staffs, for example, efficiencies may be lost as each Sales group provides some services separately. Similarly, if each division has its own Accounting department, the firm may end up with multiple, incompatible, and relatively expensive accounting systems.

This problem can be especially severe in functions like R&D that require very advanced technical knowledge. If there are multiple functional groups in separate divisions, each has to develop this technical knowledge on its own (or use some form of knowledge management system; see below). That tends to be wasteful and can lead to less effective R&D compared to combining all research efforts in one organizational unit.

The third structure, Matrix or Project, can be used to balance the desire for economies of scale in specialized functional areas against the desire for a Divisional structure. Figure 6.3 shows an example. In this case, the firm is organized into both Functional and Divisional groupings. Each employee is assigned to two groups: a functional area, and a divisional area. For example, an engineer might be assigned to software design, and to laptop computers. Notably, each employee has two bosses – one in his function, and one in his division.

A relatively permanent, formal structure of the type shown in Figure 6.3 is called a Matrix structure. The term matrix refers to the two-dimensional design where each employee has to organizational assignments. In fact, the organizational chart in Figure 6.3 looks like a $2 \times 2$ matrix.

Many firms assign functional workers to projects that span functions. Such projects are more temporary in nature than the formal divisions illustrated in Figure 6.3. Project struc-
ture refers to such relatively temporary matrix-type structures. Both use the same techniques; the difference is one of relative permanence and formality. Finally, ad hoc, short-lived cross-functional teams use the same general principles, but are even more temporary than a Project structure.

It is worth noting that in a Divisional structure an employee also has, in a sense, two organizational affiliations: Division, and Function with the Division. However, this is different from the Matrix approach. In the Divisional approach, there are different Functions (e.g., different Sales groups) for each division. In the Matrix approach, there is a single functional organization that spans divisions. In other words, in a Divisional structure functions are nested within divisions, while in a Matrix structure functions are overlaid across divisions.

The Matrix or Project structure allows for many of the benefits of both Functional and Divisional structures. By having one side of the matrix organized by function, many of the benefits of Functional structure are achieved. An employee can focus investments in human capital, becoming a subject-matter expert. The career path is clear (along functional lines, as greater functional expertise is developed). Performance evaluation by the functional supervisor is more effective.

At the same time, workers are grouped across functions within divisions. This can greatly improve coordination in several ways. First, workers in the same division have common goals. Second, each worker has a boss whose responsibility is to promote the division’s performance. This motivates workers to coordinate better. Third, workers communicate and work more directly with colleagues from other functional areas. Compare that to a traditional hierarchy, where most of the cross-functional coordination is achieved at the top of each function. In a Matrix or Project approach, an important advantage is that a great deal of coordination can occur at lower levels, closer to where most of the work is actually done. This idea will be important when we introduce the idea of an integration problem below.

Matrix structures would seem to provide the best of both Functional and Divisional structures. However, they also have substantial disadvantages. This derives from the violation of the principle of a single decision maker. In this structure each employee has two
bosses. These bosses have goals that conflict with each other. The functional boss’ goal is to maximize functional expertise. For example, the R&D boss is likely to want to emphasize more basic R&D that can be applied to many different products, rather than more applied R&D that applies to a specific product division. By contrast, the engineer’s Divisional manager will pressure the engineer to focus efforts on the specific product.

In most cases, a worker’s primary loyalty is likely to be to his functional boss, since his career will be within the functional job ladder. Therefore, matrix structures may end up with relatively more emphasis on functional expertise rather than divisional performance than a pure Divisional structure (but less than a Functional structure).

This is a very difficult situation for the engineer to be in. He will constantly be pressured by both managers to focus his work in different areas. The bosses will disagree with each other about how his performance should be evaluated and rewarded. Thus, Matrix structures tend to lead to more office politics, greater conflict, slower decision making, and more bureaucracy (for example, more time spent in meetings to resolve conflicts). Of course, to some extent that is the point: the conflict is a coordination mechanism that reveals and resolves coordination problems. The problem is that it is a complex, costly mechanism, so it should only be used if the advantages of a Matrix type structure are high enough.

**Network Structure**

A final structure that has received increasing attention in the last two decades is the Network structure, illustrated in Figure 6.4 (the circle represents the boundary of the firm). A Network structure is difficult to define rigorously. It is a structure that gives more emphasis to informal relationships between individual workers and managers inside and outside the organization. In all organizations, there is an informal structure that runs parallel to the formal structure. This informal structure is each manager’s network or set of relationships with colleagues. When an employee needs to get something done, he does not always follow the formal chain of decision making implied by the organization chart. Instead, sometimes he directly contacts a colleague somewhere else in the organization. The benefit of doing so is that decision making can be done more quickly, with less garbling and better coordination. The cost is that it can undermine the formal chain of command.

Similarly, most firms make at least some use of ad hoc cross-functional teams as described in the previous subsection. These provide some balancing of Functional expertise and Divisional focus, but more flexibility than a rigid Matrix. Some firms place particularly great emphasis on these kinds of ad hoc work groups, and on encouraging managers to use their network of relationship to directly communicate and coordinate. Such organizations are sometimes called Network structures.
In Figure 6.4, the firm is organized into three different teams. These teams might be responsible for designing specific new products, or for working with specific clients. The firm also has two other organizational units of the more traditional Functional form, R&D and Production. These are presumably organized this way to maximize gains from specialization and economies of scale. Finally, in this case there are three outside groups that one or more of the inside groups works with regularly, a major Customer, an Industry Consortium, and a key Supplier.

In this structure, employees in the teams collaborate with each other to promote the team goals. In this sense, one could easily have written the organization chart in Figure 6.4 as having two functional divisions, and three other divisions (the teams).

The idea that a Network structure is meant to emphasize is that not all work is organized into traditional hierarchies. In this example, a manager in Team 2 may need to develop good working relationships with managers in the other two teams, with R&D and Production, and also with the major Customer.

One way to think of Network structures is that they are a form of internal market mechanism for getting work done, but the asset that is used in this market is not a tangible financial one. Instead, it is a manager’s intangible personal relationships with others, often called social capital. Network analysts have found that managers who fill structural holes – fill gaps between two larger networks – have particular value in a network structure. This is because they complete the network structure, better matching supply with demand and making the influence market work more effectively. However, the use of intangible social capital means that the influence market does not have readily observable prices. For this reason, the transactions involve more negotiation and are more political in nature. Thus, while Network structures approximate a market design in quite interesting ways, they are an imperfect substitute.
Putting this all together, some firms use techniques associated with Functional, Divisional, or Matrix structures, but in relatively impermanent and fluid ways. To the extent that they do, decision making and coordination are based more on informal relationships than on formal chains of command. In practice, most firms use elements of all four approaches to structure; it is a question of degree.

Which Structure Should a Firm Use?

Which structure should your firm use? The first point is that it should not use a single structure for the entire firm. Firms use a blend of structures for different parts of their organization. Moreover, even for the same set of employees, techniques from several different structures are often blended (say, for different sets of decisions). Thus, the question is better framed as, which structure should your firm emphasize in each area?

Regardless of approach, there are two important forces that should always be remembered: the principle of a single decision maker, and the value of specialization. Consider the first of these two. Network-style structures can be costly because they are confusing. When the ultimate decision maker has not been clarified, it is more likely that employees can work at cross purposes. Therefore, even in a Network structure it is usually important to clearly specify leaders and goals of teams and other organizational units. For the same reason Network organizations usually still make extensive use of hierarchical decision making. A Network approach can be quite effective for decision management – brainstorming and creative activities – but decision control is still necessary.

A second important force is specialization. Virtually all firms make extensive use of functional structures. This allows employees to focus their skills, tasks, and careers. Therefore, functional hierarchy is generally an important part of almost every organization, and should be the starting point for most discussions of structure.

Beyond the importance of specialization and a single point of authority, other factors matter. A third factor affecting optimal structure is complexity of the business. Complexity implies that there are more areas that the firm has to master. The desire for gains from specialization therefore tends to lead to greater differentiation into smaller organization units. This drives not just functional organization, but also divisions and sub-divisions. Therefore, the more complex the business, the more elaborate should the structure be (Divisional, Matrix, or Network).

As we will argue in the next chapter, complexity also tends to imply that there is more specific knowledge at lower levels in the organization. For this reason, greater complexity generally means that a firm should make greater use of decentralization. Thus, there is a positive relationship between more complex structures like Divisional, and use of decentralization, both to and within the division. Similarly, simple functional organizations are more likely to use relatively centralized, hierarchical decision making.

Decentralization, and differentiation of the organization into sub-units, generates a greater need for coordination. Therefore, firms that use these approaches to structure have to make greater use of coordination mechanisms. We describe various coordination mechanisms below.

A fourth factor affecting optimal structure is stability of the firm’s business environment. The more stable the environment, the less important will be specific knowledge at lower levels, and the more will the firm be able to centralize decisions. In addition, stability im-
plies that the firm is more likely to have mastered its business processes, so that formal procedures can be established. When this is the case, there is less value to differentiating the firm into sub-units that focus on making related sets of decisions – because there are fewer important decisions to be made.

We saw in Chapter 5 that more hierarchical structures provide greater control, but at the cost of both slower decision making, and less creativity. Therefore, firms that need to act more quickly, have a stronger strategic need for creativity, and have less downside risk should decentralize more and make less use of hierarchy.

A fifth factor in determining the optimal structure is the nature of the coordination problem. In the next section we define two types of coordination problems. The latter, integration problems, favor Matrix or Network approaches. The former are amenable to simpler Divisional approaches.

### Coordination

#### Two Types of Coordination Problems

Coordination is needed when the work of two or more sub-units of the organization must be combined in some way to create greater firm value. A very simple example is an assembly line, where the output of one worker is given to the next, who adds some work to it, and passes it down the line. These workers need to coordinate the quantity and timing of their output with each other. They also may need to coordinate in more subtle ways so that their parts work well together and do not create quality problems.

Let us distinguish between two general types of coordination problems. One is coordination problems that do not require units to communicate with each other in order to coordinate. Call this a *synchronization problem*. The output of workers must be synchronized in some way, but they do not need to speak to each other to do so. The quantity and timing of assembly line work is an example in most cases.

Another example would be a multiproduct firm that has a strong brand name. In that case, products must be designed so that they are all consistent with the overall product image, quality, and look-and-feel that the firm is trying to establish. However, individual product line managers do not need to communicate with each other on a daily basis. A final example is a firm whose strategy is to provide uniform service at all of its retail locations. The stores do not need to communicate with each other, but they do need to be coordinated so that they all provide roughly the same type of service to their customers.

A second type of coordination problem is an *integration problem*. Consider our laptop computer product design problem. In the last chapter, we argued that a firm should ask “Who / What / Where / When / Why” to determine the specific knowledge in the organization that must be used to create firm value. We then argued that the firm should decentralize decisions to the employee who possesses that specific knowledge. Let us think about laptop computer design. What specific knowledge is necessary in order to design a profitable laptop computer?

Clearly, engineering knowledge is crucial. Electrical engineering expertise is needed to design the motherboard and other electronic parts. Software engineering expertise is needed to design the operating system and applications. Materials engineering expertise is needed to design the case and parts. All of these are specific rather than general know-
ledge, because they are costly to communicate to someone else. Therefore, according to our previous arguments we should decentralize most decisions about the laptop's design to our R&D staff.

However, there is other specific knowledge that is important to laptop computer design. One example is knowledge of customer's demand for computers. There are thousands of possible laptop computer designs. Each trades off some features against other features (e.g., computing power, battery life, weight, and price). In order to design a profitable laptop, the firm needs to decide which combination of features to produce and sell. The knowledge required to estimate which types of designs will sell best resides with the firm's sales and marketing staff (and its customers). This knowledge is also at least partly specific rather than general, because it is somewhat complex, and because much of the knowledge is likely to be qualitative. Therefore, to make best use of this specific knowledge, the laptop computer design should be decentralized to Sales and Marketing.

Other kinds of specific knowledge are also important for designing the laptop computer, including product costing, production, and distribution. Putting all of this together, we have a problem: there are multiple pieces of specific knowledge that must be combined to decide how to design the laptop computer. The firm cannot decentralize the decision to one of the groups, or the decision will emphasize that group's specific knowledge but ignore the knowledge from the other groups. This is what we call an integration problem.

There are two possible ways to solve an integration problem. One is to coordinate all of these groups at a high level. For example, the CEO could oversee product design. However, this is likely to fail, because each piece of information is very costly to communicate to the CEO, and there are multiple pieces.

The best way to solve this problem is usually to use a lateral coordination mechanism at lower levels, so that those who possess the disparate pieces of specific knowledge needed to make the decision work together. This is exactly what a Matrix or Project structure is designed to do. Thus, Matrix-type structures are best used when there are multiple pieces of specific knowledge that must be combined to make an important decision. Not surprisingly, Matrix-type structures (including cross-functional teams and Network structures) are most common for new product development.

Coordination Mechanisms

So far we have seen two coordination mechanisms. One is centralization. This works best when knowledge is easy to communicate. The second is a lateral coordination mechanism such as a cross-functional team or Matrix structure, or an informal version like a Network structure. These structures suffer from the fact that they are typically complex, confusing, and costly to manage. However, they are generally necessary evils when a firm faces an integration problem.

Synchronization problems can be coordinated much more easily, because they require little or no on-going communication between units, through a variety of mechanisms. Here we briefly describe a few.

Central Budgeting & Planning

Firms often have formal annual budgeting and planning processes. Organizational units propose plans and budgets for the next fiscal year. These are combined at the next level, and ultimately at the level of each division. The central corporate office then considers
these proposals from each division, compares likely returns on investment, and allocates resources to each division. Each division then takes this budget and allocates it to the levels below, and the process continues to the bottom of the organization.

It is striking how similar this approach is to the central planning in non-market economies. These techniques are more important in companies where the benefits of centralization, described above and in Chapter 5, are greater.

This process achieves coordination in several ways. First, it provides some control, since it limits discretion over spending by each unit. Second, the bottom-up process generates a great deal of information that is processed and accumulated at the top of the firm. This enables central decision makers to make better investment and strategic decisions through a form of decision management (to the extent that the information is not distorted). Third, the process of pushing budgets and annual plans down to lower levels provides broad coordination that divisional activities over the next fiscal year will be largely consistent with each other.

Training & Standard Operating Procedures

An excellent way to achieve consistency across employees and organizational units is to standardize practices. Thus, firms that desire a uniform customer experience in every store may require that employees wear identical uniforms. They also tend to invest more in extensive training in how to perform the job. The greater is the use of standard procedures, and the greater is the investment in uniform training across the organization, the more predictable will be behavior of employees. Predictability then provides coordination (synchronization) with no need for communication.

Corporate Culture

One benefit of a strong, consistent corporate culture is that employees tend to act in similar ways. This is quite similar to the effect of standard operating procedures and extensive training mentioned above. For example, a corporate culture in which all employees understand that cooperation is valued will foster greater coordination across units. However, a strong corporate culture does not necessarily improve coordination. Apple Computer for many years was famous for a strong culture of individualism. Because of this, employees often worked at cross-purposes and there was great conflict and poor coordination in the organization.

Communication

Another way to improve coordination is to develop communication systems that span the organization. These can include company newsletters, annual meetings, memos from central management, as well as many other examples. These improve coordination by increasing the extent to which dispersed business units have the same understanding of the goals and methods of the organization at a given point in time.

General Managers, Liaisons & Job Rotation

Much coordination happens because some managers provide an interface between specialists. In Figure 6.1, the CEO provides the coordination. The CEO will be better able to coordinate across R&D, Production, and Sales if he or she has a better understanding of each of these areas of the firm. Therefore, a manager who has a less specialized experience may be a better coordinator than one who is a specialist. For this reason, there can be great value to a firm from having a small set of managers who are generalists rather
than specialists. Most employees are specialists, so the firm enjoys all of the economic benefits from specialization that are described in this book. A smaller set then coordinates across the specialists, by developing relatively shallow knowledge, but over a broad set of functions.

One way to develop generalists is through job rotation. Firms often identify a small number of very promising junior managers and put them in formal job rotation programs. These managers move from one part of the firm to another over time, violating the principle of specialization. In the process, they develop some knowledge about different functions, but not a deep knowledge of any single function. They also gain a better big picture perspective of the firm than do specialists. Finally, they gain a network of contacts throughout the organization, which they can draw on later to improve coordination.

This coordination role is also one of the most important roles for MBAs. An MBA curriculum by definition is a general management curriculum that provides some working knowledge of the most common functional areas of a typical firm. It does not usually provide in-depth knowledge of each area. MBAs are more likely to be hired or promoted into positions where coordination is important than are specialists, all else equal.

Personality

In any position where coordination is needed, the personality of the manager matters. Coordination requires the ability to talk to, understand, and work with a diverse set of colleagues. It requires the ability to forge compromises and work within a more relationship or politics oriented setting. For this reason, MBAs, liaisons, and those who work in Matrix or Network structures need a somewhat different set of interpersonal skills than do those who work solely in a specialized area. The firm should keep this in mind both in recruiting and training specialists and generalists.

Networks

As the examples of corporate culture and personality illustrate, coordination often comes through informal mechanisms as well as formal ones. One benefit of a Network Organization is that whenever coordination is needed between two units, the relevant managers can simply communicate with each other without having to worry about formal reporting relationships. More generally, organizational sociologists emphasize the value to a manager of developing a strong network of contacts inside and outside the organization. A person with strong contacts to groups that otherwise would not be closely linked can be quite influential in an organization, filling in structural holes in the network. An effective manager is often one who is entrepreneurial about developing strong contacts, and recognizing opportunities to effect better coordination between groups.

Performance Evaluation & Incentives

We have left a very important, formal coordination mechanism for last: performance evaluation and incentives. Pay for performance inside an organization is analogous to the price system in market economies. As Hayek argued, prices generate an enormous amount of coordination (more along the lines of what we termed synchronization problems above, rather than integration problems), because prices are sufficient statistics for a great

---

1 In a small firm, generalist managers are more important, as there may not be enough staff to specialize on some tasks. We elaborate on this point in the discussion on entrepreneurship in Chapter 13.

deal of information about marginal costs and benefits of resources. Decisions made on the basis of market prices mean that decision makers are taking into account the value of resources in their alternative uses – coordinating – without needing to know what those alternative uses are.

If structured properly, a good incentive system can provide exactly this kind of coordination inside a firm. It does so by evaluating performance in a way that includes the effects of one employee’s actions on those of colleagues. That is, a good performance evaluation must measure and include the effects of any positive or negative externalities of the employee on the rest of the firm. We leave this topic for last, even though it is extremely important, because it is the subject of the entire third section of the textbook, Chapter 9-12. When performance evaluation is imperfect (and as we will see, it usually is), the firm may want to use some of the additional mechanisms described above to further improve coordination.

We now turn to some additional issues in the design and implementation of organizational structure.

**Implementation**

**Span of Control & Number of Levels in a Hierarchy**

As we have seen, virtually all firms make use of hierarchy to some extent. In Chapter 5 we discussed how more levels of decision control in a hierarchy increases the organization’s control, but at the expense of innovation. We now extend that discussion to other factors affecting the number of levels in the hierarchy.

Firms face a tradeoff in how to structure a hierarchy, illustrated in Figure 6.5. A greater number of levels increases costs in several ways, in addition to less innovation. Information has to be passed between more managers, increasing the extent of garbling described above. Information processing and decision making takes longer, as each step of communication takes a certain amount of time.

If the firm wishes to reduce the number of levels, it must expand the number of managers within each level in order to perform the same amount of work. That would be a flatter structure. A flatter structure has a greater span of control, or number of employees reporting to each manager. A steeper structure has a lower span of control.
A flatter span of control reduces the costs implied by adding layers to the hierarchy, but creates its own costs. Each manager must now supervise and direct more subordinates. This takes more of the manager’s time, reducing the time that can be spent on other tasks. In addition, it will generally reduce the effectiveness at supervision, since attention is spread more thinly. For example, the ability to monitor the work of subordinates will be weaker, so they are more likely to shirk on their duties. The ability to communicate effectively with each one, direct their activities, train them, and so on, will be reduced.

Thus, a hierarchy trades off span of control versus the number of levels. There are many factors that affect the optimal span of control and number of levels in a hierarchy. Anything that lowers the cost of supervising a subordinate generally will imply a larger optimal span of control, since the marginal cost of increasing the span of control by another subordinate is lower. Similarly, anything that lowers the cost of acquiring or passing on information will generally increase the optimal number of layers in the hierarchy, since it lowers the marginal cost of adding a new layer.

**Type of Tasks to be Performed**

The type of work to be performed affects the optimal structure of the hierarchy. The more that the work is routine, the easier is it for the manager to supervise each employee. The manager has to spend less time training the subordinate, and deciding (or helping the subordinate decide) what to do in non-standard situations. Most of the activity of supervision involves simply giving directions about general policies, and passing along information about the current situation. Therefore, more routine work implies that the span of control will tend to be larger, and the number of hierarchical layers smaller.

By contrast, work that is complex and varied requires more input from the supervisor. The subordinate is more likely to need the experience and skills of the supervisor in order to analyze the situation and decide what to do. Less of the lower level work can be specified in advance through a set of standard operating procedures. More monitoring is also necessary, since how the worker chooses to spend his time is more variable and the choice is more important. Therefore, more complex and varied work implies a smaller optimal span of control, and a greater number of levels in the hierarchy.3

**Skills of Managers & Subordinates**

Higher skills of both managers and subordinates will tend to increase the optimal span of control, and reduce the number of layers of a hierarchy. Each manager and subordinate can process more information and solve more difficult problems. More talented managers will generally be more effective at supervision and direction. More talented subordinates will generally be more effective at implementing the directions of their boss. Since almost all firms put more talented managers in higher levels (see below), this effect increases the optimal span of control at higher levels compared to lower levels.

**Incentive Problems**

Hierarchy implies delegation of some decisions. Delegation, in turn, means that the firm must ensure that employees have reasonable incentives to perform their tasks in the in-

---

3 A countervailing effect is that more complex environments may also be ones where firms need to make decisions more quickly, and where innovation is more important. These issues are discussed throughout Section II of the text. These effects would favor a hierarchy with fewer levels. In order to reduce the span of control, the firm might have to break the hierarchy into smaller units. It might also find that diseconomies of scale are particularly large in such circumstances. Innovation and fast decision making tend to favor smaller firms.
terests of firm objectives rather than personal objectives. There are two broad ways to do this. One is to develop performance evaluation and incentive systems; this is the topic of Section III of the text. The other is to monitor subordinates more closely to detect shirking. If performance evaluation is more effective, less monitoring is necessary, freeing up some of the supervisor’s time. Thus, where firms have access to better incentive plans, the span of control can be larger. Roughly speaking, performance evaluation tends to be better the closer the manager is to the top of the hierarchy. This is because the manager’s actions have a more direct effect on firm value, and are better reflected in available performance measures (especially accounting numbers). This effect tends to increase span of control at higher levels of the hierarchy.

Costs of Acquiring & Communicating Knowledge

An important determinant of optimal hierarchical structure is the cost of acquiring and communicating knowledge. Lowering the costs of acquiring knowledge increases the productivity of knowledge workers in a hierarchy. This, in turn, means that a manager can increase the span of control. For example, access to inexpensive analysis tools such as spreadsheets and statistical analysis programs increases the ability of a manager to analyze information collected from subordinates, so the manager will be able to analyze inputs from more subordinates. Similarly, subordinates will be able to solve more complex problems assigned by the supervisor. This has an effect similar to that of increasing the skills of the subordinate as described above.

Put another way, modern information technology tends to be a complement to the work of managers in a hierarchy, rather than a substitute. This raises the productivity of managers, and especially those who are highly skilled and at higher levels. This, in turn, affects wages as discussed below.

Lowering the costs of communication has very similar effects. Managers and subordinates are able to communicate with each other more effectively (with less garbling), cheaply, and quickly. This allows the manager to oversee more subordinates. Thus, the effects of advances in information technology are generally to increase the span of control, leading to flatter organizations. This is exactly what was found in the study by Rajan & Wulf described at the beginning of this chapter.

Advances in information technology may have ambiguous effects on the number of layers in the hierarchy, however. Lower cost of communication tends to increase the optimal number of layers, as communication is faster and there is less garbling at each stage. On the other hand, the greater productivity of knowledge workers, combined with higher spans of control, implies that more can be accomplished by each layer of the hierarchy, implying that fewer layers are needed for a given level of output.

Skills, Pay & Structure

Firms almost universally assign new employees to lower levels of the hierarchy, and promote the best performers upward to higher levels as they advance in their careers. This sorting, and the fact that managers increase their human capital as they gain experience, implies that firms assign their most talented managers to the highest hierarchical levels.

This pattern makes sense. Consider once more the Steep hierarchy depicted in Figure 6.5. Each manager communicates directly with his boss, and with each of his subordinates. To the extent that he makes his subordinates more effective in doing their work, this effect also increases the productivity of their subordinates, who are two levels below the
original manager. There is a cascading effect, as the productivity of a high level manager improves productivity of all employees below.

The same effect does not tend to work in reverse for several reasons. First, a lower level manager works directly with only a single higher level manager – most of the direct communications are downward. Second, by design a manager has a greater impact on the subordinate’s productivity than the opposite. The manager monitors the subordinate to prevent shirking and evaluate performance. The manager provides direction and training. While the subordinate’s work is often used as an input by the supervisor, it is combined with that of other colleagues at the same hierarchical level.

For these reasons, the effect of ability (and effort) rises with hierarchical level. This has several very important implications. The first, as we have already highlighted, is that the firm should sort employees so that the most talented rise to the top of the organization. This then implies hiring at the bottom, and an *internal labor market* of upward promotion over the employee’s career. (This system is especially well suited to functional hierarchies, of course.)

The effect also means that compensation should rise with hierarchical level. This happens because more talented managers are placed at higher levels; they must be paid more because they are more productive and have better market values. However, the effect on pay is reinforced by the fact that productivity rises more rapidly than does ability in a hierarchy. Why? For all of the reasons we have discussed in this sub-section. Therefore, in most firms not only does pay rise with hierarchical level, but it rises more rapidly at each new level, up to the very top of the organization. For an example of this, see Figure 11.1 in Chapter 11.

This and the previous section reinforce the discussion of the rising returns to skills that is threaded through the text, starting in Chapter 3. Decreases in the costs of acquiring and communicating knowledge increase the productivity of managers. In a hierarchy, these effects are generally even stronger at higher levels. That only strengthens the impact of advances in information technology on the widening of pay across different skill levels. We return to these issues in Chapter 8.

**Evolution of a Firm’s Structure**

The optimal structure of a firm is likely to evolve over time in several ways. First, consider the lessons of this and the previous chapter. Younger firms are more likely to emphasize decision management. That is, they will tend to make less use of hierarchy. It is important for such firms to innovate and take risks. The downside risk from mistakes is very low, and the upside potential from successful new projects is relatively high.

In addition, young firms tend to be small firms. The smaller the firm, the more likely is a Network structure to be feasible. Most employees will know each other, so it is easy to develop a system in which employees communicate with each other directly as necessary.

As the firm matures and grows, however, it is very likely to develop more formal structure (of a Divisional or Matrix form) because the larger number of employees makes an informal Network structure less efficient. It is also more likely to adopt more formal hierarchy, because the firm now has an established product line and brand name, so it should naturally become more conservative in decision making. A key issue facing growing companies is the ability to gradually introduce more formal structures in such a way as to not damage the existing organizational culture.
As a firm matures, it is usually possible to standardize procedures more. The firm has had time to learn effective methods. Similarly, the fact that the firm has survived tends to mean that its industry is relatively stable compared to those of new firms that failed to survive. Both effects mean that the firm can now institutionalize the knowledge that has accumulated. Standardizing procedures is a form of centralization, since it removes some discretion from lower level employees. Thus, in this way as well older, more mature organizations tend to become more formal, conservative, and hierarchical.

An interesting phenomenon is churning of organizational structure. It is common to observe firms change some elements of their structure quite often. In one extreme case, Apple Computer reorganized parts of its structure fourteen times in four years. Why might this be the case?

A plausible explanation is that no structure is perfect. Any given structure is relatively good at achieving some objectives for the firm, but trades that off with less effectiveness at other objectives. For example, Apple’s structure long favored engineering (the Advanced Technology Group), but was poor at developing products that reflected knowledge of customer preferences. Reorganizing to give greater input to Sales & Marketing alleviated these problems. Another firm might find that organization by region made the company effective at differentiating its product and marketing across regions, aiding implementation of a globalization strategy. However, this might come at the expense of some economies of scale in some areas. Once the firm has learned what to differentiate by regions, and how, it might reorganize to give greater importance to economies of scale to reduce costs. Later, it might reorganize yet again, to give more emphasize to some other objective, such as developing stronger advanced technological expertise.

**Summary**

Organizational structures in large firms can get extremely complex. As a result, firms take on a certain degree of bureaucracy. Such firms become slower at decision making, and less innovative and dynamic than smaller, younger firms. These patterns are often criticized, but this is often misguided. Directing a large complex enterprise in a coordinated fashion is a quite difficult problem. The Market Metaphor only gets us so far in designing the internal organization of the firm. After all, if a market organization would work, one should question why the firm exists as a large organization at all.

In microeconomics, the equilibrium size of a firm in a competitive market is determined by the extent of economies of scale (the range of output over which the firm’s per-unit costs fall). A firm’s product line is determined by the extent of economies of scope (whether increasing output of one product of product decreases per unit costs of producing a different product).

These bureaucratic costs of organizational structure are almost certainly one of the most important sources of diseconomies of both scale and scope. Consider first diseconomies of scale. As the firm gets larger, more employees are required. These require supervision and direction, which requires management. As the firm grows, more layers of hierarchy and more divisions or other sub-units are likely to be added. These slow down decision making, and create a more conservative firm. Lower-level employees and top management are farther from each other, increasing the extent to which information and decisions are garbled in communication. More talented, and expensive, managers are required to oversee operations.
Now consider diseconomies of scope. The more complex is the firm's business, the more will differentiation into divisions and other sub-units be necessary. Even if the firm is able to exploit modularity effectively, there are more connections between the modules (divisions) to manage. Coordination becomes a larger-scale problem, since a great spectrum of specializations need to be overseen. Coordination costs will generally be larger in more complex organizations, holding size of the hierarchy constant. Once more, more talented and expensive managers are needed. All of these effects are likely to be more severe, the more that a firm’s product lines are dissimilar to each other. In fact, the argument that firms should focus on their core competence recognizes that diversifying the firm’s operations increases organizational costs.

Virtually every firm makes extensive use of functional hierarchy, even when more modern innovations like Network structures are implemented. This is because of two important factors. The first is that it is almost always more efficient to have a single decision maker in an organization – firms are almost never organized as democracies or anything close to them. This means that hierarchy is the most fundamental and universal element of organizational structure.

The second factor is the benefits from specializing worker’s skills and tasks. Specialization of skills economizes on the costs of investments in human capital, and generally increases the returns on those investments (with the notable exception of generalists such as MBAs). As we will see in the next chapter, specialization in job design can also have large benefits. A Functional structure organizes the firm around different types of skills, maximizing those gains from specialization. It also has the additional benefits that career paths are clearly defined and easy to understand, and that supervision, direction and performance evaluation are likely to be more effective.

Even if a firm opts for simple functional hierarchy, a larger or more complex firm will need to break its structure into more manageable sub-units. The most common way to do so is to establish divisions, usually based on product line, region, customer type, or technology. Each type of division has its benefits and costs. Generally, the firm should decide which type of variation is more important to focus on in its organization (say, type of customer or region), and use that to guide establishment of divisions.

Structure is driven not only by specialization and hierarchy, but also by allocation of decisions to play off need to use local knowledge and coordination. Some decisions will be made at high levels, using classical centralized hierarchy. Others will be made lower down, if there is specific knowledge to tap into at lower levels.

When decisions are made at lower levels, those decisions makers need to be motivated to act in the interests of the firm. That will be section 3 of the text. Thus, incentives are an important coordination mechanism, when performance measures are appropriately chosen. This is analogous to prices in a market. The firm “sells the job” to the employee by giving some residual rights to profit, motivating intrapreneurship.

Since incentive schemes are imperfect, firms use a large array of other coordination mechanisms. These include MBAs, standard operating procedures, common training, corporate culture, and communication.

The most difficult structure problems arise when two or more groups possess important specific knowledge that must be combined to make a decision. This is what we called an integration problem. Here the decision cannot simply be decentralized to one group or the other. Instead, the decision must be overseen by one manager, or made together by the groups that possess the relevant knowledge. The most complex structures, such as Matrix
or Project structures, arise to address this problem. Though difficult to manage and work in, there is little alternative to these designs in firms that face important integration problems.

Just as almost every firm makes use of classical functional hierarchy, so too does every firm make use of more informal communication and coordination structures alongside the formal organization chart. Thus, a manager’s network of colleagues is an additional coordination mechanism. Firms that give particular emphasis to relational, fluid structures are sometimes said to have Network structures.

That structure is driven by specific knowledge illustrates that the product and environment of the firm drive strategy. More complex firms tend to differentiate more into different divisions. They also tend to use more decentralized approaches. As argued in the next chapter, this will affect job design as well. Such firms are also more likely to run into coordination problems, and thus adopt more complex structures.

The literature on organizational structure emphasizes that not only does strategy drive structure, but structure can also drive strategy. A given structure brings with it a set of tradeoffs. Some tasks are emphasized or performed more effectively than others. This in turn may affect the organization’s success and evolution across products, and innovation.

For another example, complex structures are quite costly to manage. There can be a value to the firm from reducing complexity to make the structure run more smoothly.

In Chapter 5 we introduced the Market Metaphor for thinking about organizational design. There, we emphasized the ideas of Adam Smith and Friedrich von Hayek, about the value of collective intelligence of a set of decentralized agents. However, just as with economies, there can be benefits from some centralized organization inside firms. This chapter emphasized those factors. Economies of scale in resource use may suggest a single organizational unit for some purposes (e.g., production). Positive and negative externalities, including the desire for consistent standards across the organization, often require some centralized decision making. This is true to the extent that market-like mechanisms, incentives, imperfectly achieve coordination with the firm.

As we have seen, organizational structure is driven by the same factors that drive economies: specialization, economies of scale, externalities, specific knowledge of time and place, and incentives. In the next chapter, move to a micro level of focus: the design of an individual employee’s job. We will see that many of the same ideas are also important determinants of good job design.

Study Questions

1. How is your firm or university organized? Explain how this reflects the principles discussed in this chapter.

2. Is a traditional functional hierarchy an effective structure for a military? Why or why not? Would such a structure be a better fit during time of war, or peace?

3. Economies of scale – falling cost per unit as total output increases – have been historically important in the automobile and steel industries. However, modern production methods have greatly reduced the importance of economies of scale in both. How do you predict this will affect the optimal organizational structure of companies in these
industries? What effect should this have on the competitive position of firms that were leaders in emphasizing economies of scale in the past?

4. Project structures first became important in the aerospace industry. These firms produced extremely complex, highly advanced products (e.g., jet airplanes or rockets). Project or matrix structures are also very common in consulting firms. Why do these structures make sense for these types of firms?

5. Project or network structures are often criticized (especially by those who actually work in them) for their complexity and the difficulty of working within them. Explain the factors that cause such criticisms.

6. Why might network structures be more popular now than twenty years ago? Explain the likely effect of each of the following: greater use of outsourcing and close strategic relationships with key suppliers; decreasing costs of information technology; more rapid technological and competitive change.

7. Companies often find that changing an organizational structure is quite difficult in practice. What kinds of costs do you foresee that a firm would face if it decided to change structure?

8. Following on the previous question, which of the general types of structures discussed in this chapter would you predict is easier to change? Why?

9. Acme Incorporate, the company used as an example in several of the chapters of this text, had eight levels in its management hierarchy, from entry level management to the CEO. From 1969 through 1988, the company tripled the number of management employees (see Baker, Gibbs & Holmstrom 1994). However, Acme never added any new hierarchical levels. Provide at least two explanations.

References


